

A T G C T G A C T C C A G T A A G A A T G C
A C T C C A G T A A G G A A T G C T A C T C
C A T C G A T G C T G A T C C A G T A A G A
G A G G A A T G G T A A G A A T G C C C A G
G T A A G A A T C G A T G C T G A C T C A G
T C G A T G C T G A C C C A G T A A G A A T
G G A A T G C T A C
A A T G G T A A G A
G A A T C G A T G C
C T C A G G A A T G
A C T C C A G T A A
T G C T G A T C C A
A T G C T G A C T C
A C T C C A G T A A
C A T C G A T G C T G A T C C A G T A A G A
G A G G A A T G G T A A G A A T G C C C A G
G T A A G A A T C G A T G C T G A C T C A G
T C G A T G C T G A C C C A G T A A G A A T
C A T C G G A A T C G A T G A A T C G A T A
T A A G A A T C G A T G C T G T G A C T C A
A C T C C A G T G A A T C G A T C T A C T C
C A T C G A T G C T G A T C C A G T A A G A
G A G G A A T G G T A A G A A T G C C C A G
G A A G E N E R A T I O N G T G T G A C T C
A T G C C C A G T A T G G E N O M E C T C A
A C T C C A G T A A G G A A T G C T A C T C
C A T C G A T G C T G A T C C A G T A A G A
G A G G A A T G G T A A G A A T G C C C A G
G T A A G A A T C G A T G C T G A C T C A G
T C G A T G C T G A C C C A G T A A G A A T
A A T G G T A A G A A T G C A G A A T G C C
C A G T A A G A A T G C G A A T G C T A C T

Annual Report of the Chief Medical Officer 2016

Generation Genome



Joan Reader

I publish my annual report on health in England as part of my statutory role. In this edition of my report, I take a detailed look at genomics, exploring how we currently utilise genomics in our health and care system and how its potential may be developed.

Leading figures from the field of genomics have contributed specialist chapters. I include topics such as the care and treatment of cancer, diagnosing rare diseases, the use of genomics in screening and 'personalised' prevention, precision medicine – the targeting of drugs to do the most good and least harm. I wanted also to consider genomics within the context of society and include a chapter considering the ethical and societal discourse around genomics. Using the evidence I make recommendations, aimed at those able to bring about change, to guide how our potential can be realised to both improve patients' outcomes and maintain the UK's leadership role in genomics.

Genomics is not tomorrow. Its here today. I believe genomic services should be available to more patients, whilst being a cost-effective service in the NHS. This is exciting science with the potential for fantastic improvements in prevention, health protection and patient outcomes. Now we need to welcome the genomic era and deliver the genomic dream!

Sally C Davies

Prof Dame Sally C Davies

Editors and authors

This report could not have been produced without the generous input of the following people.

Editor-in-Chief

Claudia Langenberg

MRC Epidemiology Unit, University of Cambridge

Policy Adviser

Mark Bale

Genomics England and Department of Health

Project Manager and Editor

Orla Murphy

Department of Health

Chapter Authors

Chapter 1 Chief Medical Officer's summary

Chapter lead
Sally C Davies

Department of Health

Chapter 2 100,000 Genome Project

Chapter lead and author
Marc Caulfield¹

Chapter authors

Marc Caulfield^{1,2}, Ellen Thomas^{1,3}, Clare Turnbull^{1,2}, Richard Scott^{1,4}, Augusto Rendon^{1,5}, Louise Jones^{1,2}, Kay Lawson^{1,6}, Nirupa Murugaesu^{1,7}, Clare Craig¹, Matina Prapa¹, Simon Thompson^{1,2}, Katarzyna Witkowska^{1,2}, Laura Shallcross⁸, Emma Baple^{1,9}, Caroline Wright^{1,10}, Katherine Smith^{1,2}, Matthew Parker¹, Ellen McDonagh^{1,2}, Eik Haroldsdottir¹, Antonio Rueda^{1,2}, Damian Smedley^{1,2}, Laura Riley^{1,2}, Suzanne Wood^{1,2}, Tim Rogers^{1,2}, Mark Bale^{1,11}, Lisa Dinh¹, Katrina Nevin-Ridley¹, Sue Hill^{1,12}, Tom Fowler^{1,2}

1 Genomics England, London

2 William Harvey Research Institute and the Barts Cancer Institute, Queen Mary University of London

3 Guy's Hospital, London

4 Great Ormond Street Hospital, London

5 University of Cambridge, Cambridge

6 University College London Hospital, London

7 St George's Hospital, London

8 UCL Institute of Health Informatics and Genomics England

9 Southampton General Hospital, Southampton

10 Wellcome Trust Sanger Institute, Hinxton

11 Department of Health, London

12 NHS England, London

Chapter 3 Genomics and therapeutics

Chapter lead
Munir Pirmohamed

Institute of Translational Medicine, University of Liverpool

Chapter 4 Developing medicines targeting severe genetic diseases

Chapter lead
David Altshuler

Vertex Pharmaceuticals

Chapter 5 Cancer diagnosis

Chapter lead
Ultan McDermott

Wellcome Trust Sanger Institute

Chapter 6 Rare diseases

Chapter lead
Patrick Chinnery

Cambridge University Hospitals NHS Trust, University of Cambridge

Chapter authors

Patrick Chinnery, Cambridge University Hospitals NHS Trust, University of Cambridge.

Alastair Kent, Genetic Alliance UK

John Bradley, Cambridge University Hospitals NHS Trust, University of Cambridge

Chapter 7 Genomics and obesity

Chapter lead
Stephen O'Rahilly¹

Chapter authors

Sadaf Farooqi¹, Vanisha Mistry¹, Stephen O' Rahilly¹

1 University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science

Chapter 8 Personalised Prevention

Chapter lead
Ron Zimmern

PHG Foundation

Chapter 9 Pathogen genomics

Chapter lead
Sharon Peacock

London School of Hygiene and Tropical Medicine

Chapter contributors

Eleanor Barnes, Nuffield Department of Medicine and the Oxford NIHR BRC, University of Oxford. Judith Breuer, University College London. Miles Carroll, Public Health England, Porton Down. Derrick Crook, University of Oxford, and Public Health England. Xavier Didelot, Imperial College London. Mark Fife, The Pirbright Institute, Surrey. Andrew Fox, Public Health England. Jane Greatorex, Public Health England. Ian Goodfellow, University of Cambridge. Paul Kellam, Imperial College London. Donald King, The Pirbright Institute, Surrey. Julian Parkhill, The Sanger Institute, Wellcome Genome Campus, Cambridge. Tim Walker, University of Oxford

Chapter 10 Risk-stratified cancer screening

Chapter lead
Douglas Easton

Department of Public Health and Primary Care, and Department of Oncology, University of Cambridge

Chapter authors

Douglas Easton, Department of Public Health and Primary Care, and Department of Oncology, University of Cambridge. Nora Pashayan, Department of Applied Health Research, University College London. Paul Pharoah, Department of Public Health and Primary Care, and Department of Oncology, University of Cambridge.

Chapter 11 Genomics in newborn screening

Chapter lead
Philip L. Beales

Chapter authors

Heidi C. Howard, Centre for Research Ethics and Bioethics, Uppsala University. Philip L. Beales, Centre for Translational Genomics, University College London.

Chapter 12 Non-invasive prenatal testing

Chapter lead and author
Peter W Soothill

University of Bristol, St Michael's Hospital, Bristol

Chapter 13 Solving Data Challenges

Chapter editor
Claudia Langenberg¹

Chapter authors

Genomic data environment; Peter Counter²
Bioinformatics and statistical analyses; Jessica Vamathevan,³
Ewan Birney³.
Clinical interpretation of genomic data; Clare Turnbull²
Lifelong electronic health records for genomic medicine;
Claudia Langenberg,¹ Harry Hemingway^{4,5,6}

- 1 MRC Epidemiology Unit, University of Cambridge
- 2 Genomics England
- 3 European Bioinformatics Institute (EBI), Wellcome Trust Genome Campus, Cambridge
- 4 Farr Institute of Health Informatics Research, University College London
- 5 Institute of Health Informatics, University College London
- 6 NIHR University College London Hospitals / UCL Biomedical Research Centre, London

Chapter 14 Economics of sequencing

Chapter lead
John Chisholm

Genomics England, Queen Mary University of London

Chapter authors

John Chisholm, Genomics England, Queen Mary University of London.
Sarah Wordsworth, Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford.

With contributions from

Katarzyna Witkowska, Genomics England.
Calum Boyd, Genomics England.
Mark Bale, Genomics England and Department of Health.
Tom Fowler, Genomics England

Chapter 15 Genomic information and insurance

Chapter lead
Mark Bale

Genomics England and Department of Health

Chapter authors

Mark Bale, Genomics England and Department of Health
Sally C Davies, Department of Health

Chapter 16 Ethics and the social contract for genomics in the NHS

Chapter leads (authors contributed equally)

Anneke Lucassen, Faculty of Medicine, University of Southampton.
Jonathan Montgomery, Faculty of Laws, University College London.
Michael Parker, Ethox, Oxford University

Contents

A single PDF download of this report is available via www.gov.uk

All of the sections of this report are also available as discrete downloads. For this reason, every section is numbered separately.

Foreword
Editors and authors
Contents
Chapter 1	Chief Medical Officer’s summary
Chapter 2	100,000 Genomes Project
Chapter 3	Genomics and therapeutics
Chapter 4	Developing medicines targeting severe genetic diseases
Chapter 5	Cancer diagnosis
Chapter 6	Rare diseases
Chapter 7	Genomics and obesity
Chapter 8	Personalised prevention
Chapter 9	Pathogen genomics
Chapter 10	Risk-stratified cancer screening
Chapter 11	Genomics in newborn screening
Chapter 12	Non-invasive prenatal testing
Chapter 13	Solving data challenges.....
Chapter 14	Economics of sequencing.....
Chapter 15	Genomic information and insurance
Chapter 16	Ethics and the social contract for genomics in the NHS
Annex 1	Glossary
Acknowledgements

Chapter 1

Chief Medical Officer's summary

Chapter author
Sally C Davies¹

¹ Chief Medical Officer, Department of Health

About Genomics

A gene is a piece of DNA with a code for a specific instruction – like whether your eyes are blue or brown. A genome is an organism's whole set of DNA. When the human genome was sequenced for the first time, scientists assumed that there would be at least 100,000 genes. In fact, there were around 20,000 – the same a starfish has. They also found that less than 5% of the genome was comprised of genes; in the past the rest was assumed to be junk. Now it is known to be incredibly important, with a vital role in controlling and regulating the way your body works. That's why the whole genome is sequenced.

The study of all the DNA in the genome together with the technologies that allow it to be sequenced, analysed and interpreted is collectively called genomics, or genomic medicine if applied to patients.

Glitches that make you, you

About 99.8% of our DNA is the same as other human beings. But the 0.2% that is different - about 3 to 4 million letters - is what makes each of us unique. Some variation between us is perfectly healthy but some is not and it is these unhealthy differences that the 100,000 Genomes Project is looking for. You can think of them as spelling mistakes or missing paragraphs and pages in your instruction manual.

Comparing the visible you with the invisible you

Information about exactly how an illness is affecting you is called phenotypic information. This includes the symptoms you have or whether your illness is controlled by a particular medicine. The 'phenotype' is an essential part of interpreting your genome. The science of genomics relies on accurate phenotypic information from the NHS and other sources. All your phenotypic information is collectively called the phenome.

This text was kindly provided by Vivienne Parry, writer and broadcaster, and Director of Engagement, Genomics England

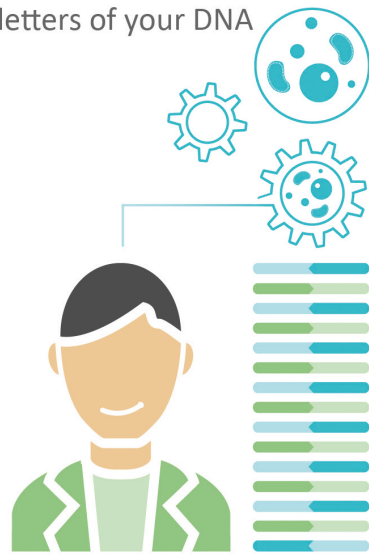
Figure 1 About your genome

Your genome

is found in almost every cell of your body and it is the instructions for making you.

It contains all your **20,000 genes**

It is all **3.2 billion** letters of your DNA



```
ATCGATGAGCTCTAGCGATCGATTGAG
CTCTAGCATCGATTGAGCTCTAGCGATC
GATCTAGCGAGCGAATCGATT
GATCTAGCGATCGATTGAGCTCTAGC
GATCTAGCTCTAGCGATCGATTGAG
GCTCTAGCGAGCGAATCGATTGAGCTCT
AGCGATCGATTGAGCTCTAGCGATCGAT
TGAGCTCTAGCGATCGATTGAGCTCTAG
CGAGCGAATCGATTGAGCTCTAGCGATC
GATTGAGCTCTAGCGATCGATTGAGCTC
TAGCGATCGATTGAGCTCTAGCGAGCG
AATCGATTGAGCTCTAGCGATCGATTGA
GCTCTAGCGATCGATTGAGCTCTAGCGA
TCGATTGAGCTCTAGCGAGCGATCGAAT
```

1 genome = 3.2 billion letters of DNA. If it was printed, your genome would fill a stack of paperback books 61m high or fill 200 telephone directories!

1 sequenced genome equals 2 billion bytes or 200 GB. That's enough to fill the memory of an average laptop.

Source Genomics England

1. Introduction

1.1 Overview

My long-standing interest and research in genomics have led me to focus my fifth annual report (advocacy volume) on genomics. I have been working with patients with genetic conditions all of my professional life, specifically those suffering from sickle cell disease and thalassemia. As a result I have been involved in the science and evolving discussion of genomics for decades and my report reflects this history.

Genomic medicine has the potential to save costs and improve quality of care by targeting treatment, maximising benefit and reducing side effects. For patients with rare diseases, it can shorten their 'diagnostic odyssey' helping to identify therapeutic options faster and improve outcomes. The new science of genomics is opening up better diagnoses for patients, better and safer treatments, opportunities for screening and the possibilities for prevention. These will all improve as we learn more about genomes and their relation with illness and treatment response.

We now talk about 'personalised' or 'stratified' medicine and what we really are trying to deliver is both diagnosis and treatment related to the genomic signature of a particular patient. Genomic medicine will bring particular improvements to the care of patients with cancer or rare diseases. This means giving the most effective drugs in cancer (Chapter 5, 'Cancer diagnosis'), drugs which will cause fewer side effects (Chapter 3, 'Genomics and therapeutics'), seeking new drugs and treatments (Chapter 4, 'Developing medicines targeting severe genetic diseases') and moving to personalised prevention (Chapter 8, 'Personalised prevention'). There will also be other applications, many of which we are not yet aware of.

Like me, my annual reports are independent of Government. They bring together experts in a field of my choosing to set out the latest evidence-based understanding. I am very grateful to the expert authors of the chapters of this report, more than I can mention here. In this summary chapter I have pulled out some of the key themes that I think are important and I have made a series of recommendations for how genomic medicine and services in England can be improved.

My report identifies the opportunities that advances in genomic technology can deliver for clinical practice and public health: the genomic dream. I make recommendations to help deliver this promise of genomic medicine and science for the UK. The chapters of this report cover a broad spectrum of topics spanning from diagnosis, screening, prevention and therapeutics to an analysis of the ethical, legal and social implications of genomics. Case studies are presented of successes and challenges, with examples of where genomic medicine has had an impact - on clinical practice, science and public health.

In the UK we have seen a massive change in our ability to diagnose, treat and support patients with genetic diseases. The UK is an international leader in this field, having already begun to bring genomics into our NHS to make it the best health system possible. Through the establishment of Genomics England and the 100,000 Genomes Project, we are ahead of the game in transforming our NHS by integrating genomics into the health service in a way other countries dream of.

This is being enabled by a technological revolution with data sciences, genomic and computer technology all evolving rapidly. They have all become cheaper, faster and easier and this has contributed to fundamental advances in genomic science and medicine.

So a lot of fantastic work is already happening in some parts of the NHS and some patients are already beginning to see the benefits. For professionals, however, it can be a perpetual battle to make this happen. Much worse, it is very complex and not easy for patients to access and understand. We all deserve the opportunity to access the best care, so now services and research must be made available to all. Only in this way can we realise the genomic dream of faster and better diagnoses and treatments. I very much welcome that the Royal College of General Practitioners is considering the key issues arising from the implementation of genomic medicine for primary care and is already working with their partners to develop the necessary resources to support this.[†]

To make this dream a reality across England and secure the vision of NHS transformation needed, as well as build on the 100,000 Genomes Project, we need to: embed national standards; streamline laboratories; and, in a secure environment, agree to use of data for our own benefit and others’.

1.2 National standards and streamlined laboratory services

We need to review and improve the way that genomic medicine is organised within the NHS. In the past, as with other specialties, genomic services developed as ‘cottage industries’ built on regional expert presence and local interests and funding. Historically, this approach met patient needs and has saved and improved many lives.

But the scale of the modern NHS and the opportunities offered by genomic medicine mean it is now time to build a first-class genomic service that is scalable, future-proof and delivers value for money. The aim must be an equitable service with higher throughput and at a lower cost than is currently achieved. This can only be done through national standards and centralised genomic laboratories and related services. I welcome the changes in commissioning plans laid out by NHS England in their Prior Information Notice on recommissioning NHS Genomic Laboratory Hubs.

This will inevitably mean fewer laboratories doing different types of work. Running fewer sequencing machines at full capacity allows sequencing to be affordable, standardised and accessible for updates. These laboratories will be different from those we had in the past, as the nature of the expertise needed in the NHS is changing. The interpretation and analysis of genomic data now involves high-powered computing, not banks of test tubes.

Under this new model DNA sequence data produced centrally would be distributed via a central database to local NHS Genomic Medicine Centres, where NHS staff, often supported by their colleagues in academia, will be responsible for the interpretation of the DNA sequencing results. The longer this system is in place the better it should become for patients; the consistent reporting will be increasingly supported by knowledge about the DNA sequence, which flows from regularly updating and analysing the central DNA database.

The corollary of this shift is that it is essential for clinicians to work with professions not traditionally considered ‘clinical’. Modern genomic science has evolved into a new concept of the ‘clinical team’ which now includes: diagnostic staff in laboratories and imaging; computer scientists; statisticians; (bio)informaticians; and data scientists who assemble, process and assess the data to advise on diagnosis and treatment. Clinical reports need to be discussed and reviewed by multi-disciplinary teams representing this new diversity of skills and expertise. This must be the expectation of the public. This is the only way patients will get the best outcomes.

This is not a fantasy future. The 100,000 Genomes Project has already shown that the reading of the entire DNA code (Whole Genome Sequencing) to clinically-accredited standards can be delivered at high throughput from a single modern sequencing laboratory in England. We are also seeing a similar process taking place, for the benefit of patients, in other services such as x-rays and imaging results.

[†] Royal College of General Practitioners’ Council Paper, 23rd September 2016

So it is now time to move on from our cottage industry, though recognising it has served us in the past. It is time to build a national, first class service that is scalable, future-proof, and delivers value for money.

1.3 Addressing data protection concerns

In this new age, more patients will agree to use of data about them. New data which is added will mean that we understand more about the genomic factors behind illnesses, and more about what we can do to help patients and improve outcomes. As a patient, I could be getting my diagnosis 'live' from a comparison of all genomic data available.

It is clear that data protection is as important as ever before. Personal data must absolutely always be stored appropriately and securely. People are right to expect this of the NHS as for any other organisation which stores sensitive information. However, this must not prevent data from being used responsibly, as doing so can bring such huge benefits to patients.

Key to advancing genomic medicine will be helping patients to understand that by agreeing to use of data about their illness, they bring direct benefits to themselves. Using live data to make diagnoses and assessments means that the latest science is brought to bear on their illness, helping them to get the best result fast. When we rely solely on data published in traditional journals, we are generally using evidence that is at least 12-24 months old. As the field of genomics is advancing so quickly this time lag can make a big difference. Where patients consent, they may also be able to be contacted if new treatments emerge for their diagnoses.

The more people have contributed their data, the better the results for any one individual. So use of data in this way benefits all patients, both as a group and individually. Doctors see the altruism of patients every day as they join clinical trials or donate blood and organs. We need to help patients to understand they have the option to agree to use of data to help themselves and others.

The field of genomics is complex, and we cannot expect patients to understand it readily. As members of clinical teams we must engage patients and the public and develop real partnerships. We need to continue and strengthen an open dialogue and make sure that the argument for joining in is heard. To achieve this we need to maintain patients' and the public's trust and make genomics everyone's business.

This does not mean dismissing concerns about data protection. Rather it means ensuring that new systems which are built are as secure as possible, and emphasising to patients that the UK is ideally placed to provide these safeguards. We have experience of using sensitive data with consent in other branches of medicine, including communicable diseases and screening programmes. Now we must build on this experience while carefully considering ethical and consent issues.

Key to this is a short, simple, understandable and workable consent process for patients to choose how confidential genomic data about them is used.

While we are doing this, though, we need to reflect on the conversation about towards 'genetic exceptionalism'; the idea that the genome is categorically different from other data types because it contains the unique information that makes us 'us'. Clearly this concept has implications for how genomic data is used, and I believe that our evolving understanding of genomics means we need to move beyond it.

There are two main reasons for this. The first is that there are many other types of commonly-used medical information which are as – if not more – sensitive. For many people, information about their mental health or sexual history could have a more significant impact on their lives if used improperly. Yet, under rightly strict data processes, this information is used on a daily basis because it is clinically justifiable as in the patient's interest. Genomic data should be treated in the same way – safeguarded properly and used in the patient's interest.

The second reason addresses the 'uniqueness' argument. As explored above, while it is true that the variations in our genome determine many things that make us who we are it is important to remember that there is much commonality between all of our genomes. The uniqueness comes from variations that make up a tiny proportion of the whole. This is not just pedantry – because as I set out above it means that the information in our genome could help our family members, and indeed in some cases everyone, get better and faster diagnoses. The missed opportunity cost of 'genetic exceptionalism' is therefore high.

While I understand the concerns that lie behind the argument of 'genetic exceptionalism', provided appropriate safeguards are in place I do not think it stands up to scrutiny. It also could prevent patients from getting access to the best quality care possible.

1.4 Other key areas

In further sections of this Summary Chapter, the individual chapters of the Report are condensed and explained. Before that, however, I wanted to pull out some key points and themes that are cross-cutting, in addition to laboratory services and data which I have discussed above.

The first of these is on **research and international collaboration**. Though our understanding of genomics has advanced rapidly in recent years, it is so far from complete. Researchers in the UK and across the world are working with patients to improve our knowledge, and the success of the UK is to a large extent made possible by the wide sharing of research data across the globe. We are strongest when we are working together.

This is closely aligned to **investment in research**. The UK is recognised for its leadership role in genomics research, catalysed by significant investment and supported by the Medical Research Council, National Institute for Health Research (NIHR), Wellcome Trust, Cancer Research UK, universities, and others. There are clear benefits to investing in 'critical mass centres' as national infrastructure, which can then stimulate further national and international collaboration. We can see this from the Genome Campus at Hinxton, which includes both the Sanger Institute and the European Bioinformatics Institute. Large scale investment will continue to be needed to realise research opportunities. The more these can be integrated into the health system the better; an approach which is developing in France.

Investment in research must be accompanied by **investment in services**, or the benefits will not be properly carried on to the patients. Our wider society will benefit not only from better health, but also by attracting new investment, creating jobs in research and the NHS, building a competitive environment that attracts world-leading researchers and clinicians, creating a genomic literate workforce, and offering cost-effective treatments sooner to those patients who can benefit the most. Early experience in genomics across the world has shown that implementation at scale, alongside decommissioning (stopping) old practices, does drive costs of WGS down. But we have never been good at stopping things in the NHS when outdated. This time we must make the changes.

I also emphasise the importance of **parity of access across England**. Patients expect, and should have, the same access to testing, diagnosis and advice regardless of where they live. The new national laboratory structure I propose above should help achieve this, but in order to deliver it we will also need national standards for defining the symptoms (the phenotype) of patients in a systematic way. National commissioning of genetic testing by NHS England needs to ensure standardised processing of samples and use of up-to-date technologies to secure equitable access to rapid, high quality results.

Key to achieving this parity of access is reforming **professional attitudes**. There is a tendency in some parts of the NHS to think of genomics as a thing far in the future, or even worse, a potential burden rather than a boon. While I understand where this view comes from, in the long run it will prevent patients from accessing the best care. This is true for all disciplines, including **public health**, where genomics can have such a big impact such as for screening and for diagnosis, and in the future, for precision medicine.

So we need a new genomic paradigm to be integrated into all **training curricula** and specialty training of all clinicians, not just doctors. Adopting genomic technologies into routine practice will require changes in the design, operation and workforce of healthcare organisations. The skill sets that are needed are mostly unfamiliar to the current workforce, and experts are in short supply both nationally and internationally.

Finally, I would particularly flag the summary of Chapter 16 of this report, 'Ethics and the social contract for genomics in the NHS'. In many ways this is the most important as it draws together the important requirements that need to be in place if genomic medicine is to be ethically and socially acceptable.

I really believe we can deliver this genomic dream for our patients, our NHS and England. But to do so everyone needs to embrace the mantra of 'patients first' and welcome both the exciting science and the necessity of NHS change.

The following chapter summaries show how we can do this, with my recommendations for action tabled after them. The recommendations are targeted at the key organisations in this field. Working together we can make such a big difference to so many people, that we may hardly recognise medicine in the future. This is an exciting and tremendous opportunity. Following the recommendations, the full chapters themselves delve into these complex issues in more detail, written by experts in each field. I am very grateful to those chapter authors for their time and thoughts – their work is so important.

2. 100,000 Genomes Project

I start by showcasing a programme that I, as Chief Medical Officer, am particularly proud of: the 100,000 Genomes Project. The whole world is watching us and we are now seeing a live realisation of the genomics dream.

Some of the challenges and benefits that are likely to accompany genomic medicine becoming part of routine clinical care can be anticipated by looking at the experience of Genomics England. In December 2012, then Prime Minister Rt Hon David Cameron, launched a challenge to sequence 100,000 whole genomes. To deliver this ambition, Genomics England was created as a company with the Department of Health as the sole shareholder on behalf of the public, funded by the National Institute for Health Research (NIHR). NHS England, Public Health England and Health Education England also contribute to the overall programme, so Genomics England is a member of the NHS family of organisations.

Genomics England has developed a partnership with Illumina, a leading sequencing technology company, and has now sequenced over 31,000 genomes, largely from patients with rare diseases and cancer. In Chapter 2 of this report, '100,000 Genomes Project', the challenges and some of the success stories encountered so far are described. This includes reengineering of the clinical pipeline, development of standards and databases alongside a platform for bioinformatics and genome interpretation. The NHS, led by NHS England, is providing consented patient samples and phenotype information/data. This programme of work, driven by Genomics England, has led to significant transformation and standardisation within the NHS, particularly for rare diseases and now for cancer patients. The 100,000 Genomes Project has been welcomed by patients. Other countries, like France, have started a centralised national genomic plan and associated delivery structure. Working with appropriate international partners, against international competition, a concerted UK effort is required to successfully deliver the realisation of our UK genomics vision, and maintain our leading position in this area, so that NHS patients are the first to reap the benefits of the genomics revolution.

The 100,000 Genomes Project

Patients who take part in the project may be able to get diagnosis.

But for most, taking part means knowing they are helping medical research for future generations.

For some, genome sequencing may mean a specific treatment can be recommended.

Research on genomes will help us understand diseases and what's causing them. It can help researchers develop treatments and new diagnosis.

100,000 genomes

70,000 patients and family members

21 Petabytes of data.

1 Petabytes of music would take 2,000 years to play on an MP3 player.

13 Genomic Medicine Centres, and

85 NHS Trusts within them are involved in recruiting participants

1,500 NHS staff (doctors, nurses, pathologists, laboratory staff, genetic counsellors)

2,500 researchers and trainees from around the world

Who is involved?

It is estimated half of all Britons will get some form of cancer at some point in their lives.



A rare disease is one that affects 1 in 2,000 or less of the UK population. There are up to 8,000 rare diseases – affecting a total of 3 million people in the UK.

8,000 rare diseases affecting 3,000,000 people in the UK



There are over 100 rare diseases included in the Project and 7 common cancers.

7 common cancers 100+ rare diseases



3. Precision medicine

Chapter 3 Genomics and therapeutics

Chapter 4 Developing medicines targeting severe genetic diseases

Chapter 5 Cancer diagnosis

Chapter 6 Rare diseases

Chapter 7 Genomics and obesity

Chapter 8 Personalised prevention

Chapter 9 Pathogen genomics

Chapter 3 Genomics and therapeutics

Chapter 4 Developing medicines targeting severe genetic diseases

In precision medicine, therapies are designed and targeted on the basis of their underlying molecular pathophysiology, directing resources to those patients who benefit the most and protecting patients from side effects. Central to the vision, the genomic revolution delivers for human health, is its potential for drug discovery.

Drug development is a long and expensive process with a high failure rate of drugs taken forward to clinical development. While conventional approaches focus on the small part of the genome coding for proteins thought to be “druggable” at that point, systematic use and integration of human genomic information is now employed to select and validate drug targets, lower the risk of drugs failing late in development, and predict off-target (and potentially harmful) effects in humans. It has been estimated that the clinical success rate for drug development could be significantly increased by selecting targets for development that have supportive genetic evidence.

Chapter 3 of this report, ‘Genomics and therapeutics’, provides recent examples of the role that genomic discovery can play for the identification of new drugs, such as anti-PCSK9 monoclonal antibodies for cholesterol lowering, or anti-sclerostin antibodies for osteopenia. Other important current and potential future applications include the repurposing of existing drugs for new indications, personalising the intensity of drug therapy and dosing, and predicting and avoiding side effects. The discussions about subgroups of patients that are targeted for treatment based on their specific cancer cell mutations, mirrors the successes seen for inherited germline mutations, for example those causing cystic fibrosis, one example of a rare, severe genetic disease discussed in more detail in Chapter 4 of this report, ‘Developing medicines targeting severe genetic diseases’.

Chapter 5 Cancer diagnosis

Recent advances in large-scale next-generation sequencing have informed our understanding of the inherited and acquired genomic causes of cancer. Chapter 5 of this report, ‘Cancer diagnosis’, takes us from the history of cancer being first discovered as a disease of the genome, to the future of cancer research and personalised cancer care.

The pace of developments in diagnostic methods, particularly the changes that are unique to a particular cancer and the ability to detect circulating cancer as cells or free DNA, demands a new approach to NHS diagnostic laboratories. In Chapter 5, we see:

“The molecular diagnostics laboratory of tomorrow will need to be able to implement complex NGS assays and create efficient informatics solutions for their analysis, all under an approved regulatory process. It will also need to keep abreast of the fast-moving field of novel diagnostic tests, which may run the gamut of DNA, mRNA, microRNA, lncRNA, methylDNA, and possibly in the near future proteomics. Arguably, a semi-centralised system of molecular diagnostics laboratories with shared SOPs for standardised NGS platforms would be the most cost effective way to deliver a genomics-based service for healthcare systems.”

A refined genomic cancer “diagnosis” can help to guide cancer treatment and now we know enough to implement these new technologies in routine clinical practice to identify and target those patients who are most likely to benefit. Many cancer drugs benefit only a small number of patients diagnosed with a specific cancer type. Using these treatments only in those who will respond can improve outcomes and reduce side-effects for patients, and thus reduces avoidable costs for the NHS.

We need new models to replace the old “one size fits all” approach for developing, prescribing and evaluating new cancer drugs. The UK has the opportunity to build the right infrastructure, to maintain its global position as a leader in running clinical trials and become the country of choice to deliver tomorrow’s complex cancer trials for the benefit of patients, their families and our society.

We need to ensure that all opportunities are taken up for NHS patients to participate in these new clinical trials. Delivering this will need partnerships between patients, clinicians, NHS commissioners, research funders, particularly NIHR and CRUK, as well as with the life sciences sector.

Chapter 6 Rare diseases

Over 7,000 different rare diseases have been identified. Although individually each is present in less than 1 in 2,000 people, collectively rare diseases are common and affect 7% of the UK population; that is around 3 million people. The majority of rare diseases have a genetic cause and no cure. When symptoms of a rare disease become apparent, often in childhood, their rare and specific syndrome or signature may go unrecognised and this can lead to delayed diagnosis and repeated referrals. This causes uncertainty and stress and expense for patients, parents and families. This so called "diagnostic odyssey" is also expensive for the health care system and prevents timely access to treatments, where they exist, and support services for patients and their carers.

Chapter 6 of this report, 'Rare diseases', reviews the challenges rare diseases can present and offers potential solutions to improve these. Individual clinicians and even large medical centres may have never previously encountered another case, so joined-up care and a national infrastructure to integrate and agree to use of detailed clinical and research data information from different sources and throughout a patient's journey is critical for patients with rare diseases.

Whole genome sequencing (WGS) provides an opportunity to substantially increase the diagnostic yield and speed to diagnosis for rare diseases within the NHS. Success stories are emerging, also illustrated in Chapter 2 of this report ('100,000 Genomes Project'), which describes how embedding WGS within the NHS through the 100,000 Genomes Project, an exemplar of seamless integration of research and clinical care, can accelerate the rate and speed of diagnosis for rare diseases.^{#,§,**} The UK is at the forefront of rare disease research and treatment, supported by government funding and our medical research charities. We are also attracting investment from the life sciences sector. The opportunities for patients to participate in clinical research and global collaboration with clinical, academic and industry partners are great. This collaboration is essential to overcome the existing challenges of WGS and realise our vision for rare disease patients, to deliver a rapid diagnosis and best care for all.

Every European Union country has a rare disease plan. As the UK updates our implementation plans for rare diseases, we should take the opportunity to develop an optimised diagnostic service that benefits patients by empowering clinicians and supporting clinicians to know when and how to initiate genetic testing.

Referral pathways should be based on appropriate multi-disciplinary teams (MDTs) at the appropriate local, regional and national level for very rare and complex syndromes, with genome sequencing commissioned nationally.

With more patients with rare diseases undergoing a WGS DNA test the number of diagnostic reports is set to sharply increase. Interpreting the results of the DNA test in the context of the clinical phenotype requires input from experts. With the large number of rare diseases, expert teams must be able to work across the NHS nationally with an adequate number of multi-disciplinary teams (MDTs) for each of the different rare diseases domains, for example, rare diseases of the heart, the blood and immune system, the eye and for neuro-developmental disorders, etc. These MDTs must be embedded in the NHS Genomics Medicine Centres and their members should be able to see the genome sequencing results of all patients with a particular group of rare diseases.

Carss et al, *AJHG* PMID: 28041643

§ Myer et al, *Nat Genetics* PMID: 27992417

** Sivapalaratnam et al, *Blood*, PMID: 28064200

The DDD study

Pioneering genome-wide sequencing for discovery and diagnosis in rare diseases

Aims

The Deciphering Developmental Disorders (DDD) Study is a nation-wide collaboration between the NHS and the Wellcome Trust Sanger Institute with the aims of applying the latest genomic technologies to

- Discover the genetic causes of developmental disorders, sharing knowledge globally through publication and innovative and responsible data sharing platforms
- Diagnose thousands of patients, addressing unmet clinical needs, transferring knowledge into the NHS and up-skilling the NHS genetics workforce.

Almost every NHS clinical geneticist across the UK recruited patients with rare diseases that were apparent at birth or in early childhood, often affecting several body systems, and for whom a genetic diagnosis had not proved possible using conventional tests. Scientists at the Wellcome Trust Sanger Institute sequenced and analysed all of the genes in these patients and their parents (~33,000 individuals).

The DDD study (www.ddduk.org), has diagnosed several thousand children, and identified >30 new genetic disorders so far, for a combined cost of ~£15 million. The combination of clinical genetics and scientific expertise, and the nation-wide scale of the project resulted in an unprecedented diagnostic yield of ~40% in patients!

Study design

The DDD study was designed to be minimally disruptive to patients and NHS services, working with the grain of current clinical practice.

Data-sharing

The wealth of variants in every human genome (4-5 million) poses a huge challenge in interpreting data from genome-wide sequencing studies to provide safe and accurate diagnoses for patients. The DDD study has set the standard for responsible data sharing to improve diagnosis globally, through the DECIPHER web portal (<https://decipher.sanger.ac.uk>).

Research

In addition to the core research conducted at the Wellcome Trust Sanger Institute, DDD has catalysed and supported additional research projects led by NHS genetics services and their academic partners. More than 200 such projects have been established and ~50 publications delivered to date.

Impact

The DDD study is having world-leading impact in the following domains:

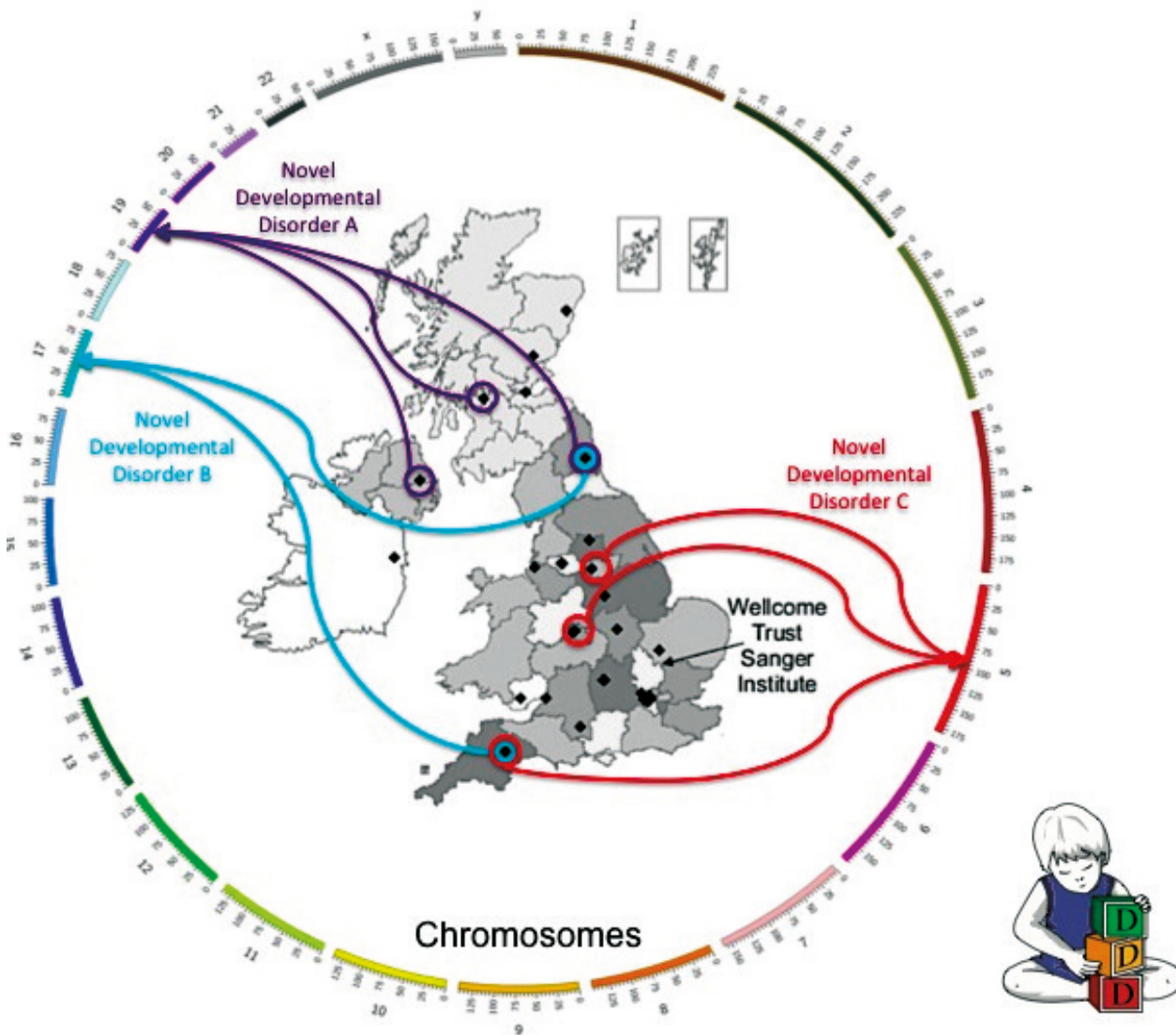
- **Science**
>60 publications, including two flagship papers in the top journal Nature.
- **NHS genetics services** – diagnoses for patients. Stimulating training and research. A strong evidence base for implementing improved services.
- **Translation** – improved diagnostic assay developed with UK SME (OGT), spun-out diagnostic software company (Congenica)
- **Data-sharing** – unequivocal evidence of the benefits of NHS genetics services acting collectively as a distributed network of expertise, and enabling NHS data sharing through DECIPHER.
- **Families** – a molecular diagnosis for their child's developmental disorder in more than a third of cases. Sharing anonymised diagnostic information globally via DECIPHER to catalyse participation in therapeutic trials. Facilitated supportive social networks through SWAN UK (Syndromes Without A Name) (www.undiagnosed.org.uk) and Unique (www.rarechromo.org).

The DDD study is an exemplar of how harnessing NHS clinical expertise, UK scientific know-how and the scale of the NHS can have world-leading impact on scientific knowledge and clinical practice, ultimately benefitting patients and their families.

The DDD study was funded by the Health Innovation Challenge Fund, a joint initiative of Wellcome Trust and Department of Health. The study website is www.ddduk.org

This text was kindly provided by Matt Hurles, Head of Human Genetics and Senior Group Leader at Wellcome Trust Sanger Institute and Helen Firth, Consultant Clinical Geneticist at Addenbrooke's Hospital, Cambridge.

Figure 1 Nationwide discovery of novel rare diseases



Source Deciphering Developmental Disorders (DDD) Study

Chapter 7 Genomics and obesity

The dramatic recent increase in obesity and associated diseases has been driven by environmental and societal changes that promote overeating and sedentary behaviour. There is a large genetic contribution to obesity, body size and shape but this has tended to be overlooked. Some rare genetic variants, for instance, one causing leptin deficiency, can result in extreme obesity. Studies of these patients have shown powerfully the way in which a genetically determined drive to eat can not only cause extreme early onset obesity, but also alter hormonal regulation of the affected brain pathways. Levels of appetite and food intake are reset, with these new higher levels becoming the norm for that person.

The importance of genetic influences on brain pathways that regulate hunger and food intake are not restricted to rare genetic syndromes. Studies of the general population have identified many genetic variants associated with body mass index and obesity. Although each has a small effect, there is a cumulative impact on obesity when they occur together. Chapter 7 of this report, 'Genomics and obesity', outlines our current knowledge and understanding of the genetic causes of obesity, how they operate and the implications for patients, clinicians, society and regulators. In the current context, it is particularly important to recognise that a large proportion of children with severe early onset obesity carry highly penetrant (meaning a very high chance that they will affect the individual) genetic variants that affect energy balance. This affects their clinical management and also the support required for families. Recognising the contribution that genetic drivers of food intake and satiety have on obesity and weight gain highlights the damaging and powerful effects of the current obesogenic environment with its omnipresent stimuli encouraging consumption of high calorie foods. Addressing this environment is an important challenge and logical public health approach to protect all, including those who are genetically predisposed to obesity.

Chapter 8 Personalised prevention

Genomics underpins health and maintenance of health as well as ill health. While the use of genomic technology for the diagnosis and treatment of disease is evident, its role for disease prevention is less obvious at this time. Personalised prevention in the context of genomics implies that genetic information is used to identify individuals at increased risk of disease, and target or tailor preventive strategies. This can apply to both primary and secondary prevention and Chapter 10 of this report, 'Risk-stratified cancer screening', provides examples for the latter by demonstrating how genomics may assist the early detection of disease. For primary prevention, with the ambition to stop the disease from occurring in the first place, this should be seen as a complementary strategy to traditional, universal public health approaches that attempt to shift the distribution of a risk factor in the whole population. The opportunities for genomic insights to inform the prevention of serious ill health should be considered as important as specialised diagnosis and intervention, as highlighted in Chapter 7 of this report, 'Genomics and obesity'. The opportunities for new models of care in an equitable healthcare system – and a new social contract – need to be embraced by policymakers, professionals and the public (see Chapter 15 of this report, 'Genomic information and insurance – background and context').

As Chapter 7 shows, the causes underlying a seemingly homogeneous disease or phenotype, such as childhood obesity, can vary widely and the same preventative intervention is unlikely to be universally effective. Stratifying preventive efforts for common diseases according to disease subtypes with a similar underlying aetiology may therefore be a useful strategy to target interventions. There is not yet sufficient evidence to support this and it is important that the public and patients understand the limitations of existing commercial tests directed at consumers looking for "targeted" lifestyle advice. While it is now possible to use genetic variation to predict people at different levels of risk for a disease in the population, this information often only adds marginally to what we already know based on established risk factors for common, complex diseases such as type 2 diabetes or heart disease. This may change in the future. It also differs according to disease type, specifically depending on whether any known good clinical or other predictors already exist. There are, therefore, potential genomic predictions and personalised approaches to prevention of disease areas of great public health importance that remain relatively understudied in the context of disease prediction but that will be amenable to preventive interventions, such as psychiatric diseases.

Chapter 9 Pathogen genomics

Diagnostic tests for infectious diseases in microbiology laboratories are an essential part of healthcare. Accuracy, cost-efficiency and turnaround times are important criteria for the adoption of new technologies that promise improvements in test performance. The introduction of sequencing technologies into diagnostic and public health microbiology has already improved patient care, for example with viruses, such as the detection of existing and emerging drug resistance for HIV, and selected bacteria. In March 2017, Public Health England announced that WGS is now used for the first time to identify different strains of *Mycobacterium tuberculosis*, the cause of TB, allowing much faster and more accurate diagnoses. Conventional (old) methods previously took up to a month to confirm the diagnosis, so implementation of genetic sequencing now significantly reduces the time until a patient can receive targeted treatment to generally just over a week. Bringing genomics into routine microbiological practice holds the promise of rapidly diagnosing both the cause of infection and its susceptibility to treatment, thus reducing the threat of antimicrobial resistance.

Chapter 9 of this report, 'Pathogen genomics', highlights some of the technical, financial and logistic barriers that limit the wider adoption of sequencing technology across different areas of diagnostic and public health microbiology. Case studies illustrate potential benefits for patients, and national and international outbreak control. Examples include moving from routine capillary to ultra-deep sequencing to increase the sensitivity for early detection of antiviral drug resistance and rapid change of treatment for high risk transplant patients, the use of viral sequence data to trace the spread of foot-and-mouth disease across farms in England, and the investigation of viral changes and transmission across African countries during the 2014-15 Ebola outbreak.

To realise the vision of extending WGS to other pathogens of public health importance, commercial development of fast, accurate WGS solutions at lower cost is required. In parallel, changes are needed to prepare the organisations delivering this, to optimise sample workflow, to develop and implement accredited software for standardised processing, collating, analysing and reporting of sequence information, and to build a workforce with the required skill set.

4. Screening

Chapter 10 Risk-stratified cancer screening

Chapter 11 Genomics in newborn screening

Chapter 12 Non-invasive prenatal testing

National screening programmes for early detection of cancer or its precursors exist for breast, colorectal and cervical cancer and targets people on the basis of age and sex. With genetic discoveries of the last decade, it is now possible to identify people at different levels of cancer risk based on common and rare inherited genetic variants independent of family history.

Cancer screening programmes in England do not yet make use of the existing knowledge of an individual's inherited (germline) genetic susceptibility to cancer, despite cancer being a disease of the genome. Chapter 10 of this report, 'Risk-stratified cancer screening', outlines how cancer screening programmes may be improved by targeting individuals on the basis of their genetic risk in combination with other factors. Genetic risk stratification is already technically feasible in the context of our rapidly increasing knowledge about genetic risk factors for cancer in the population and falling costs for sequencing.

Combining information on genetic variants with other risk factors has the potential to improve cancer prediction and would allow us to better target people for existing cancer screening. This approach also has the potential to make screening feasible for cancers where there is currently no national screening programme, such as prostate cancer.

Newborn screening in England is done using a heel prick blood spot test and presently screens for nine rare serious conditions with the aim of preventing or treating them as early as possible. Introduction of sequencing technology into newborn screening would allow testing for these as well as other genetic conditions simultaneously, and reduce the need for follow-up testing. Chapter 11 of this report, 'Genomics in newborn screening', reviews the evidence for the use of genome sequencing in the context of newborn screening, and considers challenges around its implementation, including the potential storage and use of untargeted sequence information generated at birth at later stages of life. The authors argue that implementation of sequencing needs to be driven by the interests of the child and not technology, with a focus on screening for genetic variants that confer high risk of treatable or preventable diseases.

Prenatal screening has long used genetic tests to inform reproductive choices so pregnant women in England are offered a screening test for Down's syndrome, Edwards' syndrome and Patau's syndrome between 10 and 14 weeks of pregnancy. Around 30% of women choose not to have screening in pregnancy. There are some women who do choose to have screening but receive a false negative result and the first time they know their baby has Down's syndrome is when the baby is born. These are screened women who are given a "green light" and are unaware that they will give birth to a child with Down's syndrome. In addition, over 3% of results are false positive, this is the large number of women who are offered invasive "diagnostic" testing, which carries risk of miscarriage (although low), who do not have a baby with trisomy. Chapter 12 of this report, 'Non-invasive prenatal testing', considers the implications of sequencing technology for prenatal testing and improvements it can offer over existing approaches.

Twenty years ago, methods that can obtain and assess fetal DNA from maternal blood using floating cell free DNA (cfDNA) were developed. This was a fundamental milestone in pregnancy screening and care. This now offers the possibility of genetic testing for different conditions directly using fetal DNA without an invasive procedure to obtain cells or use of proxy measures of risk, such as ultrasound measures of nuchal thickness. For Down's syndrome, the high sensitivity and specificity that can be achieved through sequencing technology present major advantages as a simple blood test can now much more confidently identify women at risk of carrying a child with Down's syndrome, and also significantly reduce the number of women requiring invasive confirmatory testing.

Women and families who can afford it already have the option of this non-invasive pre-natal testing (NIPT) approach because commercial tests exist. Within the NHS, its implementation as 'contingent' screening, as proposed, would mean that the new test is offered 'in addition' and only to women at higher risk. Using the cost of providing definitive information to a woman about her baby as a baseline, the addition of NIPT, and the consequent reduction in invasive tests, is cost neutral. But unlike 'universal' screening, this approach does not affect the number of women with a false negative test i.e. the 15-20% of children with Down's syndrome not currently detected through screening. Cost of NIPT will continue to change so it is important to continue to model the best use of these tests while carefully considering the ethical issues.

The Nuffield Council on Bioethics^{††} has recently published a comprehensive report on the ethical issues of prenatal testing. As well as raising important points about the impact of contingent non-invasive testing as part of the Downs Syndrome screening programme it also touches on the implications of future developments. The potential combination of genomic sequencing methods and non-invasive testing raises the possibility that women without a family history of a severe genetic condition may be offered testing for other genome changes that are associated with congenital abnormalities. I endorse the report's conclusions about the need for careful evaluation of such developments, especially if these are developed or marketed as commercially available testing services.

^{††} *'Non-invasive prenatal testing: ethical issues', Nuffield Council on Bioethics, March 2017.*

5. Sequencing in the NHS

Chapter 13 Solving data challenges

Chapter 14 Economics of sequencing

The primary purpose of genomic data generated for clinical care or public health is to benefit patients, their families and society. The systematic and standardised aggregation and integration of genomic data with other information for research can deliver our genomic dream. This provides enormous potential to better understand, prevent, diagnose and treat disease in the future. This strategy is cost-effective and can, in theory deliver results more rapidly, as it makes additional use of genomic sequence information generated for a clinical indication, and builds on valuable databases and resources that already exist within the NHS and other organisations. Examples from molecular biology have shown how use, integration and reuse of data can enable and accelerate scientific and clinical advances, ranging from recombinant DNA drugs, animal cloning, gene therapy, forensic science to stem cell therapy.

Integration of genomics into routine clinical practice requires systems and a workforce equipped and prepared to handle the scale and complexity of genomic data. Integration of expertise from disciplines not traditionally part of the healthcare sector will be required, from bioinformatics to process engineering.

Part of the reason that the transformative power of genomic technology can now be practically be evaluated in the clinic as part of the 100,000 Genomes Project or other efforts is due to the dramatic fall in the cost of WGS. Chapter 14 of this report, 'Economics of sequencing', considers economic aspects of sequencing in two parts. The first part outlines the wider economic context that has led to worldwide investment in sequencing technology and identifies some of the challenges of its economic evaluation considering the rapid developments in technology. The second part reviews the economic evidence base for the use of sequencing technology in a clinical setting for a range of indications including rare diseases, cancer, pathology, risk assessment and newborn screening.

The end-to-end costs of WGS are predicted to continue to fall, and the cost-benefit assessment for clinical care, the health care system, and society as a whole remains work-in-progress using different assumptions. The economic evaluation of the 100,000 Genomes Project will be instrumental in our thinking of how to best approach the complexities of assessing and modelling the costs and benefits of sequencing. For the country as a whole, economic benefits from initiatives such as the 100,000 Genomes Project include "spillovers" through the stimulation of investment in related industries. We will though, only reap the full advantages of quality, turnaround and cost effectiveness by national commissioning of services and laboratory centralisation.

<http://nuffieldbioethics.org/project/non-invasive-prenatal-testing>

6. Ethical and societal considerations

Chapter 15 Genomic information and insurance

Chapter 16 Ethics and the social contract for genomics in the NHS

The final chapter in my report, 'Ethics and the social contract for genomics in the NHS', is in many ways the most important as it draws together the important requirements that need to be in place if genomic medicine is to be ethically and socially acceptable. As we have seen in other chapters, the knowledge from genomic technologies needs patients and families to agree to use of data within a clinical and confidential framework and emphasises the importance of sharing data to benefit the patient's own care, their family and that of others. These benefits mean greater integration between healthcare and research, which also leads to a faster pace of integrating knowledge into care. If implemented properly in an integrated health and care system this raises new obligations and responsibilities for patients, clinicians, their hospitals and indeed the institution of the NHS itself. The importance of ethical reflection and patient engagement in the development of the coordinated national and international developments in genomic medicine should be recognised.

There are reasons to rethink – or at the very least reinforce – elements of the current social contract as set out in the NHS Constitution to take account of the advances in genomic medicine. One clear example is the need to revisit the narrow model of patient confidentiality in which a patient confides intimate personal information with a health professional. Whilst this is still a major element of genomic medicine we also need to explicitly acknowledge the importance of linking data on a national, and even international, scale in order to give patients with rare diseases a diagnosis or to identify rare genetic changes in common disorders. This means that the NHS – supported by the public, Government and regulators – will need to develop an arrangement for handling genomic data that is acceptable to patients.

The main requirements for this new social contract include clear, specific but routine clinical consent for data use. This is consent for research based care which is best care, not a separate academic endeavour. Patients have a right to expect the NHS to hold the data securely and to place standards in place to protect them from unauthorised disclosures. But the emphasis on confidentiality must be balanced against the interests of other family members and broader society, especially where genomic information may prevent serious disease.

There also needs to be a recognition that the duty of care that a clinician has must be recognised to now extend to include duties on the researcher, the bioinformatician and the data manager i.e. the whole broad team contributing to diagnosis, advice and related research. This is how modern science based services work. With the increase in genomic

medicine and the interdependencies between clinical care and research it will become the norm that research will produce information that has clinical significance. This responsibility will also be placed on the wider health system where new information becomes available that is relevant for a patient or family member. The ability of the NHS to earn the trust of patients and society on these issues will be an important part of the social contract of our healthy future.

Alongside this social contract there will need to be protections against unfair discrimination based on genomic findings. The commissioning and provision of healthcare, education and other welfare services is covered by general legislation such as the Equality Act 2010 and the Data Protection Act. This is also supported by guidance, professional codes of practice, legal judgements.

One of the most common areas of concern is around the impact of genetic information on insurance. This emerged as a key concern in the late 1990s and since then has been governed by an agreement between the Government and Association of British Insurers. The Concordat and Moratorium on genetics and insurance has been periodically reviewed and has been influential in international agreements and legislation. The current arrangements are summarised in Chapter 15 of this report, 'Genomic information and insurance', but the agreement is under review. Having considered the issues carefully, I support the long-standing Government policy to maintain a flexible semi-voluntary regulatory structure for this area as it is such a fast moving technology. The Concordat and Moratorium were developed to prevent individuals from being deterred from obtaining predictive genetic tests due to the fear of potential insurance consequences. Trying to legislate however, on this basis would raise questions about the use of other non-genetic information that is predictive of ill health: such as simply asking people if they have a family history of a particular disease. Unfortunately, I found the Concordat and Moratorium is complex and not widely understood even though it works. It is based on a series of positions arrived at through the application of general legislation and rules relating to the financial services sector. It also adopts a series of pragmatic measures based on the insurance industry's underwriting principles for different types of insurance.

One of the key benefits of the Moratorium that is not widely appreciated is that insurers have different requirements for underwriting insurance contracts based on the size of the sum insured. The most welcome part of the current moratorium is that no-one needs to disclose a genetic test result if the policy is worth less than £500,000 for life assurance or £300,000 for critical illness cover or £30,000 per year for income protection policies. This means that at current estimates more than 95% of insurance customers would not need to disclose genetic test results.

7. Recommendations

Systems and services

	Recommendation	Organisation
1	<p>I recommend that DH establishes a new National Genomics Board, chaired by a minister. This would facilitate collaboration and ensure effective delivery of appropriate actions, with key priorities including:</p> <ul style="list-style-type: none"> • Patient and Public Interest; • Genomic Research Coordination; • Industrial Development; • NHSE and Genomics England Partnership; • Regulation Development. 	DH
2	<p>I recommend that NHSE should, as planned, recommission all genomics services nationally to ensure a national network which enables equitable service provision across the country. This should work to national standards and processes with national centralised laboratories and regional hubs underpinned by a secure central data platform. It should enable easy access to genomic by researchers with appropriate consents. And it should reinforce the existing strong relationships between Genomics England, NHS, and industry.</p>	NHSE
3	<p>I recommend that NHSE embeds implementation research (including cost effectiveness) at all stages of service redevelopment and laboratory reconfiguration, supported by NIHR.</p>	NHSE
4	<p>I recommend that the following issues should be considered by both NHSE in its new commissioning plans for genomics and DH in the national implementation plans under the UK Rare Disease Strategy:</p> <ul style="list-style-type: none"> • A Human Phenotype Ontology (HPO) as phenotype standard should be used across the whole country, for all patients; • Clinicians should have access to regularly updated information to guide when and how to initiate genetic testing; • A national network of Multidisciplinary Teams (MDTs) should be established to review and advise on complex and ultra-rare syndromes. Regional MDTs will need to do the more common 'rare disease' reviews and local MDTs the common 'rare disease' reviews. These should be embedded in the NHS Genomic Medicine Centres; • A standing genetics committee should be established which involves clinicians and researchers to advise annually on the type and range of tests ('purchase list') and indications for their use. This should be supported by Genomics England's experience and NHSE in order to reduce the time to diagnosis for patients with rare disease; • All patients with severe childhood onset of obesity should have access to rapid genetic assessment, early diagnosis, and appropriate management. 	DH NHSE
5	<p>I recommend that NHSE and PHE explore the feasibility of integrating laboratory services for screening tests using sequencing technology into any centralised genomic service. They should review the advantages, disadvantages and cost-effectiveness of having a separate process for such screening tests.</p>	NHSE PHE

	Recommendation	Organisation
6	<p>I recommend that the National Screening Committee conducts a systematic evaluation of the opportunities offered by genomics for present and potential screening practices. These may be national, population-based programmes as well as cascade or individual. This should include evaluation of:</p> <ul style="list-style-type: none"> • cost-effectiveness; • feasibility; • acceptability; • impact on uptake. 	NSC
7	<p>I recommend that NHSE, working with the National Screening Committee, ensures that the implementation of contingent NIPT is accompanied by an evaluation of the cost-effectiveness of universal NIPT testing for Down's syndrome and other indications. This should consider different assumptions of sequencing costs, including research on its uptake and cost savings from fetal anomaly screening elements that a NIPT programme would replace.</p>	NHSE NSC

Research

	Recommendation	Organisation
8	I recommend that research funders should require health research applicants to justify any research application that does not include genomic analysis.	Research funders
9	I recommend that DH, building on the learning from the 100,000 Genomes Project, convenes a group to agree a national, simple, two-stage routine consent model, acceptable to patients, that allows re-contact for invitation to enrol in research studies and clinical trials. This group should include Genomics England, NHSE, Health Research Authority, academia, and civil society.	DH
10	I recommend that Genomics England, NHSE, and the Human Tissue Authority explore the feasibility of offering the opportunity to be enrolled in the 100,000 Genomes project to existing NHS patients, in relevant clinical trials, with stored tissue samples.	Genomics England NHSE HTA
11	I recommend that DH should ensure any future Government Life Sciences Strategy provides funding for the digital infrastructure necessary to make the UK a great place to carry out clinical trials that embed genomics, maximising the potential for learning through reanalysis and appropriate pooling of genomic information.	DH
12	I recommend that CQC should have as one of its characteristics of a well-led organisation an assessment of support for opportunities for patients to join cutting-edge research projects and clinical trials.	CQC
13	I recommend that PHE ensures researchers have easy and quick access to national pathogen, registry and screening data.	PHE
14	I recommend that when UK companies develop near-patient whole genome sequencing for serious pathogens using new sequencing technologies, then Innovate UK should ensure early clinical evaluation.	Innovate UK

Data, Standards, Regulation

	Recommendation	Organisation
15	I recommend that DH works with relevant government departments to ensure that when implementing the 'EU General Data Protection Regulation' any new UK data protection legislation does not place unnecessary restrictions on the processing of genetic data for patient care and research.	DH
16	I recommend that DH works with international partners and health authorities to support systems, mechanisms and rules for rapid sharing of research and health intelligence data especially where it facilitates control of epidemic and pandemic disease threats.	DH
17	I recommend that DH ensures that there are new coordinated approaches to standard-setting and regulation to meet developments in sequencing, bioinformatics and clinical reporting. This should be implemented as a priority alongside the Accelerated Access Review.	DH
18	I recommend that the MHRA should work closely with Genomics England and NHSE to ensure that the EU 'In Vitro Diagnostic Devices Regulation' is applied appropriately to future genomic medicine services.	MHRA

Engaging staff and patients

	Recommendation	Organisation
19	I recommend that Genomics England and NHSE should engage in an extensive public dialogue on the shared social contract between patient, public, clinicians and academics in relation to genomic medicine. This needs to be a collaborative exercise and build on the experiences of the 100,000 Genomes Project and NHS expertise in clinical genetics.	Genomics England NHSE
20	I recommend that regulators of healthcare professionals ensure that undergraduate and post-graduate training equips doctors and other clinicians for the present and future genomic eras.	Regulators of healthcare professionals
21	I recommend that relevant Royal Colleges ensure such training is complemented by a continued emphasis on Continuing Professional Development in all specialities and for revalidation in the future, in particular for those clinicians who did not receive undergraduate training on genomic medicine.	Royal Colleges
22	Health Education England should continue the work of the Genomics Education Programme, developed as part of the 100,000 Genomes Project, and ensure that this continues to provide staff with relevant data science expertise for the NHS.	HEE

Miscellaneous

	Recommendation	Organisation
23	I recommend that DH ensures the current review of the existing insurance Concordat and Moratorium carefully considers the implications of genomic medicine for insurance and the high levels of public awareness and concern. The review should take note of the need to support a new approach to equitable and integrated care that combines elements of clinical practice and research.	DH
24	Current regulatory pathways for medicines need to be revisited. This is necessary in light of new evidence of the benefits of repurposing existing drugs based on genomic insights. To secure affordability for our NHS, we also need to ensure that the Orphan Drug legislation cannot be misused in order to gain a monopoly on new indications for existing drugs.	DH

Key

CQC	Care Quality Commission
DH	Department of Health
HEE	Health Education England
HTA	Human Tissue Authority
MHRA	Medicines and Healthcare products Regulatory Agency
NHSE	NHS England
NSC	National Screening Committee
PHE	Public Health England

Chapter 2

100,000 Genomes Project

Chapter lead

Prof Mark Caulfield^{1,2}

Chapter authors

Mark Caulfield^{1,2}, Ellen Thomas^{1,3}, Clare Turnbull^{1,2}, Richard Scott^{1,4}, Augusto Rendon^{1,5}, Louise Jones^{1,2}, Kay Lawson^{1,6}, Nirupa Murugaesu^{1,7}, Clare Craig¹, Matina Prapa¹, Simon Thompson^{1,2}, Katarzyna Witkowska^{1,2}, Laura Shallcross⁸, Emma Baple^{1,9}, Caroline Wright^{1,10}, Katherine Smith^{1,2}, Matthew Parker¹, Ellen McDonagh^{1,2}, Eik Haroldsdottir¹, Antonio Rueda^{1,2}, Damian Smedley^{1,2}, Laura Riley^{1,2}, Suzanne Wood^{1,2}, Tim Rogers^{1,2}, Mark Bale^{1,11}, Lisa Dinh¹, Katrina Nevin-Ridley¹, Sue Hill^{1,12}, Tom Fowler^{1,2}

- 1 Genomics England, London
- 2 William Harvey Research Institute and the Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London
- 3 Guy's Hospital, London
- 4 Great Ormond Street Hospital, London
- 5 University of Cambridge, Cambridge
- 6 University College London Hospital, London
- 7 St George's Hospital, London
- 8 UCL Institute of Health Informatics and Genomics England
- 9 Southampton General Hospital, Southampton
- 10 The Wellcome Trust Sanger Institute, Hinxton
- 11 Department of Health, London
- 12 NHS England, London

1. Summary

The aims and goals of the 100,000 Genomes Project delivered by Genomics England are:

■ **Enhanced patient benefit in rare disease, cancer and infection**

The 100,000 Genomes Project is using the latest approaches to WGS (whole genome sequencing) to provide genomic diagnoses for those with residual unmet diagnostic need in rare inherited diseases. In malignant disease we aim to understand the genomic architecture and therefore the potential for stratified healthcare in rare and common cancers. In infectious disease we hope to gain a better understanding of the genomic features within pathogens which lead to antimicrobial resistance and within patients that govern host response to severe infection. In time, we hope these findings will lead to development of new or more effective diagnostics and treatments for NHS patients facing these diseases.

■ **Accelerating genomic medicine into healthcare**

Working with National Health Service England (NHSE) and other partners we have created 13 NHS Genomic Medicine Centres (GMCs) of excellence; providing the capability and capacity to enable these services to be of the highest calibre and widely available for clinical care of NHS patients. Alongside this, Health Education England have placed 700-person years of education to create the essential NHS capacity to deliver world-leading genomic medicine. To drive up the clinical value for participants we have created a coalition of intellects involving more than 2,600 UK and international clinicians and scientists to form the Genomics England Clinical Interpretation Partnership (GeCIP). This creates the research engine tightly connected to the NHS that will continually improve the accuracy, reliability and value of this information for patient care, and add to the knowledge base of the genetic basis of disease.

■ **New life course scientific insights and discovery**

We are creating a knowledgebase of 100,000 whole genome sequences, with the consent of patients linked to continually updated longitudinal life-course health records of participants from primary care, hospitals, national registries and outcomes. We aim to concentrate all genomic data from the NHS within this data centre for analysis at scale by healthcare professionals and researchers to enhance the potential for patient benefit. By providing access to this fully anonymised, unique data resource to industry for the purpose of developing new knowledge, methods of analysis, medicines, diagnostics and devices we can accelerate patient benefit. Initially Genomics England formed a 13-company precompetitive consortium, GENE that is helping to optimise the programme for industry collaboration. Our aspiration is to draw in new opportunities for patients and inward investment to the UK for improved diagnostics, enabling platforms and innovative clinical trials of the very latest therapies (these aspects are considered in Chapter 14, 'Economics of Sequencing').

■ **Increasing public knowledge and support for genomic medicine**

By delivering an ethical and transparent programme which has public and patient trust and confidence in genomics and its value for patient care, we can embed this transformation within our NHS and make this a routine part of 21st Century healthcare (these aspects are considered in Chapter 16, 'Ethics and the social contract for genomics in the NHS').

2. Background

In December 2012 Prime Minister David Cameron announced a programme of whole genome sequencing (WGS) as part of the UK Government's Strategy for life sciences. In July 2013 the Department of Health established Genomics England as a wholly owned (by Department of Health), limited company focused upon delivery of the 100,000 Genomes Project; combining standardised, longitudinal life-course clinical information and whole genome sequencing data in rare disease, cancer, and infection^{1,2}. These disease areas were chosen because there was evidence of tractable health benefits such as new genomic diagnoses, better targeting of therapy and understanding transmission and evolution of pathogens from the application of genomic medicine to healthcare.

2.1 Building the platform for NHS transformation in genomic medicine

The 100,000 Genomes Project is designed to produce new capacity, via 13 NHS Genomic Medicine Centres, and capability through 700-person years of education and training that will transform the application of genomic medicine in the National Health Service in England. This will also produce new capability for UK and international clinical genomics research via a secure Genomics England data infrastructure for the protection and analysis of clinical and genomic data. Diagnostic access to the data for the NHS will include identifiable data for clinical reporting. Research access will be restricted to the de-identified dataset. An Access Review Committee, which includes representation from the public and participants in the project, determines who is given access to the data and will be responsible for review of research outputs and their clinical impact. Challenges in creating this environment have included the scale of the dataset, ensuring standardised data formats across the NHS and partners in academia and industry, and managing the complexity of the authentication system, such that the required security solutions do not impede research collaboration and generation of research outputs which will ultimately benefit patients.

In addition, Genomics England has built key international partnerships in Australia and British Columbia to expand the prospects of increasing diagnoses and developing new approaches to stratified healthcare. To share best practice we have also joined the Global Alliance for Genomics and Health.

Scientific and medical progress requires the long term commitment and collaboration of researchers across the globe. We welcome other population-scale genome sequencing initiatives such as those in the US, Japan, France, Canada, the Netherlands, Estonia and Denmark, as well as industrial initiatives such as the GlaxoSmithKline Regeneron partnership to exome sequence UK Biobank, and the AstraZeneca 2 million whole genome sequencing initiative. All of these projects will enhance our ability to interpret genomic data, thereby enabling the benefits of genomics to reach patients worldwide.

2.2 Achieving the highest quality, value and transformation for the NHS

Whole genome sequencing is similar to reading a very complex book. To identify rare changes in genetic code, termed variants, within the 3.3 billion DNA 'letters' that make up our genomes can be very challenging, but just like reading a very complex book if we read and reread the genetic code up to 30 or more times (technically referred to as 'sequencing depth' of 30-fold or 30X) we can improve our understanding and confidently identify rare variants.

To ensure we achieve the best quality and value for healthcare, Genomics England undertook an open competitive test of the sequencing market. As a result we formed a partnership with Illumina to oversee and manage the whole genome sequencing for the 100,000 Genomes Project. This means the NHS gets the benefit of the latest technologies at the lowest price per genome, ensuring the project remains at the leading edge whilst providing best value. With the assistance of the Wellcome Trust and the Wellcome Trust Sanger Institute, Genomics England has based our genome sequencing centre at the Wellcome Genome Campus in Hinxton.

2.3 Stimulating a vibrant UK genomics industry

Genomics England tested 29 suppliers of genome analysis and annotation. With Innovate UK, Genomics England also stimulated this market with investment of £10 million from the Small Business Research Initiative (SBRI) in the most promising companies to mature their genomic analytics. A group of 10 companies passed this 'bake off' stage. These companies were then invited to take part in a tender to provide interpretation services in relation to the first 8,000 participants in the 100,000 Genomes Project. Genomics England then shortlisted five suppliers, three of whom now have contracts to assist Genomics England with interpretation and display of genome data: WuXI NextCODE, Fabric Genomics and Congenica.

Genomics England also created a precompetitive consortium of 13 companies, GENE ranging from small and medium enterprise to large pharmaceutical companies. These companies will help to shape the project to ensure Genomics England stimulates a vibrant UK genomics industry and bring the most advanced opportunities to patients in the UK.¹

The ambition of the 100,000 Genomes Project is to make the UK the world leader in the application of genomic medicine to healthcare.

3. Enhanced opportunities for patient benefit in rare disease

3.1 The impact of rare disease

It is estimated that there are between 6,000 and 8,000 rare diseases worldwide. Each disease or syndrome may affect less than 0.1 percent of the UK's population, but because there are so many disorders they cumulatively affect the lives of 3.5 million people in the UK^{1,3,4,5,6}. Approximately 75 percent of rare diseases affect children, often leading to disability, with 30 percent of rare disease patients dying before their 5th birthday^{1,3,4,5,6}. The majority of rare diseases have a genetic cause; many of these are due to a single gene defect, although there may be modifiers elsewhere in the genome that have implications for treatment and outcomes. Recent research such as the Deciphering Developmental Disorders (DDD) study has revealed the potential of exome sequencing to increase diagnoses for patients. Now with the 100,000 Genomes Project we have the opportunity to extend diagnostic yield^{7,8,9,10}.

3.2 The strategy for the rare disease programme in the 100,000 Genomes Project

If WGS has been used, there is evidence that diagnoses may be further augmented by 25 to 50 percent across a range of rare disease phenotypes (depending on family structure)^{11,12}. The ideal family structure to identify causative gene variants for rare disease varies according to the genetic architecture of the disorder. For highly penetrant paediatric disorders, the optimal structure is a parent-offspring trio based on two parents and one affected offspring (proband). This is important because in many cases, the causative variant will have arisen for the first time in the proband, and this can be confirmed by comparing the parents' genomes with the child's genome.

In autosomal dominant diseases that pass down through many generations of the same family, the optimal family structure includes more distantly related individuals with the same disorder. Guidelines have been produced for the 100,000 Genomes Project to describe these optimal structures for different family patterns and disorders.

There are more than 200 disorder categories, covering over >2400 conditions included in the rare disease arm of the 100,000 Genomes Project. A large number of these have been nominated by the NHS and scientists; each condition is reviewed by peer experts, then a decision is made by the Science Advisory Committee. Conditions are accepted if there is good evidence of a single-gene (monogenic) basis for the selected group of cases, as defined by a set of clinical eligibility criteria. Disorders of all types can be recruited by every NHS Genomic Medicine Centre, to promote equity of access to the programme.

The comprehensiveness of clinical characterisation (or phenotyping) has a major impact on the likelihood of successfully identifying a disease-causing variant. This has led Genomics England to create detailed standardised clinical data models for each disease using the Human Phenotype Ontology (HPO)¹³. This detailed characterisation of patient phenotypes collected using an internationally accepted ontology, as well as clinical and genetic test results, is expected to improve our prospects of returning a diagnosis.

For very rare conditions there may be very few families with a disorder in the world. In this setting it may be necessary to use interfaces such as Decipher to combine limited clinical characteristics and variants with other similar families across the world^{1,3,4,5,6,8}. Genomics England's findings may enable genomically-driven reclassification of rare diseases leading to opportunities to recall patients for deeper phenotyping through parallel initiatives such as the NIHR Rare Diseases Translational Research Collaboration. This approach may pave the way for functional characterisation of findings - thereby adding further value to datasets, improving diagnostic utility and possibly identifying new targets and therapies. Therefore we envisage that gene discovery in the 100,000 Genomes Project will create significant opportunities for scientific innovation through the focus on residual unmet need, and our emphasis on national and international collaborations.¹⁻¹³

3.3 Whole genome sequencing strategy and quality assurance for rare diseases

In the 100,000 Genomes Project Genomics England is currently sequencing germline whole genomes at a read depth of 30-fold, covering 97.3% of the genome at a read depth of 15-fold or greater. To interpret the genome we then align the reads to a reference genome and generate a portfolio of variants which can be annotated with additional information such as frequency in control cohorts. Variants in genes known to cause disease are assigned to tiers by a semi-automated bioinformatics pipeline, according to the likelihood of each variant being the cause of the family's disease.

3.4 Early results in rare disease

Results of the tiering pipeline have been returned to the NHS for over 2,000 families in the rare disease programme. The diagnostic yield to date is at least 20%; this is likely to increase as NHS GMCs and GeCIP focus in more depth on each family's data over the coming months and years. Clinicians and scientists within the NHS GMC network review the variants highlighted by tiering, carry out additional analyses, and decide which genomic variant(s) are highly likely to be causing the family's rare disease. They validate these variants using a different type of genetic test in their laboratory and report them back to clinicians and participants.

4. Enhanced opportunities for patient benefit in cancer

4.1 The impact of cancer

In 2015 there were 299,923 new cases of cancer with 139,000 deaths in the UK. Cancer is fundamentally a disease of disordered genomes where mutations (including copy number aberrations, insertion/deletion variants, complex rearrangements, and non-synonymous substitutions) lead to uncontrolled cellular proliferation^{14,15,16,17}. The clinical impact of sequencing technologies has enabled more precise definitions of disease, uncovered mechanistic insights into pathogenesis, and identified therapeutic targets based on genetic variation or aberration. Sequencing approaches have also catalogued the complex evolutionary changes that occur in an individual's cancer during treatment and over time. This has demonstrated that there are both expanded clonal populations and low frequency sub-clones, each with specific genomic architecture¹⁶⁻¹⁹.

Large-scale sequencing studies and meta-analysis across cancer types have confirmed the importance of >450 key genes in driving cancer¹⁶⁻¹⁹. 'Molecular Oncology' has emerged as a result, with clinical application of these genomic biomarkers used to predict tumour behaviour, prognosis and drug response, along with increasing administration of bespoke targeted drugs that subvert and/or switch off the oncogenes activated by particular 'driver mutations'. While molecular profiling in the clinic is typically done using single gene tests and gene panels, many emerging prognostic and therapeutic biomarkers across tumour types include "pan-genomic" "signatures" made up of small mutations, copy number changes and hypermutability - only tractable by analysis across the genome²¹⁻²⁵. However, we are in our infancy of molecular pathology and the application of genomics to cancer care. Current taxonomies of cancers are still largely defined by the organ of origin and histological description of abnormal cells and the majority of patients are treated with empiric regimens of cytotoxic chemotherapy and irradiation.

4.2 The strategy for the cancer programme in the 100,000 Genomes Project

In the 100,000 Genomes Project Genomics England is using WGS to identify novel driver mutations for cancer and understand its evolutionary genetic architecture through primary and secondary malignant disease (by multiple biopsy and WGS). To do this it is necessary to sequence the genome of the tumour, and also to sequence the patient's germline (inherited) genome. Only by comparing the two genomes can we robustly identify the somatic variants (those mutations acquired by the cancer). By partnering stratified healthcare programmes and outcome studies in patients from the NHS in England, we aim to enable understanding of WGS benefits in defining predictors of therapeutic response to cancer therapies. In cancer, the genome is only one part of the picture and it is widely accepted that multi-omic approaches including transcriptomics, proteomics, epigenetics and cell-free DNA (cfDNA) will offer additional biological insights into cancer^{18,19,20}. Genomics England is taking the significant opportunity of this programme to evaluate the value of cell-free tumour DNA in plasma to detect temporal changes in cancer and are developing non-invasive 'liquid biopsy' based on WGS²⁶. This may provide a valuable non-invasive test for disease monitoring, enable innovation in trial design and encourage industrial investment in UK clinical research. To fully realise the benefits of the 100,000 Genomes Project Genomics England has learnt from and will integrate data from the Cancer Genome Atlas (TCGA) project and the broader International Cancer Genome Consortium (ICGC), producing an inventory of genomic, transcriptomic and epigenomic changes in a wide range of different tumour types and be able to link this to clinical impact for patient care^{18, 19, 20}.

4.3 Whole genome sequencing strategy and quality assurance for cancer

Genomics England is sequencing to at least 75-fold coverage, recognising that tumour tissue may have lower cellularity and therefore greater depth of coverage is required. A major component in cancer annotation is analysing the consequence of larger scale genomic changes, such as structural variants, copy number aberrations, loss of heterozygosity and other chromosomal mutational events, evaluating the best emerging tools with which to best call these more complex variant types.

4.4 Clinical characterisation of patients with cancer in the 100,000 Genomes Project

For each cancer the Genomics England team defined a specific set of phenotypic characteristics with the NHS, clinicians and researchers with relevant expertise. This core phenotypic data set aligns with the key reporting registries for cancer such as: the Cancer Outcomes and Services Dataset, the Systemic Anti-Cancer Therapy Dataset, and the Radiotherapy Dataset²⁷. It is also being aligned with the clinical audit datasets collected for specific cancers: colorectal, lung and prostate. Data is recorded using controlled clinical terminologies, and structured ontologies based wherever possible on SNOMED-CT and other internationally accepted classifications^{27, 28}.

4.5 Molecular pathology and whole genome sequencing

Genomics England is using this programme to drive transformation of molecular pathology in the NHS. Genomics England undertook exhaustive experimental work to measure the impact of the multiple elements of the tumour handling pathway on sequence quality, from cold ischemic time through fixation conditions to DNA extraction. WGS of DNA derived from standard NHS formalin-fixed paraffin-embedded tumour (FFPE) revealed substantially impaired yield of usable DNA, highly variable quality of sequence and patterns of artefact in the WGS. Accordingly, we have emphasised the need to collect fresh or fresh-frozen tumour samples for our main cancer programme. As well as collecting tissue from surgical resection, Genomics England is also sequencing DNA from cancer biopsies, to enable inclusion of tumours that are not usually resected, such as advanced metastatic cancers, where the clinical benefits of the data may be highest. Furthermore, through using biopsies, Genomics England is able to capture samples from patients ahead of receiving chemotherapy to shrink the tumour prior to resection (neoadjuvant chemotherapy), which would otherwise cause changes in the tumour genome and complicate analyses. Attempting whole genome sequencing on biopsy samples also enables those patients who could have a resection sample an extra opportunity to achieve a successful sample. Genomics England continues to test alternative fixation methodologies and ways of preserving adequate DNA while retaining diagnostic material.

Collection of fresh tissue has presented several substantial challenges and has necessitated radical alterations in tumour handling pathways in theatre, outpatient settings (including endoscopy and radiology) as well as in pathology laboratories. Critical to enacting these pathway changes has been achieving a consensus statement between the Health Research Authority (HRA), Human Tissue authority (HTA), Royal College of Physicians and NHSE that acquisition of fresh tissue is standard of care for diagnostics in cancer. Accordingly, no special consent is required; however, while also preserving the tissue in a manner suitable for DNA sequencing, the tumour sample must therefore be available and suitable for histological examinations required to make any diagnosis. Parallel to recruitment to the cancer programme, we have conducted a programme of experimental work, evaluating alternative approaches to tissue handling, including vacuum-packing and storage at 4° C to extend cold ischemic time, alcohol-based fixative as alternatives to formalin and alternatives to liquid nitrogen to facilitate acquisition of biopsy samples.

5. Enhanced opportunities for patient benefit in infection

5.1 Impact on infection

Infectious diseases are responsible for 7 percent of deaths in the UK per annum and 8 percent of all hospital bed days²⁹. It has been estimated that they cost the UK economy approximately £30 billion per annum^{29,30,31}. WGS of pathogen genomes (both viruses and bacteria) is being adopted for routine management of infectious diseases, providing information on species taxonomy, virulence, transmission and anti-microbial resistance. In addition there are tremendous opportunities for clinical research in species determination, genotype-to-phenotype prediction, and infection control. In each case, scientific progress should lead directly to significant public health and economic benefits, as well as cost savings to the NHS^{30,31,32,33}.

In the 100,000 Genomes Project Genomics England has sequenced 3,000 multidrug resistant tuberculosis (TB) organisms with Public Health England, and other pathogens are under consideration^{30,31,32,33}. In March 2017 the Secretary of State announced a national NHS TB diagnostic sequencing service across England. WGS in severe host response to infection can be nominated for inclusion and recruited within the rare disease programme (see section 3) and a cohort of severe responders has already been approved for recruitment as a pilot of this.

6. Accelerating genomic medicine into healthcare

6.1 A national network of NHS Genomic Medicine Centres

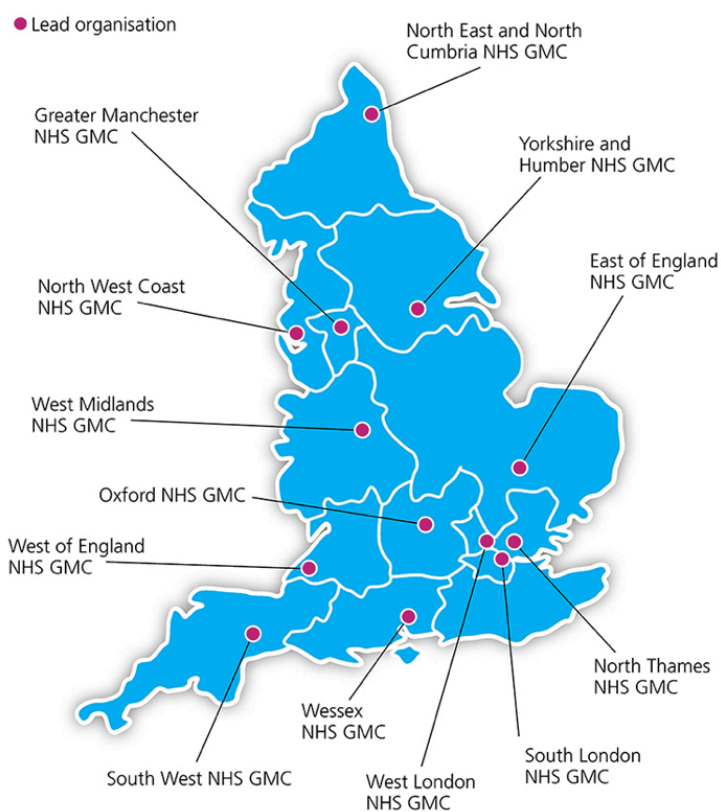
A fundamental part of the legacy of the 100,000 Genomes Project are the 13 NHS Genomic Medicine Centres (GMCs) of excellence that harness the capability and capacity of the NHS and provide geographic equity of access and coverage across England (Figure 3). These are commissioned by NHSE to identify, consent and provide clinical data and appropriate samples from participants. The NHS GMCs form part of a peer performance managed network; assessed on enrolment targets, data quality and provision of high quality samples enabling WGS. NHS GMCs are able to see their own patients' data in identifiable format and receive reports from Genomics England on the analysis of WGS from which they decide whether to validate and feedback the findings to patients.

Figure 3 Geographical location of the 13 NHS Genomic Medicine Centres.

NHS Genomic Medicine Centres



Creating a lasting legacy for genomic medicine



Source NHS England, 2017

6.2 Genomics England Clinical Interpretation Partnership

Genomics England formed the Genomics England Clinical Interpretation Partnership (GeCIP) to ensure that the UK maximises research opportunities. Following two open advertisements more than 2,600 UK and international clinicians and scientists volunteered to work within 40 self-organised and self-governing domains facing rare disease, specific cancers and offering disease-agnostic ('cross-cutting') analytical, ethical and health economic skill sets¹. This is designed to harness the talent-base in the NHS and academia to drive up the clinical interpretation of data from the 100,000 Genomes Project for greater patient benefit¹. GeCIP domains are not funded by Genomics England, but have started winning substantial grant awards on the basis of their national and international collaborations of expertise with privileged access to the genomic and clinical data from the project.

6.3 Lifelong electronic health record linkage

In partnership with NHS Digital (formerly the Health and Social Care Information Centre) the 100,000 Genomes Project is creating a longitudinal life-course electronic health database of all participants, based upon a flow of electronic health data from primary care, hospitals, outcomes, registries and social care records³⁴. These extensive records will provide the opportunity to evaluate WGS in the context of rich and extended phenotypes such as biochemical parameters, health outcomes and mortality data, and pharmacogenomics. For example, the first 22,000 participants experienced 1.4 million hospital episodes of care between 1997 and 2017, indicating the richness of this data. This will allow researchers in GeCIP domains to move beyond the primary phenotype of the rare disease, cancer or infectious disease that led to the patient's enrolment, to evaluate the WGS in the context of other continuous traits, diseases and response to therapy.

6.4 Opportunities for industry partnership from the 100,000 Genomes Project

To maximise new opportunities for patients, attract inward investment to the UK and stimulate a vibrant UK genomics industry, Genomics England has created the GENE consortium of 13 companies working in a precompetitive environment to help structure the research environment for industrial collaboration¹. We hope to undertake and attract stratified healthcare trials and bring new medicines to the UK at the earliest opportunity to enable participants to gain accelerated access to the best treatments (the potential to create a vibrant genomics industry in the UK is considered in Chapter 13 on economics).

6.5 Public engagement and patient involvement

Patient and public involvement has been an integral and vital part of the 100,000 Genomes Project. As the project began, the views of potential participants on ethical issues raised by genomic medicine were sought and fed into the development of patient literature. Potential participants were also involved in the development of consent materials. Each of the 13 NHS GMCs have a patient and public involvement and engagement (PPIE) lead who is responsible for local awareness and patient involvement. There is a national Participant Panel which has a particular responsibility with regard to data access requests. They also ensure the interests of participants are always at the centre of the 100,000 Genomes Project by:

- Making sure that the views and feedback from participants right across the project are fed back to Genomics England
- Developing programmes as appropriate to help ensure the interest of participants remain at the centre of work on the Project in a number of areas, including but not restricted to children and young people, education, information and communication and participant services.
- Acting as a responsive consultant, providing timely advice, guidance and recommendations to Genomics England and to other bodies as required.

This group acts as an advisory committee to the Genomics England Board and participants are represented on the Data Access Review Committee, the Ethics Advisory Committee and the GeCIP Board. As part of the Genomics England Engagement Strategy, a programme of activities – the Genomics Conversation – was launched to engage the general public and relevant stakeholders in key topics relating to genomic medicine. The Genomics Conversation has rolled out a broad range of activities, including debates, discussions, presentations, and outreach through social and traditional media^{35,36}.

7. Case studies

Case study 1 - A mother describes her family's experience in the 100,000 Genomes Project

When my daughter was born, everything seemed fine but by the time she was a year old we could tell that she wasn't moving around as much as other children. We could see she was behind on all her milestones and at 13 months old she had her first seizure. We were told that they may be febrile convulsions, but as the seizures became more and more frequent we could tell something wasn't right.

Doctors couldn't tell us what was wrong but said that a diagnosis didn't really matter as they could just treat Jessica's individual symptoms. We felt that it really mattered having a diagnosis.

It's the not knowing; I found that really difficult. Most children with an undiagnosed condition have developmental delay, epilepsy and a squint – all of the things that Jessica has – so if you Google that you get umpteen possibilities. After lots of invasive tests, we were told that Jessica had an undiagnosed condition. As soon as we knew this we became a part of SWAN UK (Syndromes Without A Name) who support families who have undiagnosed conditions. At a SWAN organised information day for families, a representative from the 100,000 Genomes Project spoke about what the project was and how it might help families. We were really keen to join the project as we'd already got to the end of medical testing and did not have any answers.

It was really easy taking part. We had our blood taken and that was it. We didn't need to come up to the hospital again until we got a result. We would recommend it to anyone. After a two year wait, we got a call to say that they had found a genetic error in Jessica that isn't shared by either me or her dad. The gene error and associated condition, called GLUT1 Deficiency Syndrome, means that her brain doesn't have enough glucose to function properly. We had always been worried that something preventable happened during the pregnancy or at the birth to cause Jessica's problems and it was such a relief to know there's nothing we could have done differently.

Now that we have this diagnosis there are things that we can do differently almost straight away. Her condition has a high chance of improvement on a high fat diet, which means that her medication dose is likely to decrease and her epilepsy may be more easily controlled.

Hopefully she might have better balance so she can be more stable and walk more. She's now four years old but still looks like a wobbly toddler trying to move around!

A diagnosis also means that we can link up with other families who are in the same boat and can offer support. I'm really looking forward to saying 'We are one of you, we have this problem too!'

The results are also going to be very useful for family planning. If we had had another child before, we didn't know if they would be affected but now we can say that there is only a tiny chance.

More than anything the outcome of the project has taken the uncertainty out of life for us and the worry of not knowing what was wrong. It has allowed us to feel like we can take control of things and make positive changes for Jessica. It may also open doors to other research projects that we can get on to. These could be more specific to her condition and we are hopeful that they could one day find a cure.

About Glucose transporter type 1 (Glut1) deficiency syndrome

Glucose transporter type 1 (Glut1) deficiency syndrome is a rare genetic metabolic disorder characterized by deficiency of a protein that is required for glucose to cross the blood-brain barrier. Glut1 deficiency syndrome is due to mutations in the SLC2A1 gene and is inherited as an autosomal dominant trait but may be inherited as an autosomal recessive trait. The most common symptom is seizures (epilepsy), which usually begin within the first few months of life. However, the symptoms and severity of Glut1 deficiency syndrome can vary substantially from one person to another. Additional symptoms include abnormal eye movements, movement disorders, developmental delays, and varying degrees of cognitive impairment and speech and language abnormalities. Glut1 deficiency syndrome does not respond to traditional epilepsy treatments but has been successfully treated with the ketogenic diet (high fat) because the fat is used to make glucose inside the brain.

Case study 2 - Focal Segmental Glomerulosclerosis

In his mid-twenties a patient developed high blood pressure and progressive kidney failure. His father, brother and uncle all died of this condition. A biopsy of the patient's kidney was suggestive of focal segmental glomerulosclerosis which leads to kidney failure. He was treated with dialysis and aged 29 received a kidney transplant. At the age of 57 his transplant kidney began to fail and he received a further period of dialysis and then a second kidney transplant.

His daughter also developed high blood pressure and urinary protein loss but because of good blood pressure control has not developed kidney failure. She has been desperately worried about whether her own daughter, now in her mid-teens, will develop this inherited form of kidney disease. Routine NHS genetic testing had not identified a genomic diagnosis for either the father or the daughter so they volunteered for the 100,000 Genomes Project Pilot.

The 100,000 Genomes Project identified a known pathogenic mutation for focal segmental glomerulosclerosis in the gene inverted formin 2 which affects the basement membrane in the glomerulus.

This was present in the father and daughter so segregated within the family with the disease. The NHS validated these findings and chose to return these results to the family. Importantly they went a stage further and tested the fifteen year old who does not have this variant and so will not develop this disease. Although this has not identified a targeted treatment for this family, just knowing the genetic diagnosis and receiving reassurance that the youngest family member is not at risk has been a huge relief to this family.

About focal segmental glomerulosclerosis.

Focal segmental glomerulosclerosis causes protein loss in the urine (nephrotic syndrome) in children and adolescents and is a cause of kidney failure in adults. It accounts for about a sixth of the cases of nephrotic syndrome. Minimal change disease (MCD) is by far the most common cause of nephrotic syndrome in children: MCD and primary FSGS may have a similar cause

8. Conclusions

The 100,000 Genomes Project has created a superb opportunity for the UK to become world leaders in the application of genomic medicine to healthcare. The programme has now been extended to Scotland, Wales and Northern Ireland, offering UK-wide coverage and access, and through international partnerships to other countries opening the possibility of multi-country stratified healthcare trials. The development of a longitudinal life-course dataset alongside WGS coupled with plans to concentrate all NHS genomic testing in the same data centre will create enormous long term opportunities for patients in our NHS and across the world who suffer from rare diseases, cancer and infection.

9. Suggestions for policy makers

- The increased complexity of genome-wide interpretation requires increased integration of laboratory and clinical working, both within clinical genetics and across a broad range of other medical specialties to mainstream genomics in healthcare.
- Processing of tumour samples from biopsy or resection should routinely move to DNA-preserving protocols, to allow NHS cancer patients to reap the clinical benefits of new molecular pathology and oncology discoveries and access suitable clinical trials.
- Data, informatics and bioinformatics solutions in genomic medicine are more effective and affordable if implemented centrally for the NHS; clinical control may be maintained locally as the output of the centralised pathway is returned to NHS laboratories around the country for reporting to clinicians and patients.
- Federation of clinical and scientific expertise in the rarest of diseases is required to implement effective standardised genomic medicine across the NHS. Collaboration between clinical laboratories, clinicians and academics should be facilitated and incentivised across the gamut of rare diseases.
- Widespread mainstreaming of genomic technologies in the diagnostic context requires ongoing education for health professionals, patients and the public.

10. References

1. Genomics England. The 100,000 Genomes Project Protocol v3. 2017. Available at <http://doi:10.6084/m9.figshare.4530893.v2> and www.genomicsengland.co.uk
2. Department of Health. Strategy for UK life sciences: one year on. 2012. Available at: <https://www.gov.uk/government/publications/strategy-for-uk-life-sciences-one-year-on>
3. Department of Health. UK Strategy for Rare Diseases.. 2013. Available at: <https://www.gov.uk/government/publications/rare-diseases-strategy>
4. IRDiRC Policies and Guidelines. 2013. Available at: <http://www.irdirc.org/reports-guidelines/policies-guidelines/>
5. Orphanet. Available at: <http://www.orpha.net/consor/cgi-bin/index.php>.
6. PhenoTips. Available at: <http://phenotips.org/>.
7. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet.* 2013 Oct;14(10):681-91. doi: 10.1038/nrg3555. Epub 2013 Sep 3.
8. Bragin E, Chatzimichali EA, Wright CF, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D993-D1000. doi: 10.1093/nar/gkt937. Epub 2013 Oct 22.
9. Akawi N, McRae J, Ansari M, et al. Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nat Genet.* 2015 Nov;47(11):1363-9. doi: 10.1038/ng.3410. Epub 2015 Oct 5.
10. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature.* 2015 Mar 12;519(7542):223-8. doi: 10.1038/nature14135. Epub 2014 Dec 24.
11. Glissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* 2014 Jul 17;511(7509):344-7. doi: 10.1038/nature13394. Epub 2014 Jun 4.
12. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet.* 2015 Jul;47(7):717-26. doi: 10.1038/ng.3304. Epub 2015 May 18.
13. Human Phenotype Ontology. Available at: <http://www.human-phenotype-ontology.org/>.
14. Office of National Statistics. Cancer registration statistics, England: first release, 2015. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/firstrelease2015>
15. International network of cancer genome projects. *Nature.* 2010 Apr 15;464(7291):993-8. doi: 10.1038/nature08987.
16. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013 Oct;45(10):1113-20. doi: 10.1038/ng.2764.
17. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science.* 2013 Mar 29;339(6127):1546-58. doi: 10.1126/science.1235122.
18. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014 Jan 23;505(7484):495-501. doi: 10.1038/nature12912. Epub 2014 Jan 5.
19. Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, Lawrence MS, Gonzalez-Perez A, Tamborero D, Cheng Y, Ryslik GA, Lopez-Bigas N, Getz G, Ding L, Raphael BJ. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet.* 2015 Feb;47(2):106-14. doi: 10.1038/ng.3168. Epub 2014 Dec 15.
20. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin SF, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD, Caldas C, Rosenfeld N. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature.* 2013 May 2;497(7447):108-12. doi: 10.1038/nature12065. Epub 2013 Apr 7.
21. Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med* 23, 517-525 (2017).
22. Rizvi, N.A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124-8 (2015).

23. Qiu, Z.W., Bi, J.H., Gazdar, A.F. & Song, K. Genome-wide copy number variation pattern analysis and a classification signature for non-small cell lung cancer. *Genes Chromosomes Cancer* (2017).
24. Notta, F. et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 538, 378-382 (2016).
25. Alexandrov, L.B., Nik-Zainal, S., Siu, H.C., Leung, S.Y. & Stratton, M.R. A mutational signature in gastric cancer suggests therapeutic strategies. *Nat Commun* 6, 8683 (2015).
26. O'Leary B, Turner NC. Science in Focus: Circulating Tumour DNA as a Liquid Biopsy. *Clinical oncology (Royal College of Radiologists (Great Britain))* 2016.
27. National Cancer Registration and Analysis Service. Collecting and Using Data. Available at: http://www.ncin.org.uk/collecting_and_using_data/
28. NHS Digital. Terminology and Classifications. Available at: <http://systems.hscic.gov.uk/data/uktc/snomed>
29. Chief Medical Officer's Report 2011 - Davies S. Annual Report of the Chief Medical Officer, Volume One, 2011, On the State of the Public's Health. London: Department of Health (2012).
30. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet.* 2012 Sep;13(9):601-12. doi: 10.1038/nrg3226. Epub 2012 Aug 7.
31. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, et al; COMPASS-TB Study Group. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med.* 2016 Jan;4(1):49-58. doi: 10.1016/S2213-2600(15)00466-X. Epub 2015 Dec 4.
32. Bradley P, Gordon NC, Walker TM, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun.* 2015 Dec 21;6:10063. doi: 10.1038/ncomms10063.
33. Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet.* 2011 Jun;12(6):417-28. doi: 10.1038/nrg2999. Epub 2011 May 18. Review.
34. NHS Digital. Available at: <https://digital.nhs.uk/>
35. Genomics England. Earning Trust: Public Engagement and Patient Involvement Strategy 2015-17. Available at <https://www.genomicsengland.co.uk/library-and-resources/>
36. Genomics England. A Year of Conversations about Genomics. Available at: <https://www.genomicsengland.co.uk/a-year-of-conversations-about-genomics/>

Chapter 3

Genomics and therapeutics

Chapter lead

Prof Munir Pirmohamed, University of Liverpool

1. Summary

New target identification

Elucidation of the genetic basis of rare diseases may identify novel targets which can lead to the development of compounds which have the potential to treat common complex diseases. Examples include PCSK9 inhibitors in the treatment of hypercholesterolaemia and anti-sclerostin antibodies for the treatment of osteoporosis.

Re-purposing of drugs

The discovery of the genetic basis of diseases including rare diseases may allow the re-purposing of medicines which have been widely used for different indications. This has the advantage of the use of a drug that has a known safety profile and is of lower risk for new drug development. An example is the use of riboflavin in childhood motor neurone disease.

Targeting specific mutations

Identification of the mutation(s) responsible for causing disease (including in the somatic genome of a cancer) may lead to drugs that target the mutated protein, and either inhibit its (adverse) function that drives disease or partially restore its (normal) function thereby ameliorating the effects of the mutation. Examples include the use of vemurafenib in malignant melanoma and ivacaftor in cystic fibrosis.

Stratification of intensity and type of therapy

Genomics can be used for stratification allowing individualisation of therapy. This may involve transcriptomic technologies which assess gene expression profiles prior to therapy predicting future disease course (for example, allowing for de-intensification of chemotherapy in breast cancer) or the stratification of disease through identification of disease-causing mutations (for example in diabetes allowing a switch from insulin therapy to oral hypoglycaemics).

Dose individualisation

The dose of a medicine required to produce its therapeutic or toxic effect varies in different individuals and is due to a combination of clinical and genetic factors. These can be utilised to develop dosing algorithms which can individualise the dose of the drug based on genotype and clinical factors “normalising” drug exposure in patients with different genetic variants. An example is the assessment of thiopurine methyltransferase activity prior to 6-mercaptopurine (6MP) use in patients with acute lymphoblastic leukemia (ALL) with alteration of the 6MP dose based on enzyme activity.

Improving drug safety

Serious adverse drug reactions, which can result in fatalities, may be due to genetic factors. Identification of these genetic factors prior to drug prescription and avoiding the culprit drug in patients with the susceptibility variant can prevent the adverse reaction. The best example is the use of HLA-B*57:01 genotyping prior to the use of the anti-HIV drug abacavir which has reduced the incidence of abacavir hypersensitivity from 5-7% to <1%.

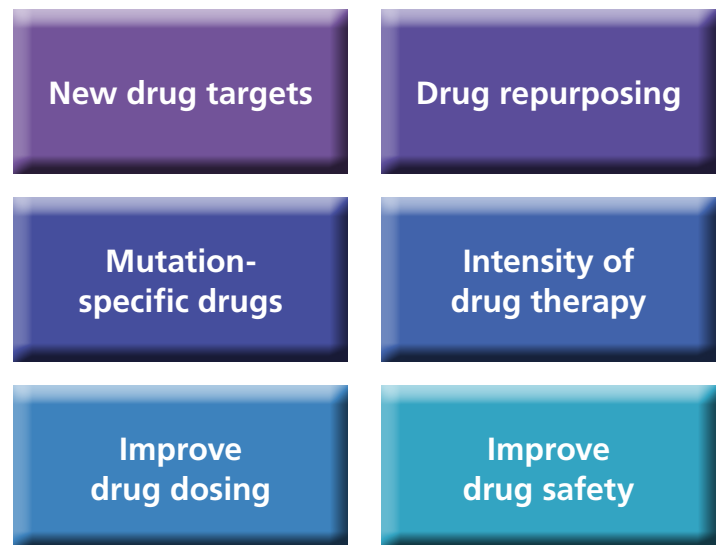
2. Background

A key issue for the genomic revolution is how it will help in improving treatment for diseases. As the number of people globally who have their genomes sequenced increases, the depth and breadth of information available will allow us to progress personalised or precision medicine, to ensure that patients get the right treatment at the right dose and at the right time. This will be crucial in ensuring that we optimise the benefit-risk ratio of all therapies maximizing efficacy and safety. However, the rate of translation from discovery to application and adoption into healthcare will vary with different therapeutic areas, and will be dependent on the quality of evidence, the unmet medical need, the cost-effectiveness of the drug/diagnostic combination, whether it has been approved by the regulators (MHRA, NICE etc), availability of the test in the NHS, whether it will be implemented in primary or specialist care, and education and training of the prescribers in the relevant specialty.

Genomics is already beginning to make an impact in the development of new drugs, and in using existing drugs better. This overall area is called pharmacogenomics, which has been defined by the international conference on harmonization¹ as “The study of variations of DNA and RNA characteristics as related to drug response”. In this chapter, we highlight areas where there have been significant advances, using clinical vignettes, that have led to patient benefit. These include the following (see Figure 1):

- Identifying new drug targets using genomic information.
- Repurposing existing drugs for new indications based on new genomic information.
- Developing drugs targeted at specific mutations.
- Using genomic technologies to stratify the intensity of drug therapy.
- Using genomic information to improve drug dosing.
- Using genomic information to prevent adverse drug reactions.

Figure 1 – Pharmacogenomics – areas which have seen significant advance



3. Genomics and impact on therapeutics

New target identification

Elucidation of the genetic basis of rare diseases can lead to the identification of novel targets that can then be used for drug development, and subsequent treatment of common complex diseases. There are two striking examples of this. Sclerosteosis is an autosomal recessive disease, first described in 1958, which affects about 100 people worldwide. The disease is characterised by skeletal overgrowth and sometimes syndactyly (“webbed fingers”), with homozygotes being severely affected, while heterozygotes have increased bone mass and rarely get fractures². The disease is caused by several inactivating mutations in the SOST gene on chromosome 17, which leads to a defective sclerostin protein. This has led to the development of anti-sclerostin antibodies which in animal models have been shown to increase bone mass, and in early phase trials in osteopenic patients, to increase both bone mineral density and biomarkers of bone formation³. Unlike bisphosphonates which are widely used in osteoporosis, anti-sclerostin antibodies are anabolic rather than anti-resorptive. Larger trials are currently ongoing to determine the effect of anti-sclerostin antibodies on fractures rates in patients with reduced mineral bone density.

The second example relates to the proprotein convertase subtilisin/kexin type 9 (PCSK9) gene⁴. Loss-of-function mutations in this gene lead to lowered plasma levels of LDL-cholesterol (LDL-C), while gain-of-function mutations lead to high LDL-C and premature cardiovascular disease. It is known that PCSK-9 levels modulate LDL receptor recycling on hepatocytes and can affect LDL-C levels. This has recently led to the development and licensing of monoclonal antibodies which in randomised controlled trials have been shown to lower LDL-C⁵. Larger trials are currently on-going to determine the effect of anti-PCSK9 therapy on cardiovascular mortality.

Both these examples show the importance of studying rare genetic conditions and how discovery of the molecular basis of the disease can be further studied using cellular and animal models, ultimately leading to development of novel therapeutics for common diseases. Indeed, a recent analysis has shown that selecting targets where there is a genetic basis for disease could double the success rate in the development of new drugs⁶, which is much needed given the concerns that have been widely raised about the high rate of attrition in drug development.

Repurposing existing drugs

The development of a new drug is a long and expensive process. Based on evidence of the genetic defect, it may be possible to utilise existing drugs for new diseases (or indications), a process which has been termed re-purposing. This has the advantage of being able to use a marketed drug with a well-known safety profile for diseases (including rare diseases) where there may be an unmet medical need. This makes the whole process much quicker, less expensive and with lower attrition rates than observed with new drug development. An example of this is the Brown-Vialetto-Van Laere syndrome (also known as Fazio Londe syndrome), an autosomal recessive condition affecting approximately 60 patients worldwide⁷ (see case study 1). The disease is characterised by difficulties in swallowing, hearing, slurred speech and weakness of limb muscles, with death being caused by respiratory problems. It is sometimes known as childhood motor neurone disease. Depending on the age of onset, survival may be between 1 to 10 years from the appearance of the first symptoms. Exome sequencing has identified that the disease is caused by mutations in the family of genes (SLC52A2 and SLC52A3) which code for a transport protein responsible for vitamin B2 or riboflavin transport across the gut and into neuronal cells⁸. Deficiency of riboflavin can affect many intracellular processes that culminate in energy production. Since the demonstration of the genetic defect responsible for this disease, patients have been treated with high dose riboflavin which has led to stabilisation of the disease, marked improvement of symptoms or complete resolution of symptoms⁷, which will also in all likelihood result in prolonged survival.

Case study 1

At the age of three years, Ava, who had previously been well and developing normally, started developing neurological symptoms which included slurred speech, slowed development and frequent falls. Between the ages of three and six years, Ava was seen by numerous doctors and in several hospitals undergoing many different investigations and treatments (including surgery) with some presumptive diagnoses. Because of the inability to arrive at a definitive diagnosis, whole genome sequencing was conducted. This led to the identification of a mutation in the riboflavin transporter, and to a diagnosis of Brown-Vialetto-Van-Laere syndrome. The parents were hugely relieved to have received a diagnosis, and importantly there was a potential for treatment. Ava was started on riboflavin as a result of the diagnosis. She has multiple disabilities which first started appearing at the age of three years but the riboflavin has resulted in a halt to her deteriorating symptoms, and there has been an improvement in some symptoms including her mobility.

Adapted from "Ava's story" - <http://undiagnosed.org.uk/archives/5333>

Developing drugs for genetic variants

Pathogenic genetic variants may be amenable to targeting through new drug development. This has been highly successful in cancer where targeting of genetic variation that drives tumour development has led to revolutionary treatments which have resulted in durable clinical responses. The most impressive example is that of imatinib (and the subsequent tyrosine kinase inhibitors) which targets BCR-ABL for the treatment of Philadelphia chromosome positive chronic myeloid leukaemia⁹. In the pivotal trial, complete cytogenetic response was seen in 76% of patients treated with imatinib, compared with 14.5% in the comparator arm (interferon- α and low-dose cytarabine). Driver mutations have also been identified for solid tumours through sequencing approaches which have led to the development of novel therapies¹⁰. For example, approximately 50% of metastatic malignant melanoma cases have activating mutations in codon 600 of the BRAF gene¹¹. This led to the development of a novel BRAF inhibitor vemurafenib, which received regulatory approval in the EU within 180 days. In the development programme, vemurafenib was shown to increase progression-free survival by about six months in both treatment-experienced and treatment-naïve patients with metastatic malignant melanoma¹². The summary of product characteristic (SmPC or drug label) for vemurafenib states that "patients must have BRAF V600 mutation-positive tumour status confirmed by a validated test", which represents an example of a drug-diagnostic combination that is likely to be commonplace in medicine in the future.

Germline mutations can also serve as drug targets. Cystic fibrosis (CF) is caused by mutations in the cystic fibrosis transmembrane regulator (CFTR) on chromosome 7. Over 2,000 mutations have been identified in the CFTR gene – these can be categorised in functional classes as group I (no protein), II (no protein trafficking), III (protein trafficked to cell membrane but no gating function), IV (less function), V (less protein) and VI (less stable protein)¹³. A drug development programme which screened approximately 600,000 compounds led to the discovery of ivacaftor, which targets the G551D mutation (functional class III) which is present in 4% of the cystic fibrosis population. Ivacaftor has been shown to be highly effective, improving respiratory function (increase in FEV1 by 10%)¹⁴, and quality of life measures (respiratory symptoms, physical and social functioning, and health perceptions)¹⁵. The license for ivacaftor has recently been extended to other class III mutations (G1244E, G1349D, G178R, G551S, S1251N, S1255P, S549N and S549R). The most common mutation in the CF gene (Δ 508) (functional class II) can also be treated by a recently licensed drug combination of lumacaftor and ivacaftor¹⁶ – the effect however is less than seen with the G551D mutation, with a FEV1 increase of about 4%.

Stratifying the intensity and type of drug therapy

The course of a particular disease can vary in different patients, yet at present we largely treat all patients in the same way. Tests which could differentiate those patients who are more likely to develop a more severe disease course from those with a more benign course would help in stratifying therapy – i.e. more intensive therapy would be given to patients where the disease course was predicted to be more malignant. In order to achieve this, there has been increasing interest in gene expression profiling. For instance, in breast cancer, chemotherapy has been offered to patients with locally invasive cancer dependent on its size, grade and whether lymph nodes are involved. Using cancer breast tissue removed at surgery, it is now possible to analyse the tissue for expression of 21 genes, with patients being given a recurrence score – the lower the score, the less likely the cancer is to recur. A recent trial showed that patients with recurrence scores between 0 and 10 can be treated with endocrine therapy only, and chemotherapy avoided, with very good outcomes at five years¹⁷. This genomic test thus allows women to avoid receiving unnecessary chemotherapy with its attendant severe side effects without any adverse consequence to their overall likelihood of cancer recurrence. There is much on-going research to determine whether such gene expression profiles may also be of value in other diseases such as sepsis, to personalise therapies based on predicted disease severity.

Identification of genetic mutations can also allow stratification of disease and therefore treatment. For example, young patients with high blood sugars are assumed to have type 1 diabetes which requires treatment with insulin. However, some of these patients have monogenic forms of diabetes, such as maturity onset diabetes of the young (MODY), where mutations in the transcription factor genes encoding hepatocyte nuclear factor 1A and 4A (HNF1A and HNF4A) can lead to a switch in treatment from insulin to low doses of sulfonylurea drugs¹⁸. Similarly, genetic testing of neonatal diabetes can now allow the identification of numerous genetic forms of diabetes¹⁹, which can, in some cases, lead to the substitution of insulin sulfonylureas (for example in neonates with mutation in genes that comprise the potassium ATP channel (KCNJ11 and ABCC8)²⁰.

Improving dosing through genomics

“Poison is in everything and no thing is without poison. The dosage makes it either a poison or a remedy.”
Paracelsus, 1493–1541

We currently treat patients on the basis of one-dose-fits-all. However, the same dose can lead to marked differences in the total amount of drug getting into the circulation or tissue where it acts (which is called exposure). This may lead to variability in efficacy with some patients not responding, while others who have high exposure develop toxicity. This variability in exposure can be due to a combination of genetic and environmental factors²¹. Individualising the dose of a drug based on a person’s genetic profile may lead to equivalent exposures in different patients, and improve the efficacy and reduce the toxicity of the drug. This has been shown with the drug 6-mercaptopurine (6MP), which is used to treat childhood acute lymphoblastic leukaemia²². 6MP is metabolised by an enzyme called thiopurine methyl transferase (TPMT). The TPMT gene has many mutations which can render it inactive. Approximately 10% of patients are heterozygotes while 1 in 300 have no enzyme, which increases their susceptibility to 6MP-induced bone marrow suppression. Personalising dose based on genotype leads to equivalent systemic exposure to 6MP in different genotype groups, reducing the risk of severe bone marrow suppression, especially in those patients who lack the enzyme.

Warfarin represents another example; it is a very commonly used drug in the UK, taken by about 1-1.5% of the population. The dose of warfarin varies between different individuals from 0.5mg/day to 20mg/day (see case study 2). This is due to various clinical factors (such as age, body mass index and the use of interacting medications) and genetic factors – the latter account for about 40% of the dose variation²³. The most important genes are CYP2C9 (responsible for the metabolism of warfarin) and VKORC1 (which is involved in the vitamin K cycle and is inhibited by warfarin). Inability to predict the dose of warfarin can predispose patients to bleeding because of over-anticoagulation, or thrombosis because of under-anticoagulation. Dosing algorithms which incorporate both clinical and genetic factors have been developed to improve the control of anticoagulation as measured by the international normalised ratio (INR). A randomised controlled trial in the UK and Sweden was able to show that genotype-guided dosing was able to improve overall anticoagulation control when compared to standard dosing used in UK and Swedish anticoagulant clinics²⁴. Genotyping for the trial was undertaken on a point-of-care machine which provided the genetic results within two hours. Furthermore, this approach has been shown to be cost-effective²⁵. Genotype-guided dosing of warfarin is currently being implemented in the UK, initially in pilot sites, with roll-out more widely if successful.

Case study 2

Eric, aged 65 years, started developing palpitations and breathlessness on exertion. He went to see his GP, who, based on a clinical examination and an ECG, diagnosed him with atrial fibrillation. Eric was treated with bisoprolol, a beta-blocker, which controlled his heart rate. The GP also referred him to the anticoagulant clinic, who assessed Eric carefully to ensure that he would be a candidate for warfarin. This assessment included measurement of his weight (82kg), whether he was on any medications that might interact with warfarin (he was not) and whether there were any risk factors for bleeding. The clinic used the HAS-BLED score to assess bleeding risk in patients being considered for warfarin treatment. The HAS-BLED score, which is approved by NICE, evaluates the risk of bleeding based on a number of factors. Eric's HAS-BLED score was low and he was therefore considered to be eligible for warfarin. Eric was given general advice about avoiding excess alcohol, ensuring that he has a stable diet, and letting any health professionals know that he was on warfarin. Eric was started on 10mg on day one, 5 mg on day two and 5 mg on day three, and was asked to attend on day four to have his INR checked (a blood test which measures the degree of "thinning" of blood). Just before attending the anticoagulant clinic to have his INR measured, Eric went to the local supermarket to get some essential food items. While shopping he knocked his left leg against a shelf but it was minor and he did not take notice of it.

However, by the time he attended the anticoagulant clinic two hours later, he had developed bruising on his left calf. Over the next hour, this swelling increased in size and became painful. The anticoagulant clinic checked his INR and found it to be raised to 6 (the aim of anticoagulation is to maintain the INR between 2 and 3). Eric was immediately referred to the acute medical unit who diagnosed a bleed into his calf muscles and the development of a haematoma. This was confirmed by ultrasound. Eric was admitted to hospital, his anticoagulation was reversed with vitamin K (which is an antidote for warfarin) and he was kept under continuous observation. The haematoma did not get any bigger, and the surgeons felt that the haematoma did not need surgical evacuation. Over the next two weeks, Eric had to stay in hospital for pain control and to receive physiotherapy. The haematoma took over two months to resolve completely.

In a follow up appointment, Eric was genotyped for genes known to affect the response to warfarin. He was found to be carrying two copies of the CYP2C9*3 allele (i.e. he was a homozygote), which is known to (a) be present in 1 in 500 of the population and (b) results in a reduction in the ability to breakdown warfarin by 90%. If his CYP2C9 status had been known prior to the start warfarin, Eric would have received lower doses of warfarin, which would have prevented the rise in INR, and the bleed into his calf.

Preventing serious adverse drug reactions through genomic testing

Adverse drug reactions (ADRs) are responsible for 6.5% of hospital admissions²⁶. Not all of these will be preventable through genomic testing, but genetic factors may be important, to a greater or lesser extent, in many ADRs. There is increasing evidence of the importance of the HLA genes on the short arm of chromosome 6 in predisposing to serious adverse drug reactions affecting the skin, liver and bone marrow²⁷. The best example of this is with the drug abacavir, an anti-HIV drug associated with serious hypersensitivity reactions (characterised by skin rash, fever, lung and GI involvement) (see case study 3). The HLA allele, HLA-B*57:01, is a predisposing factor for abacavir hypersensitivity in many different ethnic groups. Pre-treatment testing for HLA-B*57:01 has been shown to reduce the incidence of hypersensitivity from 5-7% to less than 1%, and this is cost-effective²⁸. The prescribing instructions for abacavir mandate genetic testing before the use of the drug. For carbamazepine, an antiepileptic, there is a strong association between HLA-B*15:02 and the risk of Stevens-Johnson Syndrome in South East Asian populations (Han Chinese, Thai and Malays)²⁹. Pre-prescription genotyping is again mandated in these populations, and patients in Thailand are now issued with a genetic card with their HLA-B*15:02 status. This HLA allele is rare in the Northern European population where another HLA allele, HLA-A*31:01, has been to act as a predisposing factor for carbamazepine-induced hypersensitivity. Indeed, since the beginning of this century, over 24 different HLA associations have been identified with different drug-induced ADRs involving the skin, liver, muscle and bone marrow²⁷. The challenge will now be to determine how they can be used in clinical practice to improve drug safety.

Case study 3

P, who is male and aged 28 years, was diagnosed with HIV in 2004. As part of the anti-HIV multi-drug treatment regimen, one of the drugs commenced was abacavir. This was a new drug approved by the European Medicines Agency in 1999. P started taking his drugs (as per instructions), and apart from feeling mildly nauseated, he did not have any problems. Two weeks after starting the drugs, P started to feel unwell, had a mild rash covering his body, and was feeling hot and sweaty. He self-diagnosed a viral infection and since he did not feel particularly unwell, went off to work. The next morning, the rash was still present, and he measured his temperature, which was 37.50°C. P took his anti-HIV drugs as normal, and within one hour, he felt very unwell. The rash was more extensive, his skin felt hot to touch, and he felt faint. He asked a friend to take him to the hospital, where he was seen in the A&E Department. On arrival, P had a temperature of 38.50°C, his BP was 90/60, and he had an extensive maculopapular skin rash affecting the whole of his body surface area without any mucosal involvement. The A&E doctor asked for an opinion from the HIV clinic, who immediately suspected that this was a case of abacavir hypersensitivity, and stopped all his anti-HIV drugs. P unfortunately deteriorated with BP going even lower. He was resuscitated in the A&E Department, and was then admitted to the intensive care unit (ICU). P needed intensive fluid management on the ICU together with the use of steroids and inotropes (medications to increase his BP), and all together stayed on ICU for four days. Once discharged from ICU, he went to the infectious disease ward where he stayed for another four days, while his symptoms resolved and signs improved.

Two weeks after discharge, P was reviewed in clinic, and was started on alternative anti-HIV drugs. P remained generally well for the following two years with HIV viral load controlled. In 2006, the HIV physician was fairly confident that P was allergic to abacavir, but given that P had been started on several agents at the same time in 2004, it was difficult to be absolutely sure that abacavir was the culprit drug that caused the reaction. The HIV physician therefore checked the HLA status, and found P to be positive for HLA-B*57:01, which provided further proof that the reaction suffered in 2004 was due to abacavir. Soon after this, given the increasing amount of evidence of the role of HLA-B*57:01 in predisposing to abacavir hypersensitivity, the prescribing guidelines were changed to recommend all patients to be genotyped for this HLA allele prior to the use of abacavir. Since 2006, every UK HIV clinic has been testing for HLA-B*57:01, and the incidence of abacavir hypersensitivity has dropped from about 7% to <1%.

4. Conclusions

It has been estimated that there are over 10,000 potential drug targets in the human genome³⁰. This represents a potential goldmine to develop new therapeutics that will advance the current practice of medicine, and help in developing new treatments for areas of unmet medical need. This chapter has provided some key examples of how genomics knowledge is already driving forward drug development. While there has been a focus on new drug development, it is also important to remember that the majority of drugs used in the NHS are generic or off-patent. Variability in response (efficacy, safety, dose requirement) is also seen with these generic drugs. Studying the mechanisms by which generic drugs lead to variability in response is therefore also important as it will help in understanding the reasons for variability seen with new drugs, and may also allow faster adoption of genomic technologies into the NHS for patient benefit (and at a lower cost) given the long timelines required for new drug discovery and development. However, the two areas are not mutually exclusive and both should progress in parallel.

A key opportunity is the potential to re-purpose existing drugs for diseases (including rare diseases) for which we currently have no preventive or treatment options. This has the benefit of using drugs where there is a known safety profile for diseases where treatments may not currently exist. It also overcomes the issue of expensive new drug development, particularly in rare diseases, where it may not be economically attractive for Industry to develop drugs.

Although the expansion of genomic knowledge brings about a lot of opportunities, it does also pose some challenges including how we will develop the evidence base in a timely and cost-effective manner to ensure that these advances improve patient health and create economic gains. New models of regulation, re-evaluation of what constitutes evidence, assessment of cost-effectiveness of new innovations, and development of new business models, will all be needed to ensure we maximise the gains for patients from the genomic revolution.

6. Suggestions for policy makers

- Embrace and integrate all forms of evidence rather than relying on the usual hierarchies of evidence, allowing for the development of a framework that provides a proportionate degree of regulation without stifling innovation.
- Develop educational programmes which start at school, into university education and post-qualification, for all healthcare workers to ensure that we have a well-informed public and highly skilled workforce, to facilitate adoption of these innovations into practice.
- Increased funding needs to be made available to undertake research not only on how genomics can facilitate new drug development, but also on how it can improve the use of existing or generic drugs. Understanding of the molecular basis of disease may also allow for the re-purposing of drugs, i.e. the use of existing drugs for new indications where there may be an unmet medical need.
- There needs to be an increasing emphasis on the development of intelligent decision support systems which will enable prescribers to use genomics (and other technologies) in their everyday clinical practice. It is not going to be possible for all clinicians to have the depth of knowledge to implement personalised medicine approaches. Developments in information technology that are able to integrate genomic knowledge into electronic prescribing systems are going to be vital to enable adoption into practice.

7. References

- European Medicines Agency. ICH Topic E15: Definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories 2007 November 2007.
- Larsson S. Anti-sclerostin - is there an indication? *Injury* 2016;47 Suppl 1:S31-5.
- McClung MR, Grauer A, Boonen S, et al. Romosozumab in postmenopausal women with low bone mineral density. *N Engl J Med* 2014;370:412-20.
- Chapman MJ, Stock JK, Ginsberg HN, Forman P. PCSK9 inhibitors and cardiovascular disease: heralding a new therapeutic era. *Curr Opin Lipidol* 2015;26:511-20.
- Giugliano RP, Desai NR, Kohli P, et al. Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 in combination with a statin in patients with hypercholesterolaemia (LAPLACE-TIMI 57): a randomised, placebo-controlled, dose-ranging, phase 2 study. *Lancet* 2012;380:2007-17.
- Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet* 2015;47:856-60.
- Bosch AM, Stroek K, Abeling NG, Waterham HR, Ijlst L, Wanders RJ. The Brown-Vialetto-Van Laere and Fazio Londe syndrome revisited: natural history, genetics, treatment and future perspectives. *Orphanet J Rare Dis* 2012;7:83.
- Johnson JO, Gibbs JR, Megarbane A, et al. Exome sequencing reveals riboflavin transporter mutations as a cause of motor neuron disease. *Brain* 2012;135:2875-82.
- O'Brien SG, Guilhot F, Larson RA, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med* 2003;348:994-1004.
- Relling MV, Evans WE. Pharmacogenomics in the clinic. *Nature* 2015;526:343-50.
- Davies H, Bignell GR, Cox C, et al. Mutations of the BRAF gene in human cancer. *Nature* 2002;417:949-54.
- Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 2011;364:2507-16.
- Amaral MD. Novel personalized therapies for cystic fibrosis: treating the basic defect in all patients. *J Intern Med* 2015;277:155-66.
- Ramsey BW, Davies J, McElvaney NG, et al. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *N Engl J Med* 2011;365:1663-72.
- Quittner A, Suthoff E, Rendas-Baum R, et al. Effect of ivacaftor treatment in patients with cystic fibrosis and the G551D-CFTR mutation: patient-reported outcomes in the STRIVE randomized, controlled trial. *Health Qual Life Outcomes* 2015;13:93.
- Wainwright CE, Elborn JS, Ramsey BW, et al. Lumacaftor-ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR. *N Engl J Med* 2015;373:220-31.
- Sparano JA, Gray RJ, Makower DF, et al. Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med* 2015;373:2005-14.
- Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT. Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* 2003;362:1275-81.
- De Franco E, Flanagan SE, Houghton JA, et al. The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. *Lancet* 2015;386:957-63.
- Pearson ER, Flechtner I, Njolstad PR, et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med* 2006;355:467-77.
- Pirmohamed M. Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. *Annual review of genomics and human genetics* 2014;15:349-70.
- McLeod HL, Krynetski EY, Relling MV, Evans WE. Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia* 2000;14:567-72.
- Pirmohamed M, Kamali F, Daly AK, Wadelius M. Oral anticoagulation: a critique of recent advances and controversies. *Trends in pharmacological sciences* 2015;36:153-63.
- Pirmohamed M, Burnside G, Eriksson N, et al. A randomized trial of genotype-guided dosing of warfarin. *The New England journal of medicine* 2013;369:2294-303.
- Verhoef TI, Redekop WK, Langenskiold S, et al. Cost-effectiveness of pharmacogenetic-guided dosing of warfarin in the United Kingdom and Sweden. *Pharmacogenomics J* 2016.
- Pirmohamed M, James S, Meakin S, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *Bmj* 2004;329:15-9.

27. Pirmohamed M, Ostrov DA, Park BK. New genetic findings lead the way to a better understanding of fundamental mechanisms of drug hypersensitivity. *The Journal of allergy and clinical immunology* 2015;136:236-44.
28. Mallal S, Phillips E, Carosi G, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 2008;358:568-79.
29. Yip VL, Marson AG, Jorgensen AL, Pirmohamed M, Alfirevic A. HLA genotype and carbamazepine-induced cutaneous adverse drug reactions: a systematic review. *Clin Pharmacol Ther* 2012;92:757-65.
30. Bailey D, Zanders E, Dean P. The end of the beginning for genomic medicine. *Nat Biotechnol* 2001;19:207-9.

Chapter 4

Developing medicines to prevent the development and alter the course of severe genetic diseases

Chapter lead and author

Prof David Altshuler, Vertex Pharmaceuticals

1. Summary

In order to minimise the impact of severe diseases it would be ideal to intervene before symptoms occur with therapies that modify the course of disease in order to prevent the suffering and long-term morbidity and mortality that might otherwise occur. This is the key concept underlying what has been termed 'precision medicine', in which a strong understanding of the molecular pathophysiology of disease enables the design of therapies to directly target and repair the underlying defect, and the targeting of that therapy to patients most likely to benefit. Such a paradigm requires deep understanding of human disease, methods to identify at-risk individuals before disease symptoms develop, the discovery of therapies that address the underlying biological processes that cause the disease, and the development and acceptance of surrogate disease endpoints that are predictive of disease modification. Yet for most diseases we lack the knowledge and means to intervene early with disease-modifying therapy.

Advances in human genetics have identified the biological basis of a large number of early onset, severe genetic diseases, inspiring new classes of therapies aimed at the underlying causes of these previously intractable conditions.

However, a targeted approach to prevent the morbidity and mortality associated with genetic disease requires a fundamental shift in how new medicines are developed and appraised for use in healthcare systems:

- The specific genes and pathways underlying genetic diseases are seldom amenable to traditional pharmaceutical approaches, requiring risk-taking innovation in the technological approaches to therapy.
- When genetic risk factors are highly predictive of future disease, it is possible to tailor clinical trials to individuals most likely to benefit. Genetics may predict in whom disease will develop, but not the specific timing at which symptoms occur. Natural history studies are therefore needed to establish the course of disease, and to explore how much gene activity is needed to avoid disease outcomes.

- The specific mutations that lead to genetic disease are typically large in number, rare in the population, and diverse in their molecular properties. Genetically targeted therapies will require methods to identify which mutations have shared molecular and clinical characteristics, and regulatory bodies and healthcare providers to use such categorisation in evaluating medicines and providing diagnosis and treatment.
- Many inherited diseases have early onset of symptoms and, in some cases, manifestations may not be reversible once developed. Thus, it may be necessary to start therapy early in life to prevent the onset of significant symptoms, clinical burden and long-term consequences. An early intervention approach will require paths to safely evaluate medicines in younger populations, as well as to monitor therapy and demonstrate value in individuals who have not yet developed clinically significant disease manifestations.
- Successful implementation of preventative therapy will require a healthcare delivery shift from a paradigm of symptomatic relief and rescue therapy to prevention of disease symptoms before they occur.

Keys to success will include: (a) investments in developing new therapeutic modalities that expand the range of 'druggable' targets; (b) wider use of genetic diagnosis and natural history studies to identify and accurately predict risk in individuals; (c) approaches to identify which rare mutations share molecular mechanisms; (d) validation and use of surrogate endpoints sufficient to evaluate benefit of therapy prior to disease onset; (e) establishing regulatory pathways for use in targeted prevention and; (f) development by policy-makers of appraisal processes that assess and incorporate the value to individuals, families and society of pioneering these new approaches to prevent the onset and long-term consequences of severe genetic disease.

2. Background

Severe genetic diseases are among the most devastating of all human conditions, impacting the lives of children and their families and interrupting the normal human cycle of growth, development, work and family. Until relatively recently, the underlying causes of genetic diseases were largely unknown, treatments were ineffective and outcomes poor. Over the past 30 years, the genes responsible for thousands of human diseases have been identified, laying bare their precise biological causes.¹ This knowledge provides a foundation and a guide to the development of transformative therapies.

As a method for discovering causes of diseases, human genetic mapping is unique in that it is unbiased by prior assumptions about the nature of the specific biological process causing the disease. Rather than hypothesise that a well-studied cellular or biological process may be responsible, human geneticists trace the inheritance of every human chromosome, pinpointing the specific changes in the DNA sequence that track with disease. This newly identified genetic risk factor is then studied to understand its biological functions, clinical consequences and clues to therapy. The fruits of this effort can be valuable in offering new clues in the complex journey of drug discovery: genes, and the biological processes they control, documented to play a causal role in human disease.

Where a genetic disease is rare, severe in effect, and early in onset, the causal mutations themselves are typically rare in the general population. These disease mutations are often so rare that individual patients each carry a different and unique mutation. This diversity of mutations in a given gene can be understood based on commonalities in which the functional effects of different mutations cluster into one or a few shared mechanisms. For example, a variety of different mutations in the same gene might all lead to the encoded protein not being made, with similar clinical effect. (This article focuses on rare diseases, and will not discuss further common, late-onset diseases, where the mutations are typically more common in the population and have subtle effects on the regulation of gene function.)

Knowledge of human biology remains rudimentary and disease genes are discovered in the population rather than via well-studied laboratory model systems; for these reasons, it is not surprising that the functions of many human disease genes remain poorly understood. However, dogged pursuit by biologists has led to increasing knowledge of the biological processes of many human disease genes and the development of new laboratory systems in which to study them. The search for genetically inspired medicines starts when a gene is demonstrated to have a substantial impact on human disease, a sufficient level of understanding has been obtained and tools to enable laboratory research become available.

3. Developing medicines to prevent the development and alter the course of severe genetic diseases

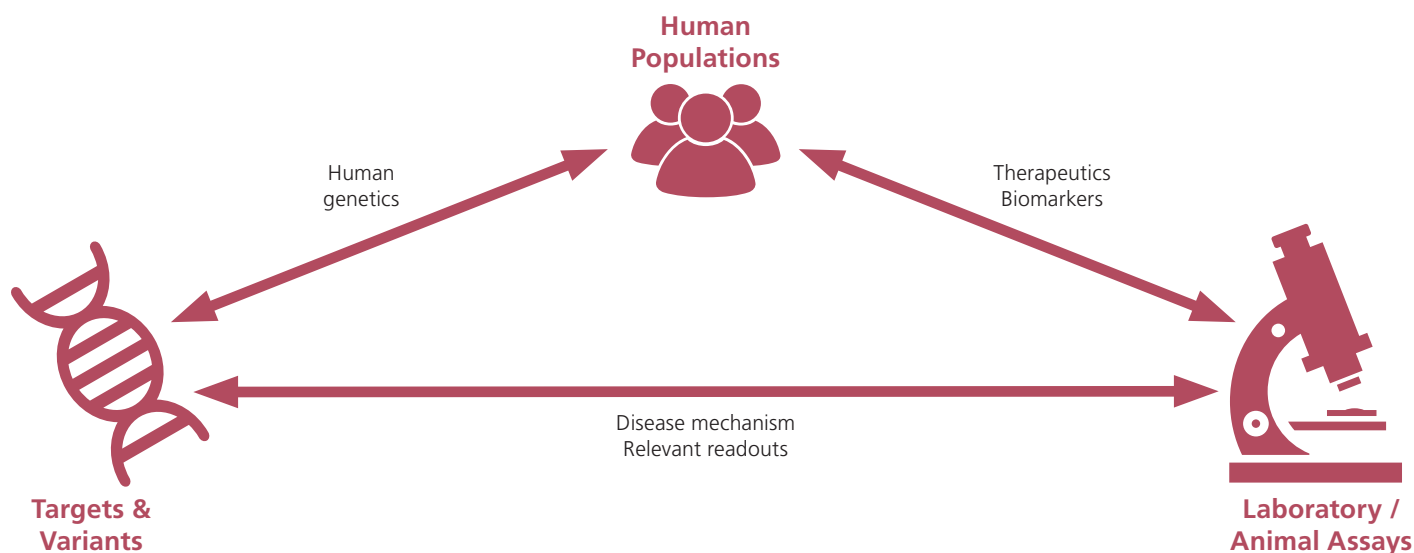
A 'human genetics first' approach contrasts with what has become the more typical pharmaceutical approach. Pharmaceutical companies tend to focus on a small subset of human proteins known as 'druggable targets'.² The definition of the term druggable target is somewhat circular: a protein that can be approached with the technologies for drug discovery available at any given time. At present, fewer than 5% of the 20,000 or more human genes are said to be druggable — if taken literally, this means that 95% of human genes cannot be addressed regardless of their importance to human disease.

Moreover, many of the efforts to address druggable targets have been focused on hypotheses obtained in cell and animal models. While such models have great utility, it is often impossible to know in advance whether any given model truly applies to patients. In fact, a frequent approach to test the disease relevance of a hypothesis generated in an animal model has been to develop a candidate medicine and test it in clinical trials. A focus on druggable targets and laboratory models made it possible to fill pipelines with many clinical candidates that served as 'shots on goal' for clinical trials. Unfortunately, it has become clear that too often these shots on goal miss their ultimate target, which is transformative benefit to patients.³ Recent analyses argue that the most important factor in determining the productivity of pharmaceutical research and development (R&D) is the probability that the therapeutic hypothesis is valid in patients, because failure in late-stage trials consumes such a large proportion of overall R&D spending.^{4,5}

An approach based on human genetics¹ puts the patient, rather than a laboratory model, at the leading edge of discovery (see Figure 1 below).⁶ A greater certainty about the role of the target in disease has been argued to lower the risk of failing to translate from the laboratory to the clinic.⁷ Challenges do not disappear, of course, but rather manifest in different aspects of the drug discovery process. As noted, most genes in the human genome are not classically druggable, and many disease genes are thus not approachable with 'off the shelf' technology. Rather, inventing a therapy for a genetic disease requires innovation and risk-taking in the technology of inventing therapeutics. Many of the mutations that cause genetic disease are individually rare and therefore the usual approach to clinical trials of studying large groups of people with a given disease is not suited to testing genetically targeted medicines. Finally, the inherited processes that trigger disease may act early in life, such that therapy may not be fully effective if applied later in the course of disease. This means that therapies for genetic diseases aimed at preventing the manifestations of disease may be most effective if therapy is begun early in the disease process.

However, realising this vision will require shifts in the type of challenge undertaken by academia, industry, regulators, healthcare providers and payers.

Figure 1 Genomics research – a patient-centred model



Source Vertex

4. Challenge 1 - developing innovative therapeutic modalities

Medicines that target the genetic causes of human disease have conceptual appeal: tackling some of the most severe diseases of childhood; addressing the underlying cause, rather than downstream symptoms; targeting at-risk individuals; and the potential to intervene before permanent damage occurs. Moreover, a focus on human biology, rather than the biology of artificial laboratory models, may reduce the rate of late-stage failures in pharmaceutical industry R&D caused by 'failure of translation' from the laboratory to the clinic.

For academia, it is critical that the effort be undertaken to understand the functions of human disease genes and reduce them to disease-relevant laboratory models that can support drug discovery. For industry, few of the best targets will be so-called druggable with conventional small-molecule chemistry and biologics, demanding innovation in the science of therapeutics. This will include new ways of using chemistry (as in the case of CFTR), as well as the development of nucleic acid therapies such as gene therapy, RNA interference, mRNA therapy and gene editing. For nucleic acid therapies, the challenge is not only in the DNA or RNA therapeutic, but also in delivering that agent to the particular cell and organs relevant to disease. Nucleic acid approaches have the potential to intervene in a manner that may be durable over longer periods of time.

Each of these therapeutic innovations will require substantial investment. It took two decades and significant investment from the discovery of monoclonal antibodies until antibody-based therapeutics became successful, and gene therapy has been studied for a longer period of time. This is because each new therapeutic modality requires that those who develop it not only solve scientific problems, but also pioneer manufacturing, delivery, safety testing, and every other step needed to demonstrate safety and efficacy. Without the ability to rely on the certainty of established technology and precedent, innovators must take on tremendous risk. Moreover, once the trail is blazed, competitors will rapidly follow, making the reward at times ephemeral.

Innovation in therapeutics also poses challenges to regulators, as each new approach raises unanticipated questions throughout the drug discovery process. Application of regulatory standards developed in a different context have the potential to slow progress to the extent that they imperfectly address the specific risks and benefits of a new approach.

5. Challenge 2 – developing preventative therapies

Developing preventative therapies for severe, early onset diseases will be particularly challenging because they will require demonstrating benefit in younger people who have not yet developed significant disease manifestations. In cystic fibrosis (CF), cystic fibrosis transmembrane regulator (CFTR) modulation demonstrated safety and efficacy in individuals with established disease. But in other genetic diseases, a medicine may not be effective once the disease manifestations are well established. In CF, the community organised patients into registries and established natural history for each class of mutations, enabling targeted clinical studies. In other diseases, where the community is not as well organised and natural history is less clear, it will be important to design and perform such studies. In CF, well-identified specific mutations such as F508del and G551D were targeted when performing the initial trials. In other diseases it may be that all mutations are individually rare, such that it will be necessary to develop methods to group mutations for study rather than treating each individually.

Despite the great promise of genetics for patients and their families suffering from these diseases, failure to address these barriers could result in long delays in providing treatments, or even limit the investment needed to develop treatments.

Steps to bridge this gap include developing surrogate markers that can be used to monitor and demonstrate the value of therapy, ways to group rare mutations based on strong understanding of the molecular defect and mechanism of action to enable clinical studies and regulatory approvals, and establishment of rare disease registries with data on genotype and natural history.

Markers that faithfully reproduce the pathophysiology of disease can be used to obtain early readout of potential benefit and to monitor therapy. Regulatory bodies, including the US Food and Drug Administration (FDA) and European Medicines Agency (EMA) have established pathways allowing for expedited, conditional or adaptive reviews that enable provisional approval based on surrogate markers and novel endpoints, with more comprehensive clinical data, as part of confirmatory trials, to follow. When combined with a strong safety database to minimise the potential for harm, such paths may offer an approach to balance the benefit of early treatment with the need to ultimately demonstrate benefit on patient outcomes.

Grouping of mutations will be key for those diseases caused by a large collection of heterogeneous mutations, each of which is individually rare. Given the small number of patients with each mutation, it will be impractical to perform trials involving each in isolation, requiring approaches to aggregate them for study and regulatory approval. A principled approach will require a strong understanding of molecular mechanisms, genotype-phenotype correlation in the population, and predictive laboratory measures to demonstrate drug responsiveness of the mutation in human disease relevant models.

Registries of patients with rare genetic diseases can be enabling for patient care, research and clinical development. Specifically, if the disease is rare and if many patients have a different genotype, then it can be challenging to understand the natural history of disease, obtain materials and data for research, and enrol clinical trials. Natural history studies that define genotype-phenotype relationships and enable comparative studies on disease modification can be invaluable. The model of CF, in which patients are cared for in specialised centres that also perform clinical research, has been enabling and should be replicated elsewhere.

6. Challenge 3 – developing a commercial model to support R&D

A high rate of failure in late-stage clinical trials represents a major hurdle to pharmaceutical innovation and healthcare improvement. Some of this failure can be attributed to the lack of human validation for many targets and the limits of the existing catalogue of druggable targets. In this regard, a turn to human validated genetic targets and to therapeutic innovation has potential to improve outputs for industry and healthcare alike.

Nonetheless, navigating this period of innovation will require substantial and risky investments to address novel targets, investments in new technologies of drug discovery, and the pioneering of new models for drug development and regulatory approval. Undertaking those projects requires that rewards be available that provide sufficient incentives for innovators to take risks, and that regulators and healthcare systems work together with innovators to establish new and effective paradigms of collaboration.

In the case of severe genetic diseases, the potential impact may be substantial for those affected, but the number of people over whom the investment can be amortised is few. This requires different pricing models than in cases of large disease areas in which the benefits may be less per person, but can be amortised over many people. In rare diseases, the calculus should take into account the magnitude of the benefit to the patients treated, number of patients who can benefit, investment and risk undertaken, and the investment needed to advance and extend therapies to additional patients who have the same disease but carry different mutations or were treated earlier in life.

Ironically, the potential for long-term and durable treatment with nucleic acid therapies creates further uncertainty regarding commercial models as compared to treatments that need be given chronically. When the disease is rare and the treatment potentially curative, new business models may be needed to provide sufficient incentive to create the technologies and treatments needed, while protecting on the downside if treatment proves less durable than hoped.

7. Challenge 4 – ensuring healthcare systems are fit for purpose

The potential for the role of genetic medicines in preventing severe diseases from manifesting presents a fundamental challenge for healthcare systems and policy-makers.

To the extent that treatment may be started early to prevent disease manifestations before they occur, traditional outcome measures that assess clinical and associated cost effectiveness based on short-term clinical trials of medicines prior to license may no longer be applicable. The burden of proof for clinical effectiveness may require consideration of the potential consequences to patients of not intervening, requiring a fundamental shift in decision making on how resources are allocated to healthcare systems. Possible solutions could include novel managed access agreements for licensed indications in which outcomes are measured on lack of disease progression from an early age, as opposed to symptomatic improvement in patients with established disease.

The long-term healthcare systems resource implications for the development of precision medicines are also significant. This is not because of the cost of the medicines themselves, as although they will likely have an initial high acquisition cost to reflect the commercial investment and low patient numbers, prices will nevertheless drop significantly over the medium term as market exclusivity is lost and competitors develop comparable products. Rather, the establishment of precision medicines to prophylactically treat genetic diseases will increase the burden on healthcare systems as previously untreatable populations succumb to common disorders associated with older age.

8. Conclusions

The vision for precision medicine is one in which genetics leads to a deeper understanding of the root causes of human disease, new therapies are developed that are highly effective and targeted to those in need, and ultimately where knowledge of genetic risk can be used to prevent onset of severe diseases. As we can see, this vision is coming to fruition in specific cases and, if the pace of innovation continues, we can expect more successes in the years to come.

We can also see that the traditional approaches used in drug discovery will be tested by this new paradigm, and innovation is needed across the spectrum of inventing, testing, approving and reimbursing new therapies for these devastating disorders.

Specifically, scientists need to redouble efforts to understand the functions of genes that cause human diseases and translate them into laboratory systems suitable to support drug discovery. Biopharmaceutical companies need to invent new ways to drug the undruggable, pioneering new approaches such as small-molecule chemistry, proteins (biologics) and nucleic acids. Patients and caregivers, together with their healthcare professionals, need to establish registries and care centres that define best practice in clinical care, enable research and define genotype-phenotype correlation and natural history. Regulators need to work with innovators to tailor pathways appropriate to the characteristics of these new therapeutics, qualify markers as surrogate endpoints and develop methods to aggregate mutations for clinical study and regulatory approval. Healthcare systems and payers need to appropriately evaluate and value the impact on families and society of preventing a relentless genetic disease before severe symptom onset. The entire system will have to evolve and collaborate more closely if we are to realise the potential of this transformative opportunity.

It would be a significant loss if the availability of technologies needed to develop transformative treatments for severe human diseases exist, and yet together we lack the ingenuity and capacity to develop the regulatory and business models needed to provide them to patients in a socially responsible and effective manner.

9. Suggestions for policy makers

The development of precision medicines that target the underlying genetic causes of disease are one of the most exciting areas of life sciences, are subject to significant investment from both the private and public sector, and will change the way disease is treated. The implications for healthcare systems are immense but largely unquantified despite or because of the investment in the science. The chapter author suggests the following occur:

- A multi-disciplinary committee established within the DH to review the broad implications of precision medicines for the NHS on an ongoing basis that can make recommendations across government and the NHS to support changes in policy that are required to ensure the expedited access of precision medicines to patients.
- NICE, DH and MHRA consider new models of criteria for the assessment of medicines that are able to assess efficacy, safety and economic value to the NHS from a baseline perspective of disease not progressing and / or mitigated over a longer-term than available clinical trial data at marketing authorisation.
- The development of precision and genetic medicine research continues to be prioritised by policy-makers throughout forthcoming “Brexit” negotiations; and that long-term funding commitments are provided to public sector initiatives supporting the fundamental challenges in the biological understanding of genes, development of surrogate markers and natural history of disease.

Case Study Cystic Fibrosis

Cystic fibrosis (CF) is a severe inherited disease with median age at death in the United Kingdom of 28 years.⁸ Worldwide, approximately 75,000 people suffer from CF.⁹ The disease was named (in 1938) for its anatomic and pathological features of mucus plugging leading to dilatation and scarring of ducts (“cystic fibrosis”) in the pancreas and other organs.¹⁰ CF runs in families, with an autosomal recessive pattern of inheritance. It was recognised early on that patients with CF had thick, sticky mucus and abnormalities in their handling of salt.¹¹ However, for decades, the underlying cause remained elusive.

In 1989, an international team identified the gene responsible for CF,^{12,13} pioneering many of the scientific methods for gene discovery that have now become standard in human genetic analysis. People with CF were found to have inherited mutations in a gene termed CFTR (cystic fibrosis transmembrane regulator). The gene was previously undiscovered, and initially its function was unknown. It was soon discovered that the normal function of the CFTR protein is to transport chloride and bicarbonate ions across the membranes of epithelial cells¹⁴ in multiple organs including the lungs, GI tract and sweat glands.

While a majority of CF patients carry a single, relatively common mutation in CFTR (F508del), a long list of individually rare mutations were then identified and characterised. These mutations were categorised according to the functional effect on the encoded protein¹⁵ — some cause a dysfunctional protein to be delivered to the apical cell surface, while others cause defects in the processing and trafficking of the protein (so the protein never makes it to the cell surface). Other mutations truncate the protein so no active CFTR is produced. Natural history studies were performed and demonstrated a strong relationship between the severity of CFTR dysfunction and disease phenotype. These data provided guidance on how much CFTR function would be needed to provide clinical benefit.

These discoveries uncovered the pathology of the disease: inherited defects in the CFTR protein lead to defects in chloride ion transport. The defect in chloride ion transport leads to dehydration of secretions of various organs. Dehydration of mucus causes plugging resulting in organ damage and, ultimately, the clinical features that lead to repeated hospitalisations and death. Progress was made in treatment by supportive measures (e.g. oxygen, pancreatic enzymes and nutrition), approaches to thin and clear the mucus, and antibiotics to treat infections. Recurrent pulmonary infections, together with chronic airway inflammation and bronchiectasis leading to progressive decline in lung function^[16], emerged as the primary cause of morbidity and mortality in CF.^{17,18} However, the inherited defect in chloride ion transport remained unaddressed.

In the course of understanding the CFTR gene and its cellular functions, laboratory systems were developed to study the function of the CFTR protein.¹⁹ These assays made it possible to attempt a new approach to CF therapy based on correcting the defect in the proteins encoded by the CFTR gene and restoring chloride ion transport. This approach was high-risk (unlikely to succeed) because CFTR was not considered to be a “druggable” protein and there was no precedent for developing a small-molecule, orally-available medicine capable of restoring the function of a mutant protein. It was thought by many at the time that only a gene-based therapy could address the underlying cause of such a genetic disease.

Through more than 15 years of work, scientists at Vertex Pharmaceuticals deployed new types of laboratory assays and chemistry to invent small-molecule drugs that address mutations in the CFTR protein. This includes one medicine to address CFTR protein that makes it to the cell surface, but does not transport chloride ions normally,²⁰ as well as a second class of medicines to address mutations in which the CFTR protein fails to reach the cell surface.²¹ To date, Vertex has discovered, invented and placed into clinical development eight different medicines, spanning three different mechanisms of action, to treat CF. The clinical evaluation of these medicines was enabled by decades of work by the CF community to develop registries for patients with CF, define genotype-phenotype relationships and establish a natural history of disease, including relationships among multiple measures of CFTR function in patients (e.g. sweat chloride, FEV1, nasal potential differences and clinical outcomes).²² Vertex collaborated with the CF community and clinical care centres to perform randomised clinical trials evaluating the efficacy and safety of these candidate medicines in patients with CF.²³⁻²⁸

The first medicine, called ivacaftor (Kalydeco®), is now approved in the U.S.²⁹ for people aged 2 years and older who carry one of 10 specific mutations in which the protein encoded by the *CFTR* gene is produced at the cell surface but does not transport chloride ions normally. Ivacaftor is also approved in the European Union for people aged 6 years and older who have nine of these *CFTR* mutations and in people aged 18 years and older with a specific *R117H* mutation.³⁰ A second drug, called lumacaftor, targets the more common F508del mutation in which the protein is processed abnormally by the cell and does not make it to the cell surface. The combination of lumacaftor and ivacaftor (Orkambi®), is now licensed in the U.S.,³¹ European Union,³² Canada, and Australia for patients aged 12 years and older carrying two copies of *F508del*. A third candidate medicine, tezacaftor, recently completed Phase 3 studies in patients carrying two copies of *F508del*, and also in patients carrying one copy of *F508del* and a second mutation with residual function. Four additional candidate medicines are being studied in Phase 1 and Phase 2 studies as triple combination therapies with tezacaftor and ivacaftor in

patients with one copy of *F508del* and a second allele with minimal function. As a class, these medicines are collectively referred to as CFTR modulators because they modulate the function of the CFTR protein.

Prior to these studies, it was not known whether treatment with a CFTR modulator would have clinically significant effects in people with CF who already had organ damage as a result of their disease. Clinical trials in patients with established CF showed that CFTR modulation does result in improvements in measures of lung function, nutritional state and quality of life, as well as reductions of pulmonary exacerbations requiring hospitalisation and use of antibiotics.^{27,28} When compared to observational data from propensity-score matched control patients in the US CFF Patient Registry who did not receive CFTR modulation, longer-term clinical trial data with both Kalydeco and Orkambi show a slowing in the rate of lung function decline over 2 to 3 years of treatment, respectively.³³⁻³⁵

Despite this significant progress, much work remains and is being done to advance care in CF. Today, only approximately one third of patients with CF are eligible for treatment with a CFTR modulator, and studies are ongoing to evaluate Kalydeco and Orkambi in younger patients and those with additional genotypes.³⁶ In younger patients and those with less-established disease manifestations, early treatment has the potential to prevent decline in lung function.³⁷ Improved methods are needed to demonstrate improvements or delay in decline in asymptomatic individuals. Many of the remaining CF-causing mutations are individually infrequent, and it is challenging to obtain the laboratory and clinical data needed to evaluate medicines for mutations that may only be known in a handful of people throughout the world. Based on genotype-phenotype correlation, higher levels of CFTR correction may enhance clinical benefit, and new agents are being developed to evaluate this possibility. Some patients carry mutations that do not produce any CFTR protein, meaning that innovative nucleic acid-based therapies may be needed to help these individuals.

More than 25 years after the CFTR gene was identified, we have a much clearer understanding of the biology underlying the disease, genetic tests that can categorise patients according to their specific mutation, and treatments that modulate (in a mutation-specific manner) the defect in chloride ion transport that causes the disease. Having crossed this key threshold, next steps include further definition of the patients who benefit from existing therapies, next-generation treatments based on the same principles of targeting the underlying cause of disease, and new types of treatments such as gene editing, particularly for those whose mutation cannot be treated by chemical modulators of CFTR.

10 References

- Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322(5903):881-888.
- Imming P, Sinning C, Meyer A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov*. 2006;5(10):821-834.
- Bunnage ME. Getting pharmaceutical R&D back on target. *Nat Chem Biol*. 2011;7(6):335-339.
- Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9(3):203-214.
- Scannell JW, Bosley J. When quality beats quantity: decision theory, drug discovery, and the reproducibility crisis. *PLoS One*. 2016;11(2):e0147215.
- Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*. 2013;12(8):581-594.
- Cook D, Brown D, Alexander R, et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov*. 2014;13(6):419-431.
- UK Cystic Fibrosis Registry. 2014 Annual Data Report. 2014; <https://www.cysticfibrosis.org.uk/registryreports>. Accessed July 10, 2016.
- Sosnay PR, Siklosi KR, Van Goor F, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet*. 2013;45(10):1160-1167.
- Andersen DH. Cystic fibrosis of the pancreas and its relation to celiac disease; a clinical and pathological study. *Am J Dis Child*. 1938;56:344-399.
- Quinton PM. Chloride impermeability in cystic fibrosis. *Nature*. 1983;301(5899):421-422.
- Kerem B, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989;245(4922):1073-1080.
- Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*. 1989;245(4922):1066-1073.
- Anderson MP, Gregory RJ, Thompson S, et al. Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. *Science*. 1991;253(5016):202-205.
- Welsh MJ, Smith AE. Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell*. 1993;73(7):1251-1254.
- Ratjen F, Tullis E. Chapter 46 - Cystic Fibrosis. In: Albert RK, Spiro S, Jett JR, eds. *Clinical Respiratory Medicine*. 3rd ed. Philadelphia: Mosby; 2008.
- Liou TG, Adler FR, Fitzsimmons SC, Cahill BC, Hibbs JR, Marshall BC. Predictive 5-year survivorship model of cystic fibrosis. *Am J Epidemiol*. 2001;153(4):345-352.
- MacDonald KD, McKenzie KR, Zeitlin PL. Cystic fibrosis transmembrane regulator protein mutations: 'class' opportunity for novel drug innovation. *Paediatr Drugs*. 2007;9(1):1-10.
- Van Goor F, Straley KS, Cao D, et al. Rescue of DeltaF508-CFTR trafficking and gating in human cystic fibrosis airway primary cultures by small molecules. *Am J Physiol Lung Cell Mol Physiol*. 2006;290(6):L1117-1130.
- Van Goor F, Hadida S, Grootenhuis PD, et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc Natl Acad Sci U S A*. 2009;106(44):18825-18830.
- Van Goor F, Hadida S, Grootenhuis PD, et al. Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809. *Proc Natl Acad Sci U S A*. 2011;108(46):18843-18848.
- McKone EF, Emerson SS, Edwards KL, Aitken ML. Effect of genotype on phenotype and mortality in cystic fibrosis: a retrospective cohort study. *Lancet*. 2003;361(9370):1671-1676.
- Davies JC, Cunningham S, Harris WT, et al. Safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2-5 years with cystic fibrosis and a CFTR gating mutation (KIWI): an open-label, single-arm study. *Lancet Respir Med*. 2016;4(2):107-115.
- Davies JC, Wainwright CE, Canny GJ, et al. Efficacy and safety of ivacaftor in patients aged 6 to 11 years with cystic fibrosis with a G551D mutation. *Am J Respir Crit Care Med*. 2013;187(11):1219-1225.
- De Boeck K, Munck A, Walker S, et al. Efficacy and safety of ivacaftor in patients with cystic fibrosis and a non-G551D gating mutation. *J Cyst Fibros*. 2014;13(6):674-680.
- Moss RB, Flume PA, Elborn JS, et al. Efficacy and safety of ivacaftor in patients with cystic fibrosis who have an Arg117His-CFTR mutation: a double-blind, randomised controlled trial. *Lancet Respir Med*. 2015;3(7):524-533.
- Ramsey BW, Davies J, McElvaney NG, et al. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *N Engl J Med*. 2011;365(18):1663-1672.
- Wainwright CE, Elborn JS, Ramsey BW. Lumacaftor-Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR. *N Engl J Med*. 2015;373(18):1783-1784.
- Ivacaftor [package insert]. Boston, MA: Vertex Pharmaceuticals Incorporated; 2015.

30. European Medicines Agency. Kalydeco summary of product characteristics. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002494/WC500130696.pdf. Accessed July 13, 2016.
31. Orkambi [package insert]. Boston, MA: Vertex Pharmaceuticals Incorporated; 2015.
32. European Medicines Agency. Orkambi summary of product characteristics. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/003954/WC500197611.pdf. Accessed July 13, 2016.
33. Konstan MW, McKone E, Moss RB, et al. Evidence for reduced rate of lung function decline and sustained benefit with combination lumacaftor and ivacaftor (LUM/IVA) therapy in patients (pts) 12 years of age With cystic fibrosis (CF) homozygous for the F508del-CFTR mutation. Paper presented at: ECRD; May 26-28, 2016; Edinburgh, Scotland.
34. McKone EF, Borowitz D, Drevinek P, et al. Long-term safety and efficacy of ivacaftor in patients with cystic fibrosis who have the Gly551Asp-CFTR mutation: a phase 3, open-label extension study (PERSIST). *Lancet Respir Med*. 2014;2(11):902-910.
35. Sawicki GS, McKone EF, Pasta DJ, et al. Sustained benefit from ivacaftor demonstrated by combining clinical trial and cystic fibrosis patient registry data. *Am J Respir Crit Care Med*. 2015;192(7):836-842.
36. <https://clinicaltrials.gov/ct2/results?term=VX661&Search=Search>. Accessed July 12, 2016.
37. Galletta LJ. Managing the underlying cause of cystic fibrosis: a future role for potentiators and correctors. *Paediatr Drugs*. 2013;15(5):393-402.

Chapter 5

Cancer diagnosis

Chapter lead and author
Dr. Ultan McDermott¹

¹ CDF Group Leader, Wellcome Trust Sanger Institute

1. Summary

Cancer: a disease of the genome

The ability to undertake large-scale next-generation sequencing across many thousands of cancers has identified the genes that when mutated in cells eventually lead to the development of a cancer.

A turning point in cancer research: sequencing the human genome

The completion of the human genome project in 2003 provided a largely complete genetic map of a human genome – without this many of the subsequent efforts to understand the causes of human cancers would have been impossible.

The technology revolution – next-generation sequencing

The ability to identify changes in the genome, transcriptome and epigenome of cancers has not only revolutionised our understanding of what makes many cancers 'tick' but more significantly the falling costs now mean that for the first time we are seeing such technology being advocated for routine use in the clinic for treatment selection.

Towards personalised cancer medicine and avoiding bankrupting the system

There are two self-evident truths in cancer treatment today – most drugs only benefit a small number of patients in any given cancer type, and the cost of new drugs is often very high. Using next-generation sequencing to identify those patients most likely to benefit from any specific treatment is thus becoming an imperative.

Cancer genomics, pharma and tomorrow's drugs

The identification of which patients will respond to a drug using these new technologies will require a significant rethink in how clinical trials are designed and executed – the old model of a 'one size fits all' for cancer drugs is no longer tenable.

2. Background

Arguably the seminal observation of a causative genetic component in cancer was made over a century ago when Theodor Boveri speculated that the malignant tumours might be the consequence of chromosomal derangements.¹ Since then incontrovertible evidence for the role of mutations in the human genome causing cancer has steadily accumulated from studies in disciplines including epidemiology, genetics, animal models and molecular biology. Key to these discoveries was the launch in 1990 of the Human Genome Project, followed by a number of large-scale international efforts (such as the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA)) to characterise the genomes of many of the commonest cancers.² Thus, today we know that all cancers arise due to the acquisition of somatic mutations in their genomes, which fundamentally alter the function of key cancer genes.³ Such mutations are responsible not only for the development of the cancer in the first instance but also in maintaining the proliferation status and evasion of cell death that are hallmarks of cancer.⁴ Increasingly, a number of these mutations are implicated in the likelihood of survival and treatment response in the clinic.⁵⁻⁷

Perhaps as transformative as these discoveries has been the development of next-generation sequencing (NGS) technologies. Initially, their role was to enable large-scale human and cancer genome research projects, but increasingly they are being used to replace or augment existing clinical assays and provide clinically useful prognostic or predictive information. The introduction of such technology, and its potential for major cost savings in the better stratification of patients for optimal therapy and the avoidance of futile treatments, will in the short term be disruptive - existing diagnostic disciplines in the healthcare setting will have to adapt to new regulatory requirements for NGS and clinical teams will race to build new mechanisms to allow such information to be incorporated into the daily decision-making process for cancer patients. Issues such as cost, availability, turnaround time, data security, research access and harmonising/sharing of cancer genomic data for the public good will all need to be resolved, although efforts such as the Global Alliance for Genomics and Health (GA4GH) are beginning to create harmonised approaches in all of these areas.*

As important as all of these will be the requirement for a robust assessment of the health economics of the large-scale implementation of a cancer genome sequencing strategy into healthcare systems. I would make the argument that such technology has the potential for huge cost-savings in the treatment of cancer patients, by building knowledge banks of clinical outcome. Such databases, containing sequenced tumours linked to such clinical outcome would allow us to better identify those patients most likely to benefit (or not) from specific therapy. This would allow healthcare providers to avoid the often high costs (and wasted time for the patient) of offering treatment to patients who are unlikely to respond. This would greatly increase the cost-benefit ratio of therapeutic interventions for NHS cancer patients at a time when across Europe and the US financial constraints are increasing.

* For more information see <https://genomicsandhealth.org/>

3. Cancer: a disease of the genome

The discovery of NGS technologies offered a unique opportunity to galvanise the international scientific community and arguably embark on the most ambitious sequencing project ever imagined, namely to sequence the exomes/genomes of 25,000 tumours across 50 tissue types. As a result of these efforts, both the TCGA and ICGC have presented to the world the mutational landscape for a host of different tumour types.^{8,9} They have helped to define the cancer genes that are complicit in tumorigenesis for each tumour type as well as identify those regions of the chromosome that are structurally deranged.¹⁰ Although some individuals are born with mutations in particular genes ('germline variants') that strongly predispose them to developing cancer, in >90% of cancers are due to mutations in cancer genes that arise during that individual's lifetime ('somatic variants'), and as a consequence of factors such as aging, environmental exposures and lifestyle.

As a consequence of cancer sequencing studies it is now feasible to create for each tumour type a catalogue of those altered pathways and processes that may be amenable to targeting therapeutically. It is clear that in addition to alterations in classical signal transduction pathways (e.g. EGFR mutations in lung cancer, BRAF mutations in melanoma), large-scale sequencing studies have revealed driver mutations in cellular processes as diverse as metabolism (IDH1 and IDH2 in AML and glioblastomas), histone modifications (MLL2, EZH2, UTX, KDM5A, KDM5C and CREBP), splicing (SF3B1, U2AF1, SRSF2), chromatin SWI/SNF complexes (SMARCA1, SMARCA4, ARID1A, ARID1B, PBRM1) and apoptosis (MCL1, BCL2A1, BCL2L1). Many of these biological processes are now the focus of intense activity in the pharmaceutical industry as potential novel targets in cancer, and have enabled partnerships between industry and academia to better realise the potential of such novel targets. In the UK, the Open Targets initiative has been developed as an open innovation public-private partnership between pharma (GSK, Biogen) and academia (the European Bioinformatics Institute and the Wellcome Trust Sanger Institute) to use sequencing to identify and prioritise new drug targets.

4. A turning point in cancer research: sequencing the human genome

Although today we take for granted the importance of genome sequencing in helping understand the processes that driver cancer, in the past there was debate as to the benefits of systematic sequencing studies and how they might detract from more focused scientific research.¹¹ Others argued that completing the sequence of the human genome would provide essential information for discovering the genes that underpin tumourigenesis.¹² The launch of the Human Genome Project in 1990 and its delivery of a near-complete sequence by 2003 marked the beginning of a series of international cancer genome sequencing studies that previously would have been impossible.¹³ It would have been almost impossible to understand the changes that make up an abnormal cancer genome without first understanding what a normal human genome is. For all such cancer studies, the normal human genome has acted as an essential foil. It is now abundantly clear that these studies have fundamentally changed how we view cancer biology, opening up new and exciting lines of enquiry into a diverse range of cellular processes and treatment strategies for patients.

5. The technology revolution – next-generation sequencing

The earliest studies of cancer genomes used a technology first developed in 1977 in Cambridge by Frederick Sanger and still used today.¹⁴ It consisted of capillary-based sequencing and each exon had to be laboriously amplified and sequenced individually in a time-consuming and expensive process. Copy number alterations were detected using probe-based arrays and chromosomal rearrangements could not be systematically identified at all. With the advent of massively parallel sequencing that all changed dramatically. This process involved generating and capturing millions of DNA molecules from a sample on a surface and then sequencing them simultaneously (thus the expression 'massively parallel').¹⁵ This enabled billions of bases to be sequenced in a single run and subsequent advances enabled specific regions of the genome (for example, the exons of all coding genes or the 'exome') to be captured using specific baits and thus sequenced more efficiently.¹⁶ These advances transformed our ability to sequence exomes and genomes in ever greater numbers and to begin to define the unique repertoire of mutational events that underpin each tumour type. These same technologies also made feasible for the first time the large-scale analysis of cancer transcriptomes and epigenomes, and thus ultimately all of the most important breakthroughs in cancer genomics since then.^{17,18}

6. Towards personalised cancer medicine and avoiding bankrupting the system

Some of the most compelling insights gained from cancer genome sequencing have been in the understanding that any given cancer type is actually composed of a number of subtypes, often with different behaviours, survival rates and likelihood of responding to treatment. Thus, layering NGS information on top of classical pathological information is changing how we manage patients, improving survival - this will only increase as we learn more of the previously hidden world of cancer biology. Broadly speaking, the molecular markers in use today can be divided into those that are diagnostic (aid in the diagnosis or sub classification of a cancer), prognostic (have an association with clinical outcome independent of treatment) or predictive (predict the likelihood of response to a specific type of therapy). Today there are already a number of markers used in routine practice that fall into each of these categories (Figure 1).¹⁹ Of note, almost all of these individual assays could be generated by a single NGS test designed to capture coding exons for mutations, relevant heterozygous SNPs for copy number events, intronic regions involved in gene fusions and even methylated promoters using bisulphite sequencing protocols. Thus the concept of one test-one readout is increasingly obsolete – NGS assays will allow a single test to detect a large range of events, saving time, money and precious tissue material but at the expense of requiring increasingly complex informatics requirements.

Figure 1 – Molecular markers of proven clinical utility in cancer

Colorectal cancer	
KRAS mutations	PCR, DNA seq
NRAS mutations	PCR, DNA seq
BRAF p.V600E mutation	PCR, DNA seq
MSI, MMR protein loss	PCR, IHC
CEA	Immunoassay

Breast cancer	
ER/PR	IHC
ERBB2 (HER2)	FISH, IHC

Non-small cell lung cancer	
EGFR mutation	PCR, DNA seq
ALK gene fusion	FISH

Acute myeloid leukaemia	
FLT3-ITD	PCR, DNA seq
CEBPA mutation	PCR, DNA seq
NPM1 mutation	PCR, DNA seq
KIT mutation	PCR, DNA seq

Glioma	
1p/19q codeletion	FISH, aCGH
IDH1/2 mutation R132H	DNA seq
MGMT methylation	MS-PCR, MS-MPLA

Prostate cancer	
PSA	Immunoassay

Note

A selection of cancers are shown where specific molecular markers (left hand column) have been shown to be

- diagnostic (green shading),
- prognostic (blue shading) or
- predictive for therapy (red shading).

The type of assay required to detect each marker is indicated in the right hand column.

Where a marker satisfies multiple purposes, two colours are present.

For example, non-small cell lung cancer (one of the most prevalent and fatal cancers in men and women in the UK) is now additionally classified according to the mutation status of the genes EGFR and ALK, and where targeting those mutations with specific drugs improves outcome.^{6,7} Increasingly, the driver for this increased complexity in the classification of cancers is evidence that specific genomic alterations affect the likelihood of response to cancer therapeutics. This is reflected in the number of Federal Drug Administration (FDA) approved cancer drugs from 2001-2015 where a specific genomic alteration is

required to be detected for patient treatment and subsequent reimbursement (Table 1).^{5,7,20-37} Arguably this move towards personalised cancer medicine should, in the long term, be cost saving as finite cancer drug budgets are spent on those patients more likely to derive clinical benefit from a particular drug treatment. However, in the short term this has meant most pathology departments have had to rapidly upgrade existing molecular diagnostics (typically FISH, cytogenetics, Sanger sequencing for small numbers of genes) to include NGS for larger panels of genes.

Table 1 FDA-approved molecular markers for therapy stratification in cancer (2001 – 2015)

Tumour	Gene (mutation)	Prevalence gene alteration (%)	Drug	FDA approved	Therapeutic target	Response rate in mutant tumours (%)	Study
Chronic myeloid leukaemia	BCR-ABL (translocation)	>95	Imatinib	2001	ABL1	>95	BJ Druker, 2001
Gastrointestinal stromal tumour	KIT (mutation), PDGFRA (mutation)	85 (KIT), 5-8 (PDGFRA)	Imatinib	2002	KIT, PDGFRA	>80	J Verweij, 2004
Non-small cell lung cancer	EGFR (mutation)	10	Gefitinib, erlotinib	2003, 2004	EGFR	70	T Mok, 2009
Chronic myeloid leukaemia (imatinib-resistant)	BCR-ABL (translocation)	>95	Dasatanib	2006	ABL1	>90	M Talpaz, 2006
Breast cancer – node +ve	HER2 amplification	15-20	Trastuzumab	2006	ERBB2	HR 0.48	EA Perez, 2011
Colorectal cancer	KRAS and NRAS wild-type	48	Panitumumab	2006	EGFR	HR 0.74	JY Douillard, 2013
Melanoma	BRAF (mutation)	40-70	Vemurafenib	2011	BRAF	>50	PB Chapman, 2011
Non-small cell lung cancer	ALK rearrangement	2-7	Crizotinib	2011	ALK	57	EL Kwak, 2010
Non-small cell lung cancer	EGFR (mutation)	10	Afatinib	2013	EGFR, ERBB2	50	LV Sequist, 2013; JC Yang 2013
Melanoma	BRAF (mutation)	40-70	Debrafenib	2013	BRAF	52	A Hauschild, 2012
Melanoma	BRAF (mutation)	40-70	Trametinib	2013	MEK1	22	K Flaherty, 2012
Breast cancer (metastatic)	HER2 amplification	15-20	Trastuzumab	2013	ERBB2	44	S Verma, 2012
Melanoma	BRAF (mutation)	40-70	Debrafenib/ Trametinib	2014	BRAF, MEK1/2	64	C Robert, 2015
Ovarian cancer	Germline BRCA1/2	14	Olaparib	2014	BRCA1/2	34	B Kaufman, 2015
Breast cancer (metastatic)	ER+ve/HER-ve	30-40	Palbociclib/ Letrozole	2015	CDK4/6	55	RS Finn, 2015
Non-small cell lung cancer	PD-L1 expression	23-25	Pembrolizumab	2015	PD-L1	41	EB Garon, 2015
Melanoma	BRAF (mutation)	40-70	Vemurafenib/ Cobimetinib	2015	BRAF, MEK1/2	68	J Larkin, 2015
Non-small cell lung cancer	EGFR T790M (mutation)	50-60*	Osimertinib (AZD9291)	2015	EGFR T790M	57-61	P Janne, 2015
Non-small cell lung cancer	ALK rearrangement	All**	Alectinib	2015	ALK	38-44	AT Shaw, 2016; SI Ou, 2016

Abbreviations BRCA1/2 – Breast And Ovarian Cancer Susceptibility Protein 1 or 2; KIT - v-Kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homologue; PDGFRA - Platelet-derived growth factor receptor; alpha polypeptide. * of patients developing resistance to EGFR TKI therapy, ** any patient with ALK rearrangement and who had disease progression on ALK inhibitor Crizotinib.

7. The cancer testing laboratory of the future

As clinicians demand more and more information from ever decreasing amounts of patient material, the ability of NGS platforms to provide multiple readouts from a single sample will become essential. This has come at a time when many laboratories are working to transition from CPA (Clinical Pathology Accreditation) to the new ISO15189 standard, and where the use of NGS platforms will undoubtedly impose additional quality assurance burdens. The additional regulatory requirements that will be required from the informatics platforms used to analyse such data will very quickly overwhelm all but the most experienced molecular diagnostic laboratories. Furthermore, there is little doubt that this is just the beginning – recent publications point to the potential benefits of NGS from circulating cancer DNA (in real-time) to detect the emergence of drug resistance or to even supplant the need for diagnostic biopsies.^{38,39} The recent excitement over the durable responses seen in some cancer patients to immune checkpoint inhibitors is fuelling the development of new algorithms to detect neoantigens expressed by cancer cells or to tease out the contribution of different immune populations in a cancer biopsy – all of this from NGS data.⁴⁰

The molecular diagnostics laboratory of tomorrow will need to be able to implement complex NGS assays and create efficient informatics solutions for their analysis, all under an approved regulatory process. It will also need to keep abreast of the fast-moving field of novel diagnostic tests, which may run the gamut of DNA, mRNA, microRNA, lncRNA, methylDNA, and possibly in the near future proteomics. Arguably, a semi-centralised system of molecular diagnostics laboratories with shared SOPs for standardised NGS platforms would be the most cost effective way to deliver a genomics-based service for healthcare systems. No doubt some assays could indeed be contracted out to specialised commercial entities with expertise in particular fields e.g. plasma DNA sequencing and analysis. Although in theory such a model would allow these laboratories to harmonise their protocols for the sequencing itself, more challenging would be setting up computational pipelines in each centre with equivalent accuracy, sensitivity and specificity as well as the ability to upgrade all systems in a coordinated fashion as software improvements and upgrades come to light (be they security upgrades, technical upgrades or improved methods).

One possible solution is to have semi-centralised laboratories sequencing clinical material using the same NGS platforms, and for the NGS data analysis to be carried out in a cloud environment through either a public cloud, for example, Amazon Web Services, Microsoft or Google, or indeed a private cloud solely managed for the NHS (although this would require significant resource to maintain and develop). Such an environment would enable a relatively small team of expert users to provide the ongoing support and improvements in the analysis of what will be increasingly complex and large datasets from cancer patients. Indeed, some of this expertise is already being developed through the Genomics England initiative.

Finally, given the past experience within many healthcare systems of the vast expense that can be incurred through individual hospitals or consortia building their own bespoke solutions for their informatics requirements, this approach could ultimately be the most cost effective means to delivering cancer genomics data to hospitals. One can imagine a dedicated web portal, with secure access, whereby physicians could visualise the NGS results for their own patients from within the oncology clinic.

8. Cancer genomics, pharma and tomorrow's drugs

The old model of clinical trial design involved selecting patients based on a particular tumour type and with limited stratification based on small amounts of genomics information. Increasingly, regulatory bodies such as the FDA and health technology assessment bodies (e.g. The National Institute for Health and Care Excellence) require some molecular stratification to be included in any submission that allows enrichment for patients more likely to benefit from treatment. This focus on clinical utility as part of the drug evaluation process will only increase the importance of cancer genomics in the process of drug development and clinical trial enrolment for pharma. However, some of these molecular markers are only found in small numbers of patients, resulting in the scenario of many hundreds of patients requiring to be sequenced to identify a small handful eligible for treatment. To illustrate the challenge facing pharma, a recent review estimated that to run a study testing a new candidate drug in a patient subpopulation selected by a molecular marker with a 2% incidence one would need to screen 78 patients

for every one patient recruited to the study. A standard 20 patient Phase I expansion in this patient subpopulation would require screening at least 1560 patients in total to find these 20 patients and at a staggering cost of US\$1.8 million (assuming \$1000 per assay).⁴¹ This type of approach is clearly not sustainable. A more efficient process is to sequence each patient for 10-100's of mutations across a panel of genes, with multiple treatment arms available depending on which gene was found to be altered. This would result in more patients being eligible for clinical trials and also make better use of limited tissue material as compared to multiple rounds of sequential testing. Indeed, a number of such 'umbrella' trials are now underway across the globe where patients with a given type of cancer are assigned to a specific treatment arm based on the molecular makeup of their cancer. For example, a number of major initiatives that have been initiated in the US and UK will allow for a molecular stratification of advanced lung cancer and assignment to different treatment arms based on this stratification (Table 2).

Table 2 Examples of umbrella clinical trials in cancer

Study	Cancer Type	Stage of disease	No of genes tested	No treatment arms	Co-ordinating centre
LUNG-MAP	NSCLC - squamous	Metastatic	>200	5	LUNG-MAP, USA
MATRIX	NSCLC	Metastatic	28	7	Birmingham, UK
BATTLE	NSCLC	Metastatic	8	4	MD Anderson, USA
FOCUS4	Colorectal	Metastatic	6	4	Oxford, UK
ALCHEMIST	NSCLC - adenocarcinoma	Adjuvant	2	2	NCI, USA

Abbreviations NSCLC – non-small cell lung cancer.

The BATTLE clinical trial test lung cancer patients for a number of genes and then allocates the patient to one of four different treatments (Table 2). Also in the US, the Lung-MAP clinical trial allows stratification of relapsed Squamous Cell Lung Carcinoma patients into one of 7 different treatment arms. Here in the UK, Cancer Research UK, AstraZeneca and Pfizer have collaborated to initiate The National Lung Matrix Trial in NSCLC and that stratifies patients into one of 8 different treatment arms.[†] Such trials are complex and can lead to various arms being added or removed to the study as new drugs become available or as it becomes clear that a particular treatment is not effective. Nevertheless, they represent an exciting departure from the existing model of clinical trial design and patient stratification.

[†] For more information see <http://public.ukcrn.org.uk/search/StudyDetail.aspx?StudyID=17746>

As we move into an era when the \$1000 genome and the \$400 exome are fast becoming a reality, and set against the costs of designing and implementing a clinical trial (and never mind the costs of developing a drug in the first place), gathering additional information in this way on the genetic profile of the tumour will be seen to make financial common sense. The ability of the NHS to deliver a cancer NGS solution with harmonised ‘wet lab’ protocols aligned to a single cloud-based informatics solution for mutation detection would be particularly relevant to these problems and in theory would increase the attractiveness of the UK for clinical trial development. It is pertinent to note that if genomics does indeed identify novel targets in cancer and this leads to the development of drugs, the clinical trial designs described above will be essential to mitigate against the cost of offering these to our NHS cancer patients.

Case Study 1 – A colon cancer patient with inoperable liver metastases

Case Study

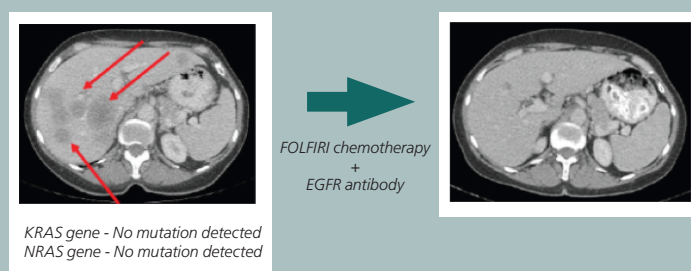
A 45-year old tree surgeon presented to his GP with abdominal pain and a palpable mass in his right groin. CT scan revealed a primary colon cancer in his caecum as well as inoperable liver secondaries (see Panel A, red arrows).

Previous studies had shown that the absences of mutations in the genes KRAS and NRAS in the tumour increased the likelihood of responding to a monoclonal antibody Cetuximab when added to standard chemotherapy. Sequencing of the these genes did not reveal any mutations and the patient was commenced on treatment with chemotherapy (FOLFIRI) combined with antibody (Cetuximab) for 3 months.

A CT scan after 3 months of treatment revealed a dramatic response – all but one of the secondaries was no longer visible. The response was so pronounced that the liver team agreed to undertake surgical resection of the single remaining secondary (see Panel B – green arrow). The patient remains alive and free of recurrence more than 2 years from surgery.

Source *TBC*

Panel A



Panel B



9. Conclusions

Next-generation sequencing has revolutionised the study of cancer genomes and dramatically increased our understanding of how cancers evolve, develop and what might be tomorrow's drug targets. Like any disruptive technology, NGS will create significant upheaval and new challenges for regulatory and ethical bodies, clinical trialists, pharmaceutical companies and clinicians. Nevertheless, the potential benefits from more efficiently stratifying patients for clinical trial enrolment to better understand which patient groups benefit most from a specific therapy are likely to be immense. In particular, at a time of increasing (and non-sustainable) healthcare expenditure the opportunity to use our treatment resources more cost-effectively should be pursued with vigour. This truly is an opportunity to move into a 'brave new world' of personalised cancer treatment and patient stratification in a way that we could not have imagined a decade ago. The UK is almost uniquely placed, with its single point-of-care healthcare system, extensive genomics expertise and strong history of clinical trials, to develop a unified platform and use genomics to transform clinical practice and clinical trials. However, particularly in the US and despite a more fragmented healthcare system, strenuous efforts are being made to capitalise on the technological advances outlined above to tackle some of the biggest questions about cancer. In particular, the US Cancer Moonshot aims to break down the barriers that keep researchers from sharing data by building central repositories that bring together cancer genomics and electronic medical records. In a similar vein, in June 2016 the NCI launched the Genomics Data Commons to promote cancer data sharing. Whether the UK becomes a partner in these types of endeavours or seeks to build its own infrastructure is an important question.

10. Suggestions for policy makers

Harmonise cancer sequencing in the NHS

Establish a pilot across 10-20 cancer centres to use the same sequencing platforms and molecular marker panels underpinned by standard SOPs. Get input from pharma as to their 'wish list' for clinical trial patients. Raw sequence data to be archived at each clinical site locally but analysis of data to be in a single cloud environment and using an agreed mutation detection pipeline (ensure data access agreements and consent to enable academic groups to use such data to try and improve the pipeline/build better algorithms). Leverage experience of Genome England in cloud solutions for large data analysis. Commercialise such data to allow pharma and biomedical industry to mine for drug targets, novel stratification groups etc. If successful, improve the model using lessons learnt and roll out across the NHS diagnostic laboratories in England.

Educating the next generation of cancer clinicians

At present most medical students spend typically 2-3 weeks learning about oncology despite cancer being the second leading cause of death in the UK. Add to that limited teaching on cancer genomics and you have a recipe for an entire generation of physicians completely unprepared for the explosion of molecular data in cancer patients over the next few years. We urgently need a greater emphasis on what tomorrow's clinician will be expected to deliver, the types of complex data that will be generated and the tools required to analyse such data to improve clinical outcomes and efficiency.

Build infrastructure for tomorrow's clinical trials

As clinical trials become more complex and require more genomics data for patient stratification, there is a real opportunity for the UK to become the best country in which to host such studies. Harmonising our cancer sequencing as described above would be a step in this direction. Investing in centralised informatics solutions for the analysis of NGS data would be another. Signalling to the relevant stakeholders (pharma, patient advocacy groups, biotech and informatics companies) that this is a priority for the UK would send a strong message about our commitment to maintaining our global position as a leader in running clinical trials.

11. References

1. Boveri, T., Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J Cell Sci*, 2008. 121 Suppl 1: p. 1-84.
2. International Cancer Genome, C., et al., International network of cancer genome projects. *Nature*, 2010. 464(7291): p. 993-8.
3. Stratton, M.R., P.J. Campbell, and P.A. Futreal, The cancer genome. *Nature*, 2009. 458(7239): p. 719-24.
4. Hanahan, D. and R.A. Weinberg, Hallmarks of cancer: the next generation. *Cell*, 2011. 144(5): p. 646-74.
5. Chapman, P.B., et al., Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*, 2011. 364(26): p. 2507-16.
6. Shaw, A.T., et al., Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med*, 2013. 368(25): p. 2385-94.
7. Mok, T.S., et al., Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med*, 2009. 361(10): p. 947-57.
8. Lawrence, M.S., et al., Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 2014. 505(7484): p. 495-501.
9. Tamborero, D., et al., Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep*, 2013. 3: p. 2650.
10. Zack, T.I., et al., Pan-cancer patterns of somatic copy number alteration. *Nat Genet*, 2013. 45(10): p. 1134-1140.
11. Weinberg, R., Point: Hypotheses first. *Nature*, 2010. 464(7289): p. 678.
12. Dulbecco, R., A turning point in cancer research: sequencing the human genome. *Science*, 1986. 231(4742): p. 1055-6.
13. International Human Genome Sequencing, C., Finishing the euchromatic sequence of the human genome. *Nature*, 2004. 431(7011): p. 931-45.
14. Sanger, F., S. Nicklen, and A.R. Coulson, DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 1977. 74(12): p. 5463-7.
15. Bentley, D.R., et al., Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 2008. 456(7218): p. 53-9.
16. Gnirke, A., et al., Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*, 2009. 27(2): p. 182-9.
17. Wang, Z., M. Gerstein, and M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009. 10(1): p. 57-63.
18. Krueger, F., et al., DNA methylome analysis using short bisulfite sequencing data. *Nat Methods*, 2012. 9(2): p. 145-51.
19. Febbo, P.G., et al., NCCN Task Force report: Evaluating the clinical utility of tumor markers in oncology. *J Natl Compr Canc Netw*, 2011. 9 Suppl 5: p. S1-32; quiz S33.
20. Druker, B.J., et al., Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med*, 2001. 344(14): p. 1031-7.
21. Verweij, J., et al., Progression-free survival in gastrointestinal stromal tumours with high-dose imatinib: randomised trial. *Lancet*, 2004. 364(9440): p. 1127-34.
22. Talpaz, M., et al., Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias. *N Engl J Med*, 2006. 354(24): p. 2531-41.
23. Perez, E.A., et al., Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *J Clin Oncol*, 2011. 29(25): p. 3366-73.
24. Kwak, E.L., et al., Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*, 2010. 363(18): p. 1693-703.
25. Douillard, J.Y., et al., Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med*, 2013. 369(11): p. 1023-34.
26. Sequist, L.V., et al., Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol*, 2013. 31(27): p. 3327-34.
27. Ou, S.H., et al., Alectinib in Crizotinib-Refractory ALK-Rearranged Non-Small-Cell Lung Cancer: A Phase II Global Study. *J Clin Oncol*, 2016. 34(7): p. 661-8.
28. Hauschild, A., et al., Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*, 2012. 380(9839): p. 358-65.
29. Flaherty, K.T., et al., Improved survival with MEK inhibition in BRAF-mutated melanoma. *N Engl J Med*, 2012. 367(2): p. 107-14.
30. Verma, S., et al., Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med*, 2012. 367(19): p. 1783-91.
31. Robert, C., et al., Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med*, 2015. 372(1): p. 30-9.
32. Kaufman, B., et al., Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. *J Clin Oncol*, 2015. 33(3): p. 244-50.
33. Finn, R.S., et al., The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. *Lancet Oncol*, 2015. 16(1): p. 25-35.

34. Garon, E.B., et al., Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med*, 2015. 372(21): p. 2018-28.
35. Larkin, J., et al., Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N Engl J Med*, 2014. 371(20): p. 1867-76.
36. Janne, P.A., et al., AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med*, 2015. 372(18): p. 1689-99.
37. Shaw, A.T., et al., Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. *Lancet Oncol*, 2016. 17(2): p. 234-42.
38. Thress, K.S., et al., Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med*, 2015. 21(6): p. 560-2.
39. Siravegna, G., et al., Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med*, 2015. 21(7): p. 827.
40. Van Allen, E.M., et al., Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*, 2015. 350(6257): p. 207-11.
41. Hollingsworth, S.J., Precision medicine in oncology drug development: a pharma perspective. *Drug Discov Today*, 2015. 20(12): p. 1455-63.

Chapter 6

Rare diseases

Chapter lead

Prof Patrick Chinnery¹

Chapter authors

Prof Patrick Chinnery¹, Mr Alastair Kent², Prof John Bradley³

1 Cambridge University Hospitals NHS Trust, University of Cambridge

2 Rare Diseases UK, Genetic Alliance UK

3 Cambridge University Hospitals NHS Trust, University of Cambridge

1. Summary of key points

- Although individually rare, collectively rare diseases affect ~3 million of the UK population. They have a major impact on people's lives, and form a large part of the work of our national health and social care services.
- Advances in medical research continue to identify new rare diseases and their causes. Many are genetic, often come to light in children, and have no cure.
- Patients and families often describe a 'diagnostic odyssey' in which many referrals and tests are arranged before reaching a delayed diagnosis. This can cause stress and inconvenience, waste resources, and prevent early access to treatments and support services.
- The UK is at the forefront of rare disease research and treatment. This reflects an active biomedical research community within our Universities, working in partnership and often embedded within the NHS.
- Government investment through the National Institute for Health Research (NIHR) and the Medical Research Council (MRC) has consolidated UK leadership and delivery of rare disease research. This benefits patients, and is attracting investment from the life sciences industry.
- The NIHR BioResource Rare Diseases and NIHR Rare Diseases Translational Research Collaboration (RD-TRC) helped us to understand how these diseases cause debility and affect people's lives, and how they progress over time. This knowledge underpins the development of new diagnostics and treatments, in collaboration with the life sciences industry.
- Embedding whole genome sequencing within the NHS through Genomics England and the 100,000 Genomes project is accelerating the rate of diagnosis for rare diseases.
- Increased awareness amongst health care workers and improved diagnostic tests will have a positive impact on patient care.

2. Background

Rare diseases are defined as disorders that affect less than 1 in 2,000 of the population. Over 7,000 different rare diseases have been identified, each with its own particular problems that affect different systems in the body. Most rare diseases have no cure. Although individually each disease is rare, together they affect 7% of the UK population – some 3 million people.¹

Most (>80%) rare diseases are genetic, and are caused by a known error in the genetic code, called a 'mutation'. The mutations can be inherited, and often affect more than one family member.² Sometimes the mutation is only present in an affected individual (called de novo mutations), which reduces the chance of the disorder affecting more than one individual. Other rare diseases that are not inherited are not caused by severe mutations in a single piece of genetic code (called a gene). These diseases include rare infections, reactions to drugs and toxins, and inflammation affecting specific tissues and organs. However, even in these conditions our genes may affect our susceptibility and how we respond to a rare disease.

Rare diseases are often chronic life-long conditions that cause significant disability and limit employment opportunities, thus having a major impact on people's lives. Collectively, rare diseases generate a lot of work for the NHS and social services.

Simply by being rare, each rare disease presents a particular challenge for health and social services. Most general practitioners, and even most secondary care doctors, will only look after a family with a specific rare disease once, or perhaps twice in their entire career. This has several consequences:

■ **Diagnosis**

Lack of prior experience contributes to a delay in diagnosis. Doctors may not recognise a rare disease that they have not seen before, either because it is exceptionally rare, or because the disorder has only recently been described. Many families experience a 'diagnostic odyssey' before their condition is finally diagnosed. This process can take many years, or even decades, during which unnecessary referrals and investigations may be arranged. Without a diagnosis patients and families face considerable uncertainty about the future, and find it difficult to access the treatments and support that they need. Without a genetic diagnosis it can also be very difficult to advise families about the risks of having another child with the same disorder, and it may not be possible to offer genetic testing to prevent another child being affected.

■ **Personalised or precision health care**

Different rare diseases require different treatments. For each rare disease, treatment needs to be tailored to the individual patient. This may involve targeted medicines delivered through a specific prescription, monitoring for specific complications, or the provision of custom-made aids and appliances. In many instances the best way to treat patients is not known. In part, this is because no single centre has sufficient expertise to advance our understanding of a specific disease and its therapy. Even if the best treatments are known, a lack of experience with a particular rare disorder can lead to sub-optimal care for an individual patient.

Having a national health care system free at the point of use, the UK has been at the forefront of rare disease research. This has benefited patients globally, and is driving investment in the UK by the life sciences industry. Collaboration between the patients, NHS, universities, and charities has established a unique infrastructure, but harnessing these resources to deliver advances in patient care - linked to economic growth - remains a challenge.

Box 1 - The importance of diagnosing rare diseases

If you, or your child has a serious, possibly life threatening, disease then you would hope to receive an accurate diagnosis, delivered in a timely manner. This would enable you to understand what has happened to you or your child, to receive any appropriate treatment, make decisions about having future family, and to plan any changes that you might be facing as the disease progresses. A diagnosis makes it possible to link up with others in the same situation, to learn from their experiences. If you have a rare disease this is much more difficult.

Rare Disease UK, in their report 'Rare Disease UK: The Rare Reality – an insight into the patient and family experience of rare disease' (2016), has shown that an average patient with a rare disease consults with five different specialists, receives three misdiagnoses and has to wait four years before they find someone who can give them a definitive diagnosis, and begin to put together a coordinated care plan for them. This delay, and the detours up blind alleys because of misdiagnoses, means that they may miss out on interventions that might help them. Because ~80% of rare diseases are genetic, and ~75% affect children, they may have another child with the same condition which they could otherwise have avoided. They may receive treatment for a condition they don't have, which is unlikely to help them and which may do them harm. They are denied the opportunity to learn, to plan, to benefit from expert advice and support and the practical help and emotional support that can come from others in the same boat.

Despite recent rapid advances in our knowledge there are still many families who have a child who is currently undiagnosable. As many as 8,000 children born in Britain every year may currently be unable to benefit from a diagnosis, and about half of children referred to clinical genetics centres do not get diagnosed because there is not yet a definitive test that will give a clear answer. The majority of rare diseases have no specific treatment available. Since the adoption of the Orphan Medicinal Product Regulations by the European Union in 2000 over 130 new drugs for rare diseases have been licensed. This is more than ever before, but there are over 8,000 different rare diseases currently identified, with the list growing by around five a week, so there is clearly a long way to go before every patient and every family can be confident of receiving a timely diagnosis, triggering access to coordinated care and effective therapy.

Without sustained investment in high quality biomedical research into causes and cures for rare diseases then this situation will not change. Today's undiagnosable patients will remain undiagnosed tomorrow and in perpetuity. There will be no new therapies addressing unmet medical needs, and the majority of rare diseases will remain poorly understood, untreatable, incurable and frequently life shortening. 30% of children born with a rare disease will not live to see their fifth birthday. Rare Disease UK report that, in a survey, 80% of patients with rare diseases were eager to take part in research, and there was almost universal recognition of its value in generating new knowledge.

For this to benefit rare disease patients this new knowledge must be translated into new ways of working that enable better management of complex conditions, novel therapies that will modify disease trajectories, or possibly even cure them. The creation of new diagnostics that will give every patient the expectation that they can find out what has happened to them, their child or loved one, and what can be done to help.

Research and development work in rare disease is a substantial component of the National Institute for Health Research's programme of work. As a result UK researchers and clinicians are often global leaders in their field. These initiatives are making a substantial contribution to improving the lives of patients and families with rare diseases in Britain and around the world through leadership, academic excellence and by partnership with others, including patients, regulators and industry.

3. Challenges

3.1 Overview

Here we highlight some of the challenges presented by rare diseases, which often arise simply because the disorders are rare, limiting progress on many fronts. Recent advances in technology provide solutions to some of these difficulties, and closer working with patient groups will ensure that efforts are directed towards the needs of patients and families. The NHS worked to support rare diseases research through the NIHR BioResource Rare Diseases and NIHR Rare Diseases Translational Research Collaboration (RD-TRC), and to transform the use of genomics in the management of rare diseases through Genomics England Ltd. Although these developments focussed on research advances, the overall aim was to develop new clinical services to improve the care available for families from the NHS.

3.2 Shortening the diagnostic odyssey

In a survey of eight rare diseases, a quarter of families described a gap of between 5 and 30 years between the onset of their first symptoms and receiving a definitive diagnosis.³ This causes considerable uncertainty, adding to the anxiety experienced by families at a vulnerable time (see Case study A).

There are several reasons for this 'diagnostic odyssey', including a lack of awareness of a particular diagnosis by health care professionals, and the complex diagnostic approach required for many rare diseases. With many new rare diseases being identified every year, even recently trained doctors will find it difficult to remain informed about all rare diseases.

Box 2 - They are now superseded by the NIHR Translational BioResource for Common and Rare Diseases'. This is relevant for: The importance of NIHR infrastructure

The National Institute for Health Research (NIHR) BioResource - Rare Diseases and NIHR Rare Diseases Translational Research (RD-TRC) Collaboration were established to improve diagnosis and enable studies to develop and validate new treatments for rare diseases. The NIHR RD-TRC was rolled out across over 40 NHS Trusts, and with Genomics England led the pilot for the 100,000 Genomes Project. Nearly 10,000 participants have had whole genome sequencing to date. In parallel, detailed clinical and laboratory information (phenotypic data) were captured using Human Phenotype Ontology (HPO) terms, which were deposited in a national Open Clinica database developed by the RD-TRC.

The NIHR RD-TRC was built around 14 themes (Cancer, Cardiovascular, Dementia & Neurodegenerative Disease, Eye disease, Gastrointestinal disease, Non-malignant Haematology, Immunological disorders, Metabolism and Endocrinology, Musculoskeletal disease, Neuromuscular disorders, Respiratory disease, Skin disease, Renal disease and Paediatrics). Each theme was led through a NIHR Biomedical Research Centre, Biomedical Research Unit or Clinical Research Facility. The primary aim of the NIHR RD-TRC was to provide resources and build capacity for the detailed description (in depth phenotyping) of patients with rare diseases. The collaboration has supported in excess of 55 rare disease projects, 12 clinical fellowships in rare disease research, and established an MPhil in rare disease.

The NIHR BioResource Rare diseases and NIHR Rare Diseases TRC are now superseded by the NIHR Translational BioResource for Common and Rare Diseases.

Case study A Segmental Overgrowth Study

Turning Lives Around – Segmental Overgrowth Study

Metabolic and endocrine research supported by the Rare Diseases Translational Research Collaboration (RD-TRC) is breaking new ground – and tangibly making a difference for patients. The Segmental Overgrowth Study is based at the Institute for Metabolic Science in the University of Cambridge, and in the last four years has grown from treating one patient – Mandy Sellars – to several hundred, along the way attracting industry attention and collaboration.

Scaling up

Chief Investigator Dr. Rob Semple said: “We started off with some advantages in that the genetic defect we identified was an activating mutation, was ‘drug-able’ and could draw on pharmaceutical developments in cancer drugs, but nevertheless there was a big gap between that scientific discovery and the progressive scaling-up to wider populations.

“That’s the gap that RD-TRC has filled very effectively and has allowed us to move from meeting one patient [Mandy Sellars] and identifying a rare defect, to coordinating a population of several hundred patients in the UK and making a strong case to the Department of Health for specialised service commissioning and for clinical trials, which we believe have an excellent hope of addressing some components of the conditions we are dealing with.”

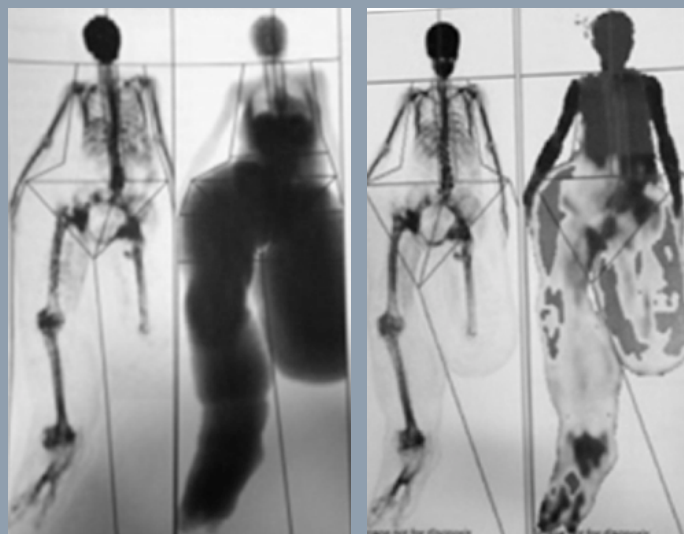
Low doses effective

Dr. Semple works closely with Dr. Vicki Parker, who described how the team’s research led to a treatment for Mandy: “We thought Mandy may be affected by a mosaic condition and we performed a powerful genetic screen known as whole exome sequencing on skin samples taken from Mandy’s arm and leg. We quite quickly found an underlying genetic mutation, which we thought was responsible for the overgrowth in Mandy’s legs, in the gene *PIK3CA*.”

It was also very well known that this particular mutation is a hotspot in cancers, and the team began to look into some of the treatments currently available for cancer to see if they could be beneficial in Mandy’s condition.

The team ran a series of dose-response studies in cells derived from affected patients, funded by the NIHR RD-TRC, and recognised that treatment with very low doses of drug were likely to be effective. Dr. Parker said: “In 2012 Mandy’s legs were continuing to grow and on the basis of studies we’d done, we thought it reasonable to trial sirolimus in an off-licence capacity in Mandy.”

So, Mandy started with a small daily dose (1mg) of sirolimus three years ago and this treatment has to date been very successful, with sustained loss of tissue from her legs. (See DEXA images, taken before and 6 months after treatment, below.) Dr. Parker said: “We’ve very much benefited from RD-TRC support, we’ve had a massive increase in recruitment in the past year and we simply wouldn’t have been able to cope without the resources the RD-TRC has provided.”



DEXA scan imaged before (left) and after (right) treatment with sirolimus, showing a reduction in the soft tissue in both legs.

Dr. Semple further noted that Mandy’s role has been pivotal: “She’s been the key recruiting sergeant in our efforts to find more patients, and by her tireless efforts personally and on social media, she’s probably accounted for a significant proportion of the patients we now have in our cohort.”

Waiting decades for a diagnosis

James is in his thirties and lives near Cardiff. James was just five when he was initially diagnosed with Proteus Syndrome, at Great Ormond Street Hospital: “At the end of all [the tests] the best that the doctors could come up with was Proteus Syndrome...they just sent us on our way, as there was no community to go to for support. So that’s what we settled with and that’s what my parents lived with.”

From a young age James has had to undergo regular tests and monitoring – not just on the parts of his body that are overgrowing (fingers, leg, toes, ear etc.) but also his organs, his hearing and the varicose veins in his legs. Interestingly, these tests revealed that the bigger parts of his body were growing at the same rate as the non-affected parts – unlike Proteus, which is progressive.

After being told aged 20 by a Proteus specialist that he didn’t have Proteus, James made it his mission to find out what exactly he did have – along the way finding what he describes as his ‘inner peace’. Life was difficult for James and he felt at times incredibly isolated: “Looking different, not being able to walk far, finding stairs hard some days, seemed to be not good enough for the other teenagers. If teenagers don’t know or don’t understand then they throw judgment.

“All this made me develop many insecurities that I’ve fought over the years, mostly on my physical looks [but also on] confidence, trust, showing my legs, wearing shorts, talking about my overgrowth, taking up new opportunities.

“My peace with myself eventually came; I was 27. I finally felt comfortable in my own skin...I had to be happy in life and only I could make that choice.” But then in February 2013, James had no idea things were going to change... After viewing the Channel 5 documentary *Shrinking My 17 Stone Legs*, James got in touch via Twitter with Mandy Sellers, who forwarded his email address to Dr Victoria Parker at Cambridge University Hospitals. A subsequent visit to Cambridge and DEXA scan revealed that the overgrowth had created more tissue in James’s body – but his skeleton was not affected. In August 2014, Dr Parker called James and told him that he had a gene change in PIK3CA – and that his condition had a name, Segmental Overgrowth Syndrome (SOS). James said: “I can’t describe how I feel through all this, with what the team at Addenbrooke’s has done for me. All I can say is I’m grateful. I am feeling a lot happier and comfortable in myself. I’ve taken up swimming and I wear those shorts on holiday!”

The pace of discovery means that it will not be possible for health professionals to recognise every rare disease. However, increasing awareness of rare diseases, and how to refer to appropriate specialists, will go a long way to address the issue. This can be achieved through undergraduate and postgraduate training of health care professionals, through ongoing professional development, and through the concerted work of patient support groups and charities (see Figure 1).

Figure 1 Rare disease day 2015 - increasing awareness of rare diseases by engaging future health care professionals in educational activities



A major recent development has been the implementation of whole genome sequencing (WGS) within the NHS by Genomics England through the 100,000 Genomes Project. From 2015, WGS can be initiated through one of 13 geographically distributed NHS Genomics Medicines Centres (GMCs) throughout England, with linked initiatives in the three devolved nations.⁴ Given that 85% of rare diseases are genetic,² this provides a nationwide opportunity to achieve a rapid diagnosis, even for patients with disorders that have been characteristically difficult to diagnose in the past – such as children with developmental delay.

The faster diagnosis provided by genome sequencing may also save NHS resources. Before the advent of genome sequencing, families often attended hospital appointments for years before a clear diagnosis was possible, undergoing expensive and invasive tests that were repeated each time the technology improved. The more rapid diagnosis provided by whole genome sequencing can mitigate the need for numerous hospital appointments and investigations. Individual case studies have shown that, in many instances, the cost of WGS is much less than the traditional approach to investigation, often saving thousands of pounds in individual families.

Box 3 - The challenge of diagnosing rare diseases

In a 2008 review of GPs Eurodis' 'Survey of the delay in diagnosis for 8 rare diseases in Europe ('EurordisCare2'), four of the most difficult aspects of diagnosis included: unusual presentations, non-specific presentations,

very rare conditions, the presence of more than one disease at a time (co-morbidity). All four aspects are common in rare diseases.

The DDD study: Pioneering genome-wide sequencing for discovery and diagnosis in rare disease

Aims

The Deciphering Developmental Disorders (DDD) Study is a nation-wide collaboration between the NHS and the Wellcome Trust Sanger Institute with the aims of applying the latest genomic technologies to:

- Discover the genetic causes of developmental disorders, sharing knowledge globally through publication and innovative and responsible data sharing platforms
- Diagnose thousands of patients, addressing unmet clinical need, transferring knowledge into the NHS and up-skilling the NHS genetics workforce.

Almost every NHS clinical geneticist across the UK recruited patients with rare diseases that were apparent at birth or in early childhood, often affecting several body systems, and for whom a genetic diagnosis had not proved possible using conventional tests. Scientists at the Wellcome Trust Sanger Institute sequenced and analysed all of the genes in these patients and their parents (~33,000 individuals).

The DDD study (www.ddduk.org), has diagnosed several thousand children, and identified >30 new genetic disorders so far, for a combined cost of ~£15 million. The combination of clinical genetics and scientific expertise, and the nation-wide scale of the project resulted in an unprecedented diagnostic yield of ~40% in patients.

Study design

The DDD study was designed to be minimally disruptive to patients and NHS services, working with the grain of current clinical practice.

Data-sharing

The wealth of variants in every human genome (4-5 million) pose a huge challenge in interpreting data from genome-wide sequencing studies to provide safe and accurate diagnoses for patients. The DDD study has set the standard for responsible data sharing to improve diagnosis globally, through the DECIPHER web portal (<https://decipher.sanger.ac.uk>).

*Text kindly provided by
Matt Hurler, Head of Human Genetics and Senior Group Leader Wellcome Trust Sanger Institute, Cambridge
Helen V Firth, Consultant Clinical Geneticist, Cambridge University Hospitals*

Research

In addition to the core research conducted at the Wellcome Trust Sanger Institute, DDD has catalysed and supported additional research projects led by NHS genetics services and their academic partners. More than 200 such projects have been established and ~50 publications delivered to date.

Impact

The DDD study is having world-leading impact in the following domains:

- **Science**
>60 publications, including two flagship papers in the top journal Nature.
- **NHS Genetics services** – diagnoses for patients
Stimulating training and research. A strong evidence base for implementing improved services.
- **Translation** – improved diagnostic assay developed with UK SME (OGT), spun-out diagnostic software company (Congenica)
- **Data-sharing** – unequivocal evidence of the benefits of NHS genetics services acting collectively as a distributed network of expertise, and enabling NHS data sharing through DECIPHER.
- **Families** – a molecular diagnosis for their child's developmental disorder in more than a third of cases. Sharing anonymised diagnostic information globally via DECIPHER to catalyse participation in therapeutic trials. Facilitated supportive social networks through SWAN (www.undiagnosed.org.uk) and Unique (www.rarechromo.org).

The DDD study is an exemplar of how harnessing NHS clinical expertise, UK scientific know-how and the scale of the NHS can have world-leading impact on scientific knowledge and clinical practice, ultimately benefitting patients and their families.

3.3 Understanding the disease

The falling costs of gene sequencing and its widespread application has led to the discovery of many new rare diseases. Many of these disorders are so rare that even large centres may only identify a few patients with the same disorder. In the early stages it is often not clear what range of problems a patient with a new rare disease will face. This is an important gap in knowledge for both families and health care workers.

Understanding the range of clinical features (called the 'phenotype') alerts health care professionals to the possibility of the diagnosis in other patients. Without a clear understanding of how the disease changes over time, it is not possible to develop informed care pathways using existing health care services, nor to develop new treatments tackling the most important aspects of the disease (Case study A).

The key to addressing these issues is the secure collection of clinical information on patients with rare diseases. Unlike many other countries, in the UK the NHS provides a national infrastructure enabling clinical data collection from the largest number of patients possible. Work by Public Health England will provide the first understanding of the prevalence and distribution of many rare disease registry.⁵ More detailed clinical information (the 'deep phenotype') will enable a greater understanding of how specific rare diseases affect patients and families. If measured over time, the 'deep phenotype' will give insight into the prognosis for patients and families (See Case study B).

Establishing compatible electronic health care records will greatly facilitate the national data collection. An important step is the categorisation of new diseases using standardized terms set out by expert groups. The Human Phenotype Ontology (HPO) provides one widely adopted example.⁶

The use of digital media also presents exciting opportunities for patients and carers to collect real-time clinical information through questionnaires, rating scales, or wearable devices which monitoring activity, and can be used to study the effects of new treatments.

Using state-of-the-art imaging, biochemical, or metabolic measures (metabolomics) it is possible to measure the progression of a rare disease in a few months. NIHR supported this work through the NIHR Rare Diseases Translational Research Collaboration (RD-TRC).⁷ This was a critically important step in the development of new treatments. Understanding the 'natural history' of the disease, makes it possible to test new treatments to determine whether they can halt its progression.

Case study B

The importance of deep phenotyping

Serum samples collected between 2000-05 from 914 patients across the UK with a rare kidney disease called IgA Nephropathy (IgAN). The in depth phenotyping has produced some exciting results. The work was carried out by Chief Investigator Dr. Jonathan Barratt and his team at the NIHR Leicester-Loughborough Diet, Lifestyle and Physical Activity Biomedical Research Unit and Dr. Daniel Gale, UCL funded by the Rare Diseases Translational Research Collaboration (RD-TRC).

Discoveries

Dr. Barratt and his team started the Deep phenotyping of the UK Glomerulonephritis DNA Bank IgA nephropathy cohort study in March 2014. As of August 2015 it is ongoing and the observations made to date have led to further work. One of the findings saw the team identify alleles (different versions of the same gene) associated with IgA1 O-glycosylation in IgAN. This discovery is likely to increase our understanding of the pathogenesis of IgAN and be a target for future therapeutic intervention, and is now the subject of ongoing evaluation.

What is IgA?

Dr. Barratt and his team started the Deep phenotyping of the UK Glomerulonephritis DNA Bank IgA nephropathy cohort study in March 2014. As of August 2015 it is ongoing and the observations made to date have led to further work. One of the findings saw the team identify alleles (different versions of the same gene) associated with IgA1 O-glycosylation in IgAN. This discovery is likely to increase our understanding of the pathogenesis of IgAN and be a target for future therapeutic intervention, and is now the subject of ongoing evaluation.

Impact

RD-TRC funding provided the background for a successful bid for funds from a major pharmaceutical company that has committed to invest £1.5m in this IgAN research programme. This investment will secure future deep phenotyping of this cohort, and establish a new cohort of 3,000 IgAN patients from across the UK for future studies.

RD-TRC funding also supported the team's application for IgAN to be adopted into the National Registry of Rare Kidney Diseases (see RaDaR: rarerenal.org).

Spreading the results

The team has shared some of the results with the clinical and scientific community at international meetings and in peer-reviewed publications. The research findings have also been shared with the newly-established UK IgAN Patient Support Group (updates are available in their newsletters).

3.4 Preventing rare diseases

Defining the precise genetic cause for a rare disease also enables accurate and reliable genetic counselling. This is usually based on the known inheritance pattern for the mutation causing the disorder. With this information, families are empowered to make reproductive decisions about future offspring.

Although it is possible to estimate the recurrence risks without a genetic diagnosis, this is less reliable, and it may not be possible to offer specific genetic tests to prevent the disease from happening again in the family. For example, once a family has a genetic diagnosis, it is usually possible to offer prenatal diagnosis (when a pregnancy is tested at an early stage to see whether the disorder has been passed on), or even pre-implantation diagnosis (when an embryo created by in vitro fertilization, IVF, is tested before being placed in the womb).

Thus, achieving a genetic diagnosis allows families to access existing NHS services which offer several different approaches to prevent further family members becoming affected by a rare disease.

3.5 Integration of clinical care

The management of rare diseases is often complex, and requires close working between the health services and social care providers, bridging across traditional 'care boundaries'. Effective communication is the key, particularly between primary, secondary, tertiary and quaternary healthcare services; health and social care; voluntary sector services; communication across all age groups and also between different medical disciplines.

A key role of the specialist centres is the development of care pathways to share their expertise with local services. Research to develop the best care pathways for rare diseases is being carried out across Europe, supported by the EU RARE-Bestpractices programme. The UK continues to contribute to and support projects such as the European Project for Rare Diseases National Plans Development (EUROPLAN)¹² and International Rare Diseases Research Consortium (IRDiRC), including its goal to develop 200 drugs for rare diseases by 2020.⁸

Orphanet is one source of information on rare diseases and drugs to help improve the diagnosis, care and treatment of patients with rare diseases.¹⁰

3.6 Patient, family and public involvement

Patient, family and public involvement is particularly strong in the field of rare diseases. This informs and improves the quality and effectiveness of research, and should be encouraged at all stages of the research process. Many patient groups have engaged the power of social media, including Facebook, Twitter and PatientsLikeMe to empower patient support and advisory groups in collaboration with the voluntary sector (See Case study C).

Rare disease support is provided by Genetic Alliance UK,^{*} Rare Disease UK,^{16†} and a very large number of disease specific patient organisations. Public awareness has also been increased through national, England-wide, engagement activities. The global annual Rare Disease Day (Figure 2) is a powerful tool raising the profile of rare diseases.¹¹

NIHR supports the patient organisation INVOLVE, which sets standards for best practice for patient and public involvement in research.¹²

Figure 2 Rare Disease Day 2015 - school children learning about rare diseases



* <http://www.geneticalliance.org.uk>

† <http://www.raredisease.org.uk>

Case study C

Patient, family and public engagement in research

ADAPT (Alpha-1-antitrypsin Deficiency Assessment and Programme for Treatment)

Why do some patients have lung disease and others have skin, vascular and liver disease? In future researchers may be closer to understanding why, thanks to a cohort of patients with AATD who have been deeply phenotyped as part of the ADAPT programme.

ADAPT includes a 2-year prospective assessment (where researchers watch for outcomes, such as the development of disease) of patients with Alpha-1-antitrypsin deficiency (AATD). Patients are deeply phenotyped and disease progression over 18 months determined (observed to allow for best prognosis and treatment) and matched to several novel biomarkers (measurable indicators of the severity or presence of a disease state). The study has recruited 194 participants with more than 120 from Professor Robert Stockley and his team at University Hospitals Birmingham, who focussed on recruiting patients with mild/moderate forms of the disease able to travel to the centre. The remaining severely affected patients were recruited closer to home from seven satellite centres.

What is AATD?

Alpha-1-antitrypsin (AAT) deficiency is a rare genetic lung disease which may lead to chronic obstructive pulmonary disease (COPD) and emphysema, as well as other rare conditions such as panniculitis and vasculitis. Patients with AAT deficiency are at increased risk of developing COPD (especially if they smoke). But the disease is highly variable, and the lack of understanding about which patients develop lung impairment or liver disease and why, has held back the design of new treatments.

Bigger picture

With the help of RD-TRC funding, this network (consisting of the Birmingham central hub and satellite centres) has provided detailed phenotyping of patients with AATD who had been referred to secondary care. The data collected on both the severe and milder forms of the disease together provide a comprehensive picture of its natural history and impact, while the 'hub and spoke' network of central hub and satellite centres provides a clear referral mechanism for patient management and monitoring.

Future Impact

Detailed phenotyping, sample collection and biomarker development will help guide future AATD management and treatment strategies.

The network is in a strong strategic position to deliver future studies. It can provide rapid access to patient phenotypes, while the availability of expert centres and patient contributors within the network should facilitate faster patient recruitment to trials. These could include a number of phase 2 studies currently under discussion with University Hospitals Birmingham and pharmaceutical companies.

Patient, carer and public Involvement

Patients have been involved from early on, providing patient input into the Steering Committee and, through the patient support group Alpha-1, using a 'Traffic Light' system to review the research proposal.

3.7 Developing new treatments in partnership

Developing new treatments for rare diseases is challenging. Drug development costs are expensive, and treatments for rare conditions may not be financially appealing to pharmaceutical companies. Recruitment to clinical trials can be difficult for rare disorders, many sites may need to be involved, and there will be set-up costs and local approval requirements for studies. However, in a growing number of cases, successful treatments have been developed for rare diseases. A close working partnership between patients and families, health care providers, patient support groups, charities and the life sciences industry is the key to success (See Case study D).

Although many rare diseases need completely new treatment approaches, there are several examples where existing and often safe drugs can be repurposed to treat rare diseases. This can provide a 'short cut' to new therapies.

In recent years, national regulatory bodies, such as the Health Research Authority (HRA),¹³ have streamlined the bureaucracy, and orphan drug regulations in Europe have encouraged both small and large pharma companies to develop new treatments in niche areas.¹⁴ Regulation (EC) 141/2000 (of the European Parliament and of the Council) sets out a procedure for the designation of medicinal products as orphan medicinal products, and provides incentives for the development and marketing of designated orphan medicinal products. The legislation recognises that 'some conditions occur so infrequently that the cost of developing and bringing to the market a medicinal product to diagnose, prevent or treat the condition would not be recovered by the expected sales of the medicinal product', and stated that 'patients suffering from rare conditions should be entitled to the same quality of treatment as other patients'.

Designation as an orphan medicinal product offers a number of incentives for industry, and perhaps most notably market exclusivity. Specifically the regulations state that the Community and the Member States shall not, for a period of 10 years, accept another application for a marketing authorisation, or grant a marketing authorisation or accept an application to extend an existing marketing authorisation, for the same therapeutic indication, in respect of a similar medicinal product. Examples of drugs granted orphan designation by the European Commission in recent years include everolimus for the treatment of tuberous sclerosis, and eculizumab for the treatment of atypical haemolytic uraemic syndrome and infection-associated haemolytic uraemic syndrome.

For many rare diseases, international collaborations between researchers is the key to rapid progress, with European organisations facilitating multi-national trials (such as TREAT-NMD).²⁰

Case study D

Developing new treatments in partnership

Congenital Hyperinsulinism in Infancy (CHI): Unravelling the Phenotypic Diversity in CH

Researchers have identified a number of genetic associations in 40% of all cases of CHI – but in the remaining 60% of children, no mutations have been identified.

So this study, funded by the NIHR Rare Diseases Translational Research Collaboration, aimed to find out more about the phenotypic diversity in CHI which could potentially lead to better and more targeted and personalised treatment.

What is CHI?

Congenital Hyperinsulinism in Infancy (CHI) is a heterogeneous condition of severe hypoglycaemia (low blood sugar). The estimated incidence of CHI is one in every 40,000 to 50,000. It is much more common in communities where marriage between blood relatives occurs. Hypoglycaemia due to CHI can be unpredictable and severe. This has particularly concerning effects on the central nervous system, with 30% of patients showing adverse neurode CHI?

Currently treatment of CHI can be complex and difficult, with the first- and second-line medical therapies linked to significant adverse effects and often failing to manage the condition. In extreme cases subtotal pancreatectomy (where part of the pancreas is removed) may have to be undertaken.

This study therefore looked at alternative ways to examine the phenotype in CHI in order to develop novel diagnostic and prognostic tests. The research team worked on: establishing a tissue library (or Digital Tissue Atlas) from patients with CHI and age-matched controls; initiating a cross-platform clinical database; and recruiting patients for phenotype profiling and long-term neurological outcomes and development. Samples from 33 patients (the original target was 10) were collected, giving the team and their collaborators a unique resource to define the phenotypic basis of CHI. The team also recruited 15 patients (initial target 10) to capture the phenotype, genetic and epigenetic profiles of CHI patients for profiling and biomarker (something that can be objectively and accurately measured) development.

The final objective of the programme, to capture the metabolomic profiles of CHI patients and correlate with long-term neurological outcomes, with a target of four patients, started in December 2014 and is ongoing.

Collaboration with biotech

A further benefit from the study was that the team saw how phenotype profiling could contribute to the development of novel approaches to treatment. This included the development of drugs that block the action of insulin (called antagonists). To achieve this, the team submitted a £2.3m application to Innovate UK through a partnership with the biotech firm Heptares Therapeutics to develop an antagonist for the treatment of hyperinsulinaemic (recurring or persistent) hypoglycaemic conditions including, but not limited to CHI. This application was successful and the new project began in May 2015.

4. Conclusions

The sheer number of different rare diseases, and the relative infrequency of each disease, present major challenges for health and social care providers and make research difficult. However, increasing awareness of the importance of rare diseases – both nationally and internationally – is leading to progress across a range of rare diseases.

The UK continues to play a leading role internationally, in discovering new rare diseases, improving diagnostics, and developing new treatments. This position has been consolidated by new NHS investment and substantial Government support for rare disease research, built around the UK strategy for rare diseases (2013).²¹

Recent technological advances – particularly in genome sequencing and health care informatics – provide a solution to the traditional hurdles presented by rare diseases. Close working between patient and family groups, the NHS, universities and the life sciences industry should have major impact over the next 5 years, reducing the burden of rare diseases and contributing to economic growth.

5. Suggestions for policy makers

5.1 Rapid diagnosis for all, using state-of-the-art technologies

The widespread adoption of whole genome sequencing (WGS) within the NHS provides an opportunity to substantially increase the speed of diagnosis for rare diseases. Interpreting WGS is challenging, but gathering experience over the next 5 years should move the UK to a position where the majority of rare diseases are diagnosed by the NHS within 6 weeks of a suspected diagnosis. Ultimately a diagnosis should be possible within primary care, enabling direct referrals to appropriate specialists. A genetic diagnosis also enables reliable genetic counselling and access to services aimed at preventing the disease from occurring again.

5.2 Leading international efforts to find new treatments for rare diseases

Every patient with a rare disease should be offered the opportunity to participate in clinical research aimed at developing new treatments for their disease. Ideally this would involve a multi-centre clinical trial with a new medicine. However, if there is no specific treatment available, patients should be able to participate in a natural history study. This will increase knowledge about specific rare diseases and lead to improved clinical management through the development of best care guidelines.

5.3 Seamless integration of clinical care and research

Every clinical episode provides valuable information about a rare disease and how it changes over time. Research assessments should be incorporated into the routine clinical follow-up of patients. This should be available to all patients with rare diseases within the NHS. With appropriate consent and safeguards in place, this data gathering and analysis will underpin the development of new treatments in collaboration with global academic and industry partners.

6. References

1. <http://www.raredisease.org.uk/what-is-a-rare-disease/>
2. <http://www.eurordis.org/>
3. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/AnnualReports/DH_113912
4. <http://www.genomicsengland.co.uk/taking-part/genomic-medicine-centres/>
5. <https://www.gov.uk/guidance/the-national-congenital-anomaly-and-rare-disease-registration-service-ncardrs>
6. <http://www.human-phenotype-ontology.org/>
7. <http://rd.trc.nihr.ac.uk/>
8. <http://www.europlanproject.eu/>
9. http://www.irdirc.org/?page_id=1035
10. <http://www.orpha.net/consor/cgi-bin/index.php>
11. <http://www.rarediseaseday.org/>
12. <http://www.invo.org.uk>
13. <http://www.hra.nhs.uk>
14. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000029.jsp
15. <http://www.treat-nmd.eu>
16. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pdf

Chapter 7

Genomics and obesity

Chapter lead

Prof Stephen O’Rahilly¹

Chapter authors

Prof Sadaf Farooqi¹, Dr Vanisha Mistry¹, Prof Stephen O’ Rahilly¹

¹ University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science

1. Summary

Obesity is a state of chronic energy imbalance where energy intake exceeds energy expenditure over a sufficient period of time to result in the accrual of adipose tissue mass that exceeds an amount deemed to be healthy. Recent increases in the prevalence of obesity in populations are driven by environmental/societal factors that promote food intake and discourage energy expenditure. Additionally, there is an increase in the proportion of children and adults with severe obesity, a fact that is often overlooked when looking at changes in obesity prevalence. Studies comparing identical and non-identical twins indicate that the major difference between those susceptible and those resistant to the obesogenic environment lies in genetic inheritance.

The first genetic variants that cause human obesity were discovered in the late 1990s in children who developed obesity very early in life. Since then, the progress in uncovering the specific genetic variants that underpin obesity of early onset as well as susceptibility to common adult obesity has progressed rapidly. Particularly germane to those concerned with the health of the public are the following, largely uncontested, facts:

- 1) If a child develops severe obesity at a young age, there is a high likelihood that s/he will carry highly penetrant variants in genes involved in the control of energy homeostasis. Although the field is still young it is already possible to attribute obesity to defects in a single gene in over 10% of such children.
- 2) This observation is highly relevant to clinical management as there is already one identified genetic subtype which responds dramatically to specific therapy and there are other subtypes for which targeted clinical trials are ongoing. Importantly, the knowledge that any child with severe early onset obesity is likely to have a powerful biological driver for their condition challenges those who advocate approaches which include, for example, removing the children from the care of their families.
- 3) Genetic factors also play a major role in determining susceptibility to overweight and obesity in the general population with almost 100 common variants contributing to that variance identified to date. However the effect size of each allele is small and even the cumulative genetic risk score is unlikely to become a useful tool for disease prediction.
- 4) Studies of patients carrying rare variants that contribute to severe, early onset obesity or common variants that contribute to weight gain indicate that the majority of such alleles act by influencing food intake and satiety. Thus, public health policies which reduce exposure to stimuli encouraging consumption of high calorie food and help to limit portion sizes represent a logical public health approach to protect those who are genetically predisposed to obesity.
- 5) Perhaps the most powerful impact of human genetics on obesity in the long term will occur via its contribution to the understanding of the “wiring diagram” of human energy balance and the identification of control points that are amenable to therapeutic modulation with drugs and/or specific nutritional or behavioural strategies.

2. Background

2.1 Public health impact

Obesity is defined as an increase in fat mass that is sufficient to adversely affect health.^{1,2} Obesity is associated with an increased relative risk of type 2 diabetes, hypertension, cardiovascular disease, liver disease and some forms of cancer³. Additionally, obesity-associated orthopaedic and respiratory problems impact adversely on quality of life and ability to work.^{4,5} An often neglected issue is the considerable social stigma, associated with severe obesity in particular, which can affect educational attainment, job opportunities and mental health.^{6,7}

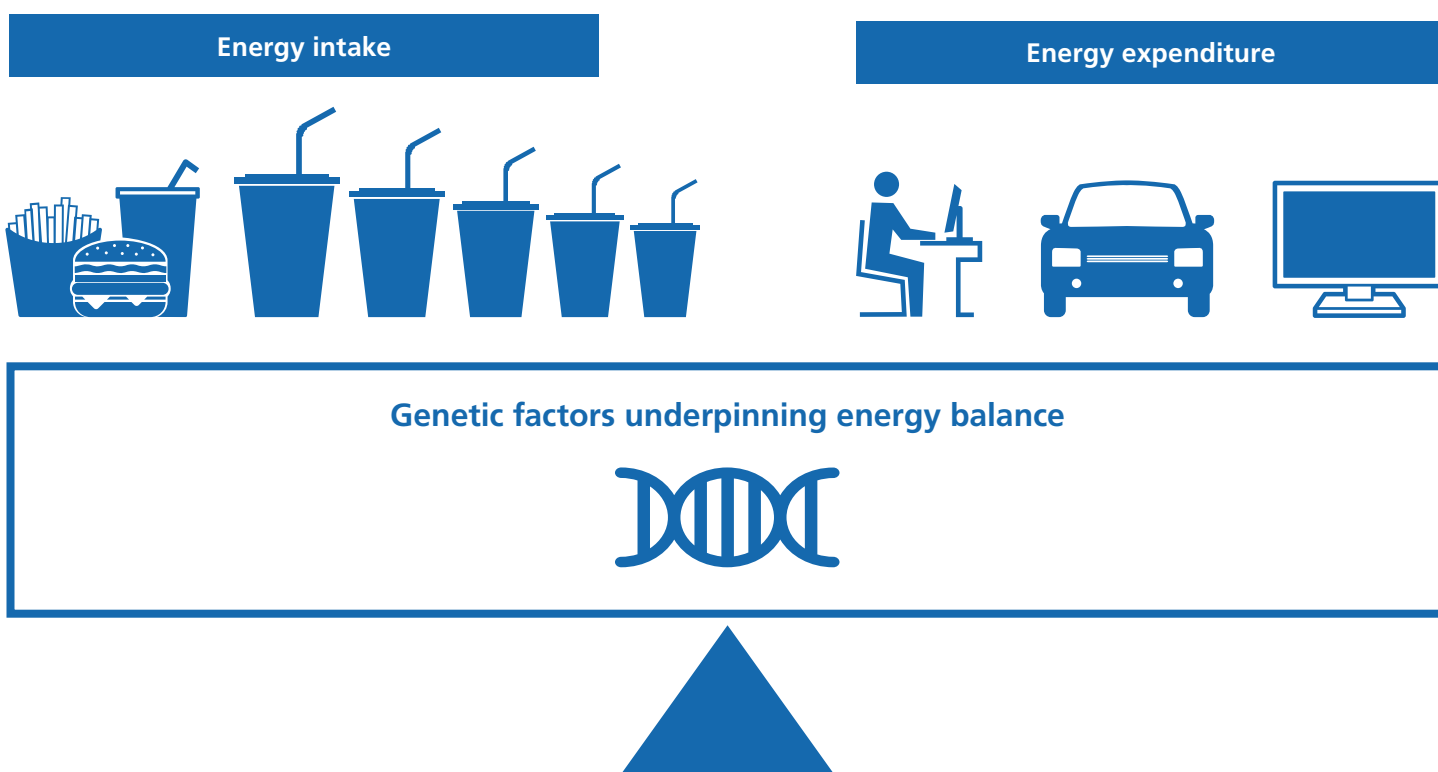
Estimates of obesity prevalence are based on body mass index (BMI; weight in kg/height in metres²), a surrogate marker of fat mass that can readily be used in population based studies. Most recent data from England and Wales (2013) demonstrates that 26% of men and 24% of women are classified as obese (BMI more than 30kg/m²). Therefore, obesity related disorders represent a significant public health concern, account for substantial health care costs and have a broader socio-economic cost to society as a result of lost productivity. Although there is limited data to inform predictions regarding the future public health burden associated with childhood obesity (11-17% in Europe depending on definitions used for classification), it is plausible that the rising prevalence of childhood obesity may impact on health in adolescence and potentially on morbidity and mortality in the future.⁸

3. Causes of obesity

3.1 The obesogenic environment

The rising prevalence of obesity is driven by the persistent imbalance between the amount of energy consumed and the amount of energy burned which results in net positive energy balance and weight gain. Thus, major contributors to the rising prevalence of obesity are factors that promote an increase in energy intake (for example, an abundance of inexpensive, easily available, energy-rich, highly palatable foods) and factors which contribute to a decrease in energy expenditure such as sedentary lifestyles (television watching, driving to work), reduced physical activity at work (office work rather than manual work) and in leisure time (Figure 1).

Figure 1 - Genetic and environmental factors influence the balance between energy intake and energy expenditure

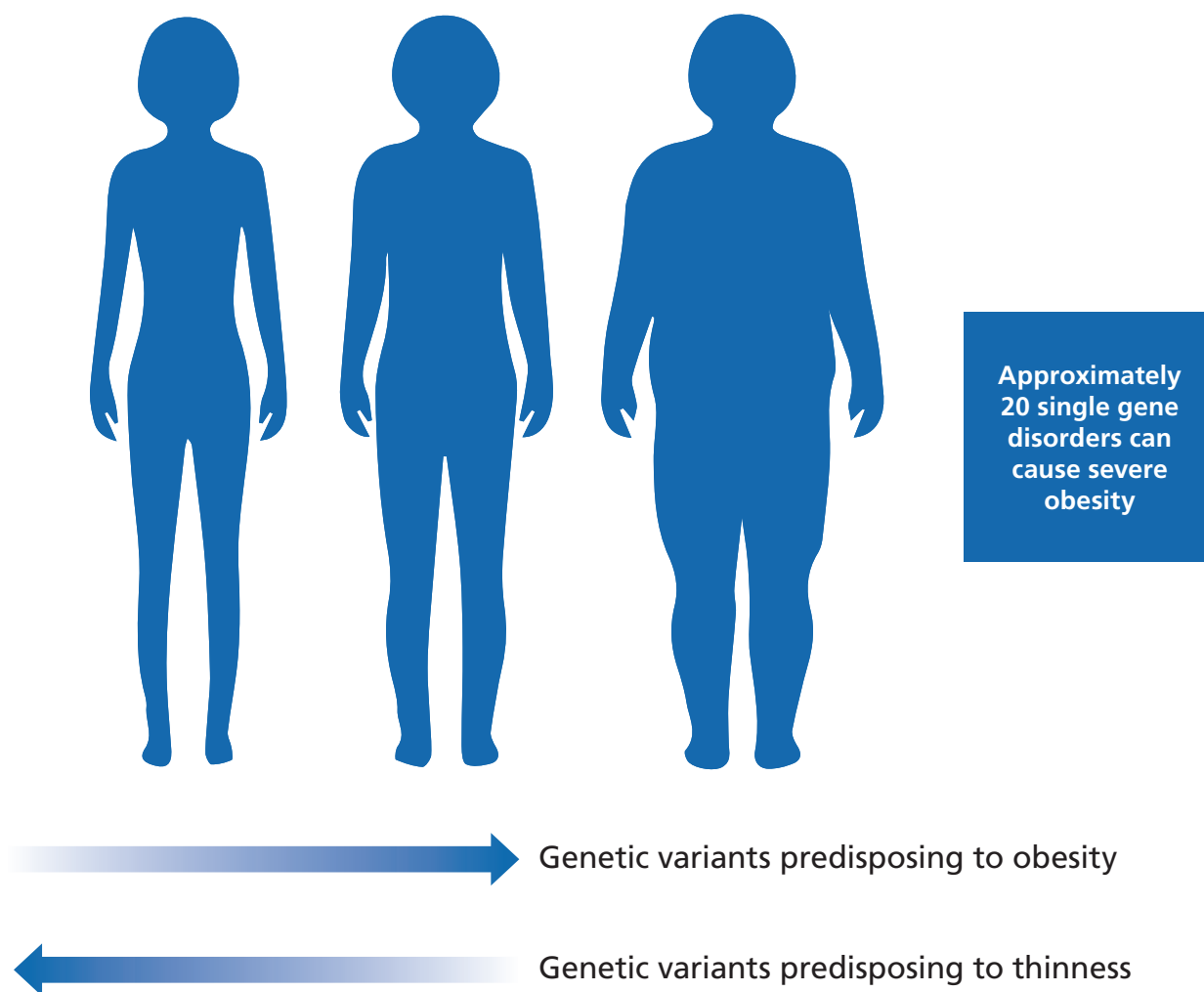


3.2 Biological factors modulate weight gain in an obesogenic environment

However, despite this obesogenic environment, there is considerable variation in body weight and fat mass between individuals.⁹ Some people are much more likely to gain weight than others and some people remain lean. This variability between individuals is influenced by complex interactions between environmental factors and biological (genetic, developmental and behavioural) factors which influence where an individual lies on the BMI distribution (Figure 1). It is important to note that as well as an increase in the mean BMI, population data show an increase in the proportion of people at the top end of the distribution – children and adults with severe obesity.^{10,11}

Indeed, twin studies indicate that heritability estimates are greatest at both ends of the BMI distribution (severe obesity and leanness)^{12,13} (Figure 2). It is plausible, and indeed has been shown in experimental clinical studies, that the response to weight loss interventions is also highly variable¹⁴; some people are more responsive to changes in diet and/or physical activity than others. Therefore, understanding the mechanisms that underpin the variability in BMI in the population and in the response to interventions is an important component of strategies to prevent and treat obesity and related disorders.

Figure 2 - Genetic variants that increase the susceptibility to obesity are more prevalent in those with severe obesity.



4. Genetic contributors to obesity – the evidence

The biological factors that influence variability in body weight between people are to a large extent, inherited¹⁵⁻¹⁷. Evidence that there is a genetic contribution to body weight comes from twin studies which cumulatively demonstrate that the heritability proportion of the total phenotypic variance of a quantitative trait attributable to genes (in a specified environment) of body mass index (BMI) is between 0.71-0.86¹⁸. Studies in 5,092 UK twins aged 8-11 years growing up during a time of dramatic rises in obesity, support the substantial heritability of BMI (~77%). In this study by Wardle and colleagues, there was a very modest effect of the shared-environment (which can inflate heritability estimates); the remaining environmental variance was largely unshared¹⁹. Similar heritability estimates were found when studying identical twins who were reared together and apart²⁰ and in large studies of adopted children whose body weights tend to be similar to those of their biological parents compared to their adopted parents²¹. It is likely that genetic factors influence how much weight people gain and how much they lose in response to changes in the amount of food consumed and the amount of exercise undertaken. As is true of all complex human traits, studies that base estimates of heritability of adiposity on family rather than twin studies generally report a somewhat lower figure for heritability. A detailed discussion of the reasons for this are beyond the scope of the current review. Carefully controlled experimental studies of identical twins conducted under direct supervision have shown that the amount of weight gained in response to a fixed amount of excess calories and the amount of weight lost after a fixed amount of physical activity is very similar between twins, but varied considerably across different sets of twins²²⁻²⁴. These studies demonstrate that the variability in body weight and in the physiological response to energy intake (food consumed) and energy expenditure (calories burned) is very strongly influenced by genetic factors²⁵⁻²⁷.

5. Genetics of obesity – a spectrum from rare to common genetic variation

5.1 Genetic discoveries in severe obesity

Genetic studies focussing on people with severe childhood obesity from a young age have led to the identification of multiple genes in which rare, highly penetrant variants cause obesity²⁸⁻³². Whilst individually these disorders are rare, cumulatively, at least 10% of children with severe obesity have rare chromosomal abnormalities and/or highly penetrant genetic mutations that drive their obesity³³. This figure is likely to increase with wider accessibility to genetic testing and as new genes are identified from exome and genome sequencing through research programmes and through partnerships and national initiatives such as 100,000 Genomes Project.

Some genetic obesity syndromes are associated with learning difficulties and developmental delay (Prader-Willi syndrome) and major clinical problems (for example, visual loss/renal abnormalities in Bardet-Biedl syndrome) which mean that children come to medical attention at a young age. However, over the last 15 years there has been increasing recognition that there is a large and increasing group of genetic disorders where severe obesity itself is the presenting feature^{34,35}. These children are often identified as a result of early and marked weight gain but the lack of other clinical features often mean that a genetic diagnosis is not considered in many, and may only be offered when they reach secondary care. There are currently limited specialist facilities to support the assessment, early diagnosis, and appropriate management of such patients in the NHS. This poses a challenge as some current and several new clinical guidelines (including those from the Endocrine Society) for childhood obesity will recommend that genetic testing is performed in patients with severe obesity where clinical features of genetic obesity syndromes exist and/or where there is a family history of severe obesity.

5.2 Impacts of rare genetic obesity syndromes

The diagnosis of a genetic obesity syndrome can provide information that has diagnostic value for the family to whom genetic counselling can be provided. There is particular value of a genetic diagnosis in severe obesity which, unlike other clinical disorders is often not recognised as a medical condition by some health care professionals, educators, and employers. The making of a genetic diagnosis can help children and their families deal with the social stigma that comes with severe obesity and in some instances, where the persistence of severe obesity despite medical advice has been considered a reason to invoke parental neglect, the making of a genetic diagnosis has prevented children from being taken into care.

A genetic diagnosis can inform management (many such patients are relatively refractory to weight loss through changes in diet and exercise) and can inform clinical decision making regarding the use of bariatric surgery (feasible in some; high risk in others). Importantly, some genetic obesity syndromes are treatable.^{36,37} There are a number of drugs in Phase 1b/2 clinical trials targeted specifically at patients with genetic obesity syndromes (www.rhythmtx.com). Mechanisms are therefore needed to provide genetic diagnoses and enable stratification of patients for appropriate treatment within the healthcare system.

5.3 Obesity as a neurobehavioral disorder

Building on work conducted in animal models, the discovery of these rare but highly penetrant clinical disorders presenting with severe obesity alone has proved that genetic/biological mechanisms influence weight.³⁸ These studies have paved the way for understanding the molecular mechanisms and physiological pathways that regulate energy intake³⁹ and expenditure in humans. Ultimately, understanding how these pathways are disrupted in people with weight problems will inform strategies to target these pathways for prevention and treatment.

To date, many of the genes which cause severe obesity affect the leptin-melanocortin pathway.⁴⁰ Leptin is a hormone made by fat, which circulates in the bloodstream as a signal reflecting energy stores.^{41,42} In the brain, leptin regulates neuronal circuits which act to increase/decrease energy intake in response to changes in energy balance.⁴³ Genetic disruption of these circuits (several components constitute the melanocortin pathway) can cause severe obesity.⁴⁴ The most consistent feature associated with genetic disruption of this pathway is hyperphagia, an increased drive to eat.⁴⁵ These genetic studies and others have shown the importance of the brain pathways that regulate the drive to eat,^{46,47} which from an evolutionary perspective is critical to defend against starvation and ensure sufficient energy stores for survival.⁴⁸

5.4 Genetic variants influencing metabolic rate

Energy expenditure (how you burn calories) consists of Basal Metabolic Rate (BMR), the fundamental energy needed for cells, tissues and organs to work, energy expended during voluntary physical activity such as exercise, non-exercise activity thermogenesis (NEAT), spontaneous movement such as fidgeting, with a small fraction due to diet-induced thermogenesis (energy needed to digest and absorb food).⁴⁹ A number of large family based population studies have addressed the contribution of genetic vs environmental factors to energy expenditure including physical activity.²⁶ For example, exercise participation within families is entirely accounted for by shared family environment. However, BMR, which is the major determinant of energy expenditure (70%), is highly heritable.⁵⁰ Whilst most of the genes associated with severe obesity do so by affecting appetite, the recent finding that obese people harbouring genetic variants in KSR2 (Kinase Suppressor of Ras2) have reduced BMR demonstrates that genetic variation in energy expenditure can contribute to weight gain in some individuals⁵¹.

5.5 Genetics as a tool for drug discovery and validation

Genetics approaches can be powerful strategies for drug target discovery and validation (or invalidation). Adding genetic evidence to support the role of a drug target in disease can increase the chances of success in drug development.⁵² This in turn reduces the timeline from target discovery to clinical trials and the cost burden of failed clinical trials due to ineffective drugs that lack efficacy. For example, the discovery of rare loss of function variants in the melanocortin 4 receptor (MC4R) provided human evidence to support the development of agonists targeting this receptor³⁴. The characterisation of the phenotype associated with MC4R mutations revealed a role for MC4R mediated signalling in modulating blood pressure.⁵³ These studies predicted that MC4R agonists might lead to an unfavourable increase in blood pressure, an effect that was observed with some compounds which were not taken forward in Phase 3 trials for this reason. Second generation compounds targeting this compound now exist and appear to be effective at inducing weight loss without concomitant increases in blood pressure.⁵⁴ Genetic studies in obesity and indeed in thinness (or obesity resistance)⁵⁵ may have an important role to play in drug development.

6. Common genetic variants – relevance for prediction

6.1 Genetic discoveries in common obesity

Genome-wide association studies (GWAS) seek to identify common genetic variants (those present in at least 5% of people) that contribute to common diseases (obesity) or traits (BMI). By comparing very large numbers of people on whom BMI data is available, more than 100 genetic loci associated with BMI and body fat distribution (often measured by waist-to-hip ratio) have been identified; many of these have been replicated in different ethnicities.⁵⁶ GWAS associated loci are often identified by the name of the nearest gene (although this may not be the gene in which variation contributes to variation in BMI). Some of the obesity GWAS loci encompass genes previously shown to play a role in energy balance in animals and in people with severe obesity (e.g. *LEPR*, *SH2B1*, *MC4R*, *BDNF*).⁵⁷ Other loci contain genes that seem to be plausible biological candidates, or have suggested genes for which there was no previous evidence.⁵⁶ A large proportion of obesity/BMI loci contain genes that are expressed in the brain.⁵⁶ Where well-powered studies for the more frequently occurring variants have been performed, obesity-associated variants seem to be associated with increased energy intake rather than decreased energy expenditure (although energy expenditure measurements have only been performed in a few cohorts).⁵⁸⁻⁶⁰ This contrasts with the loci associated with body fat distribution, which are enriched for genes expressed in adipose tissue and overlap with loci for insulin resistance which has a shared pathogenesis.⁶¹ These findings, whilst not conclusive, suggest that the common genetic variants identified in these studies modulate physiological processes that ultimately impact on energy balance.

6.2 Potential relevance for public health

There are several challenges to interpreting the potential impact of such studies. Cumulatively, the common variants identified in obesity/BMI GWAS are characterised by modest effect sizes and the proportion of variance of BMI explained by GWAS-identified loci to date remains relatively modest (<5%). At the most, single variants predict that a person's weight may increase by a few kilograms over a 10 year time frame. Given the large number of common genetic variants identified, one way of estimating their cumulative burden is to use Genetic Risk Scores (GRS) which aggregate information from multiple GWAS to summarise risk-associated variation across the genome.^{62,63} Some studies have shown that people with a high GRS consume more food based on dietary records and have a higher BMI⁶⁴; some have shown that higher GRS are associated with reduced levels of physical activity.^{65,66} However, the effects of these associations remain modest. Although GRS are statistically robust at the population level they do not take into account gene-gene or gene-environment interactive effects⁶⁷ which may in part explain why they have poor predictive power for any given individual and at present, are unlikely to have a direct impact on personal healthcare. Therefore, genotyping healthy individuals to identify those who carry a high burden of "obesogenic alleles" is not currently warranted, nor, at this time, would such genetic information usefully guide therapy for people with common forms of obesity.

Case study - Impact of a genetic diagnosis in a child with severe obesity

JF is an 11 year old boy living in Dorset – he weighs 112kg. He is generally well and in the top 20% of his class at school. He has always been heavy. Following referral by his GP, he saw a dietician at his local hospital where he sees a paediatrician. His mother is very concerned about his weight as she can see that he is now struggling to play football which he enjoys; she is concerned about his health and about him being bullied when he starts secondary school.

Despite the support of his mother and health care professionals, JF doesn't manage to lose weight. He starts to resent attending clinic as there is nothing the doctors can do to help him lose weight apart from repeating the advice about diet and exercise. The dietician questions whether he really is following the recommended diet as she feels he should be losing weight. JF and his mother find the encounters difficult and cancel some appointments.

On one occasion, JF attends the clinic with his younger brother, TF, who the physician notes is also heavy. Concerns are raised and social services become involved. There are regular visits from social workers to the family home. They talk about mealtimes, what food is provided and look in the kitchen cupboards. Mrs F explains that she tries to cook every night although it can sometimes be difficult as she is a single parent and works full-time; her husband died three years ago from a heart attack aged 45 (he had type 2 diabetes and had been obese all his life).

The paediatrician offers an appointment to see JF and TF together for a further assessment. At this point she sends blood samples to Cambridge for genetic testing as part of a further workup for the assessment of severe obesity. In parallel, social services move forward with a case conference as there are concerns that Mrs F is causing the children's obesity through neglect – this is now a Child Protection Issue.

Some months later, the genetic tests reveal that both brothers are heterozygous for a mutation in MC4R. There is a discussion between the physician in Cambridge, the local paediatrician and Mrs F to go through the implications for the family. Mrs F is tested; she does not carry the MC4R mutation (she has always been slim). It is likely that the boys inherited the mutation from their obese father. The physicians communicate this information to social services making clear that this result establishes a cause for the severe obesity in both brothers which cannot be attributed to neglect. Child Protection proceedings are halted. Mrs F is hugely relieved. She can work to support the boys with help from her local paediatrician who remains in contact with the Cambridge team.

7. Conclusions

Genetic factors contribute to a significant proportion of the variability in BMI in the population. The finding of multiple common variants in GWAS studies currently has limited utility in predicting weight related problems and the potential impact of interventions. As many of the GWAS signals identified, to date map to non-coding regions of the genome that may potentially be involved in gene regulation (rather than by directly disrupting particular genes), further experimental work will be needed to understand the mechanisms that underlie these associations. Exome sequencing of cohorts of obese, normal weight and lean people is well underway and is likely to lead to the identification of additional rare variants in new genes whose functions will need to be explored in cells, model organisms and humans. Establishing the functional relevance of rare variants (which outnumber common variants in the human genome) has diagnostic value, can inform drug development and provides opportunities for the development of precision/stratified medicine. Cumulatively, work in the genetics of obesity has shown that the variability in BMI in the population has a large genetic component. Recognition that genetic factors influence the susceptibility to weight gain is vital to the development of informed preventative and therapeutic strategies to address the public health impact of obesity and related disorders.

8. Suggestions for policy makers

- Facilitate the assessment, early diagnosis, and appropriate management of patients with severe early onset childhood obesity, including that due to single gene disorders within the NHS.
- Take into consideration the fact that genetic factors influence the susceptibility to weight gain and response to weight loss when designing and testing interventions.
- Recognise that, based on having bad luck in the genetic lottery, many people have a strong hard-wired drive to eat. It is those who suffer most when the environment provides stimuli such as food cues, constant food availability outside meals, cheap high energy dense food and large portion sizes. Efforts targeted at reducing those environmental stimuli are likely to disproportionately benefit those with high intrinsic susceptibility to obesity.
- Education alone, without major changes in the environment will be insufficient to reduce obesity prevalence. Obesity is largely not a “knowledge deficiency” disorder.
- The drive to obesity in some individuals is very strongly biological, and the adverse consequences for serious morbidity of continued nutritional overload on conditions such as diabetes can be severe. Bariatric surgery should be seen as an important and helpful option in the therapeutic armamentarium for these patients rather than being portrayed, as it is by many commissioning authorities, as a drastic, last-ditch and even somewhat dubious procedure.

9. References

1. Whitlock G, Lewington S, Sherliker P, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* 2009;373:1083-96.
2. Sperrin M, Marshall AD, Higgins V, Buchan IE, Renehan AG. Slowing down of adult body mass index trend increases in England: a latent class analysis of cross-sectional surveys (1992-2010). *Int J Obes (Lond)* 2014;38:818-24.
3. Berrington de Gonzalez A, Hartge P, Cerhan JR, et al. Body-mass index and mortality among 1.46 million white adults. *The New England journal of medicine* 2010;363:2211-9.
4. Peeters A, Barendregt JJ, Willekens F, Mackenbach JP, Al Mamun A, Bonneux L. Obesity in adulthood and its consequences for life expectancy: a life-table analysis. *Annals of internal medicine* 2003;138:24-32.
5. Kottke TE, Wu LA, Hoffman RS. Economic and psychological implications of the obesity epidemic. *Mayo Clinic proceedings Mayo Clinic* 2003;78:92-4.
6. Puhl R, Brownell KD. Bias, discrimination, and obesity. *Obesity research* 2001;9:788-805.
7. Fontaine KR, Redden DT, Wang C, Westfall AO, Allison DB. Years of life lost due to obesity. *JAMA : the journal of the American Medical Association* 2003;289:187-93.
8. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA : the journal of the American Medical Association* 2014;311:806-14.
9. Friedman JM. A war on obesity, not the obese. *Science* 2003;299:856-8.
10. Skinner AC, Skelton JA. Prevalence and trends in obesity and severe obesity among children in the United States, 1999-2012. *JAMA pediatrics* 2014;168:561-6.
11. Garnett SP, Baur LA, Jones AM, Hardy LL. Trends in the Prevalence of Morbid and Severe Obesity in Australian Children Aged 7-15 Years, 1985-2012. *PloS one* 2016;11:e0154879.
12. Moll PP, Burns TL, Lauer RM. The genetic and environmental sources of body mass index variability: the Muscatine Ponderosity Family Study. *American journal of human genetics* 1991;49:1243-55.
13. Magnusson PK, Rasmussen F. Familial resemblance of body mass index and familial risk of high and low body mass index. A study of young men in Sweden. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity* 2002;26:1225-31.
14. King NA, Caudwell P, Hopkins M, et al. Metabolic and behavioral compensatory responses to exercise interventions: barriers to weight loss. *Obesity (Silver Spring)* 2007;15:1373-83.
15. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behavior genetics* 1997;27:325-51.
16. Perusse L, Tremblay A, Leblanc C, et al. Familial resemblance in energy intake: contribution of genetic and environmental factors. *The American journal of clinical nutrition* 1988;47:629-35.
17. Friedman JM. Obesity in the new millennium. *Nature* 2000;404:632-4.
18. Silventoinen K, Magnusson PK, Tynelius P, Kaprio J, Rasmussen F. Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. *Genetic epidemiology* 2008;32:341-9.
19. Wardle J, Carnell S, Haworth CM, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *The American journal of clinical nutrition* 2008;87:398-404.
20. Allison DB, Kaprio J, Korkeila M, Koskenvuo M, Neale MC, Hayakawa K. The heritability of body mass index among an international sample of monozygotic twins reared apart. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity* 1996;20:501-6.
21. Sorensen TI, Price RA, Stunkard AJ, Schulsinger F. Genetics of obesity in adult adoptees and their biological siblings. *Bmj* 1989;298:87-90.
22. Bouchard C, Tremblay A, Despres JP, et al. The response to long-term overfeeding in identical twins. *The New England journal of medicine* 1990;322:1477-82.
23. Bouchard C, Tremblay A, Despres JP, et al. Overfeeding in identical twins: 5-year postoverfeeding results. *Metabolism: clinical and experimental* 1996;45:1042-50.

24. Poehlman ET, Tremblay A, Despres JP, et al. Genotype-controlled changes in body composition and fat morphology following overfeeding in twins. *The American journal of clinical nutrition* 1986;43:723-31.
25. Elder SJ, Neale MC, Fuss PJ, et al. Genetic and Environmental Influences on Eating Behavior - A Study of Twin Pairs Reared Apart or Reared Together. *The open nutrition journal* 2012;6:59-70.
26. Perusse L, Tremblay A, Leblanc C, Bouchard C. Genetic and environmental influences on level of habitual physical activity and exercise participation. *American journal of epidemiology* 1989;129:1012-22.
27. Fabsitz RR, Sholinsky P, Carmelli D. Genetic influences on adult weight gain and maximum body mass index in male twins. *American journal of epidemiology* 1994;140:711-20.
28. Montague CT, Farooqi IS, Whitehead JP, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* 1997;387:903-8.
29. Yeo GS, Farooqi IS, Aminian S, Halsall DJ, Stanhope RG, O'Rahilly S. A frameshift mutation in MC4R associated with dominantly inherited human obesity [letter]. *Nature genetics* 1998;20:111-2.
30. Ramachandrapappa S, Raimondo A, Cali AM, et al. Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J Clin Invest* 2013;123:3042-50.
31. Jackson RS, Creemers JW, Ohagi S, et al. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene [see comments]. *Nature genetics* 1997;16:303-6.
32. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Gruters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature genetics* 1998;19:155-7.
33. Bochukova EG, Huang N, Keogh J, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 2010;463:666-70.
34. Farooqi IS, Keogh JM, Yeo GS, Lank EJ, Cheetham T, O'Rahilly S. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *The New England journal of medicine* 2003;348:1085-95.
35. Farooqi IS, Wangensteen T, Collins S, et al. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *The New England journal of medicine* 2007;356:237-47.
36. Farooqi IS, Jebb SA, Langmack G, et al. Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *The New England journal of medicine* 1999;341:879-84.
37. Farooqi IS, Matarese G, Lord GM, et al. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest* 2002;110:1093-103.
38. O'Rahilly S, Farooqi IS. Human obesity as a heritable disorder of the central control of energy balance. *Int J Obes (Lond)* 2008;32 Suppl 7:S55-61.
39. Farooqi IS, Bullmore E, Keogh J, Gillard J, O'Rahilly S, Fletcher PC. Leptin regulates striatal regions and human eating behavior. *Science* 2007;317:1355.
40. van der Klaauw AA, Farooqi IS. The hunger genes: pathways to obesity. *Cell* 2015;161:119-32.
41. Maffei M, Halaas J, Ravussin E, et al. Leptin levels in human and rodent: measurement of plasma leptin and ob RNA in obese and weight-reduced subjects. *Nature medicine* 1995;1:1155-61.
42. Ahima RS, Prabakaran D, Mantzoros C, et al. Role of leptin in the neuroendocrine response to fasting. *Nature* 1996;382:250-2.
43. Farooqi IS, O'Rahilly S. Leptin: a pivotal regulator of human energy homeostasis. *The American journal of clinical nutrition* 2009;89:980S-4S.
44. Farooqi IS, O'Rahilly S. 20 years of leptin: human disorders of leptin action. *The Journal of endocrinology* 2014;223:T63-70.
45. O'Rahilly S, Farooqi IS. Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions. *Diabetes* 2008;57:2905-10.
46. Morton GJ, Meek TH, Schwartz MW. Neurobiology of food intake in health and disease. *Nature reviews Neuroscience* 2014;15:367-78.
47. van der Klaauw AA, von dem Hagen EA, Keogh JM, et al. Obesity-associated melanocortin-4 receptor mutations are associated with changes in the brain response to food cues. *The Journal of clinical endocrinology and metabolism* 2014;99:E2101-6.
48. Schwartz MW, Woods SC, Seeley RJ, Barsh GS, Baskin DG, Leibel RL. Is the energy homeostasis system inherently biased toward weight gain? *Diabetes* 2003;52:232-8.

49. Goldberg GR, Prentice AM, Davies HL, Murgatroyd PR. Overnight and basal metabolic rates in men and women. *European journal of clinical nutrition* 1988;42:137-44.
50. Bouchard C, Tremblay A. Genetic effects in human energy expenditure components. *International journal of obesity* 1990;14:49-55; discussion -8.
51. Pearce LR, Atanassova N, Banton MC, et al. KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* 2013;155:765-77.
52. Nelson MR, Johnson T, Warren L, et al. The genetics of drug efficacy: opportunities and challenges. *Nature reviews Genetics* 2016;17:197-206.
53. Greenfield JR, Miller JW, Keogh JM, et al. Modulation of blood pressure by central melanocortinergic pathways. *The New England journal of medicine* 2009;360:44-52.
54. Kievit P, Halem H, Marks DL, et al. Chronic treatment with a melanocortin-4 receptor agonist causes weight loss, reduces insulin resistance, and improves cardiovascular function in diet-induced obese rhesus macaques. *Diabetes* 2013;62:490-7.
55. Bulik CM, Allison DB. The genetic epidemiology of thinness. *Obesity reviews: an official journal of the International Association for the Study of Obesity* 2001;2:107-15.
56. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;518:197-206.
57. Wheeler E, Huang N, Bochukova EG, et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nature genetics* 2013;45:513-7.
58. Timpson NJ, Emmett PM, Frayling TM, et al. The fat mass- and obesity-associated locus and dietary intake in children. *The American journal of clinical nutrition* 2008;88:971-8.
59. Wardle J, Carnell S, Haworth CM, Farooqi IS, O'Rahilly S, Plomin R. Obesity associated genetic variation in FTO is associated with diminished satiety. *The Journal of clinical endocrinology and metabolism* 2008;93:3640-3.
60. Cecil JE, Tavendale R, Watt P, Hetherington MM, Palmer CN. An obesity-associated FTO gene variant and increased energy intake in children. *The New England journal of medicine* 2008;359:2558-66.
61. Shungin D, Winkler TW, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015;518:187-96.
62. Loos RJ. Genetic determinants of common obesity and their value in prediction. *Best practice & research Clinical endocrinology & metabolism* 2012;26:211-26.
63. Belsky DW, Israel S. Integrating genetics and social science: genetic risk scores. *Biodemography and social biology* 2014;60:137-55.
64. Qi Q, Kilpelainen TO, Downer MK, et al. FTO genetic variants, dietary intake and body mass index: insights from 177,330 individuals. *Human molecular genetics* 2014;23:6961-72.
65. Zhu J, Loos RJ, Lu L, et al. Associations of genetic risk score with obesity and related traits and the modifying effect of physical activity in a Chinese Han population. *PloS one* 2014;9:e91442.
66. Ahmad S, Rukh G, Varga TV, et al. Gene x physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. *PLoS genetics* 2013;9:e1003607.
67. Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genetics in medicine: official journal of the American College of Medical Genetics* 2007;9:528-35.

Chapter 8

Personalised prevention

Chapter lead and author

Dr Ron Zimmern¹

¹ Chairman, PHG Foundation

1. Summary

- Personalised prevention is now with us, particularly in the fields of rare diseases, infectious diseases and cancer. The stratification of populations into subgroups will provide the means for a more personalised approach to the common complex disorders. Genomics will play a major role, but other biomarkers may, in time, be even more important when risk predictive algorithms are constructed.
- Essential to this approach will be the need to understand the role of biological variation in populations, the nature of the combined effects of genetic and environmental factors in health and disease, the potential for subtyping existing disease categories for prevention and treatment, and the importance of distinguishing between preventive strategies that rely on reducing disease risk from those aimed at the early detection of disease.
- The purely top down, collectivist approach to prevention, which has in the past served public health so well, may not (on its own) be the most effective way to prevent disease. This must be complemented by a more personalised approach to the delivery of disease prevention. Health systems of the future will need to take this into account and place the individual citizen or patient at its core, empower them to take greater responsibility for their health and have greater regard to their personal values and wishes. The development of direct to consumer diagnostic services must be taken into consideration. For both these and conventional physician determined laboratory tests, society needs to seek an appropriate regulatory balance between safety and the need to encourage innovation.
- The issue of how personalised prevention is to be delivered, and by whom, must be determined. Given that the clinical interaction is at the heart of such delivery, it may be questioned whether the public health workforce is best placed to undertake personalised prevention or whether this should be a task explicitly given to clinical professionals.

2. Introduction

Personalised prevention in the context of genomics implies that genetic information is used to identify individuals at increased risk of disease, and target or tailor preventive strategies. This can apply to both, primary and secondary prevention and Chapter 10 of this report, 'Risk-stratified cancer screening', provides examples for the latter by demonstrating how genomics may assist the early detection of disease. For primary prevention, with the ambition to stop the disease from occurring in the first place, this should be seen as a complimentary strategy to traditional, universal public health approaches that attempt to shift the distribution of a risk factor in the whole population.

As Chapter 07 of this report, 'Genomics and obesity' has shown, the causes underlying a seemingly homogeneous disease or phenotype, e.g. childhood obesity, can vary widely and the same preventative intervention is unlikely to be similarly effective. Stratifying preventive efforts for common diseases according to disease subtypes with a similar underlying aetiology may therefore be a useful strategy to target interventions. However, there is not yet sufficient evidence to support this and it is important that the public and patients understand the limitations of existing commercial tests directed at consumers looking for "targeted" lifestyle advice. While it is now possible to use genetic variation to predict people at different levels of risk for a disease in the population, this information often only adds marginally to what we already know based on established risk factors for common, complex diseases such as type 2 diabetes or heart disease. However, this may change in the future and also differs according to disease type, specifically depending on whether any known good clinical or other predictors already exist. Hence, there is potential genomic prediction and personalised approaches to prevention for disease areas of great public health importance that remain relatively understudied in the context of disease prediction but are amenable to preventive interventions, such as psychiatric diseases.

3. Background

The potential for genomics to improve the treatment of disease is already with us. The management of cancers and the diagnosis of individuals with rare inherited disorders have been revolutionised as a consequence of learning from the Human Genome Project and subsequent scientific work. Clinical medicine is changing rapidly to accommodate genomics within its scope.¹ Much of this has been documented in this report. By contrast the potential of genomics and post genomic science for the prediction of risk, the prevention of common complex diseases and the promotion of health has been relatively neglected. To the extent that such potential exists, there has not been much opportunity within the public health community to use such knowledge as a tool for improving population health.

One reason may be the nature of classical public health practice itself.² Our Victorian forefathers epitomised this approach using legislation or directing interventions at the external environment, acting to ensure sanitation, clean water and decent housing for its citizens. A second reason is the tension between social and biological models of disease.³ Public health practitioners have focused on the former rather than the latter, emphasising the more distal and structural causes of disease while eschewing more proximal biological factors. A third stems from their underlying science, epidemiology. This is premised on the idea that populations be treated as if they were homogenous, concentrating on comparing average effects while ignoring the variation between individuals in a heterogeneous population. Genomic science makes that premise now much more difficult to accept. Population heterogeneity can no longer be ignored.

The main thrust of public health action has been to look at the aetiology of disease incidence, what Geoffrey Rose called the 'causes of incidence' as distinct from the 'causes of cases'. It has chosen also to focus on the wider environmental and structural determinants of disease.⁴ Personalised prevention seeks to focus on 'causes of cases' and to target more specifically the more proximate biological pathways to disease in individuals. Both are necessary and the placing of the one against the other is a false antithesis.

The role of genomics in the diagnosis and management of rare diseases and cancer has already been covered in some detail. This chapter will explore the personalised approach to prevention for common complex diseases.

4. Introduction to personalised prevention

Personalised prevention is an attempt to refocus attention to preventing disease at the individual level. It stems from the belief that disease prevention can be made more effective and efficient when used as a strategy to complement the classical public health population based approach.⁵ Many terms have been used when speaking of this personalised approach to medicine, including precision and stratified, and some have tried to distinguish between them.

We suggest that pragmatically we should use these terms interchangeably, and that personalised prevention should be thought of as preventive strategies and interventions for individuals that:

- manage their care in accordance with their biological characteristics and risk
- treat them as whole persons, and empower them to take greater responsibility for their own health
- have regard to their values and wishes

In this chapter we will:

- make the case for preventive strategies and interventions that take into account both environmental exposure and biological variation
- argue that the changing nature of society which places great value on individual autonomy should embed the individual citizen at the centre of health system design
- discuss how services be best organised to allow optimal implementation of such personalised interventions

We emphasise that, although the term we choose to use is personalised, the reality is that, for most common diseases, specific interventions will be tailored to stratified groups rather than individuals themselves.

4.1 The use of biological variation in preventive strategies

All who have studied public health will be aware of Geoffrey Rose and his distinction between population and high risk prevention.⁴ By so doing he in effect stratifies the population into two groups, those at very high risk of, or with, disease and the remainder. Prevention may be directed at the whole population, or by identifying those with disease or high risk of disease and treating those specifically. Using blood pressure as an example, high risk prevention will require us to diagnose those with hypertension and to treat them, whereas population preventive strategies might seek to lower the sodium consumption in that population by persuading or legally obliging food manufacturers to put less salt in their products.

A consideration of biological heterogeneity in populations allows us to extend Rose's concept by suggesting that populations may be further stratified into more than two groups. Biomarkers, genomic or others, such as proteomic or epigenetic, may be used as the tool for stratification, either on their own or combined in a risk predictive algorithm that may include environmental factors. The different strata may be dealt with differently, thus making stratified prevention another, albeit more sophisticated, form of high risk prevention; in effect a strategy of risk based prevention. Proof of principle for this approach has been shown by using genetic variants to compute breast cancer risk and how this might influence the age of onset at which they enter the mammographic screening programme in order to ensure that they do so at the same level of absolute risk.⁶

At present, the most compelling evidence for using genomics in disease prevention is to be found in rare high penetrance inherited disorders. By contrast there is little direct evidence so far for the effectiveness of such strategies in common complex disorders. However, a strong theoretical basis provides reason for optimism, based on the fact that all human traits and disease come about because of the combined effects of genetic and environmental factors and a belief that a better understanding of this interaction, using if necessary concepts derived from epigenetic considerations, will eventually bear fruit. Lack of evidence of benefit is not the same as evidence of no benefit. It is likely that in time evidence will accrue to show that the stratification of populations and more precise targeting of interventions will have a part to play in disease prevention. Sceptics, on the other hand, are not convinced that such knowledge will necessarily change behaviour.⁷ The riposte to this has to be that it is likely that while some citizens will respond positively, others will not.⁸

4.2 Use of algorithms in disease risk

There is now a very clear understanding that, rare inherited disorders excepted, a single genetic variant contributes very little to disease risk. In order for there to be clinically useful information to allow populations to be stratified, we will need to rely on a combination of genomic, environmental and other biological markers in a risk prediction algorithm. Their construction and subsequent testing in populations to determine the sensitivity, specificity and predictive value, and their ability to discriminate accurately the boundaries between strata, should not be considered trivial. Yet, as and when this can be done, a huge gain in efficiency may result. Population prevention as conceived by Geoffrey Rose is likely to give the greatest payback for the population as a whole. But it is for those who are at highest biological risk for whom a more precise and more directed intervention specific for each subtype of disease or for each risk stratum may be more effective. Conversely those at an a priori low risk may be spared from having unnecessary tests or treatments.

4.3 The combined effect of genetic and environmental factors

To understand the interaction between genetic and environmental factors is crucial. If we use obesity as an example it is likely that certain individuals may respond to some types of diet (low fat or low carbohydrate, for example) but not to others. Some evidence already exists; but with most such studies based on small sample sizes and of variable quality, it will be some years before the evidential base is sufficient to allow us to advise individuals accordingly.

Recent work in Type 2 diabetes has shown the interaction between BMI and genetic risk, but because BMI has so strong an effect, even the subgroup at highest genetic risk of diabetes has a relatively low absolute risk unless the individuals are also obese. In other diseases, the relationship may well be different, but from a population perspective it is in groups with high absolute risks that interventions will have the greatest effect.⁹ This is one of the reasons why it is likely that the use of biomarkers other than genetic variants will in the future be of greater utility in disease prediction and prevention.

4.4 Use of algorithms in disease risk

Biomarkers may also be used to categorise disease into different sub-types. At its most basic, there is clear evidence for most diseases that a high risk sub-category caused by a single gene defect exists. For breast cancer, these include individuals with BRCA1 or BRCA2 variants; for colo-rectal cancer, those with variants in the genes that cause familial polyposis coli (FPC) or hereditary nonpolyposis colon cancer (HNPCC).

In future years, it is likely that evidence will show similar success in the stratification of common complex disorders. Because, for these disorders, the contribution of each genetic variant to disease risk is small, genomic variants themselves might be less efficient components than using other biomarkers for differentiating subtypes of disease. While genomics might be used to inform and understand the disease pathogenesis, the tests that might more efficiently stratify disease risk or subcategorise disease, may well be based on some other form of biomarker, for example, proteomic or metabolomic.

The categorisation of disease into different sub-types has implications for management. Lung cancer patients who are EGFR +ve will be managed in a different way to those who are not. The prognoses of these subtypes will also differ, with EGFR +ve patients likely to respond to drugs such as gefitinib or erlotinib. In many cancers we can now also distinguish between those whose progression is likely to be swift from others where a more optimistic prognosis can be given to the patient. Oncotype Dx (<https://www.breastcancercare.org.uk/oncotype-dx>) is a test that can do this for breast cancer ER +ve Stage 1 or 2 patients.

4.5 Reducing the risk of disease versus early detection of disease

The term disease prevention may mean one of two things. It may refer to the reduction in the risk of the development of the disease, by acting on a risk factor to lower the probability that the disease develops in the first place. It may, alternatively, be used to refer to early detection of disease and delaying or mitigating its clinical effects. In public health circles these have been called respectively primary and secondary prevention. We suggest that this is a useful heuristic even though strict scientific considerations would see disease pathogenesis as a continuum. Both of these activities have the risk of overdiagnosis.¹⁰

4.6 Autonomy and the changing nature of society

The autonomy of the individual patient is now regarded as a basic tenet of medical law and ethics. Yet in conventional public health practice a paternalistic approach may prevail where attempts to change the behaviour of the individual citizen are thought to be both acceptable and desirable. Some now question this paradigm, believing that, as with clinical care, autonomy should prevail and that behaviour change should be a negotiated and more individually directed endeavour. The personalised medicine approach suggests how this might be achieved. Yet public health has traditionally been a collective enterprise, with emphasis on population values and solidarity. Such a shift will not necessarily be welcomed by some of its practitioners; even though it may well be that what we propose by way of personalised prevention to improve the health of our population might better fit the spirit of our age. Health protection activities, by contrast, should continue to be mediated through classical public health interventions.

We do not deny that individuals will need to change their behaviour if we are to improve the health of our population. The issue is whether this should be carried out through generalised messages from central government, or through a more nuanced programme of advice individually tailored to and following discussion with individual citizens. We must have strong regard to the importance of distinguishing those interventions directed at the outside world (tax on alcohol or reducing salt content in processed foods) from those mediated (and intended to be mediated) directly through the behaviour of individual citizens, by exhorting them to drink less or to reduce their salt consumption.⁵

4.7 Provision of personalised prevention

The pressing question, therefore, is through which mechanism will personalised prevention be brought to the citizen. We suggest that it is not helpful to directly equate all disease prevention with public health practice. Personalised is perhaps to be best viewed as a clinical sub-specialty. The two sets of activities may be conceived as interacting circles on a Venn diagram. Public health practitioners, or at least the majority of them, are more concerned with the distal and structural determinants of health, whereas clinicians are perhaps best placed to discuss personalised prevention within the clinical encounter. By definition, personalised prevention must take place through the agency of the individual. He or she is expected to change behaviour, or undergo screening, or in some other way to respond behaviourally to information.

With that understanding, how then can we best reap the benefits of personalised prevention and the genomics revolution? Should we build prevention into the training of every specialty including primary care? Should we develop a new specialty or subspecialty of preventive medicine under the auspices of either the Royal College of Physicians or of General Practitioners? Should personalised prevention be a core activity of primary care physicians? The Royal Colleges (medical and nursing), the Faculty of Public Health and the Academy of Medical Sciences should all be involved in debating how best to establish these activities within the UK health system. The thesis we advocate is that future health care must embrace such interactions between individuals and health care professionals. The exact mechanism by which they are to be delivered must be determined in other places.¹¹

4.8 Big Data

But just because personalised prevention focuses on the individual, the population cannot and should not be neglected because the population is no more than a set of individuals, and because it is only through studies of a large number of individuals, populations, that evidence can be adduced to support personalised interventions. In future years it is likely that evidence may not emerge through epidemiological controlled studies as we envisage today, but by collecting large amounts of real world data and using computing power and sophisticated statistical techniques to reclassify disease and stratify populations into separate sub-categories on the basis of a wide variety of biomarker data.

4.9 Direct to Consumer Services

One last point that needs to be made concerns the consumer and the commercial sector. It is our view that they, together, rather than governments or public health services are likely to be the prime drivers of innovation and activity in these areas, and must be embraced as part of the wider health system. Much of the recommendations from these sources will at present not have a sufficient degree of evidence behind them to substantiate their claims; but it is likely that in future years evidence will emerge to distinguish the clinically valid from the invalid.

The company 23andMe is a case in point. It has been perhaps the most successful of those selling tests direct to the public. However, its activities have not been entirely without controversy. In 2013 the FDA notified them and other similar companies that they were not compliant with their regulations (following 2 to 3 years of negotiation) and the company had to withdraw their health related products in the USA, but at the same time they began to market the product in the UK. In 2015, the company with FDA approval have started selling a modified health product again in the USA. These issues are complex and cannot be dealt with in detail here. Suffice it to say that several regulatory components are at play with any direct to consumer genetic testing: the regulation of laboratories, the regulation of test kits and the regulation of advertising claims. Regulatory concerns are primarily around the clinical validity (whether the test results are meaningful) of the tests marketed directly to the public, and the extent to which the evidence supports the claims made by a company in marketing the tests. The exact balance of regulatory effort is as yet to be fully worked out both in the USA and in Europe, where the new In-Vitro Diagnostic Device Regulation (2017/746) was adopted in April 2017 by the EU.

The responsibility of health policy is to ensure that citizens are given the necessary information to respond appropriately to commercial pressures and to regulate in a proportionate manner so as to provide a balance between ensuring the safety of citizens and allowing innovation to proceed without an unacceptable burden of bureaucracy.¹² We are as yet at the start of a long journey and it is more than likely that it will take time before a balance is achieved.

5. Case study

This case study was kindly provided by Prof Aroon Hingorani, Dr Marta Futema, Prof Steve Humphries (June 2017)

Case study - Familial Hypercholesterolaemia

An elevated concentration of low density lipoprotein (LDL)-cholesterol in the bloodstream is causally linked to development of coronary heart disease (CHD) in later life. In the general population, LDL-cholesterol concentration is influenced diet, lifestyle and by the additive effects of common, small effect, largely independently inherited genetic sequence variants (single nucleotide polymorphisms) located in >50 positions throughout the genome.¹

However, a substantial number of individuals throughout the UK and other countries are affected by a condition known as autosomal dominant familial hypercholesterolaemia (FH) that is caused heterozygous mutations in one of four genes (see Box 1).

Box 1 - Genes responsible for familial hypercholesterolaemia

Autosomal dominant FH is caused by a single mutant allele in the low-density lipoprotein receptor (LDLR: OMIM #143890), apolipoprotein B (APOB; OMIM #144010), or the gene for proprotein convertase subtilisin/kexin type 9 (PCSK9; #603776)^{2,3,4} Homozygous FH (HoFH), which is much rarer but also more severe, arises in offspring of parents who are both heterozygous for an FH mutation. An extremely rare autosomal recessive form of FH has also been described due to mutations in LDRRAP1 (OMIM #603813).⁵

Around 93% of UK FH patients have a mutant allele in the LDLR gene, 5% in APOB and 2% in PCSK9.⁶ Although other novel genes have been proposed,⁷ none has yet been independently confirmed.

Patients with autosomal dominant FH have an elevated concentration of LDL-cholesterol from the first year of life, with some affected individuals having lipid deposition in the skin (xanthomas) and around the eyes (xanthelasma). Untreated, FH patients exhibit an approximately 13-fold excess risk of CHD compared to the general population; men with the condition typically developing CHD in their 50s and women in their 60s.

The challenge

Although the incidence of autosomal dominant FH is frequently cited as being 1 in 500, recent reports indicate that it may be closer to 1 in 250.^{8,9,10,11} Cascade screening (Box 2) of first-degree relatives of affected individuals, as carried out in several countries in Europe, including Holland, and shown to be feasible in the UK, followed by high dose statin treatment of affected individuals, as recommended in England and Wales by the National Institute for Health and Care Excellence (NICE), has the potential to avert a substantial number of CHD events.

Box 2 - Cascade screening for FH

Cascade screening involves genetic testing of the relatives of a patient with mutation confirmed FH, using a blood or saliva sample. Cascade screening has been successfully implemented in Wales and Netherlands, is recommended by NICE, and is supported by the British Heart Foundation.

However, the initial identification of affected individuals is not straightforward. Ascertainment of index cases in the UK is currently opportunistic rather than systematic, usually relying on a patient presenting with early symptoms of heart disease or even a heart attack, or the incidental finding of an extreme LDL-C value during a health check. One problem is that LDL-C values in adults with FH exhibit substantial overlap with values observed among individuals from the general population, where a higher than average burden of common, small-effect cholesterol raising alleles can mimic the biochemical features of FH.¹² Thus biochemical screening for monogenic FH in adulthood can be inaccurate. For this reason, according to recent surveys and a national audit, the 15,000-20,000 FH patients currently treated by lipid clinics in the UK likely represents < 15% of the estimated 126,000 FH patients in the UK (based on an incidence of 1 in 500) or <7.5% of an estimated 252,000 FH patients (based on an incidence of 1 in 250).

An opportunity for genomic medicine

It was recently proposed that routine, population-wide biochemical screening for FH in childhood might provide a more accurate route to identifying index cases, the affected parent, and to then seed a cascade-screening programme of first-degree relatives.¹³ This is because the distributions of lipid values in FH cases and unaffected individuals are more widely separated in childhood than they are in adult life. An LDL-C concentration greater than 1.84 multiples of the median (MoM) for children aged 1-9 years was estimated to be diagnostic of FH with a detection rate (DR) of 85% for a false positive rate (FPR) of 0.1%. However, in large prospective outcome study of this strategy,¹⁴ the detection rate (DR) for mutation positive FH cases was lower at 54% and the false positive rate (FPR) was higher than predicted at 0.7%. Similar estimates were found in a retrospective analysis in the Avon Longitudinal Study of Parents and Children¹⁵ (see Figure 1).

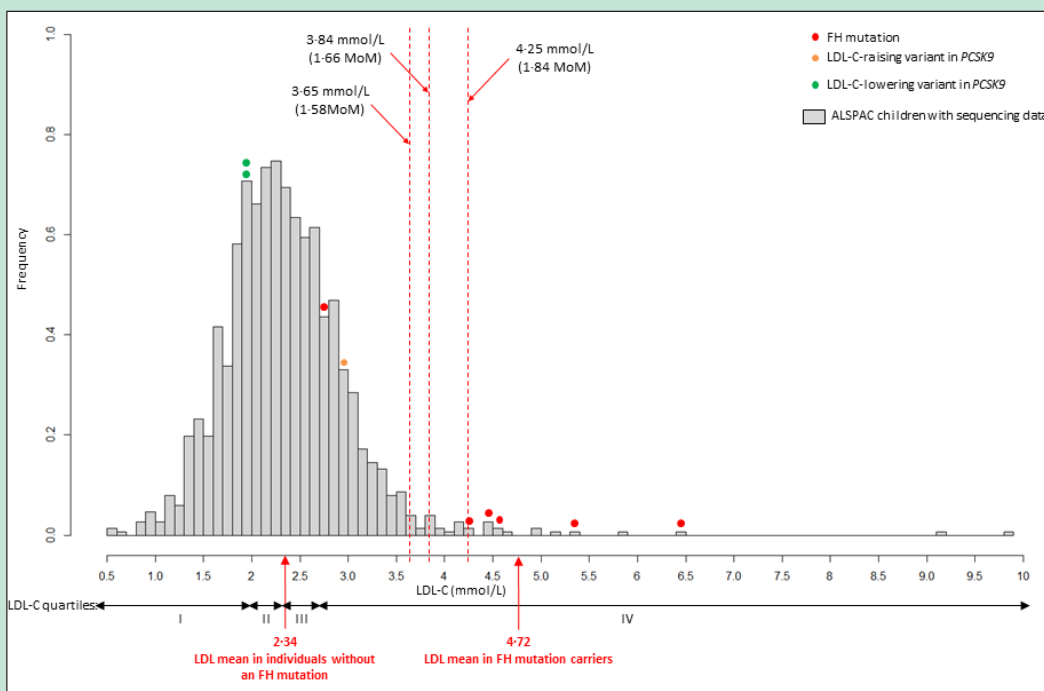
One way of addressing this issue might be to undertake next generation sequencing of FH genes in children whose LDL-cholesterol exceeds the screening threshold, to weed out false positives who are likely to have a high burden of common LDL-cholesterol alleles but not monogenic FH. Such a two-stage screen (low-cost, widespread biochemical screening followed by targeted sequencing of FH genes in biochemical screen positive samples) might be expected

to detect about half of the FH-mutation carriers in the population, a substantial improvement over the status quo. It would also detect FH patients with the highest cholesterol (and coronary risk), at a negligible overall false positive rate because of the very high accuracy of the sequencing step, and permit mutation-based testing of first-degree relatives. In theory, lowering the stage 1 cholesterol screening threshold would increase the number of people who would need to be sequenced but would also capture more individuals with FH mutations, without increasing false positives. Such a two stage screen, biochemistry followed by sequencing could also be evaluated following cholesterol measurement undertaken as part of NHS vascular health checks in adulthood.

The future

Systemic screening for FH based using a two-stage approach involving biochemical screening (stage 1), followed by targeted next generation sequencing (stage 2) would be expected to increase ascertainment of 'missing' FH patients, and should offer enhanced opportunities for primary prevention with statins or the newly-developed PCSK9 inhibitor drugs that effectively lower LDL-cholesterol.¹⁶ However, to test if such a screening approach were cost effective, whether in childhood or adulthood, will require a clinical trial.

Figure 1 - LDL-cholesterol values among 1512 children, mean age (SD) 9.9 years (4 months) from the Avon Longitudinal Study of Parents and Children.



Note Red-dashed lines indicate previously evaluated LDL-cholesterol cut points for biochemical screening for FH. Individuals with mutations are marked with dots.

Source Futema et al. *Atherosclerosis* 2017, 260 47-55).

Case study references

1. Willer CJ et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 2013. doi:10.1038/ng.2797
2. Sudhof TC, Goldstein JL, Brown MS, Russell DW. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science.* 1985; 228(4701):815-22.
3. Innerarity TL, Weisgraber KH, Arnold KS, Mahley RW, Krauss RM, Vega GL, et al. Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. *Proc Natl Acad Sci U S A.* 1987; 84(19):6919-23.
4. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet.* 2003;34(2):154-6.
5. Garcia CK, Wilund K, Arca M, Zuliani G, Fellin R, Maioli M, et al. Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science.* 2001; 292(5520):1394-8.
6. Humphries SE, Whittall RA, Hubbart CS, Maplebeck S, Cooper JA, Soutar AK, Naoumova R, Thompson GR, Seed M, Durrington PN, Miller JP, Betteridge DJ, Neil HA; Simon Broome Familial Hyperlipidaemia Register Group and Scientific Steering Committee. Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk. *J Med Genet.* 2006 Dec; 43(12):943-9.
7. Fouchier SW, Dallinga-Thie GM, Meijers JC, Zelcer N, Kastelein JJ, Defesche JC, Hovingh GK. Mutations in STAP1 are associated with autosomal dominant hypercholesterolemia. *Circ Res.* 2014 Aug 29;115(6):552-5
8. Benn M, Watts GF, Tybjaerg-Hansen A, Nordestgaard BG. Familial hypercholesterolemia in the Danish general population: prevalence, coronary artery disease, and cholesterol-lowering medication. *J Clin Endocrinol Metab.* 2012;97:3956–64
9. Do R., et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature.* 2015 Feb 5;518(7537):102-6
10. UK 10k Consortium. The UK10K project identifies rare variants in health and disease. *Nature.* 2015 Oct 1;526(7571):82-90
11. Abul-Husn NS, Manickam K, Jones LK, Wright EA, Hartzel DN, Gonzaga-Jauregui C, O'Dushlaine C, Leader JB, Kirchner L, Lindbuchler DM, Barr ML, Giovanni MA, Ritchie MD, Overton JD, Reid JG, Metpally RPR, Wardeh AH, Borecki IB, Yancopoulos GD, Baras A, Shuldiner AR, Gottesman O, Ledbetter DH, Carey DJ, Dewey FE, Murray MF. Genetic identification of familial hypercholesterolaemia within a single US health care system. *Science* 2016, 354 DOI 10.1126/science.aaf7000
12. Talmud PJ, Shah S, Whittall R, Futema M, Howard P, Cooper JA, Harrison SC, Li K, Drenos F, Karpe F, Neil HA, Descamps OS, Langenberg C, Lench N, Kivimaki M, Whittaker J, Hingorani AD, Kumari M, Humphries SE. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet.* 2013 Apr 13;381(9874):1293-301
13. Wald DS, Bestwick JP, Wald NJ. Child-parent screening for familial hypercholesterolaemia: screening strategy based on a meta-analysis. *BMJ* 2007; doi: <http://dx.doi.org/10.1136/bmj.39300.616076.55>
14. Wald DS, Bestwick JP, Morris JK, Whyte K, Jenkins L, Wald NJ. Child-Parent Familial Hypercholesterolemia Screening in Primary Care. *N Engl J Med.* 2016 Oct 27; 375(17):1628-1637.
15. Futema M, Cooper JA, Charakida M, Bousted C, Sattar N, Deanfield J, Lawlor DA, Timpson NJ, UK10K Consortium, Humphries SE, Hingorani AD. Screening for familial hypercholesterolaemia in childhood: Avon Longitudinal Study of Parents and Children (ALSPAC). *Atherosclerosis.* 2017 May; 260:47-55. doi: 10.1016/j.atherosclerosis.2017.03.007. Epub 2017 Mar 8
16. Hingorani AD, Futema M, Humphries S. Child-parent familial hypercholesterolaemia screening in primary care (Letter). *N Engl J Med* 2017 Feb 2;376(5):499. doi: 10.1056/NEJMc1615365

6. Conclusion

Personalised prevention is now with us, primarily in the fields of rare and infectious diseases and cancer. It will in time impact on the common complex disorders by providing a complementary approach to classical public health interventions. The stratification of populations into subgroups will be the initial means by which personalised disease prevention will be delivered, through more specific advice tailored to an individual's risk or a better understanding of their subtype of disease. Genomics will play a major role in this, but it is unlikely to be effective on its own unless it is used in conjunction with risk predictive algorithms in which multiple genomic variants are combined with other biological and environmental parameters within the model.

The importance of moving from a generalised top down model of health promotion to a more specific and personalised model is predicated on the reality that societal changes now pay much greater importance to individual autonomy. Health systems of the future must have regard to this fact. They must place greater importance on empowering citizens and patients to take greater responsibility for their own health and they must have greater regard for their values and wishes.

Public health practitioners interested in the organisation of health systems must take these matters into account, even though they do not deliver the service themselves. Should such activities be an explicit part of the workload of general practitioners or should it be imbedded across a variety of other clinical specialties? Whatever decision is made, policy makers should be under no illusion that the genomics revolution has made personalised prevention a reality which must be embedded as an activity within future health systems.

7. Suggestions for policy makers

- Ensure that the design of a health system for the 21st century places citizens and patients at its centre, allowing their care to be managed in accordance with both their biological risk and their personal wishes and values, and enabling priority to be given to preventive strategies and interventions.
- Establish facilities and systems that allow individual data to be collected, shared and used so as to be able to stratify them into groups in accordance with their biological and environmental risk.
- Consider how personalised prevention can be most effectively and efficiently delivered, and the respective roles of the public health and the clinical workforce in this endeavour, in particular whether the formal establishment of a sub-specialty or specialty of preventive medicine is needed.

6. References

1. Human Genomics Strategy Group (2012). Building on our Inheritance: Genomic Technology in Healthcare. Department of Health.
2. McKeown T (1976). The Role of Medicine: Dream, Mirage or Nemesis? Rock Carling Lecture. Nuffield Provincial Hospital Trust.
3. Kerr White (1991). Healing the Schism. Epidemiology, Medicine and the Public's Health. Springer
4. Rose G. Sick individuals and sick populations (1985). *Int J Epidemiol* 14: 32-38
5. Zimmern RL (2011). Genomics and Individuals in Public Health Practice: Are We Luddites or Can We Meet The Challenge? *J Public Health* 33:477-482
6. Burton H, Sagee GS, Pharoah P Zimmern RL (2012). Time to Revisit Geoffrey Rose: Strategies for Prevention in the Genomic Era. *Italian JPH* DOI: <http://dx.doi.org/10.2427/8665>
7. Hollands GJ, French DP, Griffin SJ, Sutton S, King S, Marteau T (2016). The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis . *BMJ*. 352: 1102 doi: <http://dx.doi.org/10.1136/bmj.i1102>
8. Burton H, Lueshi L, Zimmern R, Hall A (2016). Complex Diseases need Complex Solutions- including Genomics. <http://www.phgfoundation.org/blog/17195/>
9. Langenberg C, Sharp SJ, Franks PW et al (2014). *PLOS Medicine*. <http://dx.doi.org/10.1371/journal.pmed.1001647>
10. Welsh HG (2012). Overdiagnosis. Beacon Press
11. Burton H, Zimmern R (2015) Personalised Prevention and Public Health: An Urgent Agenda. <http://www.phgfoundation.org/blog/16582/>
12. Wright CF, Hall A, Zimmern RL (2011) Regulating Direct-to Consumer Genetic Tests: What is all the Fuss about? *Genet Med* 13: 295-300

Chapter 9

Pathogen genomics

Chapter lead

Prof Sharon Peacock¹

Chapter contributors

Eleanor Barnes, Nuffield Department of Medicine and the Oxford NIHR BRC, University of Oxford

Judith Breuer, Division of Infection & Immunity, University College London

Miles Carroll, National Infection Service, Public Health England, Porton Down

Derrick Crook, Nuffield Department of Medicine, University of Oxford, and Public Health England

Xavier Didelot, Department of Infectious Disease Epidemiology, Imperial College London

Mark Fife, The Pirbright Institute, Surrey

Andrew Fox, National Infection Service, Public Health England

Jane Greathorex, Public Health England, Cambridge

Ian Goodfellow, Department of Pathology, University of Cambridge

Paul Kellam, Imperial College, London

Donald King, The Pirbright Institute, Surrey

Julian Parkhill, The Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge

Tim Walker, University of Oxford

¹ London School of Hygiene and Tropical Medicine

1. Overview

Whole genome sequencing (WGS) promises to be transformative to the delivery of microbiology and public health functions involving the characterisation and surveillance of pathogens causing communicable disease. For the first time, their entire genetic information can be recovered in one step and shared, processed and analysed. Crucially, this can be linked to epidemiological and other related data, yielding complex multidimensional information which can be analysed as 'Big Data'. The benefits of exploiting such an innovative development in microbiology has been recognised in the UK and elsewhere, and a growing number of initiatives are defining the relevant applications and opportunities offered by WGS.

In the diagnostic laboratory, partial sequencing of HIV is already integrated into routine care for the detection of gene mutations associated with drug resistance. Newer sequencing technologies that undertake virus 'ultra-deep sequencing' (sequencing of numerous copies of the virus from the same individual) can detect when a small percentage of the HIV have developed such mutations, providing early warning of impending treatment failure. Looking ahead, the challenge for virology is to identify where the HIV model of testing should be replicated for other viruses. At the present time, WGS is not generally justified for bacterial identification and susceptibility testing. An important exception is *Mycobacterium tuberculosis* (the cause of tuberculosis), for which there is a strong case for the routine use of WGS to predict drug resistance and so guide treatment, and to support outbreak investigations.

Beyond the diagnostic pathway, the proactive implementation of WGS for foodborne pathogens will enable the early recognition of outbreaks. WGS could also bring a step-change to infection control investigations in hospitals, where the combination of pathogen WGS and epidemiological information can confirm or refute new outbreaks with much greater resolution than was previously possible. WGS can also demonstrate pathogen transmission between hospitals and countries and will be increasingly used during the management of international outbreaks, including those associated with novel or re-emerging pathogens.

Adopting this technology into routine practice will require changes in the design, operation and workforce of laboratories. In particular, the skill sets needed are mostly unfamiliar to the current workforce, and experts are in short supply nationally and internationally. The speed at which WGS could be implemented is tempered by several obstacles that need to be overcome. These include the optimal preparation of samples; fast, accurate and cheap sequencing platforms; and fully integrated software for processing, analysing and reporting the sequence information. Each of these requires a major development programme in its own right, necessitating commitment analogous to a major infrastructure project that is mostly beyond the resources of a single laboratory or commercial enterprise.

For implementation, a number of requirements will need to be met. An integrated suite of software is needed that supports the analysis of each microbial species. These will require thorough validation to demonstrate performance that is at least equivalent to current methods, and accreditation, at least by the United Kingdom Accreditation Service (UKAS). Management and operation of the software will need to be within a standardised production unit and separate from the software development environment. Finally, all sequence data with limited linked metadata will need to be deposited in publically accessible databases.

2. Overview of diagnostic microbiology

Diagnostic tests for infectious diseases are an essential part of contemporary medical care. The overwhelming majority of testing is undertaken in microbiology laboratories, who are required to provide an accurate, cost-efficient service with the shortest possible turnaround times. Bedside diagnostic tests represent an ideal in terms of speed and ease of use, but are not available for most infectious diseases.

Diagnostic microbiology laboratories regularly adopt new technologies that offer improvements in test performance and/or benefit patient care. This has resulted in laboratories whose component parts are made up of a patchwork of different methods and technologies that work together as a whole. No two laboratories in England are identical, but all provide a common diagnostic pathway. This begins with testing to detect the presence of pathogenic organisms in a wide variety of specimen types (e.g. blood, urine, pus, respiratory secretions, cerebrospinal fluid and stool). The majority of specimens either fail to have an organism identified, or are reported as 'no significant growth'. For example, around 90% of all blood cultures taken to investigate suspected bloodstream infection do not grow a pathogen associated with true bloodstream infection.^{1,2} For the remaining samples that are positive for a putative pathogen, these undergo identification, and may be tested for their susceptibility to antimicrobial drugs, typed if suspected to be part of an outbreak, and more rarely sent to reference laboratories for additional specialist testing. Samples that are positive for viruses are not routinely tested for susceptibility to anti-viral drugs because of the complexity of such assays, with a few notable exceptions (e.g. HIV).

The diagnostic microbiology laboratory also provides important public health functions, including the isolation of organisms that may be linked with foodborne outbreaks. Infection control is inextricably linked to microbiology laboratories, with surveillance processes in place that routinely monitor positive results to detect patterns that are consistent with outbreaks. There are also a range of specialist laboratories that provide extended diagnostic testing, or that have specialist responsibilities (e.g. food and water microbiology, emergency response). Dedicated veterinary laboratories provide diagnostic services to animal patients.

3. Current and future role of genomics in the diagnostic pathway

Most testing for bacteria and fungi is currently performed using culture-based methods for isolation followed by biochemical or other tests for identification, with a minority of molecular-based tests used for specific purposes. Culture-based tests have predominated because most pathogenic bacteria grow readily in rich laboratory media, and these assays are relatively inexpensive and can be mechanised to provide efficient workflows. By contrast, molecular testing has been the mainstay of diagnostic testing for viral diseases for several decades, largely because viruses are difficult to grow in the laboratory.

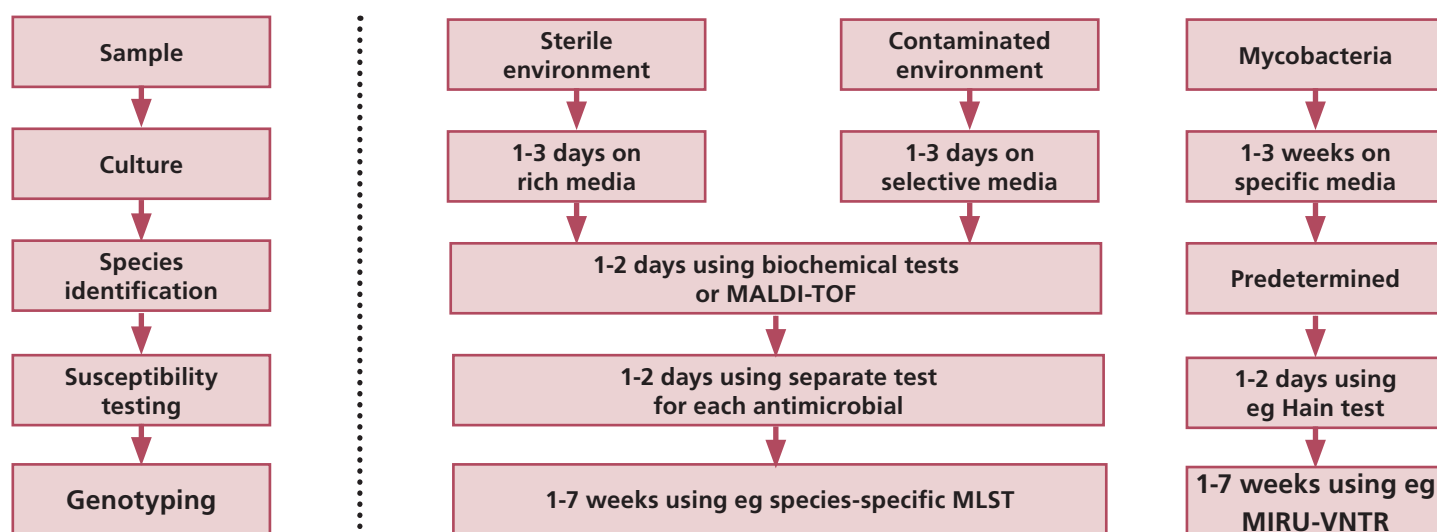
3.1 Bacteria

An illustration of the principles of current processing of bacterial pathogens adapted from work published elsewhere³ is shown in Figure 1. This simplifies the complexity of workflows in a standard diagnostic laboratory, but sets the scene for consideration of the utility of WGS.

Figure 1 is a simplified representation of the current workflow for processing samples for bacterial pathogens, showing a typical timescale. This highlights the main steps in the workflow, and is not intended to be a comprehensive or precise description. Samples that are likely to be normally sterile (e.g. blood) are often cultured on a rich medium that will support the growth of any culturable organism. Samples that are invariably contaminated with colonizing bacteria as well as the potentially infecting pathogen (e.g. stool, sputum) are grown using selective media that favours the growth of the suspected pathogen and may suppress common 'bystanders'. Identification and antimicrobial susceptibility testing is commonly performed if the culture grows a putative pathogen, but genotyping is limited to use during the investigation of suspected outbreaks. Mycobacteria identification is achieved before the end of the culture step by microscopic detection of growth. MALDI-TOF refers to a method of bacterial identification called matrix-assisted laser desorption/ionization-time of flight mass spectroscopy. MLST, multi-locus sequence typing. MIRU-VNTR, mycobacterial interspersed repetitive unit-variable number of tandem repeat.

Using WGS for bacterial detection and identification alone is not justified at the present time as this information can be determined more rapidly and cheaply using existing methods. Current bioinformatics methods to determine the species from a genome are also not perfect.⁴ Routine susceptibility testing is largely performed using culture-based methods that take around 16-24 hours to complete. Prediction of antimicrobial susceptibility based on genomes has been reported to be in good agreement with standard methods for the bacterial pathogens *Staphylococcus aureus*,⁵ *Escherichia coli* and *Klebsiella pneumoniae*.⁶ However, the current turnaround time of sequencing and data interpretation and lack of standardised software prediction tools means that WGS is not currently a competitive technology for most susceptibility testing. An important exception is *Mycobacterium tuberculosis*, the cause of tuberculosis (TB). This organism grows very slowly in the laboratory, and although rapid molecular tests are available which can be performed directly on sputum (e.g. Xpert MTB/RIF) to provide identification together with genetic prediction of resistance to rifampicin (a marker for multidrug resistance), susceptibility testing against a panel of drugs may take several weeks to complete and is complex and laborious when extended testing is required for drug resistant strains. Furthermore, recent advances allow DNA extraction, purification, and sequencing from early positive liquid TB cultures within one to two weeks of sample receipt.^{7,8} Case Study 1 makes a strong case for the introduction of WGS of *M. tuberculosis* as a matter of routine.

Figure 1 – Principles of current processing of bacterial pathogens



Case Study 1 – New approaches to the treatment and control of tuberculosis

Case Study

A patient presented to his local hospital complaining of a cough. A sputum sample was taken to investigate the patient for TB. Anti-TB therapy was delayed pending confirmation of the diagnosis and drug susceptibility results. Mycobacterial growth was detected in liquid culture after 11 days of incubation and the culture sent both to the reference laboratory for routine investigation, and for WGS (see Outline of a diagnostic work-up for *Mycobacterium tuberculosis*). The sequence-based report confirmed the culture as *M. tuberculosis*; predicted drug resistance to isoniazid, rifampicin, ethambutol, streptomycin and other aminoglycoside drugs; and linked it to a patient isolate sequenced in 2010 that was highly related (differed at only 7 of >4.4 million nucleotide positions in the genome). Standard drug susceptibility testing results for the 2010 isolate were the same as those predicted for the new isolate, with the addition of resistance to fluoroquinolone. One of the 7 nucleotide variants (gene mutations) accounted for this difference. Although the two patients lived 300 miles apart, both originated from the same country with a high incidence of drug-resistant TB. These drug susceptibility predictions were fed back to the reference laboratory, which fast-tracked the use of other methods (PCR probes for common drug resistance mutations, and traditional testing methods) to corroborate them. The patient was admitted to hospital and commenced on appropriate therapy, making a good recovery and minimising further chances of onward transmission.* The sequence data provided the opportunity to make informed judgements on whether transmission was likely to have occurred in the UK or in the country of origin, with direct impact upon the scale of public health investigation required.

* Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW; COMPASS-TB Study Group. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med.* 2016;4:49-58.

Underlying developments

WGS accurately diagnosed this patient and led to earlier initiation of therapy than would otherwise have been the case. It was able to:

- identify the mycobacterial species cultured in liquid medium by comparing its sequence to a catalogue of previously sequenced strains
- predict drug susceptibility based on a catalogue of genome mutations previously associated with resistance
- link the isolate to one previously sequenced case to make inferences about its origin

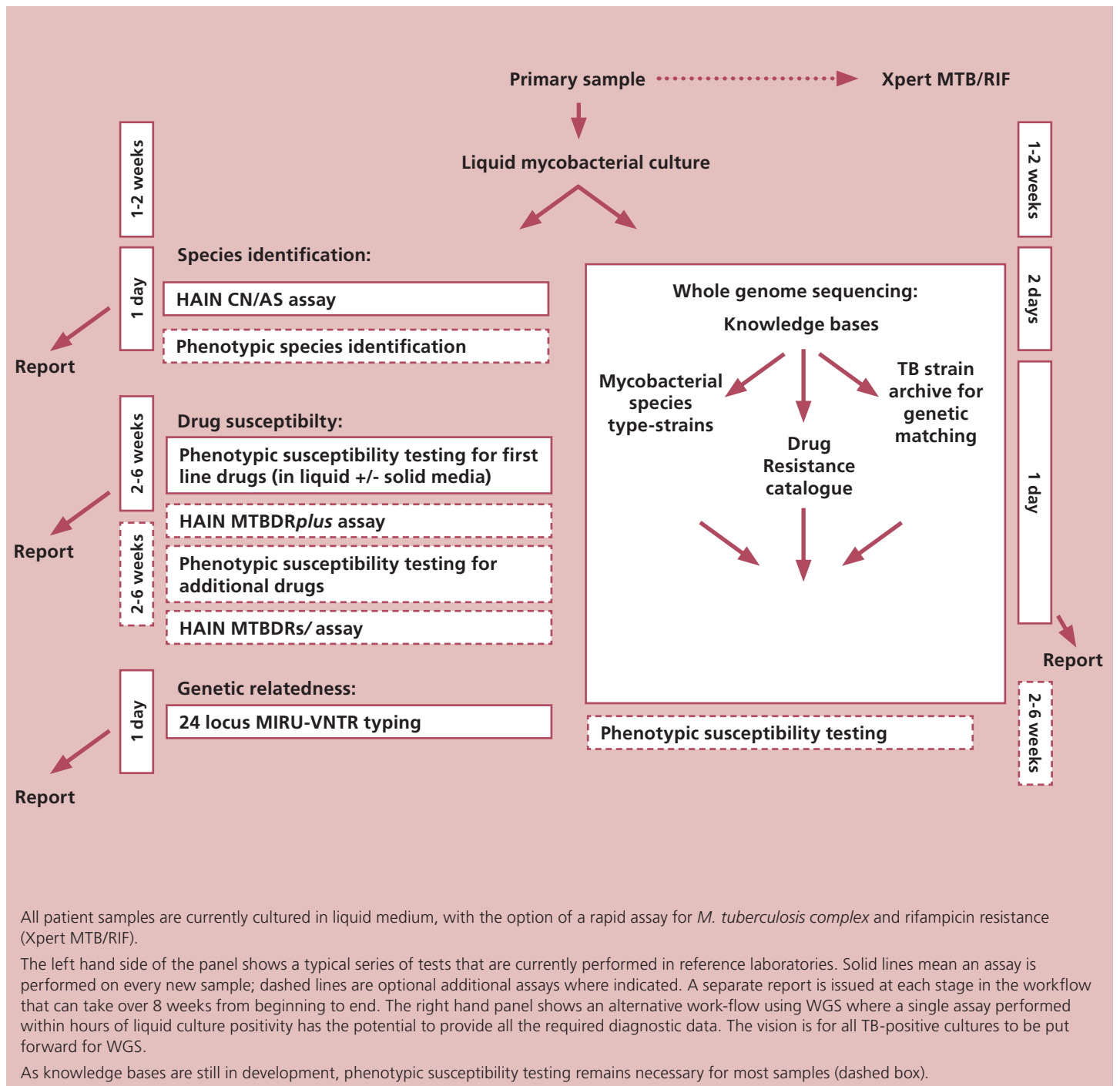
Looking ahead

Public Health England aims to introduce WGS nationwide as a routine diagnostic test to identify mycobacterial species, predict drug susceptibilities, and link genetically related samples and so guide public health investigations. The knowledge base for each diagnostic element will continue to evolve as mycobacterial species and sub-species are discovered and/or re-defined; as our understanding of the genetic determinants of drug resistance and susceptibility improves; and as the back-catalogue of historical isolates against which to search for potential transmission links grows. The current service model involves distributed sequencing (for example, in hospitals around the country) and centralised analysis. The process of analysing sequence data has until recently been labour intensive, but work is in progress on automating the analytic pipeline to manage the high sample-throughput demanded of a nationwide service. Almost all of the steps required to analyse the sequence data is already fully automated, with a residual manual step to access the sequence data before analysis.

Paradigm shift

Sequencing technology is leading to a change in laboratory workflow, but also to a paradigm shift in the way we approach epidemiology and drug resistance more generally. This Case Study demonstrates how sequencing now generates epidemiological hypotheses (as opposed to merely confirming them), and how drug-susceptibility data for past isolates can be used to corroborate drug susceptibility predictions for new isolates that are sufficiently genetically related (as opposed to inferring genetic relatedness from susceptibility patterns).

Figure 2 – Outline of a diagnostic work-up for *Mycobacterium tuberculosis*



3.2 Viruses

Key viral pathogens sought in the diagnostic laboratory are the blood borne viruses such as HIV, hepatitis C and hepatitis B for which anti-viral treatment is available, gastrointestinal pathogens such as norovirus and rotavirus, respiratory pathogens such as influenza and para-influenza, and viruses that can affect the central nervous system such as herpes viruses. In addition, specialist laboratories may be required to diagnose cases of viral illness imported from overseas, in particular the viral haemorrhagic fevers and various emerging viruses (e.g. Ebola, Zika and Middle East respiratory syndrome coronavirus (MERS-CoV)).

Capillary sequencing (the first commercially available sequencing technology) was introduced into clinical microbiology to test for HIV drug resistance in the early 1990s, a few years after the first cases were reported of people infected with HIV that had become resistant to one of the major anti-viral drugs (zidovudine).⁸ This method is based on the generation of sequence information from specific gene regions, rather than the entire genome. By the start of the 21st century, the availability of public HIV resistance databases⁹ meant that clinical reporting could be performed in a timeframe that was useful for patient management. Commercial systems were subsequently developed.¹⁰ The UK currently has around 15 clinical laboratories performing routine HIV resistance testing. Available databases are internationally curated, giving assurance that these contain accurate and regularly updated information.

Viruses evolve rapidly and may develop resistance in an infected person after drug treatment has started. The advantage of newer sequencing technologies is that they can undertake 'ultra-deep sequencing' (sequencing of numerous copies of the virus present in the patient sample). This can detect a small percentage of the viral population that have developed resistance, providing early warning of impending treatment failure following expansion of the resistant sub-population. Currently, a single laboratory in the UK uses ultra-deep sequencing to perform HIV resistance testing and reports these 'minority variants' to clinicians when present in more than 2% of the population. There is growing evidence for the utility of minority variant reporting,¹¹⁻¹³ and this is likely to be adopted elsewhere.

Looking ahead, the challenge for virology is to identify where the HIV model of testing should be replicated for other viruses. For most viral infections, full-length genome sequencing currently offers little advantage over existing molecular methods such as PCR (polymerase chain reaction) with or without targeted sequencing of fragments of the genome to detect resistance mutations or the genes in which they are found. The latter is standard practice for HBV, HCV, and in some laboratories for resistance testing of influenza and herpes viruses. The exception to this may be HCV, where new methods that sequence whole genomes have been shown to be superior to existing methods for multigene resistance testing and the detection of mixed infections.¹⁴ Research is required to define further areas where WGS of viruses could bring benefit to patients. One example is described in Case Study 2, which outlines the utility of cytomegalovirus resistance testing using ultra-deep sequencing in multi-visceral transplant patients, in whom early detection of developing resistance may allow rapid switching of anti-viral treatment.

Case Study 2 – Early detection of antiviral resistance in multi-visceral transplant patients could support precision anti-viral drug prescribing and lead to improved outcomes

Case Study

A multi-visceral (multi-organ) transplant patient infected with cytomegalovirus (CMV) had serial blood samples taken and tested to detect antiviral drug resistance using routine capillary sequencing (see Time course of CMV infection in a patient after multi-visceral transplantation). This methodology is the current standard for sequencing throughout the UK, but compared to sequencing by ultra-deep methods (sequencing of numerous copies of the virus from the same individual) is relatively insensitive. It will only detect resistance to anti-viral agents once the particular resistance mutation is present in around 20% or more of the total viral population. In this patient, despite repeated testing, resistance was only detected 156 days after transplantation. By this time, the virus was replicating exponentially, as can be seen by the rising blue line in the image. Despite switching to a different effective anti-viral agent which led to a fall in viral load, the patient died.

Underlying developments

The retrospective use of ultra-deep sequencing on stored samples from this patient detected resistance mutations in 3% of viruses in a blood sample taken 55 days after transplantation. This was some 101 days earlier than the routine capillary sequencing method. Early use of ultra-deep sequencing would have facilitated a more rapid switch to effective treatment and could have improved the outcome of this patient.

Looking ahead

More than 4,400 people in the UK had their lives saved or improved by organ transplantation during 2014-15. However, over the same period of time the number of people who donated organs fell (by 5%) for the first time in more than a decade.[#] An important added complication is the use of organs from donors who have had past infection with CMV. This infection is common in the general population (50-80%), and is often mild or asymptomatic in healthy people but results in long-term carriage of the virus. The likelihood that a donor is CMV positive is relatively high (28-68% in the UK depending on age).[~] All transplant patients who have no immune response to CMV are placed on prophylactic anti-viral treatment after transplantation surgery to help them suppress the virus. If the virus develops resistance, as in the case above, there may be serious consequences for the patient. The combination of shortage in donors and high background rates of CMV in the healthy population means that the need to manage CMV infection will be on-going and may increase alongside new transplantation advances.

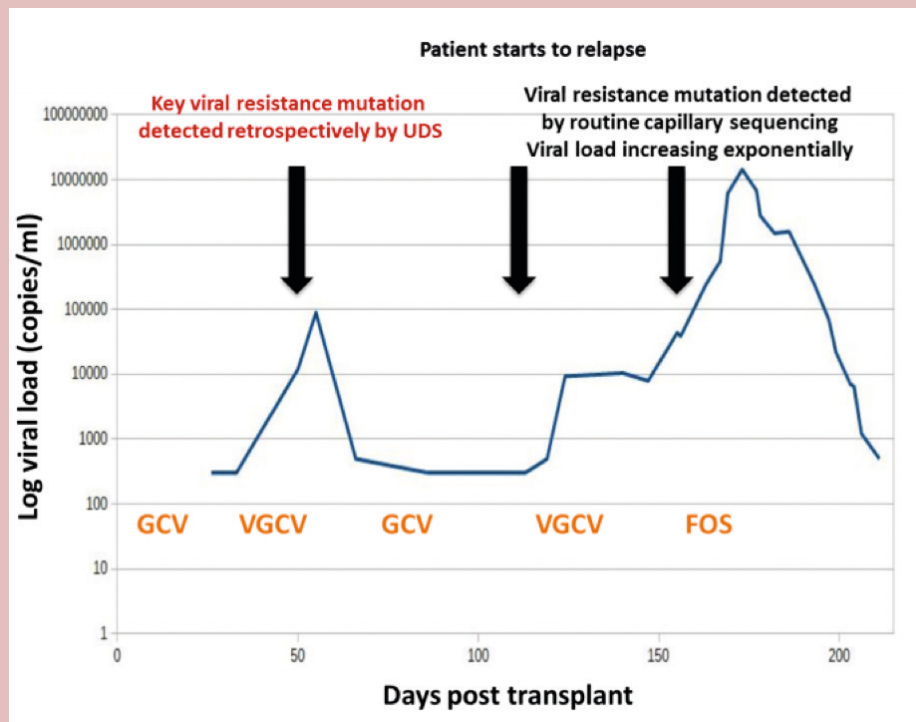
Paradigm shift

Use of more sensitive sequencing technologies such as the ultra-deep sequencing described here could help detect CMV resistance as it starts to develop and allow clinicians to change treatments, potentially improving patient outcome.

[#] HS Blood and Transplant, 2015. Organ Donation and Transplantation, Activity Report 2014/15.

[~] Johnson RJ, Clatworthy MR, Birch R, Hammad A, Bradley JA. CMV mismatch does not affect patient and graft survival in UK renal transplant recipients. *Transplantation*. 2009;88:77-82.

Figure 3 – Time course of CMV infection in a patient after multi-visceral transplantation



- The viral load (amount of virus in blood) is shown as a blue line.
- Orange text indicates the type and timing of anti-viral therapies given to the patient (GCV, prophylactic treatment; VGCV, first line treatment; FOS, second line treatment, commenced when the patient developed symptoms suggesting relapse of CMV infection).
- Black arrows indicate the points at which sequencing of the virus was performed. In red, results of retrospective detection (testing stored samples) using ultra-deep sequencing (UDS) indicates the point at which this methodology was capable of detecting the presence of a resistance mutation associated with drug resistance in the virus.
- Clinical symptoms of viral infection began on day 50 after transplantation. These were initially suppressed by drugs but recurred on day 119, and illness worsened at day 155. Routine capillary sequencing was only able to detect the viral resistance mutation at day 156.

4. WGS Beyond the diagnostic pathway

4.1 Epidemics and pandemics

In the early stages of an epidemic or pandemic, surveillance that is informed by microbial sequence data could rapidly identify transmission of the epidemic agent between countries or continents. Case Study 3 illustrates the process by which WGS analysis can demonstrate sources and directions of inter-country and inter-continental transmissions, in this case for foot-and-mouth virus. It is often unclear in the early stages of an outbreak how quickly spread is happening and so how much effort should be deployed to contain the outbreak. Sequencing can be used to inform this, as exemplified by the 2009 H1N1 flu pandemic when WGS was used to make an early assessment of transmissibility and severity based on an analysis of the outbreak in Mexico,

and provide early data on international spread and viral genetic diversity.¹⁶ WGS can also be used to assist in outbreak management, as recently demonstrated during the Ebola outbreak where this guided contact tracing in the final stages of the epidemic (see Case Study 4). WGS-based surveillance could also be used to detect and track newly emerging antimicrobial resistance of high clinical importance. This was recently demonstrated by the identification of novel transmissible colistin resistance in China, with immediate confirmation that this resistance element had already arrived in Europe using WGS data from the UK.¹⁷

Case Study 3 – Genomic analysis of the spread of foot-and-mouth disease virus

Case Study

Foot-and-mouth disease (FMD) is an infectious viral disease of cattle, sheep, pigs and goats. This occurs in many countries in Africa and Asia, from where it continuously threatens the livestock industries in FMD-free countries. Such spread can have dramatic impacts upon agriculture and the wider economy of an affected country, as experienced during the 2001 and 2007 outbreaks in the UK. The genome of foot-and-mouth disease virus (FMDV) is highly variable and evolves rapidly during replication, which allows the use of sequence data to reliably reconstruct the relationship between viruses recovered from different locations, or at different times. Complete genome sequences can also be used to define transmission routes at the farm-to-farm level using samples collected within outbreak clusters.

Sequencing was used in real time to understand and control the FMD outbreaks that occurred on eight farms in the UK during 2007.[±] Complete genome sequences were rapidly generated within 24-48 hours and were used to track FMDV movement from farm to farm in real time (see Use of viral nucleotide sequence data to trace the spread of FMD viruses). More recently, the Pirbright Institute has monitored the spread of an exotic FMDV lineage that has emerged from the Indian sub-continent, causing outbreaks in North Africa and the Middle East,^{±±} a dynamic situation that may relate to mass migration of people and political instability in the region. Elsewhere, the recent circulation of unfamiliar FMDV genotypes (termed topotypes) in East Asia has been attributed to increased demand for animal protein associated with rapid economic development. These unexpected long-range movements reinforce concerns about how readily FMDV can pass across international borders and raise questions about the heightened risks to the UK.

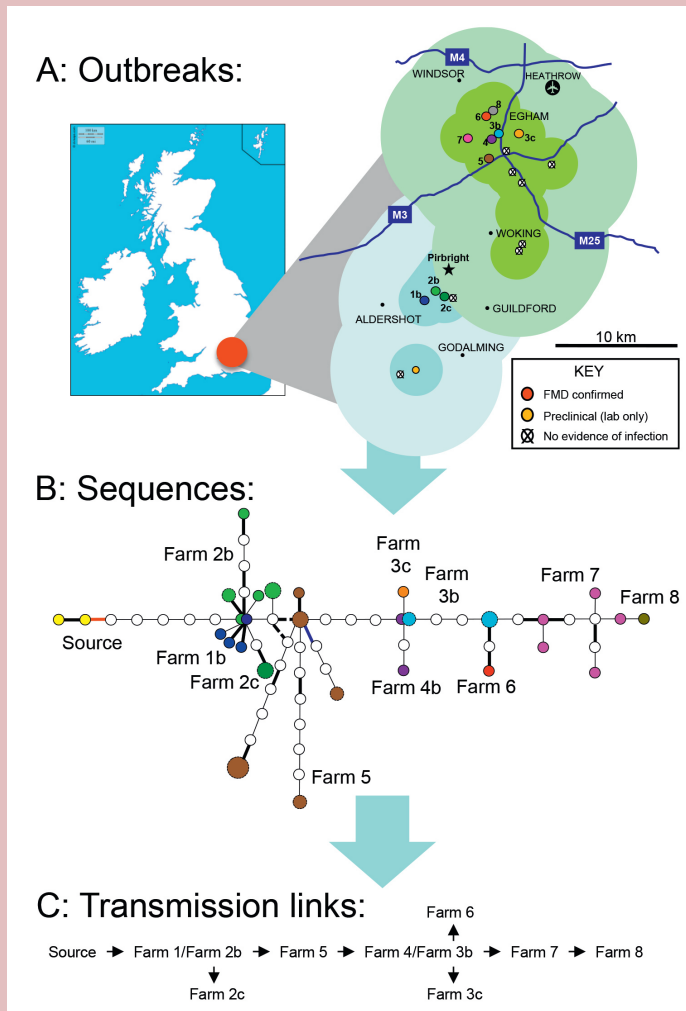
Looking ahead

Analyses of viral sequences can be used to predict the potential threat of incursion of novel FMDV lineages into the UK. Complete genome sequences will play a central role in the event of a future FMD outbreak in the UK and bespoke analytical, and statistical methods will continue to be refined to support high-resolution tracing of FMD outbreaks. There is now impetus to develop practical and inexpensive methods that allow viral sequences to be more reliably recovered from field cases of FMD, particularly for surveillance purposes in Asian and sub-Saharan African countries where exposure to modern sequencing technologies is currently limited.

± Knowles NJ, Bachanek-Bankowska K, Wadsworth J, Mioulet V, Valdazo-González B, Eldaghayes IM, Dayhum AS, Kammon AM, Sharif MA, Waight S, Shamia AM, Tenzin S, Wernery U, Grazioli S, Brocchi E, Subramaniam S, Pattnaik B, King DP. Outbreaks of Foot-and-Mouth Disease in Libya and Saudi Arabia During 2013 Due to an Exotic O/ME-SA/Ind-2001 Lineage Virus. *Transbound Emerg Dis*. 2016 Oct Vol 63 Issue 5, pages e431–e435, doi: 10.1111/tbed.12299.

±± Valdazo-González B, Kim JT, Soubeyrand S, Wadsworth J, Knowles NJ, Haydon DT, King DP. The impact of within-herd genetic variation upon inferred transmission trees for foot-and-mouth disease virus. *Infect Genet Evol*. 2015;32:440-8

Use of viral nucleotide sequence data to trace the spread of FMD viruses



This example highlights the application of complete genome sequences to reconstruct farm-to-farm transmission links during FMD outbreaks that occurred on eight farms (highlighted in different colours) in southern England during 2007.

A: Two waves of outbreaks occurred affecting two farms during August (3 and 10 km control zones highlighted in blue) and 6 farms in September (3 and 10 km control zones highlighted in green).

B: Sequences were recovered from 33 infected animals and were used to reconstruct the most-likely transmission links (C) between the farms.

Source Peacock S et al Composite figure based in part from figures in Cottam EM et al, Transmission pathways of foot-and-mouth disease virus in the United Kingdom (2007) and Valdazo-González B et al, The impact of within-herd genetic variation upon inferred transmission trees for foot-and-mouth disease virus. *Infect Genet Evol.* 2015;32:440-8.

Case Study 4 – Implementing sequencing technologies during the West Africa Ebola virus outbreak

Initial application of WGS to monitor virus mutation

The West African Ebola virus (EBOV) outbreak caused international alarm due to its rapid and extensive spread, and resulted in a significant death toll and social unrest within the affected region. There were several immediate questions asked by global health authorities relating to the virus mutation rate and the possible development of novel molecular changes which may result in pathogen adaptation to humans (and changes in biological patterns such as rates of spread). However, due to logistical and ethical issues transport of samples out of the region to advanced molecular laboratories was delayed. In July 2014, a selection of 76 samples from individuals in northern Sierra Leone who were infected during the last week of May and the first week of June were transported to North America where they underwent sequencing. Initial results published in late 2014 suggested a mutation rate that was higher than the rate previously estimated and possibly approaching that of influenza, which enhanced the concerns of international health agencies.

International application of WGS to accurately assess virus genetic changes and transmission

EBOV from the first year of the outbreak were subsequently sequenced by laboratories in Europe, North America and China, which indicated a slower mutation rate than initially thought. Furthermore, although mutations had occurred in the genetic code for the viral outer coating (the glycoprotein, which is the target of new vaccines and immunotherapeutic agents undergoing clinical assessment)¹ these were uncommon. Advanced molecular epidemiological analysis of the many hundreds of recently sequenced EBOV genomes revealed an accurate picture of how the virus spread throughout the three key West African countries. Most of the data generated was rapidly shared with international health agencies.

¹ Carroll MW et al. Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature*. 2015;524(7563):97-101.

² Quick J et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016;530(7589):228-32.

Innovative real-time WGS methods lead to actionable information

Retrospective analysis does not provide real time information and so limits the impact that WGS can have on outbreak control. To address this, several groups working in Guinea and Sierra Leone used real time WGS approaches to rapidly generate viral sequences from new cases. A novel experimental mini-sequencing device (the MinION, developed by Oxford Nanopore Technologies) was deployed to Guinea. British scientists established a MinION platform in Conakry Guinea on April 13th 2015 and within 48 hours, the first full genome of EBOV was sequenced on site.² The technology was rapidly transferred to the European Mobile Laboratory at a nearby Ebola Disease Treatment Centre. The laboratory was then able to provide a complete genetic fingerprint of new positive EBOV cases within as little as 18 hours. From this, maps of virus spread were constructed from genomic data, which were used to assist front line workers to identify and break transmission chains. At the same time, UK scientists performed sequence analysis of new Ebola cases in Sierra Leone using a more conventional but less mobile platform, generating a large reference database of up-to-date viral sequences from recent cases. Importantly, the data generated by these laboratories was shared in real time, enabling them to identify ongoing cases of cross boarder movement of Ebola positive cases. This information was shared with the local authorities, making them aware of the failures to monitor all crossing points.

Real time sequencing also proved invaluable during the final stages of the epidemic. In one example of many, hours after Sierra Leone was declared EBOV free a positive case was identified who had no known links to any known previous cases. Within 48 hours, scientists in Sierra Leone had used real time sequencing and produced a report to the local authorities that the new case was not a new introduction from another affected country or animal source, nor was it linked to an unmonitored transmission chain. Instead, WGS data confirmed that the new case was linked to a persistently infected survivor who become infected in November 2014. Such advances in sequencing platforms and translation of their outputs for field epidemiologists will be incorporated into the response to future outbreaks.

4.2 Foodborne outbreaks

Foodborne illness results from the consumption of food contaminated by one of a range of different pathogens. An outbreak investigation is undertaken by public health officials when two or more cases appear to be linked by time and/or place. Evidence used to confirm or refute an outbreak may include typing of pathogens that have been recovered from affected individuals, leftover food or other material. Typing information is compared to determine the degree of relatedness between strains. WGS will ultimately supersede other methods because of its vastly superior discriminatory power. WGS of the large outbreak of *E. coli* O104:H4 centred in Germany in 2011 was one of the first examples of its use, and drove innovation including 'crowd sourcing' to exploit the technology to help identify and control the situation.¹⁸⁻²⁰ Outbreaks are currently managed responsively (that is, investigated when a link between two or more potentially related cases comes to light). This means that recognition may occur late (as occurred with the Germany outbreak), and probably grossly underestimates the frequency of smaller outbreaks. A solution to this is the proactive use of WGS. For example, several national reference centres are implementing routine sequencing of major foodborne pathogens such as *Salmonella* species to enable early recognition of highly related organisms, which trigger an investigation. As genome databases increase in size, the precision of source attribution will improve and provide increasingly targeted investigations.

4.3 Spread of transmissible pathogens in hospital

WGS has been shown to be effective in tracking the spread of transmissible infectious pathogens at different scales, from person-to-person through to country-to-country.²¹⁻²⁴ WGS-based monitoring of transmissible pathogens in hospitals could rapidly identify new outbreaks earlier than traditional approaches, and can identify complex transmission networks (e.g. between multiple wards) that may not be readily apparent by standard infection control methods. The higher resolution of sequencing can also refute putative hospital outbreaks much more accurately than existing less discriminatory typing tools, which could save time and money spent on managing non-existent outbreaks. In an extension of hospital-based outbreak analysis, this can be used to demonstrate transmission of nosocomial pathogens between hospitals and countries. While most bacterial WGS to investigate putative outbreaks has taken place retrospectively so far, microbial sequencing to link cases has been shown in a few instances to be of clinical benefit.^{24,24}

4.4 Vaccines and biological insights

WGS-based surveillance is likely to inform vaccine development for specific pathogens, initially by monitoring the effect of newly introduced vaccines on the pathogen population. This is of particular importance when vaccines are only effective against particular 'types' of a given pathogen, examples of this being influenza, *Streptococcus pneumoniae* (a common cause of pneumonia), and *Neisseria meningitidis* (an important cause of meningitis). The introduction of such vaccines can lead to the selection and emergence of 'types' that are not targeted by the vaccine,²⁵⁻²⁷ which may then become predominant and lead to a reduction in vaccine efficacy in the population. The impact of this can be counteracted by surveillance WGS to monitor circulating 'types' over time, predict future prevalent strains, and alter vaccine formulations to maintain efficacy. Sharing of WGS data performed for clinical and public health purposes with the research community would also lead to its application to the study of pathogen biology, bringing new mechanistic insights to transmission, drug resistance and virulence.

Genome sequencing has also led to new insights that combined with genetic techniques promise to provide new approaches to disease control. An important example of this is the control of malaria through genetically modified mosquitos (see Box 1).

Box 1 Malaria

The ability to make genetically modified (GM) mosquitoes has opened up avenues for new methods of controlling the malaria parasite by controlling the Anopheles mosquito vectors that exclusively transmit it between humans when female mosquitoes take a blood meal. Genome sequencing has revealed in the mosquito several novel genetic targets with putative roles in processes such as reproduction, parasite susceptibility and insecticide resistance that together determine the mosquito's capacity to transmit disease. GM technologies for the mosquito have moved forward with the advent of the CRISPR gene editing tool allowing the precise mutation of key genes and thereby confirmation of function. Moreover, the same gene editing tools have recently been repurposed as a gene drive, wherein the mutation in the key gene is disproportionately inherited among the offspring. These gene drives could therefore rapidly increase in frequency, spreading the resulting altered phenotype into a mosquito population over a relatively short timeframe and ultimately reducing its capacity to transmit malaria.

Text kindly supplied by Andrea Chrisanti and Tony Nolan, Imperial College, London.

5. Sequencing and Public Health England

Cognizant of the challenges associated with clinical and public health use of microbial WGS, Public Health England (PHE) has focussed on developing solutions for pathogens of high public health value and where implementation is at least cost neutral. The lead pathogens for routine implementation of fully automated WGS processing, analysis and reporting solutions are for Mycobacterial reference diagnostics, which is in line with the recommendations of the 100,000 genome project, and for *Salmonella* spp. outbreak investigation. A production software has been developed that automates processing of TB WGS and determines species, predicts anti-tuberculosis drug resistance and identifies nearest genomic matches for cluster identification.²⁸ Similarly, *Salmonella* spp. WGS has been automated for outbreak detection. For new outbreaks, investigations are undertaken on an exploratory basis ahead of having the necessary population genetic context for the offending pathogen. Optimising the high performance hosting computer environment is being done in collaboration with Genomics England. Substantial collaborative working with academia and other agencies such as the UK Animal and Plant Health Agency (APHA), CDC Atlanta, British Columbia Centre for Disease Control and Saw Swee Hock School of Public Health, National University of Singapore, is facilitating development of joint solutions of public health importance.

The vision is to progressively extend WGS to further pathogens of public health importance. These will be developed sequentially based on public health priority and reusing the processing workflows optimised for TB and *Salmonella* spp. The following bacterial pathogens are already being initially evaluated for development of automated production solutions: *Streptococcus pneumoniae*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Listeria monocytogenes*, *Escherichia coli*, *Staphylococcus aureus* and multi-resistant *Enterobacteriaceae*. In parallel, next generation sequencing solutions are being developed for viruses. Methods have been developed as proof of principle for WGS of hepatitis C virus, HIV, Ebola, influenza virus and many other species. The current limitation is cost, but with new sequencing platforms coming to market this will change and with it, an anticipated reduction in cost and much improved speed of sequencing. This will facilitate implementation of WGS as a routine method for characterising viruses. PHE is with other major public health agencies seeking to release all WGS data into publically accessible repositories such as NCBI and EBI.

While PHE fulfils the role of protecting the public from infectious diseases and, thus focusses on public health priorities, the impending revolution in diagnostic microbiology in the NHS will be determined by commercial development of whole genome sequencing solutions and not through developments in PHE. None are emerging in the market place yet. When they do, there will be an imperative for PHE to source the data from patient clinical record systems linked to genomic sequence data for ongoing national surveillance. The architecture of such pathogen related data-flows and its analysis will emerge as sequencing technologies develop and commercial products appear. PHE has sufficient data storage to host such repositories and analysis resources to process and analyse the data.

6. Barriers to implementation

Introducing WGS into routine practice requires changes in the design, operation and workforce of an agency. In particular, the skill sets needed are mostly unfamiliar to the current staff, and experts are in short supply. Such a change also imposes substantial requirements for planning and the implementation of major organisational changes.

The speed at which the technology can be implemented into routine practice is tempered by structural, but soluble obstacles. There are three classes of obstacle. First is the optimal preparative and nucleic acid extraction workflows for samples. Bacterial sequencing is generally carried out on DNA extracted from a pure, overnight culture, which adds as much as one day to processing. This is required because sequencing directly from a clinical specimen step is insensitive (because of the low number of bacteria in the specimen), expensive (only one sample can be run at a time rather than running several in the same reaction), and does not currently provide data of sufficient quality. Research and development is required to give the option to apply sequencing directly to clinical samples. Second is the need for fast, accurate and affordable sequencing. Third is the need for fully integrated software for processing, analysing and reporting the sequence information. New standards will also be needed for the quality of the sequence data itself and its use for clinical purposes, together with International Organization for Standardization (ISO) accredited software and pipelines. Each of these obstacles is a major development programme in its own right and requires commitment analogous to a major infrastructure project mostly beyond the resources of a single agency or commercial enterprise. Thus, unsurprisingly, there are no well validated 'off-the-shelf' WGS diagnostic solutions available for routine use.

With some notable exceptions (e.g. TB resistance testing, *Salmonella* reference identification), sequencing is additive to current practice and increases costs. This could be justified based on improvements in the quality of care, and speed and effectiveness of public health investigation and intervention. For example, the cost of rapidly confirming an outbreak of food poisoning and identification of the source could lead to fewer cases and more rapid resolution, as well as accurate attribution of source. The detection of hospital outbreaks caused by nosocomial pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile* could also highlight areas for targeted infection control interventions that further reduce transmission and infection rates. More work is required to quantify the economic benefit of sequencing so that informed decisions can be taken about when and where to deploy this new methodology.

7. Conclusions

The introduction of sequencing technologies into diagnostic and public health microbiology could improve patient care, rapidly detect outbreaks and bring them to a close, and enhance the surveillance of major health threats such as newly emerging pathogens and antimicrobial resistance. The challenges now are to overcome the technical, financial and logistic barriers to its wider adoption, to coordinate services so that provision is equitable across the country, and ensure that data brings value to national and international surveillance programmes. This includes the need for considerable progress in sequencing directly from samples to reduce turnaround time to a minimum. Unlike other technologies where a mature instrument becomes a stable part of the technological capabilities of a diagnostic laboratory at a single point in time, there will also be an on-going need to continuously evaluate, and as necessary adopt new sequencing platforms that have better performance or cost characteristics. Unlike some previous methods such as molecular typing, any changes that take place have little or no implications for interchangeability and data sharing since the output (sequence data) remains unchanged. To deal with limited financial resources, healthcare providers will be required to prioritise how and where to apply the technology and the economics of doing so. There are also major unanswered questions in diagnostic microbiology for which future sequencing technologies have the potential to provide the solution. For example, distinguishing between viral and bacterial infection in an accurate and rapid way is one of the most important outstanding questions in diagnostic microbiology, and underpins better use of antibiotics and improved patient care. Increasing portability of sequencing technologies will also make this more accessible to all, and could become the ultimate in near-patient testing.

9. Suggestions for policy makers

9.1 Coordination and evaluation

Sequencing is being introduced at regional and national levels in a fragmented way. Clarity is required on optimal model(s) for the equitable implementation and delivery of pathogen sequencing in clinical practice across the UK. This includes planning for leadership and resources to generate, interpret, curate, standardise and quality control public sequence databases across the full catalogue of pathogens of public health importance. There is also a need to provide evidence for the cost-effectiveness and health benefit of WGS to individuals and public health. This could be achieved through increased engagement by health economics groups, supported by focused funding opportunities that facilitate this.

9.2 Sharing

Data sharing is essential to gain the greatest benefit to human and animal health. To maximize their usefulness, sequence data deposited in public databases needs to be linked with metadata about the corresponding origin of samples. Data sharing combined with national and global connectivity is essential to capture events as they happen, and to ensure that findings from genome data are converted to rapid and effective action. Sequencing efforts in the face of an outbreak of international importance have typically been undertaken by academic groups, and requires improved coordination and data sharing. This is not always the case; for example, some sequence data was concealed prior to publication during the Ebola epidemic, and valuable samples were not always shared with people with the resources and/or expertise to analyse them. This will require international standardisation of approaches for data generation and sharing.

9.3 Future freedom

Solutions to the implementation of microbial sequencing for diagnostic and public health should be generic, and not become locked into proprietary vertically-integrated systems from commercial providers. For example, an integrated system in which the sequencing, interpretation tools and genome databases are all provided together would seem superficially beneficial, but would prevent the use of alternative (cheaper) sequencing technologies as they become available, or the use of other analysis tools and databases. Avoiding this can be achieved by mandating the use of standard data exchange formats and access to raw data, and will ensure ease of update to new technologies in a rapidly moving field, allow data availability for broader use, and enable the transparent use of the best interpretation tools, whether public or private.

10. References

1. Kirn TJ, Mirrett S, Reller LB, Weinstein MP. Controlled clinical comparison of BacT/alert FA plus and FN plus blood culture media with BacT/alert FA and FN blood culture media. *J Clin Microbiol.* 2014;52:839-43.
2. Pien BC, Mirrett S, Crews BR, Reller LB, Woods CW. Controlled clinical comparison of BacT/ALERT standard aerobic and standard anaerobic blood culture bottles inoculated directly or after transport in sodium polyanethol sulfonate tubes. *J Clin Microbiol.* 2007;45:1357-9.
3. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet.* 2012;13:601-12.
4. Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Pontén T, Aarestrup FM, Ussery DW, Lund O. Benchmarking of methods for genomic taxonomy. *J Clin Microbiol.* 2014;52:1529-39.
5. Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon B, Young B, Wilson DJ, Llewelyn MJ, Paul J, Peto TE, Crook DW, Walker AS, Golubchik T. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol.* 2014;52:1182-91.
6. Stoesser NE, Martin J, Mawer D, Eyre DW, Walker AS, Peto TE, Crook DW, Wilcox MH. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *Clin Infect Dis.* 2013;56:1680-1.
7. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker TM, Quan TP, Wyllie DH, Del Ojo Elias C, Wilcox M, Walker AS, Peto TE, Crook DW. Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. *J Clin Microbiol.* 2015;53:1137-43.
8. Larder BA, Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science.* 1989;246 (4934):1155-8.
9. Shafer RW. Rationale and uses of a public HIV drug resistance database. *J Infect Dis.* 2006;194 Suppl 1:S51-8.
10. Erali M, Page S, Reimer LG, Hillyard DR. Human immunodeficiency virus type 1 drug resistance testing: a comparison of three sequence based methods. *J Clin Microbiol.* 2001;39:2157-65.
11. Li JZ, Paredes R, Ribaudo HJ, Svarovskaia ES, Metzner KJ, Kozal MJ, Hullsiek KH, Balduin M, Jakobsen MR, Geretti AM, Thiebaut R, Ostergaard L, Masquelier B, Johnson JA, Miller MD, Kuritzkes DR. Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. *JAMA.* 2011;305:1327-35.
12. Todesco E, Mercier-Darty M, Jung M, Bottero J, Boyd A, Marcelin AG, Calvez V, Morand-Joubert L. Ultradeep sequencing in the therapeutic management of HIV-1 infection at treatment initiation. *J Antimicrob Chemother.* 2015;70:1919-20.
13. Nishizawa M, Matsuda M, Hattori J, Shiino T, Matano T, Heneine W, Johnson JA, Sugiura W. Longitudinal Detection and Persistence of Minority Drug-Resistant Populations and Their Effect on Salvage Therapy. *PLoS One.* 2015;10:e0135941.
14. Bonsall D, Ansari MA, Ip C, Trebes A, Anthony Brown A, Klenerman P, Buck D, STOP-HCV Consortium, Piazza P, Barnes E, Bowden R. ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Research.* 2015;4:1062.
15. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ, Jombart T, Hinsley WR, Grassly NC, Balloux F, Ghani AC, Ferguson NM, Rambaut A, Pybus OG, Lopez-Gatell H, Alpuche-Aranda CM, Chapela IB, Zavala EP, Guevara DM, Checchi F, Garcia E, Hugonnet S, Roth C; WHO Rapid Pandemic Assessment Collaboration. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science.* 2009;324:1557-1561.
16. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu LF, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu JH, Shen J. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet Infect Dis.* 2016;16:161-8.
17. Cottam EM, Wadsworth J, Shaw AE, Rowlands RJ, Goatley L, Maan S, Maan NS, Mertens PP, Ebert K, Li Y, Ryan ED, Juleff N, Ferris NP, Wilesmith JW, Haydon DT, King DP, Paton DJ, Knowles NJ. Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. *PLoS Pathog.* 2008;4:e1000050.
18. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, Fitzgerald M, Godfrey P, Haas BJ, Murphy CI, Russ C, Sykes S, Walker BJ, Wortman JR, Young S, Zeng Q, Abouelleil A, Bochicchio J, Chauvin S, Desmet T, Gujja S, McCowan C, Montmayeur A, Steelman S, Frimodt-Møller J, Petersen AM, Struve C, Krogfelt KA, Bingen E, Weill FX, Lander ES, Nusbaum C, Birren BW, Hung DT, Hanage WP. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc Natl Acad Sci USA.* 2012;109:3065-70.

19. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, Chen W, Pu F, Peng Y, Li J, Xi F, Li S, Li Y, Zhang Z, Yang X, Zhao M, Wang P, Guan Y, Cen Z, Zhao X, Christner M, Kobbe R, Loos S, Oh J, Yang L, Danchin A, Gao GF, Song Y, Li Y, Yang H, Wang J, Xu J, Pallen MJ, Wang J, Aepfelbacher M, Yang R; E. coli O104:H4 Genome Analysis Crowd-Sourcing Consortium. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med*. 2011;365:718–724.
20. Grad YH, Godfrey P, Cerquiera GC, Mariani-Kurkdjian P, Gouali M, Bingen E, Shea TP, Haas BJ, Griggs A, Young S, Zeng Q, Lipsitch M, Waldor MK, Weill FX, Wortman JR, Hanage WP. Comparative genomics of recent Shiga toxin-producing *Escherichia coli* O104:H4: short-term evolution of an emerging pathogen. *MBio*. 2013;4:e00452–12.
21. Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen transmission. *Curr Opin Microbiol*. 2015;23: 62–67.
22. Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis*. 2013;13:130–6.
23. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TE, Walker AS. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med*. 2013;369:1195–205.
24. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, Reacher M, Haworth CS, Curran MD, Harris SR, Peacock SJ, Parkhill J, Floto RA. Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet*. 2013;381:1551–60.
25. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet*. 2013;45:656–63.
26. Ibarz-Pavón AB, Maclennan J, Andrews NJ, Gray SJ, Urwin R, Clarke SC, Walker AM, Evans MR, Kroll JS, Neal KR, Ala'aldien D, Crook DW, Cann K, Harrison S, Cunningham R, Baxter D, Kaczmarek E, McCarthy ND, Jolley KA, Cameron JC, Stuart JM, Maiden MC. Changes in serogroup and genotype prevalence among carried meningococci in the United Kingdom during vaccine implementation. *J Infect Dis*. 2011;204:1046–53.
27. Ampofo WK, Azziz-Baumgartner E, Bashir U, Cox NJ, Fasce R, Giovanni M, Grohmann G, Huang S, Katz J, Mironenko A, Mokhtari-Azad T, Sasono PM, Rahman M, Sawanpanyalert P, Siqueira M, Waddell AL, Waiboci L, Wood J, Zhang W, Ziegler T; WHO Writing Group. Strengthening the influenza vaccine virus selection and development process: Report of the 3rd WHO Informal Consultation for Improving Influenza Vaccine Virus Selection held at WHO headquarters, Geneva, Switzerland, 1–3 April 2014. *Vaccine*. 2015;33:4368–82.
28. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TE, Crook DW, Iqbal Z. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun*. 2015;21;6:10063.

Chapter 10

Risk-stratified cancer screening

Chapter lead

Prof Douglas Easton¹

Chapter authors

Prof Douglas Easton^{1,2}, Dr Nora Pashayan³, Prof Paul Pharoah^{1,2}

1 Department of Public Health and Primary Care, University of Cambridge

2 Department of Oncology, University of Cambridge

3 Department of Applied Health Research, University College London

1. Summary

- Cancer screening programmes may be made more effective by targeting the screening at individuals most likely to benefit and least likely to be harmed.
- Risk stratification based on genetic markers coupled with other risk factors provides real opportunities for improving the efficiency of the screening programmes and reducing their adverse consequences.
- The effectiveness, cost-effectiveness, acceptability, accessibility and feasibility of implementing risk-stratified screening programmes need to be evaluated.
- Decision aids for the public and training of the health care workforce would be needed.
- Multi-disciplinary efforts are needed to overcome the implementation challenges.
- Breast, prostate, colorectal and lung cancer provide the most immediate opportunities.

2. Background

Screening for early detection of cancer, or precursors of cancer, is a well-established component of cancer control strategy. Early detection could result in less invasive treatment with potentially better outcomes, and can significantly reduce cancer mortality. Currently, there are established national screening programmes for breast, colorectal and cervical cancer, available to all individuals on the basis of age and sex (see Table 1). However, screening also has significant disadvantages: it can lead to increased anxiety, false reassurance, unnecessary biopsies and overdiagnosis of disease that need not have been treated.¹ As a result, the benefits of universal screening remains controversial.

In principle, screening programmes could be more effective if they could be targeted on those individuals that are most likely to benefit.^{2,3} In contrast to a “one-size-fits-all” approach, a risk stratified screening programme would involve offering different screening approaches to individuals based on their level of risk, with the expectation of improving balance between the benefits and harms of screening. Tailored screening could involve varying the age at which screening is started or stopped, the frequency of screening, or the screening modality.⁴

The concept of enhanced surveillance for individuals at higher risk of cancer has in fact long been accepted, in context of a strong family history of the disease or because of carrying a genetic variant known to confer a high disease risk (<https://www.nice.org.uk/Guidance/CG164>). Women with a strong family history of breast cancer may be offered screening at an earlier age (typically annual from age 40). Higher risk women (for example those with mutations in the BRCA1 or BRCA2 genes) are offered screening by Magnetic Resonance Imaging (MRI), a screening modality which has higher sensitivity than mammography.⁵ Individuals at high genetic risk of colorectal cancer are offered regular colonoscopy.^{6,7} However, the current identification of individuals at high risk is based on self-referral, the stratification is relatively crude and only captures a small fraction of individuals who might benefit.⁸ Extending risk assessment to the whole population would require strong evidence that such enhanced surveillance and screening will do more good than harm. The ethical implications of case-finding, where one approaches their physician with health concerns, are different from those of population-based screening where risk assessment and screening are offered to individuals who have not expressed health concerns.

Table 1 Current protocols for cancer screening in the UK, for the general population and in high-risk individuals

	Current UK National Screening Programme	Protocols for high-risk individuals	Approximate number of common genetic markers identified	Known risk genes	Approximate risk to highest 5% of population, based on common variants	Approximate proportion of cases in 5% of population at highest risk
Breast	3-yearly mammography from 50-69 years	Annual mammography age 40-49 (moderate risk) Annual MRI from age 30-49 (high risk) ^(a)	94	BRCA1, BRCA2, ATM, CHEK2, PALB2, TP53, PTEN, STK11, CDH1, NF1, NBN	2.3 ^(b)	11%
Colorectal	2-yearly FOB, 60-74 Single sigmoidoscopy, age 55	Colonoscopy (18-24 monthly) from age 25 ^(c)	58	MSH2, MLH1, MSH6, PMS2, APC, MUTYH	2.7x ^(d)	14%
Prostate	None	Annual PSA screening 40-69 ^(e)	112	BRCA2, BRCA1, HOXB13	2.5x ^(f)	12%
Ovary	None	Ultrasound + CA125 ^(g)	183	BRCA1, BRCA2, BRIP1, RAD51C, RAD51D	1.8x	9%

(a) See <https://www.nice.org.uk/guidance/CG164> for full details and definition of moderate and high-risk

(b) Adapted from Mavaddat, N. et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 107(2015).

(c) Guidelines from British Society of Gastroenterology (BSG) and the Association of Coloproctology for Great Britain and Ireland (ACPGBI). See also Vasen, H.F. et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut* 62, 812-23 (2013).

(d) Adapted from Paci, E., Warwick, J., Falini, P. & Duffy, S.W. Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen* 11, 23-7 (2004).

(e) No national guidelines, protocol used in ongoing IMPACT study (<http://www.impact-study.co.uk>)

(f) Adapted from Amin Al Olama, A. et al. Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer Epidemiol Biomarkers Prev* 24, 1121-9 (2015).

(g) No national guidelines, protocol used by UKFOCSS (<http://www.instituteforwomenshealth.ucl.ac.uk/womens-cancer/gcrc/ukfocss>)

Data sourced from <https://www.gov.uk/topic/population-screening-programmes>

Box 1 - Overdiagnosis

One of the major drawbacks of cancer screening is overdiagnosis: the detection of a cancer as a result of screening that would not have been diagnosed in a person's lifetime had screening not taken place.^a For example, in case of screening for breast cancer, the Independent UK Panel on Breast Cancer Screening estimated that for every breast cancer death averted, three cases are likely to be overdiagnosed.^b Many of these cases are non-invasive breast tumours (carcinoma-in-situ) that need to be treated but may not have led to invasive disease.^c

The problem of overdiagnosis is even more significant in case of prostate cancer screening. In the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow up, for every prostate cancer death averted, 12 to 36 excess cases were detected.^d Thus, early detection of prostate cancer by screening with PSA testing could prevent cancer death for a subset of men,^e but at substantial cost of overdiagnosis and overtreatment.^{e,f} Concerns regarding overdiagnosis have led the UK National Screening Committee and the US Preventive Services Task Force to adopt an 'all or none' approach, recommending against population based screening for prostate cancer. Abandoning screening would eliminate overdiagnosis, but at a cost of failing to prevent avoidable cancer deaths in a subgroup of men.^g Stratified screening may reduce the burden of overdiagnosis.

-
- a) Paci, E., Warwick, J., Falini, P. & Duffy, S.W. Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen* 11, 23-7 (2004).
 - b) Marmot, M.G. et al. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 108, 2205-40 (2013).
 - c) Independent, U.K.P.o.B.C.S. The benefits and harms of breast cancer screening: an independent review. *Lancet* 380, 1778-86 (2012).
 - d) Auvinen, A. et al. Absolute Effect of Prostate Cancer Screening: Balance of Benefits and Harms by Center within the European Randomized Study of Prostate Cancer Screening. *Clin Cancer Res* 22, 243-9 (2016).
 - e) Schroder, F.H. et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 384, 2027-35 (2014).
 - f) Andriole, G.L. et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst* 104, 125-32 (2012).
 - g) Gulati, R. et al. Expected population impacts of discontinued prostate-specific antigen screening. *Cancer* 120, 3519-26 (2014).

Recent advances in genomics, together with parallel advances in imaging, have provided opportunities to define an individual's risk of cancer much more precisely, and to develop risk-based screening strategies that can be applied to the whole population.

Most of the discussion of stratified screening has involved stratification according to the level of risk. However, the benefits of screening also depend on the ability to detect tumours at an early stage and treat them effectively. In principle, therefore, it may also be possible to tailor screening strategies according to likely aggressiveness of the tumours that might occur. The efficacy of a screening programme would also be improved by better genomic approaches to identify those cancers that require treatment, another area of active research.

While the discussion here focuses on the use of genetic information for targeted screening, the same principle applies to other interventions – these include risk-reducing surgery, risk-reducing medication⁹ and programs focusing on lifestyle changes (for example weight reduction).

3. Risk assessment

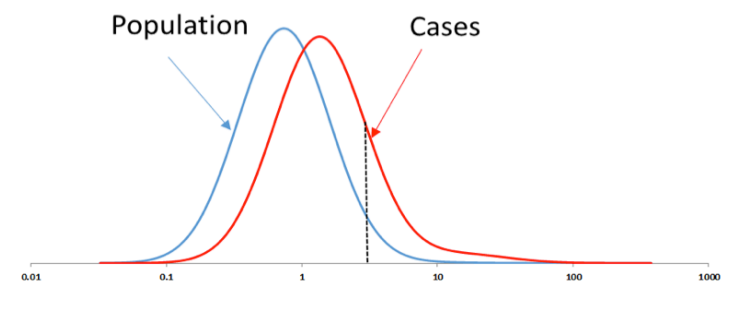
Risk assessment is an essential step for tailoring screening strategies to individual risks. Both genetic and non-genetic factors could be used for risk assessment.

3.1 Genetic risk scores

Over the past decade, technological advances have made it much easier to identify many genetic variants (typically so-called single nucleotide polymorphisms, or SNPs) that are associated with disease risk. The main advance has been through so-called “chip genotyping”, which allows many thousands of genetic variants to be tested simultaneously in large studies. Through these “genome-wide association studies” (GWAS), germline genetic variants associated with most diseases have been found, including several hundred associated with different types of cancer.¹⁰ Each variant is typically only responsible for a small change in risk, but, because large numbers of these variants have been found, it is possible to combine them together to generate a Genetic Risk Score (GRS) that is more predictive of cancer risk and enables stratification of the population into low- and high-risk groups. Such risk scores have been defined for many of the common cancer types, including breast, prostate, ovarian and colorectal cancer.

GRS is simply a weighted average of the number of risk variants carried by an individual (see Figure 1). As chip genotyping is relatively inexpensive and can be carried out in large population-based studies, it is possible to estimate the cancer risks associated with different levels of GRS quite precisely. For the same reason, it is possible to develop affordable genetics tests based on the GRS, and several companies offer them.

Figure 1 - Distribution of risk in the population and patients

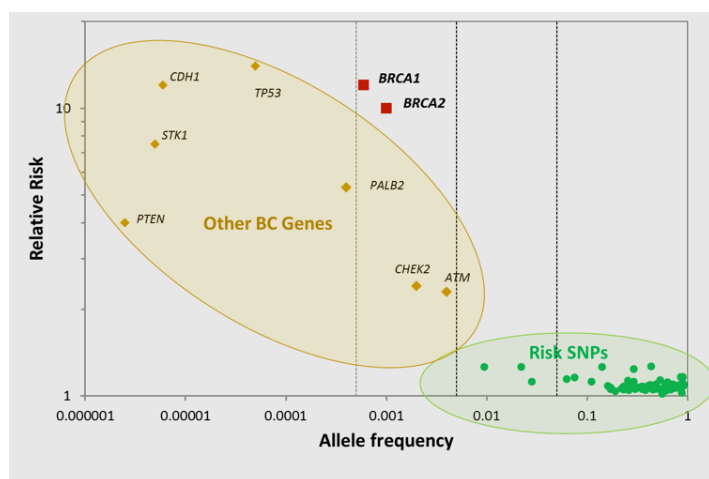


Note The genetic risk of cancer can be summarised as a genetic risk score (GRS), which is a weighted sum of the number of risk variants that an individual carries. Since the number of variants is large, the GRS follows a normal distribution as shown. Cancer cases tend to occur among individuals with higher GRS. The larger the number variants, the larger the variance of the GRS, and the better the risk stratification.

3.2 High risk genes

Another way to identify individuals at high risk of cancer is through sequencing of specific genes to identify mutations associated with cancer. This approach enables rarer genetic variants that may confer substantially increased risks of cancer to be identified. Such genes have been identified for many cancer types. Figure 2 presents the genetic variants associated with breast cancer. DNA sequencing has been in routine clinical practice for the past two decades, for specific genes such as BRCA1 and BRCA2, usually for individuals with a significant family history of cancer. More recently, the falling costs of sequencing have made it possible to extend this approach to sequencing larger numbers of genes simultaneously, a technology known as “gene panel testing”. This approach is already widely used in clinical practice, particularly in the US, with many companies offering such tests. However, the clinical validity of many of the genes on commercial panels remains to be established¹¹ Currently interest in gene-panel testing is largely limited to clinical genetics centres counselling individuals with a family history of cancer, but as the cancer risks associated with rarer variants in these genes become more firmly established, gene panel testing could also be incorporated into risk stratification in cancer prevention and screening programmes at the population level. This could improve the overall effectiveness in terms of reducing cancer mortality, but might lead to significant increases in costly interventions including MRI screening, colonoscopy and prophylactic surgery.

Figure 2 - Genetic susceptibility to breast cancer



Note Breast cancer risk is determined by many common genetic markers (SNPs), each conferring a low risk of disease. In addition, rare genetic variants in several genes predispose to more substantial risks of disease, the most important genes being BRCA1 and BRCA2.

3.3 Epigenetic markers

While much of the attention has focused on risk prediction using heritable genetic markers, other markers might also be incorporated into risk prediction: these include epigenetic markers (specifically DNA methylation) and microRNA expression.¹² Epigenetic changes, which may be triggered by environmental exposures or endogenous factors, lead to changes in gene expression, and are implicated in tumour initiation and progression. However, compared with genetic markers, these are more challenging to measure, change over time and are tissue dependent. Epigenetic changes are reversible and can be modulated by drugs, diet and other environmental factors. This reversibility provides opportunity for cancer prevention strategies.¹³ At the research level, there is considerable interest in measuring epigenetic markers in the target tissue. However, some blood markers have also shown promise and might become usable for risk stratification on a population level.¹²

3.4 Other risk factors

For some cancers, cancer risk can be defined on the basis of other risk factors, including family history, lifestyle (e.g. alcohol intake, smoking) and reproductive history (e.g. age of menarche, number of births, hormone replacement use).¹² For breast cancer, breast density is a powerful quantitative risk factor that can be measured using mammograms or MRI.¹⁴

4. Risk stratification

4.1 Breast cancer

To date, genome-wide association studies have identified 94 breast cancer susceptibility variants.¹⁵ The GRS based on these variants defines a distribution of risk such that the highest 1% of the population have a 3-fold risk relative to population average, while the 1% of the population at lowest risk have a risk that is ~1/4 the population risk. This GRS has better risk prediction than a risk score based on non-genetic risk factors; much better prediction can, however, be achieved using a combination of genetic and non-genetic factors, and breast density (Figure 3).

4.2 Prostate cancer

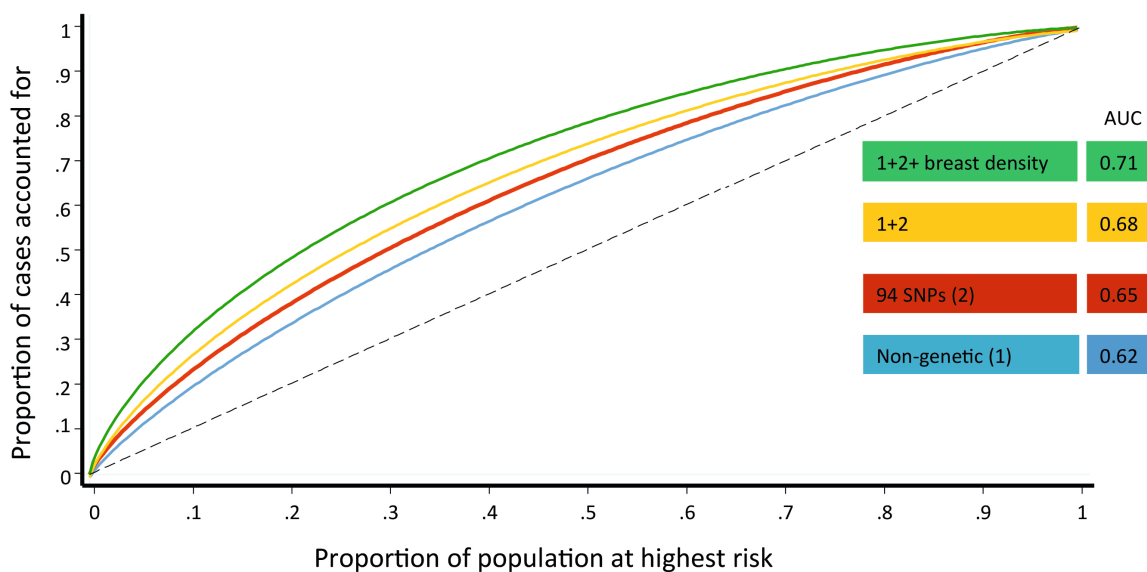
To date, 112 prostate cancer susceptibility variants have been identified.¹⁶ These variants define a genetic risk profile such that 1% of men have a risk that is more than 4-fold higher than the population average.

This risk estimate is comparable to that conferred by deleterious mutations in BRCA2.¹⁷ PSA-based targeted screening is being offered to male BRCA1 and BRCA2 carriers as part of an ongoing multi-national prostate cancer targeted screening trial, the IMPACT (Identification of Men with a genetic predisposition to Prostate Cancer) study,¹⁸ and such an approach could be extended more generally to men at high risk by virtue of their risk profile. Half of the population at highest risk for prostate cancer based on the genotype of the known 112 variants accounts for 76% of all cases of prostate cancer.

4.3 Colorectal cancer

Based on the 45 known colorectal cancer susceptibility variants¹⁹ individuals with GRS within the top 1% have around 3-fold increased risk for colorectal cancer compared to the population average.²⁰

Figure 3 - Proportion of breast cancer explained by the proportion of the population at highest risk for the disease



Note Estimates based on non-genetic risk factors* (1), known 94 breast cancer susceptibility variants (2), risk score combining 1 & 2; and risk score combining (1), (2), and breast density. The graph shows that half of the population at highest risk for breast cancer based on the genotype of the known 94 variants accounts for 70% of all cases of breast cancer. The respective proportion for the model combining breast density to non-genetic factors and the 94 variants would be 79%.

* Non-genetic risk factors include age of menarche, number of births, age of first live birth, oral contraceptive use, body mass index, alcohol, smoking, personal history of benign breast disease, family history of breast cancer in first-degree relatives. AUC – Area under the receiver operator characteristic curve. AUC is a measure of the discriminatory accuracy of a risk assessment tool.

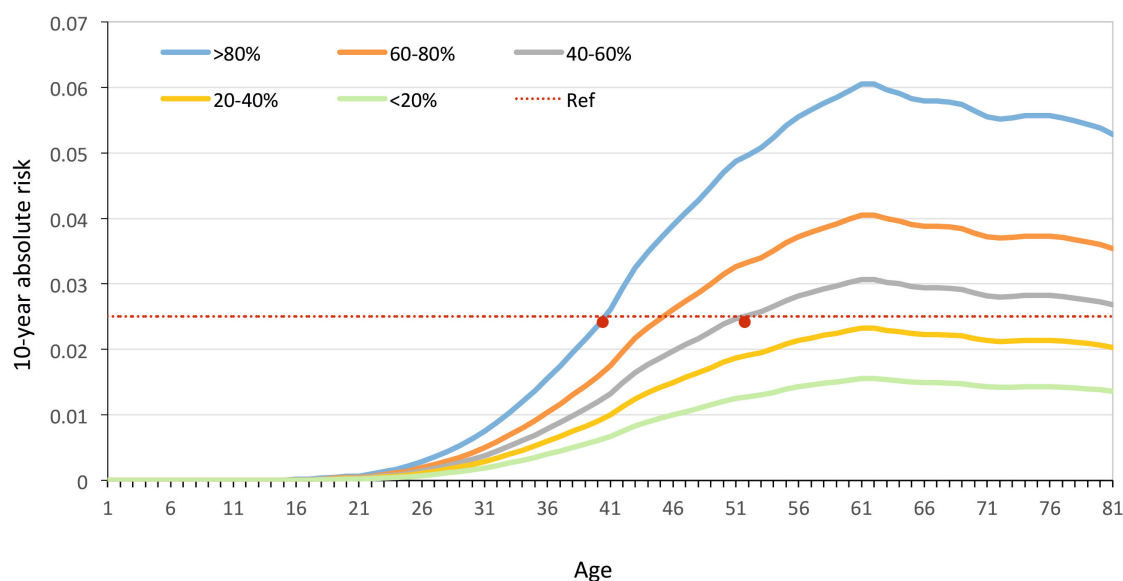
Source *tbc*

5. Risk-stratified screening

The NHS Breast Screening Programme (NHSBSP) currently offers the same screening package to all women aged 50-70 years, with an ongoing randomized extension of the programme to women aged 47-73 years. An alternative risk-based approach would be to offer screening to all women who reach a certain level of risk (thus offering to some younger women at high risk but not older women at lower risk; Figure 4).

If this level was, for example, set at 2.5% 10-year risk (the current average level of risk in the screened population), risk-based screening has been estimated to result in 31% fewer women being screened, whilst only 2% fewer cases would be detected²¹ (see Figure 5). In England, this is equivalent to 2.3 million fewer screens while detecting 420 fewer breast cancers per screening round. The cost-effectiveness of such approaches needs still to be evaluated.

Figure 4 - Age of invitation to screening for breast cancer by quintiles of risk (dependent on age, 94 SNP profile and non-genetic risk factors)

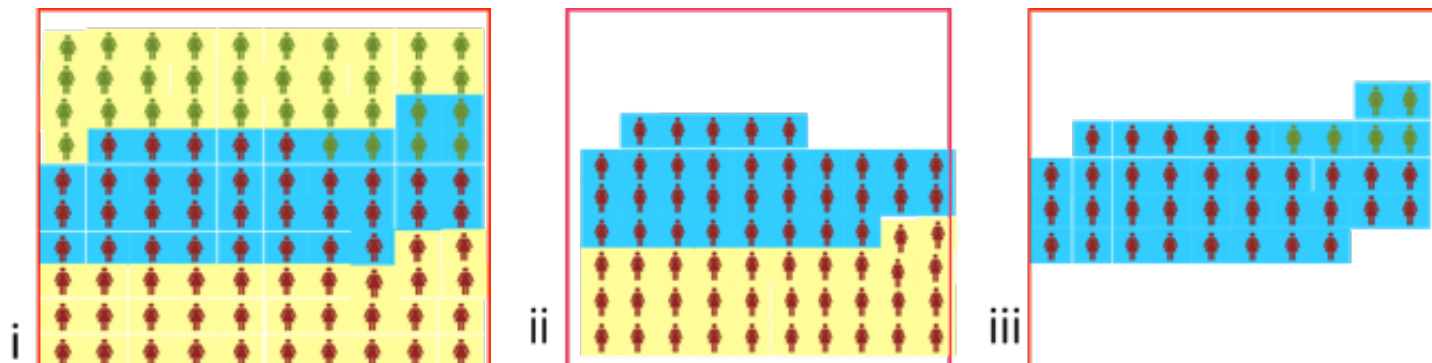


Note Women at highest risk quintile will reach the risk threshold by age 40 and could be offered screening by age 40; whereas women in the lowest risk quintile could avoid undergoing screening. Reference refers to 2.5% 10-year absolute risk for developing breast cancer corresponds to risk of UK women aged 47, i.e. age of invitation to the UK NHSBSP.

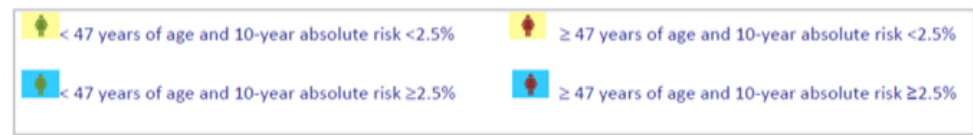
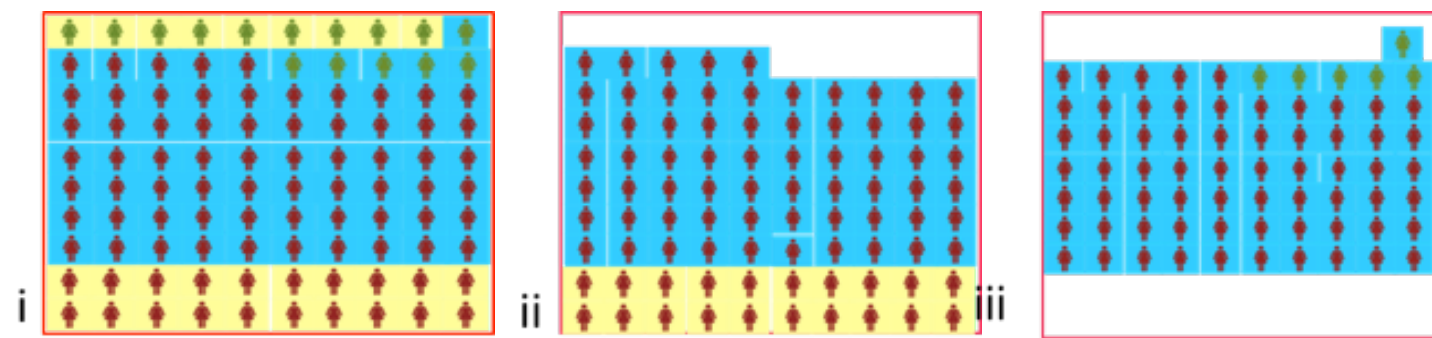
Source Pashayan, N. et al. *Public health genomics and personalized prevention: lessons from the COGS project.* *J Intern Med* 274, 451-6 (2013).

Figure 5 - Reclassification of women into different risk groups

A. Eligibility for screening for breast cancer



B. Potentially screen-detectable breast cancers



Note

A. Eligibility for screening: a population of 100 women, 35–79 years of age, by age group (<47 and ≥47 years) and risk threshold (10-year absolute risk of being diagnosed with breast cancer of 2.5%)(i), eligible for screening based on age alone (ii), or on age and risk score (based on the 94 known breast cancer susceptibility variants and non-genetic risk factors) (iii)

B. Potentially screen-detectable breast cancers: 100 women with breast cancer, 35–79 years of age, by age group (<47 and ≥47 years) and risk threshold (10-year absolute risk of being diagnosed with breast cancer of 2.5%) (i), potentially detectable following screening based on age alone (ii), or on age and risk score (based on the 94 known breast cancer susceptibility variants and non-genetic risk factors) (iii)

Similar modelling exercises for colorectal cancers²⁰ and prostate cancer²² have shown improvement in efficiency of screening programmes using stratified approaches. Screening for colorectal cancer is currently offered as faecal occult blood (FOB) screening to men and women aged 60-74, while one-off flexible sigmoidoscopy at age 55 is being introduced. Screening for bowel cancer using sigmoidoscopy has been robustly associated with a mortality reduction of approximately 30%, but has more significant side effects. Individuals at high genetic risk of colorectal cancer are typically offered colonoscopy, predicted to improve sensitivity and hence provide a larger reduction in mortality as the proximal colon can also be screened, but whose side effects are more significant.⁶ Stratified screening could involve changing the entry age, varying the frequency of screening, or extending the use of colonoscopy to a wider group of high-risk individuals.

There is currently no national screening programme for prostate cancer; while there is some evidence from trials that PSA screening can reduce mortality,^{23,24} its effectiveness, in terms of the number of prostate cancer deaths that can be prevented, is low. Therefore, any screening programme is likely to be risk-based. Two studies have shown that the probability of overdiagnosis is lower in men at higher genetic risk.^{25,26} Restricting screening to men with GRS above the population average would reduce the screening episodes by half, while detecting 80% of the non-overdiagnosed cancers and reducing overdiagnosed cancers by 38% at a cost of missing 20% of the non-overdiagnosed cancers. That is, for every non-overdiagnosed cancer not detected through screening, almost two overdiagnosed cases could be avoided.²⁶

5.1 Other opportunities

Targeted screening for ovarian cancer, through a combination of ultrasound and serum CA125, has shown some promise in early detection, but the evidence on mortality is limited.²⁷ A more viable option for stratified prevention would involve identification of women at sufficient risk to warrant prophylactic oophorectomy. This is already standard of care for carriers of BRCA1 and BRCA2 mutations, but the same approach could be extended to women at high risk on the basis of GRS and other ovarian cancer susceptibility genes. Screening for lung cancer using low dose CT has been shown to reduce mortality. Yet there is no screening programme for lung cancer in the UK – we are awaiting the findings of screening trials in Europe (e.g. NELSON trial in the Netherlands) and of pilot studies (e.g. UK Lung Cancer Screening trial).

Stratified screening has been proposed;^{28,29} stratification for lung cancer screening would most likely be based largely on smoking history, but could also incorporate genetic and/or epigenetic markers. Other opportunities include targeting screening for Barrett's oesophagus, a precursor for oesophageal cancer.

6. Challenges

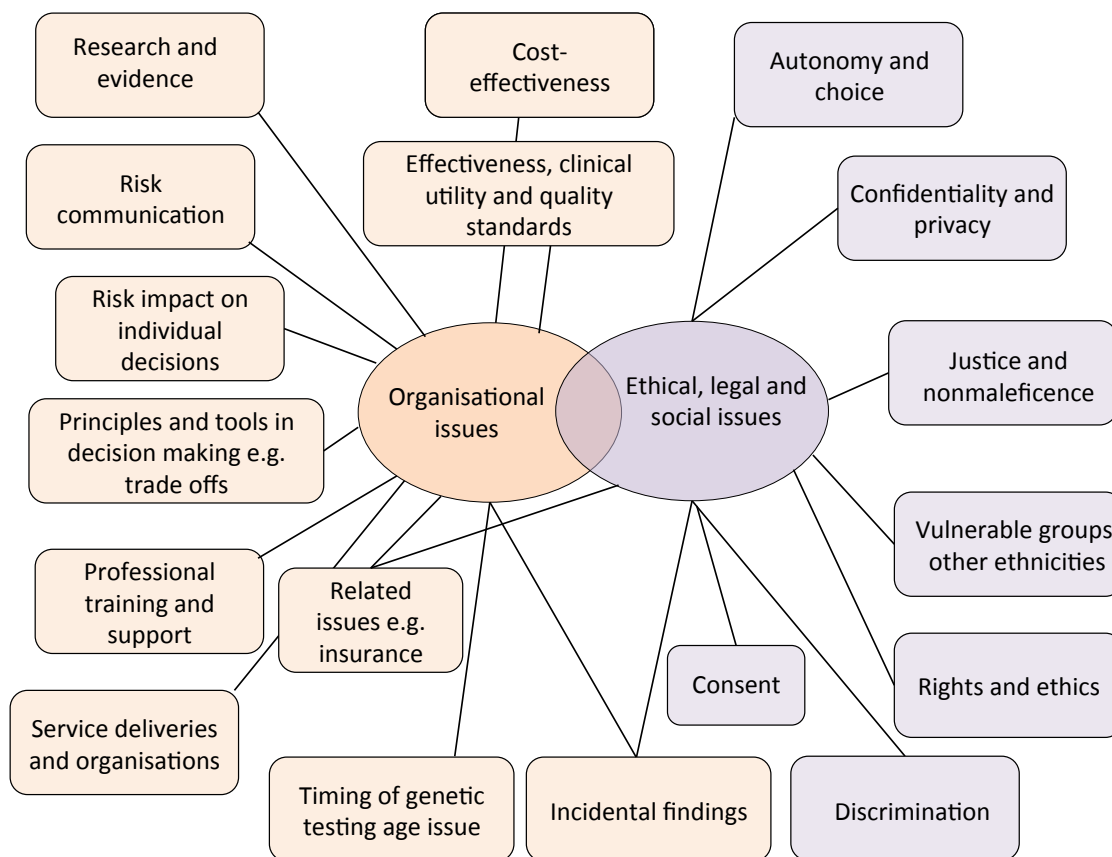
There are many questions that need to be answered before risk-stratified screening could become standard practice (see Box 2). Evidence would be needed on the effectiveness in improving the benefit to harm balance of screening, cost-effectiveness, acceptability, equity of access, and feasibility of implementing such programmes. The effectiveness of a risk-stratified screening strategy should ideally be addressed by randomized screening trials. However, as there are too many potential changes to investigate empirically in trials, mathematical modelling approaches will be needed to identify the most promising options that then could be investigated in controlled trials.

The implementation of risk-stratified screening programme is more complex than a programme with eligibility based on age alone.^{30,31} There are organisational, ethical, legal, regulatory, social, and policy implications to be considered (see Figure 6). Legislation to protect genomic data, developing the IT infrastructure, preparing the workforce,³² identifying clinical pathways, developing decision tools, and engaging with the public early on are among the steps needed to overcome the implementation challenges.

Box 2 - Proposals to address evidence gaps

1. Support research to generate robust evidence on whether risk-stratified screening does more good than harm at an affordable cost to the NHS. Study the case of the most frequently diagnosed cancers - breast, prostate, colorectal, and lung cancers.
 - Develop risk-prediction models that can be used at population level for each of those cancer types.
 - Study whether and how the natural history of cancer, cancer-specific mortality reduction following screening, proportion of false findings and probabilities of overdiagnosis, and overtreatment following screening vary by absolute risk levels.
 - Develop pragmatic randomised controlled trials to study the effectiveness of risk-stratified screening programmes.
 - Economic evaluation of the cost-effectiveness of risk-stratified screening programmes compared to the default status (no screening or existing screening programme).
2. Support research to generate robust evidence on the optimal screening strategies and how best to implement such programmes.
 - Decision modelling to study different screening strategies by varying the frequency of screening, the start and stop age of screening, screening test modality, screening test cut-off point, and risk thresholds for eligibility for screening programme. The optimum screening strategy for each cancer type studied will be the one that gives the best benefit to harm balance and is the most cost-effective.
 - Do prepare for the implementation of risk-stratified screening programmes:
 - Prepare the health care infrastructure (e.g. IT frameworks, data storage, care pathways).
 - Set policies in place to mitigate any ethical, legal, social implications of risk-stratified screening programmes.
 - Prepare the healthcare workforce.
 - Develop decision tools to communicate risk and support informed decision by the public.

Figure 6 - Issues to be considered to enable implementation of risk-stratified screening programmes



Source
 Adapted from Chowdhury, S. et al. Incorporating genomics into breast and prostate cancer screening: assessing the implications. *Genet Med* 15, 423-32 (2013)

Creative commons license :<https://creativecommons.org/licenses/by-nc-nd/3.0/>

7. Conclusions

From a technical standpoint, defining individual cancer risk is already feasible and could improve the efficiency of cancer prevention programmes, including screening, and could potentially reduce the negative consequences of screening. Indeed, some feasibility studies are already underway, both in the UK and elsewhere. Some changes, for example rationalising risk based screening for individuals with a family history, could be implemented quite quickly, but changes to national screening programmes would clearly take longer. In particular, pilot studies to assess the acceptability of stratified screening to the public, modelling to determine the changes and health economics analysis will all be needed. Most importantly, any changes would require education of health care providers and the general public.

8. Suggestions for policy makers

- Policy makers need to be open to the possibility of introducing risk-stratified national cancer screening programmes.
- Support funding for research on risk-stratified to meet the evidence gaps.
- Use national screening data to facilitate modelling approaches to identify the most promising risk-stratified screening options that then could be investigated in controlled trials.

9. References

1. Paci, E., Warwick, J., Falini, P. & Duffy, S.W. Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen* 11, 23-7 (2004).
2. Pashayan, N. et al. Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. *Br J Cancer* 104, 1656-63 (2011).
3. Hall, P. & Easton, D. Breast cancer screening: time to target women at risk. *Br J Cancer* 108, 2202-4 (2013).
4. Marmot, M.G. et al. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 108, 2205-40 (2013).
5. Leach, M.O. et al. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *Lancet* 365, 1769-78 (2005).
6. Cairns, S.R. et al. Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). *Gut* 59, 666-89 (2010).
7. Vasen, H.F. et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut* 62, 812-23 (2013).
8. Pharoah, P.D. et al. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 31, 33-6 (2002).
9. Cuzick, J. et al. Tamoxifen for prevention of breast cancer: extended long-term follow-up of the IBIS-I breast cancer prevention trial. *Lancet Oncol* 16, 67-75 (2015).
10. Available from: <http://www.nature.com/icogs>
11. Easton, D.F. et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 372, 2243-57 (2015).
12. Pashayan, N., Reisel, D. & Widschwendter, M. Integration of genetic and epigenetic markers for risk stratification: opportunities and challenges. *Per Med* 13, 93-95 (2016).
13. Verma, M., Khoury, M.J. & Ioannidis, J.P. Opportunities and challenges for selected emerging technologies in cancer epidemiology: mitochondrial, epigenomic, metabolomic, and telomerase profiling. *Cancer Epidemiol Biomarkers Prev* 22, 189-200 (2013).
14. Boyd, N.F. et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 356, 227-36 (2007).
15. Mavaddat, N. et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 107(2015).
16. Amin Al Olama, A. et al. Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer Epidemiol Biomarkers Prev* 24, 1121-9 (2015).
17. Breast Cancer Linkage, C. Cancer risks in BRCA2 mutation carriers. *J Natl Cancer Inst* 91, 1310-6 (1999).
18. Mitra, A.V. et al. Targeted prostate cancer screening in men with mutations in BRCA1 and BRCA2 detects aggressive prostate cancer: preliminary analysis of the results of the IMPACT study. *BJU Int* 107, 28-39 (2011).
19. Jenkins, M.A. et al. Quantifying the utility of single nucleotide polymorphisms to guide colorectal cancer screening. *Future Oncol* 12, 503-13 (2016).
20. Frampton, M.J. et al. Implications of polygenic risk for personalised colorectal cancer screening. *Ann Oncol* 27, 429-34 (2016).
21. Pashayan, N. et al. Public health genomics and personalized prevention: lessons from the COGS project. *J Intern Med* 274, 451-6 (2013).
22. Burton, H. et al. Public health implications from COGS and potential for risk stratification and screening. *Nat Genet* 45, 349-51 (2013).
23. Schroder, F.H. et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 384, 2027-35 (2014).
24. Andriole, G.L. et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst* 104, 125-32 (2012).
25. Pashayan, N. et al. Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genet Med* 17, 789-95 (2015).
26. Pashayan, N. et al. Reducing overdiagnosis by polygenic risk-stratified screening: findings from the Finnish section of the ERSPC. *Br J Cancer* 113, 1086-93 (2015).
27. Jacobs, I.J. et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* 387, 945-56 (2016).

28. Marcus, M.W., Raji, O.Y. & Field, J.K. Lung cancer screening: identifying the high risk cohort. *J Thorac Dis* 7, S156-62 (2015).
29. Duffy, S.W., Field, J.K., Allgood, P.C. & Seigneurin, A. Translation of research results to simple estimates of the likely effect of a lung cancer screening programme in the United Kingdom. *Br J Cancer* 110, 1834-40 (2014).
30. Dent, T. et al. Stratified cancer screening: the practicalities of implementation. *Public Health Genomics* 16, 94-9 (2013).
31. Chowdhury, S. et al. Incorporating genomics into breast and prostate cancer screening: assessing the implications. *Genet Med* 15, 423-32 (2013).
32. Chowdhury, S. et al. Do Health Professionals Need Additional Competencies for Stratified Cancer Prevention Based on Genetic Risk Profiling? *J Pers Med* 5, 191-212 (2015).

Chapter 11

Genomics in newborn screening

Chapter lead

Prof Philip L. Beales

Chapter author

Dr Heidi C. Howard

Prof Philip L. Beales

1. Summary

Objectives of newborn screening

The purpose of newborn screening (NBS) is the identification of a risk of developing a disease that is preventable or treatable. These treatments need to start in the neonatal or early-childhood years. The use of next generation sequencing (NGS) in NBS should focus on variants of high penetrance and for which there are effective and accepted preventive therapeutic interventions available. NGS should focus on targeted panels or targeted analysis to limit unsolicited findings. All NBS programmes in the UK are recommended by the UK National Screening Committee on the above criteria and cost-effectiveness.

Outcomes

There needs to be evidence of quality of life improvements and impacts on individuals and families undergoing NBS. Furthermore, the impact of using new technologies on the health care system and society must be defined and there needs to be public acceptability.

Costs

The cost of delivering new NGS technologies for NBS needs to be carefully calculated and includes consideration of developing new infrastructure, education, counselling, interventions and sample and data storage requirements.

Engagement

An open discussion with stakeholders should include harm and benefits of screening, development of government policies and delivery methods. New models to inform and counsel parents will need to be developed to maximise participation rates, inform of unsolicited results, inform of storage of data and potential use of data for research and commercial uses.

Professional and Public Education

The education and training needs of professionals involved in NBS should be assessed and new public education programmes should be developed.

2. Background

Newborn screening (NBS) involves the identification of a baby's risk of developing a disease that is preventable or treatable; it is not designed to diagnose inherited diseases. The current newborn screening programme in the UK based on the blood spot test (heel prick, dried onto a piece of filter paper) screens for nine rare but serious conditions; *sickle cell disease*, *cystic fibrosis*, *congenital hypothyroidism*, *phenylketonuria (PKU)*, *medium-chain acyl-CoA dehydrogenase deficiency (MCADD)*, *maple syrup urine disease (MSUD)*, *isovaleric acidaemia (IVA)*, *glutaric aciduria type 1 (GA1)* and *homocystinuria (pyridoxine unresponsive) (HCU)*. The introduction of tandem mass spectrometry made it possible to screen for many conditions at the same time using a single test, and led to a huge expansion in the number of conditions that could be included. NBS programmes vary in the number of conditions included around the world. In the UK, consent must be obtained verbally from the parents and noted in the child's health care records.¹ In the US for example, there is no unified national screening programme consensus, although a core panel of 29 conditions, recommended by the American College of Medical Genetics, has been adopted by many states with some offering over 50 conditions.²

The increasing use and availability of NGS in different areas of health care has led stakeholders to question whether this technology should or could be used in NBS programmes. Goldenberg and Sharp predicted that it is likely that the earliest applications of whole gene sequencing will be restricted to settings in which genetic testing is already a routine part of clinical or public health practice, such as state newborn screening programs.

The adoption of NGS approaches could allow all genes associated with the genetic conditions evaluated through newborn screening to be sequenced at once, thus greatly reducing the need for follow-up testing. However, such tests will need to be conducted in a time and cost-efficient manner, which at the time of writing is still too expensive. NGS is not able to detect all tested conditions, for example metabolite concentrations that are out of range might reflect rare conditions not detectable by mutation analysis, but that require prompt treatment, such as vitamin B12 deficiency of the newborn. The reality is that DNA testing is currently not a routine part of NBS and only a few babies have a DNA test of any kind following birth, whereas all babies undergo biochemical tests as part of screening.

3. Genome sequencing and newborn screening

Traditionally, NBS programmes have focused on identifying neonates who have a high risk of developing particular serious disorders, which have accepted treatments, recognisable latent or early symptomatic stages, an understood natural history, an appropriate test, a satisfactory cost-benefit evaluation, and are, to some degree, accepted by the public. While cited and considered by most health care systems, these parameters may be translated into concrete NBS programmes in very different ways by different countries or regions. Currently, in the European Union, different countries aim to identify between 1 and 30 different conditions.³ The large variation may be a consequence of a mixture of national and/or regional factors, including fundamental differences in health care systems' structure and functioning, available funds, politics, and involvement and input from different stakeholder groups such as health care professionals, parents, patients and the general public. Despite these differences, one trend has emerged in recent years which affects many different countries and regions in both North America and Europe: the (proposed) expansion of the list of diseases included in NBS screening panels. These expanded panels often challenge the interpretation of the original parameters or criteria for inclusion of diseases on the panel. Evidently, with the different ways in which NGS could be used in an NBS programme, the question of expanding disease panels further is also raised.

A key question that must initially be addressed is how will NGS be implemented in the NBS programme? Will it be used as a first step and largely replace current biochemical screening? Or, will it be used as follow up testing to confirm results of biochemical screening? In either scenario, another question needs to be answered: how much of the genome will be sequenced and analysed? All of it, or only targeted regions whereby only a subset of pre-determined genes are analysed to answer specific clinical questions? This approach could mean following a similar strategy to the current NBS programme but using a different tool (i.e. NGS) in the confirmatory testing.

Furthermore, is the idea that the sequence generated at birth be referred to throughout a person's life, not only during the neonate period? If so, this would cause a paradigm shift in NBS programmes. The answers to these initial fundamental questions will strongly guide many of the consequent decisions and issues that need to be managed.

4. Focused screening approaches

If, like the current NBS programme, using NGS is meant to investigate highly penetrant disease-causing variants for treatable or preventable conditions, then it is likely that the disease panel will continue to be for rare disorders. This would imply the exclusion of common complex disorders for which genetic contributions are not well understood and appear to contribute only a small factor. Therefore, with this approach, there would be a selection of variants and genes to be investigated that are clearly pathogenic, have high penetrance, and for which the condition in question has an accepted therapeutic or preventive intervention.⁴ Thus, the targeted sequencing or analysis of specific genes and variants would be appropriate for these aims. Moreover, this approach may reduce unwanted consequences of sequencing and/or analysing the entire genome, such as unsolicited findings, the time and (bioinformatics, human etc.) resources needed for the interpretation and safe storage of larger volumes of sequencing data, which may never be referred to or used again.

The question of whether to sequence and analyse the entire genome or conduct targeted sequencing or analysis is obviously related to many other values and principles surrounding NBS programmes, including the ultimate goals of the programme, the criteria for disease inclusion and whether to expand panels, as well as what are considered acceptable costs. The costs of a single WGS had finally fallen to the projected \$1,000 mark and continues to decrease further. This cost reduction and the improved coverage achievable has now shifted the argument toward WGS approaches for rare disease testing compared with whole exome sequencing (WES) and panels. New tools are available for analysing WGS data with the improved ability to reveal copy number changes and it is feasible to filter only for NBS variants. The discussions about expanding disease panels has been ongoing for some years now; it is not specific, *per se*, to the potential use of NGS and there could be as many different views on this topic as there are health care systems. While it is not the focus of this chapter and we will not discuss it further here, we highlight that the debate over expanding the disease panel is germane to most situations where new technologies allow for the testing of more loci with a decrease in cost per loci.^{3,5,6}

Should the sequences generated at birth be used to inform future medical scenarios of each individual, the questions of who will take the responsibility for storing and reanalysing and interpreting the sequence data, where it will be stored (i.e. in the medical records) as well as exactly what data will be included will have to be addressed. In this way, if genomics is to take on a larger role in medical health management, there will be a need for better education and training of health care professionals who will be referring to sequencing data. Genomics England Ltd with Health Education England are at least beginning to address these issues. In tandem, a better network of communication between different specialties and genetic/genomic experts could truly maximise the informational content of sequence data without necessarily having to (re)train all specialties in genomics. Indeed, the amount of knowledge and training needed to properly interpret sequence data is not trivial and may best be supported throughout by genomic experts. If sequence data generated at birth are to be stored in medical files, an explicit and clear protocol for safe storage and access should be elaborated. All sequencing results (raw or analysed) should be treated like other clinical information retained in a patient's medical file, including adequate protection of privacy and confidentiality.

5. Evidence

In a study using NGS to sequence 126 genes that together comprise the majority of the genes currently implicated in most newborn screening conditions (except hearing loss, critical congenital heart defects and severe combined immunodeficiency), NGS was evaluated for “second tier” NBS.⁸ This study was not population-based, but instead used samples from children known to be affected by specific newborn screening disorders. Furthermore, while the technique worked well to identify potentially harmful sequence changes in the targeted 126 genes, this information alone only provided the correct diagnosis for 75% of the cases; for the rest, information about the child’s clinical condition was needed to complete the diagnosis. The study did however, address the feasibility of using dried blood spots for NGS but better protocols remain to be determined. Further evidence regarding the use of NGS in NBS programmes should be collected from existing ongoing programmes and as needed from carefully planned pilot studies which properly manage the ethical, regulatory and logistical challenges of conducting such population-based prospective research.² 25 million USD over five years to 2015 were spent by the National Institutes of Health in the United States of America on various pilot projects to study the implications, challenges and opportunities associated with the possible integration of genome sequencing in newborns.⁷ Evidence will have to be amassed regarding the development of a suitable test (sensitivity, specificity, positive and negative predictive value, utility), including the identification of variants targeted and the related conditions; the clinical procedure, including counselling and/or the communication of information and consent; the treatment or prevention pathway; the (potential) impacts on test recipients and the health care setting; the acceptability to the general population; and the development of an analysis/calculation of costs and the actual study and analysis of costs.⁸ Particular attention should be paid to the context of, (for example, country, organisation) and methods (for example, quantitative

or qualitative) used to generate evidence in order to properly interpret and transfer the information from such evidence to a different milieu. Furthermore, an established monitoring scheme for such a programme would have to be in place and the ethical, legal and social issues should be discussed before as well as while a programme is in place. At the time of writing, the joint statement issued in 2015 by The Public and Professional Policy Committee of the European Society of Human Genetics, the Human Genome Organisation Committee on Ethics, Law and Society, the PHG Foundation, and the P3G International Paediatric Platform recommends the adoption of a targeted sequencing or targeted analysis approach.⁸ The interpretation of DNA data in a population of healthy newborns is a challenge. The genotype–phenotype relationship in metabolic conditions is often complicated. In the case of Pompe disease, for instance, there is a large clinical diversity among patients with the same genotype. Furthermore, the sensitivity of sequencing analysis for specific disorders in each target population should be carefully considered, as it may be lower compared with present metabolic testing for some disorders. An example is given by the screening strategy for cystic fibrosis (sensitivity of over 95%) where not all disease-causing CFTR variants are known. A genotype first approach might have a lower sensitivity, with a wide variance in different populations. Also, the sequencing first approach would identify not only affected children but also the carriers of a combination of variants that might never cause a significant disease. In this situation, there would be a risk that these children would be considered ‘affected’ with the potential consequences of overtreatment.

6. Cost effectiveness

Specifically regarding financial costs for revised NBS, the following should be considered: i) given the high throughput nature of NGS, a programme may appear more cost effective if more variants/genes/conditions are included, however, this fact should not drive the decision to expand the disease/variant panel; ii) should sequence data be stored and used for medical purposes throughout the lifetime of an individual, the costs and benefits of this should also be included in economic analyses; iii) the following costs should also be considered: those of safely storing the data (potentially for a lifetime), ongoing or future accurate and efficient clinical interpretation, validation and communication of results, the potential need for additional sequencing of family members, and follow-up or confirmatory health care testing. Indeed, when considering all of the costs, it is clear that the so-called “1,000 USD” genome greatly underestimates the true costs of using WGS or WES in a clinical setting.⁹

Indeed, it has been highlighted that *“genome-scale information challenges traditional economic assessment much as it challenges approaches of traditional health care delivery.”*¹⁰ Many issues contribute to this challenge, including: i) communication and understanding gaps between different stakeholders needed to develop adequate economic assessment approaches; ii) a lack of systematic evidence base showing that genomic level data results in positive health outcomes and a lack of accepted thresholds for clinical utility and value for money; iii) the lack of existing economic assessment methods that are adequate to deal with the dynamic and rapid pace of genomic discovery; and iv) the fact that there is no formal approach to capturing personal utility (of patients and families) in assessing the value of WGS or WES approaches, which is being taken increasingly more seriously in the era of more patient-centred approaches.¹⁰

7. Informed consent and the interests of the child

While many authors agree that presumed consent (and/or different versions of consent that may not meet the criteria for traditional genetic testing) is appropriate for current NBS programmes, this is not likely to be the case if (targeted) genome sequencing is introduced. Invariably, a new model of informed consent for the use of NGS in NBS will have to be elaborated. It will have to inform parents adequately while not taking so much time and resources that the programme is limited. Many professional societies have emphasised that the decision to offer genetic or genomic testing or screening, including the use of WGS, to children should be guided by the best interest(s) of the child.^{8,11} This is an important and respected ethical concept in pediatric medicine¹² which has international legal recognition in the United Nations Convention on the Rights of the Child (article 3).¹³ In the context of genetic or genomic testing this has been interpreted as meaning that only scientifically valid and clinically useful information about conditions that manifest during the neonatal or childhood periods should be offered.¹¹ Indeed, it is still an open question whether the use of WGS or WES may offer many (more) benefits to the neonate or child versus the use of current methods. However, it may lead to useful information for the parents (for example, incidental findings that could help reproductive decision making) as well as help prevent a future diagnostic odyssey for relatives or the individual as an adult.¹² Nevertheless, the guiding principle in decisions in NBS should be the best interests of the child. Therefore, any incidental findings for serious conditions that would develop in the minor and are actionable should be reported.

Given the public health benefits of NBS, that it is conducted for the best interests of the child, and that it is considered as routine pediatric care, unlike many other genetic tests conducted outside of screening programmes, NBS is often conducted without explicit informed consent.¹¹ Currently, in the UK, consent must be obtained verbally from the parents and noted in the child’s health care records.¹ Regarding informed consent, of note is that expert stakeholders are currently attempting to agree on what informed consent procedures and informed consent forms should look like in the context of WGS and WES for adults (for example EuroGentest Committee and The Public and Professional Policy Committee of the ESHG). Presumably once a form of consensus or area of agreement is reached for this, further modifications for different specific contexts could be easier.

8. Sequence data handling

The NHS has a current code of practice for the retention and storage of residual newborn blood spots.¹⁴ It “sets out arrangements for the retention, storage, use and release of residual newborn blood spots and related information and communication requirements”. Presumably “related information” could mean data generated from an analysis of the blood spot. This code of practice should be reviewed and assessed for its adequacy to cover the amount of data generated by NGS. Alternative or additional procedures should be established as needed.

It has been suggested that with the fast pace of improvement in the technology, as well as with decreasing costs, it may be more efficient to simply re-sequence individuals as needed, instead of storing sequence data for future use. In this way, much of the challenges of storing the data could be avoided. On the other hand, the notion of maximising the health care and/or research potential of any sequence generated is also an option. In this case, careful consideration will have to be made regarding how, which type of data file(s), where, and for how long the data will be stored. What will be the full computational requirements of secure and ethically acceptable data processing, storage and retention as well as what will the costs of this be? Who will have access to the data? For example, will (some of the) data be stored in medical files? Will both for-profit and non-profit researchers have access to the data? Will each individual have access to his or her own sequencing data? How could this be done? Furthermore, the exact future use(s) of the data will have to be thought through carefully - such as for research use and/or for future health care decisions of each individual - and the data should be handled accordingly. Such long-term storage and future uses could only be possible with the implementation of informed consent procedures as well as with adequate privacy and confidentiality protection. Education and communication regarding the risks and benefits of such storage could be included in Public Health communications to the public regarding genomics.

9. Direct to consumer genomics and impact on Public Health

Commercial companies, functioning outside of the traditional health care system are currently advertising and sometimes offering genome-wide testing as well as WGS or WES direct-to-consumer. In this way, parents may be able to have their children sequenced at any stage of life regardless of what the public health care system offers or what professional societies recommend. EU guidance recommends that all testing (private sector or otherwise) is physician-led and that adequate post-test discussion and counselling be offered, preferably in face-to-face consultation.¹⁵ Even if there is a health care professional involved in the actual offer, yet the testing is advertised direct-to-consumer (DTC) by commercial companies, the potential power of these messages to heavily sway public preferences and ultimately impact the public health care system should not be underestimated. Given this potential impact on the publically funded health care system, the ESHG has stated that such commercial DTC GT companies should also have to abide by the same recommendations as are set out for the public health care system.⁸

10. Conclusions

While the existence of such powerful technology such as NGS naturally prompts us to question where it can be of use in the health care system, decisions about implementation should not be technology driven. The best interest(s) of the child should be the driver for any decisions made regarding the NBS programme. The main objective of an NBS programme should be to identify genetic variants that confer a high risk of a treatable or preventable disease, for which action must be taken in the newborn or early childhood. It is likely that NGS will eventually be integrated into NBS programmes, however for now, the European Society of Human Genetics advocates a targeted sequencing or targeted analysis approach.⁷

Pilot studies involving NGS in newborns are ongoing in the USA and are generating empirical evidence that may help other health care systems better judge the actual challenges and benefits of using NGS in NBS. That being said, different national and regional NBS programmes across the globe currently differ enormously based on differing values, differences in interpretation of inclusion criteria for screening panels, and differences in (financial) resources, hence results from different settings will not be applicable to all contexts. Furthermore, many current studies are obtaining the views and opinions of stakeholders using hypothetical scenarios, which is an important exploratory step, however the results of these studies should be carefully interpreted and contextualised, especially regarding generalisability of results and meaningfulness given different stakeholder knowledge and understanding of the relevant issues.^{16,17} The role that different stakeholder groups' opinions and preferences should have in NBS programme decision-making should also be carefully considered and made explicit where possible. Some authors have clearly voiced concerns that we are not currently ready to implement population based newborn screening using a WGS approach.^{9,17,18,19}

11. Suggestions for policy makers

- The implementation of NGS in NBS should not be technology-driven and should always be for the health interests of the child.
- When considering whether genome sequencing should be incorporated into NBS programmes the following issues should be addressed: is there a gap in current NBS, that implementation of NGS would fulfill? Alleles should be highly penetrant and conditions should be treatable or preventable starting in early childhood. There needs to be robust evidence for variant disease association with high levels of sensitivity and specificity of assays. The clinical impact of tests should be monitored and there should be public acceptability of new NBS programmes. Information and consent should reflect the needs of parents. New methods of returning information to parents about disease risk and variants of unknown significance must be designed. Robust policies over the future use of generated sequence data for later medical decision-making and/or for use in research need to be agreed. The costs of implementation of new NGS-based tests need to be evaluated as does the cost of educating patients and the public, providing counselling and data and sample storage.
- Given the lack of evidence for implementing NGS approaches in NBS, targeted genome sequencing or targeted analysis approach is advocated at the current time. It is likely that NGS-driven NBS will need to be combined with biochemical approaches (for example, for congenital hypothyroidism).

12 References

1. <https://www.gov.uk/guidance/newborn-blood-spot-screening-programme-overview>
2. Botkin, J.R., et al., Parental permission for pilot newborn screening research: guidelines from the NBSTRN. *Pediatrics*, 2014. 133(2): p. e410-7.
3. Loeber, J.G., et al., Newborn screening programmes in Europe; arguments and efforts regarding harmonization. Part 1. From blood spot to screening result. *J Inherit Metab Dis*, 2012. 35(4): p. 603-11.
4. Bhattacharjee, A. et al., Development of DNA Confirmatory and High-Risk Diagnostic Testing for Newborns Using Targeted Next-Generation DNA Sequencing. *Genetics in Medicine* (2015) 17,337–347
5. Burgard, P., et al., Newborn screening programmes in Europe; arguments and efforts regarding harmonization. Part 2. From screening laboratory results to treatment, follow-up and quality assurance. *J Inherit Metab Dis*, 2012. 35(4): p. 613-25.
6. Moyer, V.A., et al., Expanding newborn screening: process, policy, and priorities. *Hastings Cent Rep*, 2008. 38(3): p. 32-9.
7. Kaiser J NIH Studies explore promise of sequencing babies' genomes 2013, <http://news.sciencemag.org/biology/2013/09/nih-studies-explore-promise-sequencing-babies%E2%80%99-genomes>.
8. Howard, H.C., et al., Whole-genome sequencing in newborn screening? A statement on the continued importance of targeted approaches in newborn screening programmes. *European Journal of Human Genetics*, 2015. 23(12): p. 1593-1600.
9. Beckmann, J.S., Can We Afford to Sequence Every Newborn Baby's Genome? *Human Mutation*, 2015. 36(3): p. 283-286.
10. Feero, W.G., C. Wicklund, and D.L. Veenstra, The economics of genomic medicine: insights from the IOM Roundtable on Translating Genomic-Based Research for Health. *JAMA*, 2013. 309(12): p. 1235-6.
11. Knoppers, B.M., et al., Whole-Genome Sequencing in Newborn Screening Programs. *Science Translational Medicine*, 2014. 6(229).
12. Senecal, K., Vears, D.F., Bertier, G. Knoppers, B.M., Borry, P., Genome-based newborn screening: a conceptual analysis of the best interests of the child standard. *Personalized Medicine*, 2015. 12(5): p. 439.
13. (CRC), U.N.C.o.t.R.o.t.C., CRC General Comment No. 14 (2013) on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1). CRC /C/GC/14 (2013). U.N.C.o.t.R.o.t.C. (CRC), Editor. 2013.
14. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/415981/Code_of_Practice_for_the_Retention_and_Storage_of_Residual_Blood_Spots.pdf
15. European Parliament, Directorate General for Internal Policies, 2013
16. Joseph, G., et al., Parental Views on Expanded Newborn Screening Using Whole-Genome Sequencing. *Pediatrics*, 2016. 137 Suppl 1: p. S36-46.
17. Reinstein, E., Challenges of using next generation sequencing in newborn screening. *Genetics Research*, 2015. 97.
18. Nardini, M.D., et al., Genomic Counseling in the Newborn Period: Experiences and Views of Genetic Counselors. *Journal of Genetic Counseling*, 2014. 23(4): p. 506-515
19. 20 Ulm, E., et al., Genetics Professionals' Opinions of Whole-Genome Sequencing in the Newborn Period. *Journal of Genetic Counseling*, 2015. 24(3): p. 452-463.

Chapter 12

Non-invasive prenatal testing

Chapter lead and author
Prof Peter W Soothill¹

¹ Emeritus Clinical Professor, University of Bristol

1. Summary of key points

One of the first applications into medical clinical care of the results of genomic research was the use of genetic tests to inform reproductive choices. Initially testing was by the use of invasive procedures but recently non-invasive prenatal testing (NIPT) using fetal cell-free DNA (cell-free DNA) in maternal blood has provided a major advance in antenatal care and screening.

Latrogenic miscarriage of wanted pregnancies is being reduced by greatly reducing the number of invasive procedures in pregnancy. For Down's syndrome screening (Down's screening), and to guide the use of anti-D, we can expect the need for hundreds of thousands of cfDNA tests per year. Challenges within the NHS will include making savings from previous screening approaches to fund the new tests and making decisions about the reorganisation of the number and location of large genetic laboratories.

2. Background

A national policy to offer all pregnant women (regardless of their age) a screening test for Down's syndrome was announced in 2001¹ and subsequently implemented in the NHS in England by the Fetal Anomaly Screening Programme of the National Screening Committee. Screening in pregnancy for fetal abnormality involves the woman carefully opting in and these tests should not be regarded as "routine". Screening is a staged process including a) a policy for who would be offered screening, b) the screening test itself which separates groups into lower or higher risk groups and then c) diagnostic testing. Each of those steps comes with specific considerations in terms of number of tests, timing, false positive and false negative results, and unwanted consequences such as miscarriage after invasive testing. False negative screening tests will result in the birth of a baby affected by a condition that the pregnant woman wanted to avoid. False positive results generate anxiety and further testing for no clinical benefit. The term 'contingent' testing is sometimes used to describe serial screening tests.

The screening tests available have improved significantly, moving from the use of maternal age alone in the 1990s, to second trimester protein biochemical testing (for example the 'Quadruple test') and then tests earlier in gestation using the nuchal translucency scanning measurement and serum biochemistry (the 'Combined test'). The improvement in sensitivity and specificity led to a fall in the false positive rate from over 5% (and so 1 in 20 of pregnant women screened) to about 2% and that resulted in a dramatic fall in the number of invasive procedures without a significant change in the detection rate (see Table 1). The better tests meant that women who wanted to know about affected pregnancies were better informed, while much fewer normal pregnancies were exposed to the risks of invasive procedures.

Table 1 Extract from the 2011-2012 Annual Report of the Fetal Anomaly Screening Programme

Type of invasive procedure	Number undertaken 2003/2004	Number undertaken 2004/2005	Number undertaken 2005/2006	Number undertaken 2006/2007	Number undertaken 2007/2008	Number undertaken 2008/2009	Number undertaken 2009/2010
Amniocentesis	28,700	24,349	22,625	14,733	12,932	12,145	9,894
Chorionic villus sampling	8,268	7,980	7,819	4,781	4,681	3,520	3,701
Total	36,968	32,329	30,444	19,514	17,613	15,665	13,595

Note Improvements in screening resulted in 923,373 less invasive procedures per year and a rise in the detection rate with no significant rise in postnatal diagnosis

Source 2011-2012 Annual Report of the Fetal Anomaly Screening Programme

Everyone, including non-pregnant people, has large amounts of free-floating, non-cellular fragments of cell-free DNA (cfDNA) within their blood plasma. The recognition that cfDNA from pregnant women includes a mixture of maternal DNA and fetal DNA opened the potential of using this phenomenon in various aspects of non-invasive prenatal testing (NIPT) and screening.² At first the fetal cfDNA was thought to come from fetal cells in the maternal circulation but it was then recognised that the rapidly dividing cells in the placenta release fetal DNA fragments into pregnant women's blood (for example "blighted ovum" pregnancies without a fetus have normal cfDNA levels³).

Since total cfDNA in maternal blood includes a significant proportion (about 5-20%) of fetal cfDNA (with significant individual variation and increasing with gestational age), the ability to detect paternally inherited DNA sequences that are not present in the mother's genome by relatively simple tests such as real-time PCR was rapidly implemented in the care of patients in special circumstances. For example Y-chromosome signals to exclude or confirm risk in pregnancies potentially susceptible to sex-linked disease⁴ or determine fetal blood group in pregnancies with allo-immunisation⁵ were rapidly implemented as clinical services. In contrast, the use of this biological phenomenon for the non-invasive detection of chromosomal abnormalities such as Down's syndrome was much more difficult. That is because pregnant women have their own (maternal) chromosome 21 sequences and so instead of identifying an additional signal not present in the mother's genome, the challenge was to detect an additional "dose" of chromosome 21 sequences only present within the fetal fraction of a sample containing a mixture of maternal and fetal DNA.

To many people's surprise, DNA sequencing technologies do allow detection of the extra chromosome 21 signals because the amount of signal DNA can be determined with sufficient accuracy to assess a changed ratio of 21-chromosome DNA signals to control sequences from other chromosomes. There are two approaches, one is "targeted" sequencing when the sequencing analysis is restricted to target signals from chromosomes (such as 21, 13 and 18) or "shot-gun" /whole genome sequencing when signals from all chromosomes are assessed. Using either of those approaches have resulted in excellent sensitivity and specificity results – both over 99%.^{6,7,8}

The high sensitivity and specificity that can now be achieved through sequencing technology present a very major advance that offers important opportunities for improved care of pregnant women but also raises questions that need to be decided and consequences that need to be managed. The implications we now face will be discussed in this chapter.

Table 2 Results of non-invasive prenatal testing (Targeted - Norton et al 2015 with Combined test as control) and whole genome sequencing (Biachi et al 2012). Test performance for Trisomy 21 in the Primary Analysis Cohort, according to maternal and risk*

Variable	Standard Screening All Patients (N-15,841)	Cell-free DNA Testing		
		All Patients (N-15,841)	Maternal Age <35 Yr (N-11,994)	Low Risk (N-14,957) [†]
True positive - no	30	38	19	8
True negative - no	14,949	15,794	11,969	14,941
False positive - no	854	9	6	8
False negative - no	8	0	0	0
Sensitivity (95% CI) -%	78.9 (62.7-90.4)	100 (90.7-100) [‡]	100 (82.4-100)	100 (63.1-100)
Specificity (95% CI)-%	94.6 (94.2-94.9)	99.9 (99.9-100) [§]	99.9 (99.9-100)	99.9 (99.9-100)
Positive predictive value (95%CI)-%	3.4 (2.3-4.8)	80.8 (66.7-90.9) [§]	76.0 (54.9-90.6)	50.0 (24.7-75.3)
Negative predictive value (95%CI)-%	99.9 (99.9-100)	99.9 (99.9-100) [¶]	99.9 (99.9-100)	99.9 (99.9-100)
Positive likelihood ratio	14.6	1755.9	1995.8	1868.6
Negative likelihood ratio	0.22	0	0	0

* P values are for the comparison between standard screening and cell-free DNA screening in the primary analysis cohort.

[†] Low risk was defined as a mid-trimester risk of trisomy 21 of less than 1 in 270 on standard screening

[‡] P=0.008

[§] P<0.001

[¶] P=0.005

Massively Parallel Sequencing Performance	Sensitivity (%)	95% CI	Specificity (%)	95% CI
Trisomy 21 (n-493)	100.0 (89/89)	95.9-100.0	100.0 (404/404)	99.1-100.0
Trisomy 18 (n-496)	97.2 (35/36)	85.5-99.9	100.0 (460/460)	99.2-100.0
Trisomy 13 (n-499)	78.6 (11/14)	49.2-99.9	100.0 (485/485)	99.2-100.0
Female (n-433)	99.6 (232/233)	97.6 to more than 99.9	99.5 (199/200)	97.2 to more than 99.9
Male (n-433)	100.0 (184/184)	98.0-100.0	100.0 (249/249)	98.5-100.0
Monosomy X (n-433)	93.8 (15/16)	69.8-99.8	99.8 (416/417)	98.7 to more than 99.9

CI, confidence interval.

3. Discussion of cfDNA screening

3.1 What do we want to screen for?

Although the policy decision was made to offer all pregnant women a screening test for Down's syndrome specifically, some of the tests could detect other problems by "accident". For example a very increased nuchal translucency can indicate an increased risk for other chromosome disorders such as Edwards (trisomy 18) and Patau's (trisomy 13) syndromes and complications such as miscarriage and heart malformations. While superficially it might appear that women who want to know about Down's syndrome would also want to know about more severe conditions, that is not necessarily the case when the prognosis is very different - a newborn baby with a very short life is different from a child with long-term disability. Over time, a consensus has now developed that screening for trisomy 21, 13 and 18 should be available if requested but women should be able to select rather than face an "all or none" approach.⁹

The potential information available from a cfDNA sequencing test could extend to fetal blood group, fetal sex in pregnancies at risk of sex-linked disease or looking for chromosomal deletions or duplications (some of uncertain significance). Some test services have decided to use all the sequencing "power" to look at the three trisomies listed above, while others have chosen whole genome sequencing. To a large extent the information obtained is determined by the information analysis chosen because the information is within the sequencing data but the depth of the information analysis is a choice. Some information could have unwanted consequences, such as paternity information or sex prediction, with uncertainty on how this will be used. Furthermore, newer approaches are moving towards single gene diseases such as skeletal dysplasia and cystic fibrosis. The longer the list of conditions looked for, the higher will be the false positive rate which is the basis of a current debate around screening for micro deletions and duplications during whole genome sequencing NIPT. The higher the screen positive rate the higher the number of women with invasive procedures, some of which would be of debatable value.

3.2 Pre-test counselling

Over the last 15-20 years a great deal of effort has been taken to ensure Down's screening tests do not become "routine" and are instead chosen in an informed way by those pregnant women who want to opt-in to screening. So patient information leaflets and training of clinical staff including midwives and obstetricians have been a focus of the Fetal Anomaly Screening Programme for many years. There has been some concern that cfDNA technology, being a "simple blood test", might remove a medical consenting process required for invasive procedures acting as a barrier to women choosing a test without fully understanding the consequences of the possible results. However, a NIHR funded research programme (RAPID) has reported that NIPT was widely welcomed by staff and pregnant women.

Also, since NIPT is a screening test, further counselling is still required before an invasive procedure. In 2011 about 540,000 of the 723,000 pregnancies in England and Wales choose to have screening¹⁰ but the work, usually by midwives, in counselling the whole group should not be underestimated. Those numbers indicate the numbers who will take up cfDNA universal screening if available but there will be more who choose fetal rhesus D grouping without Down's screening.

3.3 Prenatal screening or prenatal diagnosis?

With sensitivities and specificities for Down's syndrome of both over 99%¹¹, superficially those results would appear to be reaching the levels of performance required for a diagnostic test. However, that is not the case because the positive predictive value is still only about 50% (i.e. if the result suggests Down's syndrome about half of those women will be carrying a fetus with normal chromosomes). The reason for that low positive predictive value is partly because of the low prevalence and that the placental mass sometimes has a mixture of normal and trisomic cells (often called "mosaic") and that can also occur due to a miscarried trisomy 21 twin, sometimes not visible on ultrasound. The presence of some trisomic cells in the placenta when the fetal cells are 100% normal is not very rare (about 1% of pregnancies), and for that reason it is essential this technology is still used as a screening test, and with invasive diagnostic testing confirmation before a choice about pregnancy outcome is made after a positive result. Some women will use cfDNA to be prepared and informed while avoiding the risks of invasive testing even when termination of pregnancy was not an option for them.

3.4 Unsuccessful tests

In addition to the about 50% positive predictive value, a result is not always available from a cfDNA test. This is usually because the sample has not achieved the quality criteria, often because the amount of fetal cfDNA in the sample is unusually low. Low fetal cfDNA fraction is more likely when these samples are taken at too early a gestational age or in obese women – it seems a large BMI results in a higher blood volume and dilution of the signal. However, in most of these cases simply taking a second sample when the pregnancy is 1-2 weeks more advanced and repeating the test will then be successful in about 60-70% of cases. cfDNA tests are available in twins but there is much less data relating to the performance than in singletons, and the complex issues of zygosity/chorionicity and subsequent management decisions in multiple pregnancies mean that many think these tests should be managed within a Fetal Medicine centre.

3.5 Universal or Contingent screening

The National Screening Committee has recommended contingent screening be introduced¹² and so offering cfDNA testing when the Combined Test is high risk (more than 1 in 150). That approach has the advantage of getting the maximum possible reduction in the number of invasive procedures undertaken in unaffected pregnancies, while keeping the number of sequencing tests down to about 2% of the pregnant population who choose screening and so in the order of 10,000/year in England. Also, this approach will allow the assessment of NIPT in a national programme, especially the performance of trisomy 13 and 18 screening, uptake and failed tests. The main disadvantage is that the false negative rate of the overall programme is that of the initial screen (the Combined Test), currently about 15%. That means that in women who choose screening, about 15% of pregnancies with Down's syndrome will not be detected as higher risk and so the condition will be diagnosed after birth. There are approximately 1,840 continuing pregnancies with Down's syndrome per year. So if about 1,350 of these opt for screening, about 200 children will be born per year with Down's syndrome following the woman's choice for screening and having a low risk (false negative) result. This contingent approach also means no savings are available from the costs of Combined Screening including scanning and biochemistry because they are still needed. However, there will be significant savings from the reduction of invasive procedures from about 30,000 to perhaps 500/year in England.

3.6 Timing of test, number of laboratories, sample and result transfer and quality assurance or Contingent screening

Pregnancy care is already complex and over the last 20 years the number of clinical visits within routine antenatal care has been reduced.¹³ Therefore it is important to consider how this approach could be added to other aspects of pregnancy care within the NHS. If a contingent approach is used the result of the Combined Test must be completed first and so the cfDNA test would be after 13 weeks' gestation. With universal screening the NIPT test could be taken as early in pregnancy as it is effective and that gestational age may change as the technology improves. However, there must be time for the woman to receive and understand the information and make her choice, and also the miscarriage rate of pregnancies with chromosomal abnormalities is very high, and it might not be helpful to undertake this test in a pregnancy that is about to miscarry. Others may feel that obtaining an abnormal result may be of some comfort to women who miscarry and provide an explanation. At present it seems this would fit best into an existing 12 – 14 week visit. Early cfDNA testing has the advantage that if the fetus is rhesus D negative the result could be linked to avoiding administration of anti-D (see later), which generally starts from about 12 weeks. Other aspects are that amniocentesis should not be done before 15 weeks' gestation, so following a positive NIPT result for trisomy 21 or trisomy 18 or 13 with abnormalities on a scan, chorionic villus sampling will be offered, but for trisomy 18 or 13 on NIPT with a normal scan, waiting for an amniocentesis may be suggested. The availability of a surgical approach to termination of pregnancy is also important and may reduce as the pregnancy becomes more advanced.

A universal cfDNA screen offer would mean that all women who want screening (about 500,000/year) would have sequencing with the associated costs but the detection rate would be expected to be close to 100% provided the published data are reproduced in NHS service laboratories. That higher detection rate would be in addition to the clinical benefit of the decrease in invasive procedure numbers. The other advantage is the resources now being spent in Combined Screening could be saved and contributed to the costs. During implementation of this new technology the decision to use contingent screening first seems correct, but it also seems likely that universal screening will be the subsequent goal. Indeed, pregnant women who can afford testing are already using sequencing as the primary screen within the private sector in order to have the very low false negative rate. But if such self-funding patients add NIPT to existing approaches – i.e. have NIPT after a dating scan and then have the Combined Test that would not be a sensible approach within the NHS.

Another aspect of pregnancy care is the offer of anti-D immunoglobulin to the 15% of pregnant women who are blood group rhesus D negative. National fetal RhD cfDNA testing programmes to direct antenatal anti-D prophylaxis have been introduced successfully in other countries^{14,15,16} and this has led to the suggestion that the continuing practice of giving anti-D (a blood product pooled from multiple donors to healthy pregnant women) when no benefit can result because the fetus is rhesus D negative, is ethically unreasonable.¹⁷ The policy¹⁸ of giving anti-D antenatally to all RhD negative women, means that almost 40% of RhD negative pregnant women who will carry a RhD negative fetus (approximately 40,000 women per year in England and Wales) have been receiving anti-D unnecessarily. Giving anti-D unnecessarily is in contrast to the management at the birth of the baby, when cord blood is sent for D-grouping and the mother is offered postnatal anti-D only if the baby is RhD positive.

Maternal blood samples for cfDNA testing can be sent by the midwife to the usual hospital pathology laboratory with a form documenting the purpose of the sample and giving the estimated date of delivery (EDD) based on the dating scan. The EDD is essential and needs to be reported on the result of the test to identify the pregnancy in order to avoid the potential risk that a filed/stored result could be incorrectly ascribed to a possible future pregnancy. For example, a result from a pregnancy that miscarried could be misinterpreted. For several years samples have been flown to other parts of the world and these samples have been found to be very robust to transport and delays.

Such samples need to be recorded on the existing pathology computer systems as a maternal sample/test and then transferred as a "send-away" sample from each of the hospital pathology laboratories to the testing laboratory. Test results must be sent electronically to hospital laboratories. It is important that NHS computer systems be able to link these results in the maternal records with the subsequent child's NHS number/NHS records.

In view of the 500,000 or more tests that may in the future be requested, economies of scale need to be balanced by the security of having more than one laboratory to continue the service in the event of a laboratory problem. Also, the remarkably good published results of sequencing are probably the result of very high standards in the laboratory and quality assessment and monitoring will be essential. If the NHS decided to develop two laboratories nationally for these tests, the local sample processing and subsequent transfer will need to be agreed. Also the information technology needs to ensure both ready and reliable access to the results combined with confidentiality for what can be emotionally sensitive information.

3.7 Funding

Internationally, the cost per test of sequencing has fallen remarkably over the last few years. Savings that could help fund these tests could be made by stopping serum biochemistry (most commonly pregnancy associated plasma protein A and human chorionic gonadotropin assays) and stopping the nuchal translucency measurements. However both those tests have been considered to have other potential values:

- a) Such savings from the serum biochemistry costs would probably only be possible if the protein assay antenatal screening laboratories were closed. Some have suggested that low PAPP-A and high HCG results may be a useful way of screening for fetal growth restriction/placental disease.¹⁹ However, sequencing has the potential to indicate the total level of fetal cfDNA and this could be an indication of placental disease.²⁰
- b) The nuchal translucency measurement of over 3.5mm has been useful for conditions other than trisomy, including cardiac defects and dysmorphic syndromes. It seems possible that the nuchal translucency could be imaged rather than measured on everyone and the measurement takes some time (perhaps 5 minutes per NHS dating scan). However the early pregnancy scans would still be needed for other clinical reasons (such as pregnancy dating and very severe anomalies) and so the total savings would not be as great as if this scan could be abandoned completely.

The number of amniocentesis tests undertaken in England and Wales are expected to fall even further - from perhaps 2% to 0.1% of the pregnant population – equivalent to a drop from 10,000 to 500 per year. This will generate a significant saving but releasing that from the regional genetics laboratories may not be easy with the acknowledged increase in the need for genetic testing in other clinical specialties.

4. Clinical vignettes

Vignette 1

A 39-year old woman was in her first pregnancy after some difficulty becoming pregnant. She had some early pregnancy bleeding but she chose Down's screening by the Combined Test. The result was higher risk (1 in 110) as a result of her age and a high beta-HCG level. She opted for amniocentesis at 15 weeks and the Down's syndrome was excluded but she miscarries 10 days later.

In counselling afterwards, her obstetrician points out that the high HCG was a risk factor for miscarriage and it is impossible to tell whether the miscarriage was caused by the amniocentesis or would have happened anyway (perhaps as a result of a placental problem that caused the high HCG and so the Down's risk). The woman is angry and she still believes that the amniocentesis caused the loss of her pregnancy. The invasive procedure would almost certainly NOT have been offered had a cfDNA test been done.

Vignette 2

A 25 year old woman books rather late for care in her third pregnancy at 14 weeks' gestation but requests screening for Down's syndrome and so is offered the quadruple test. The result is low risk (1 in 310) and so she is reassured and then has an uncomplicated pregnancy and labour. Soon after delivery the midwife is concerned that the baby may have Down's syndrome and that is then confirmed.

Her obstetrician explained that the screening tests have an about 1 in 5 (20%) false negative rate. She is never able to understand or accept that a 309 out of 310 chance of a normal fetus can include a 15-20% "missed" rate.

The sensitivity of cfDNA is almost 100% and so false negatives are extremely unlikely.

Vignette 3

In 2020 (by when the NHS has introduced cfDNA screening), a 31 year old woman opts for Down's screening. A whole genome-sequencing test is used and a low risk result is obtained, which she is told makes an affected pregnancy very unlikely indeed. When 18 months old, the child is noted not to be meeting their "milestones" and developmental delay is diagnosed. A micro-array test shows a chromosomal deletion (not present in either parent) and the prognosis is very poor.

On direct questioning, her obstetrician agrees that the whole genome sequencing test that was done in her pregnancy could have detected this serious problem, but the National Screening Committee's policy is to interrogate the data to exclude trisomy 21, 13 and 18 only, and so this problem was not looked for. The woman cannot understand why, since the test she chose could have detected the problem, which has more serious consequences than Down's syndrome, and the analysis was not done.

5. Conclusions

Despite NIPT by cfDNA being a recent development, this is without question a major advance in pregnancy care. It seems likely that the technology will improve further but already the number of invasive procedures in pregnancy is falling markedly and are expected to fall much further. This is a good example of genomic technology reducing medical intervention.

After an opt-in decision following good information, reducing the false negative rate of about 15-20% for the Combined Test will be important to pregnant women, and so initial screening by cfDNA (universal screening) is likely to be the way forward. It also seems likely that relevant additional information such as fetal sex (Duchenne muscular dystrophy) and blood group (rhesus D group) will be gained using the same sequencing test.

The NHS will need to reconfigure prenatal laboratory genetic services with sequencing NIPT in mind.

6. Suggestions for policy makers

Following the existing national recommendations, policy makers should implement the recommended contingent cfDNA screening. At the same time, the consequences for laboratories if universal cfDNA screening is recommended should be considered/explored.

Robust and defensible criteria will need to be used to decide what conditions should be looked for within the sequencing data and what should not.

The consequences for NHS genetics laboratories will be significant. The number of invasive procedures in pregnancy as a result of screening tests will fall by about 95%. However the number of sequencing tests using either screening model will be very large. Policy makers should consider how many laboratories undertake these tests, balancing economies of scale with service security.

A universal cfDNA screen offer would mean that all women who want screening (about 500,000/year) would have sequencing with the associated costs but the detection rate would be expected to be close to 100% provided the published data are reproduced in NHS service laboratories. That higher detection rate would be in addition to the clinical benefit of the decrease in invasive procedure numbers. The other advantage of a universal screening offer is the resources now being spent in Combined Screening could be saved and contributed to the costs. During implementation of this new technology the decision to use contingent screening first seems correct, but it also seems likely that universal screening will be the subsequent goal.

7. References

- Chief Medical Officer Update 31. October 2001. Page 7. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4013572.pdf
- Royal College of Obstetricians and Gynaecologists. Non-invasive Prenatal Testing for Chromosomal Abnormality using Maternal Plasma DNA. Scientific Impact Paper 15 2014. <https://www.rcog.org.uk/en/guidelines-research-services/guidelines/sip15/>.
- Alberry M, Maddocks D, Jones M, Abdel Hadi M, Abdel-Fattah S, Avent N, Soothill PW. Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast. *Prenat Diagn.* 2007 May;27(5):415-8.
- Hyett JA, Gardener G, Stojilkovic-Mikic T, Finning KM, Martin PG, Rodeck CH, Chitty LS. Reduction in diagnostic and therapeutic interventions by non-invasive determination of fetal sex in early pregnancy. *Prenat Diagn.* 2005 Dec;25(12):1111-6.
- Finning KM, Martin PG, Soothill PW, Avent ND. Prediction of fetal D status from maternal plasma: introduction of a new noninvasive fetal RHD genotyping service. *Transfusion.* 2002 Aug;42(8):1079-85.
- Royal College of Obstetricians and Gynaecologists. Non-invasive Prenatal Testing for Chromosomal Abnormality using Maternal Plasma DNA. Scientific Impact Paper 15 2014. <https://www.rcog.org.uk/en/guidelines-research-services/guidelines/sip15/>
- Norton ME, Jacobsson B, Swamy GK, Laurent LC, Ranzini AC, Brar H, Tomlinson MW, Pereira L, Spitz JL, Hollemon D, Cuckle H, Musci TJ, Wapner RJ. Cell-free DNA analysis for noninvasive examination of trisomy. *N Engl J Med.* 2015 Apr 23;372(17):1589-97. doi: 10.1056/NEJMoa1407349. Epub 2015 Apr 1.
- Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Maternal Blood IS Source to Accurately diagnose fetal aneuploidy (MELISSA) Study Group. *Obstet Gynecol.* 2012 May;119(5):890-901.
- National Screening Committee recommendation on Trisomy 18 and 13. 2014. <https://www.gov.uk/government/news/earlier-screening-for-serious-abnormalities-in-pregnancy>
- 2011-2012 Annual Report of the Fetal Anomaly Screening Programme. Page 15.
- National Screening Committee statement on NIPT. 2016. http://legacy.screening.nhs.uk/policydb_download.php?doc=602
- NICE Antenatal Care for Uncomplicated pregnancies 2008. <http://www.nice.org.uk/guidance/cg62/chapter/1-Guidance>
- Clausen FB, Christiansen M, Steffensen R, Jørgensen S, Nielsen C, Jakobsen MA, et al. Report of the first nationally implemented clinical routine screening for fetal RHD in D- pregnant women to ascertain the requirement for antenatal RhD prophylaxis. *Transfusion* 2012;52:752-8.
- Wikman AT, Tiblad E, Karlsson A, Olsson ML, Westgren M, Reilly M. Noninvasive single- exon fetal RHD determination in a routine screening program in early pregnancy. *Obstet Gynecol* 2012; 120:227-34.
- de Haas M, van der Ploeg CPB, Scheffer PG, Verlinden DA, Hirschberg H, Abbink F, et al. A nation-wide fetal RHD screening programme for targeted antenatal and postnatal anti-D. *ISBT Sci Ser* 2012;7:164-7.
- Kent J, Farrell AM, Soothill P. Routine administration of Anti-D: the ethical case for offering pregnant women fetal RHD genotyping and a review of policy and practice. *BMC Pregnancy Childbirth* 2014;14:87.
- NICE. Routine antenatal anti-D prophylaxis for women who are rhesus D negative. Technology appraisal guidance 156. 2008. [www.nice.org.uk/guidance/ta156]. Accessed 7 August 2014.
- Royal College of Obstetricians and Gynaecologists. The Investigation and Management of the Small-for-Gestational-Age Fetus. Green-top Guideline No. 31 2nd Edition February 2013.
- Alberry MS, Maddocks DG, Hadi MA, Metawi H, Hunt LP, Abdel-Fattah SA, Avent ND, Soothill PW. Quantification of cell free fetal DNA in maternal plasma in normal pregnancies and in pregnancies with placental dysfunction. *Am J Obstet Gynecol.* 2009 Jan;200(1):98. e1-6.

Chapter 13

Solving Data Challenges

Chapter editor

Dr Claudia Langenberg¹

Authors

Genomic data environment

Mr Peter Counter²

Bioinformatics and statistical analyses

Dr Jessica Vamathevan,³ Prof Ewan Birney³

Clinical interpretation of genomic data

Dr Clare Turnbull²

Lifelong electronic health records for genomic medicine

Dr Claudia Langenberg,¹ Prof Harry Hemingway^{4,5,6}

1 MRC Epidemiology Unit, University of Cambridge

2 Genomics England

3 European Bioinformatics Institute (EBI), Wellcome Trust Genome Campus, Cambridge

4 Farr Institute of Health Informatics Research, University College London

5 Institute of Health Informatics, University College London

6 NIHR University College London Hospitals / UCL Biomedical Research Centre, London

1. Introduction

1.1 Overview of data challenges

The previous chapters have focussed on a range of benefits and applications of genomic medicine, here we outline some of the challenges arising from genomic data. The legal and ethical framework, evolution of patient and public understanding and the nature of consent are central in decisions about how data are obtained, shared, stored and used and reused in genomic medicine. As the National Data Guardian points out in her third report,¹ in genomic medicine traditional models of consent and information governance are not always appropriate because the longstanding distinction between research and clinical care becomes blurred and because of the identifiable nature of genomic data. Electronic mechanisms are needed to replace the current practices for researcher access to data which include paper-based agreements between users, institutions and data access committees. Current challenges and opportunities in “information governance” and the proposed need for a new national “conversation” to better inform citizens are presented in Chapter 02 of this report ‘100,000 Genome Project’.

In this chapter the data challenges we focus on include:

1. The large size of genomic data necessitates strategies for efficient storage and processing. The usual NHS approach to making copies of data from direct care for research is not practical for genomic data. A central repository is needed together with a substantial High Performance Computing environment.
2. Data from different sources are required to interpret the human genome, spanning diverse genomic and other “omic” data, and including the “phenome”, i.e. all phenotypic clinical and other health data, largely captured from clinical records.
3. The integration of data from different sources for genomic analysis and interpretation requires standards, tools and analytic pipelines that are rapidly evolving: these have been better established in bioinformatics than in health record informatics.
4. The quality and benefits of these data depend on scale and national concentration of effort and excellence, but this needs to be balanced with health system organisation.
5. The need to equip the NHS workforce with the leadership and skills in health informatics, bioinformatics and data science required to deliver the potential of genomic medicine.

2. Genomic data environment

2.1 Data size

Genomic datasets are big. Genomics England's systems, designed to process 100,000 genomes, has 16 petabytes of storage (= 16,000 terabytes). Compare this with NHS Digital's systems which process millions of transactions but hold data only in the low 100s of terabytes. The computing environment needed to process genomic data is also big. A complete service to process whole genomes at scale for the whole country will require a significant High Performance Compute environment which would be in the top few hundred systems in the world. The requirement for very large data and computing means that any arrangements for genomics data should be made very carefully to minimise any copying, and would naturally focus around a single shared repository.

2.2 NHS data environment

There is a sharp divide in the NHS between data for direct clinical care and data for secondary uses of commissioning, regulation and research. There are major differences in their legal and Information Governance frameworks, funding arrangements, ethics, consent requirements and patient and public expectations. Data for primary uses (direct care) is largely local and distributed across the service because it is generated and held in different point of care settings. Data for secondary uses is typically a subset copied out of direct care systems and brought together in national systems.

The approach of making copies of data for research works well for clinical data, and the NHS is unique in the world to have a comprehensive national set of health data collected over many years. Data available in current collections can be used for genomic research, either by linking individuals consented as part of separate research studies (e.g. UK Biobank), or patients participating in research as part of their clinical care (e.g. 100,000 Genome Project). Data for research is made available through NHS Digital who are planning to consolidate many disparate systems into the single data services platform. The Clinical Practice Research Datalink (CPRD) also collect NHS clinical data and makes it available to researchers.

The copying approach is, however, impractical for large genomic data sets. Genomic data needs to be held in a single, secure environment which allows for both identifiable clinical care use and for de-identified research use. The reasons for this are not only the sheer size of the data sets which have significant storage and networking costs, but also the very substantial computing environment needed to process manage them which cannot be easily replicated.

2.3 Data and process standards

There are already international standards in place for sequence data and variant definitions. But there are many other standards needed for safe, reliable interchange of interpretation information between different vendors and implementations. Genomics England has made a start on such definitions² but these standards need additional development and formal NHS approval. We must also participate fully in international standards development in genomic data sharing, since several countries are developing plans to build substantial genomic databases. There is significant value in sharing information from such large cohorts, but it will not be practical to combine them. There is a further need for standardisation around clinical data definitions, capture and communication and we also need to address quality and completeness issues. Far more can be interpreted from a genome sequence when an accurate patient record is available.

In addition, a national genetic processing and interpretation pipeline is required. This must be as automated as we can make it. Current genetic services are sometimes described as a 'cottage industry' as we do not yet have the maturity of practice, tools, standards and linkages needed to make interoperation work. At some stage in the future, when such maturity has grown, it may be possible to build more diverse local or regional systems that nevertheless can work effectively together.

2.4 Genomic data sharing

The normal process of copying the data out of primary direct care systems for secondary purposes is impractical, but there is another driver for keeping all the genomic data in one place: sharing and exploration is a key part of the way genomics specialist clinicians work. Genomic medicine is really still in its infancy and findings are often not clear. Was that variant really there? Is it really the cause of the symptoms? So clinical genetic scientists will often want to look across all the data for similar examples of the same disease or the same genomic profile. Clinicians, geneticists and data scientists with highly specialised skills come together to diagnose and manage conditions which are not amenable to standard medical approaches. This is research, but entirely in the cause of managing a patient, as part of direct care. It is particularly important to have the widest pool of data for rare disease which is a primary target for genomic medicine.

However, clinical genetic scientists in the UK have not clearly approved the way they can share data. They are unwilling to put findings and knowledge into international databases due to concerns over patient level data leaving the country and being publicly available.

They need a way to discuss cases across geographic and organisation boundaries but many are unsure what is legal to hold, what can be used for which purposes, and what can be passed on. This has driven a climate of caution, since there are no sanctions for not passing on what should be shared, but there is the threat of legislation, press stories and the Information Commissioner's Office for passing on what is not sharable.

2.5 National and regional approaches

There has been recent discussion of a proposed regional structure not just for clinical use but for secondary uses too. While this may appear superficially attractive, it ignores two fundamental issues: the legal framework to hold such data, and the issues around analysis of dispersed data. The legislation is open to interpretation, but NHS Digital is possibly the only organisation that can hold patient-level data without consent for purposes other than direct care. Temporary rights can be obtained based on advice from the Confidentiality Advisory Group (CAG), but any long term regional solution may need primary legislation. This is problematic in both certainty and timescale. Analysis normally requires that data be held together for searching and comparison. It is possible to do analysis across distributed systems, but one of two conditions must exist for this to be effective: either the distributed systems must be largely identical in data structure and infrastructure software, so that analysis applications can run across the systems; or there must exist a mature set of standards, and a range of sophisticated heterogeneous software systems that implement the standards. The former is not really practical in the NHS given the diversity of funding and governance arrangements. The maturity of standards and software required for the latter does not yet exist.

For the UK to deliver optimally for patients and to continue to have a leadership role in genomic medicine and research, we need to rapidly focus on finding practical solutions for data sharing and collaboration at the National level. There is a clear need to establish a single country-wide database and infrastructure for handling whole genomes. We should include the ability to ingest other genetic tests for single genes, panels of genes, and exomes, although this needs significant work to make all the data reliable and comparable. To build these national arrangements, the data and service assets created through the investment in Genomics England and the 100,000 Genome Project must be used and developed together with the expertise that has been brought together and the experience they have acquired in the process. This will provide the best value for money for the NHS and will at the same time protect our UK assets of expertise, systems and data for patients.

2.6 100,000 Genomes Project

As discussed in more detail in chapter 13, the 100,000 Genomes Project is the first major programme worldwide to integrate data for both clinical and research use. The programme runs under a fully consented research "ethics consent", but is now returning results for individual patients. These results are then validated by clinical teams in the NHS Genomic Medicine Centres before returning reports to treating clinicians. The programme has already sequenced over 30,000 genomes, and it is now less than two years until the completion of the 100,000. In the process, a semi-automated pipeline has been established for receiving data about participants, their families and their DNA samples and then for receiving and interpreting their genomic sequence data. Genomics England has a very large scale comprehensive data environment and associated knowledge bases, and a substantial High Performance Compute environment. Data analysis for clinical care or research is carried out within the environment without individual data leaving the secure environment, the so-called 'reading library' rather than 'lending library' model. Data is de-identified to ensure privacy. The experience of the 100,000 Genomes Project has highlighted many of the challenges that genomic medicine faces, and shown the transformative power that a concerted, national effort has to overcome these.

2.7 The way forward

As genomic medicine continues to move from a research activity to mainstream clinical care, we need to establish clear guidelines for how clinicians and scientists can share all types of data and knowledge and to build databases and systems to make this work. The lead we have established in the UK for genomic research and patient care is dependent on our ability to find practical solutions for collaboration.

We should establish a single country-wide database and infrastructure for handling whole genomes and associated clinical data. The data and service assets already created through the investment in Genomics England and the 100,000 Genomes Project must be used and developed together with the expertise that has been brought together and the experience that has been acquired in the process. This will provide the best value for money for the NHS and will at the same time protect our UK assets of expertise, systems and data.

3. Bioinformatics and statistical analyses

Consent is always the basis for data sharing, either implicit or explicit. The clearest and safest route is to get as explicit a consent as possible. Ideally, as part of normal care, informed consent to data uses should be routinely sought. However, consent is a complex issue due to the wide range of things which could be consented to ranging from tissue sample handling and retention, through research by academics and commercial companies, to where data can be held. There are also consent-related questions about 'additional findings' which are findings that may emerge in the course of genome analysis relating to family relationships, potential drug reactions, recessive conditions being carried and risk factors for susceptibility to diseases. Currently, each organisation and project develops their own approach to these issues which makes the processing of gathered datasets subject to a complex set of rules. We need to establish a clear, nationally agreed consent process which finds the acceptable balance between being simple to understand and sufficiently comprehensive.

3.1 Bioinformatics in healthcare

The use of genomics as a healthcare diagnostic tool is becoming increasingly common due to the desperate need to understand the underlying causes of diseases and provide more cost-effective medicines. Genomic medicine is now made possible by the precipitous and continuing drop over the last decade in the costs of generating molecular measurements – most notably DNA sequencing but also transcriptomes, proteomes and metabolomes.

The disease areas that are likely to significantly benefit from the use of genomic diagnostics are those where identification of causative gene mutations is more straightforward such as in rare diseases, cancers and infectious diseases as well as the detection of chromosomal abnormalities in non-invasive prenatal testing. Indeed pilot projects in rare diseases such as The Deciphering Developmental Disorders (DDD) study, jointly funded by the Wellcome Trust and the UK Department of Health, utilised whole exome sequencing to diagnose 27% of 1,133 previously investigated yet undiagnosed children with developmental disorders.³ Most of the diagnostic variants identified in known genes were novel and not present in current databases of known disease variation.

When identifying causative gene mutations, the first step is to catalogue all the nucleotide differences or variation in a patient's genome compared to a reference genome in rare disease cases or between healthy versus tumour genomes in cancer. The more complex and naturally the more valuable next step is to understand the clinical significance of each variant, their inheritance patterns and the strength of their association to the disease or phenotype.⁴

Box 1 Understanding the clinical and functional significance of each variant

Understanding the clinical and functional significance of each variant requires complex bioinformatics analyses and the integration of numerous other data types, including clinical outcomes:

- gene structure information to determine whether the variant lies in the coding or non-coding portion of the genome;
- for coding variants, protein structure and functional data to determine the impact of the mutation on protein function;
- transcriptomics and proteomics data to determine cell and tissue expression profiles;
- mutation experimental data from human cell or model organisms and disease variation information to understand linked phenotypes;
- protein interaction network and biological pathway knowledge to learn more about function and relationship to other proteins;
- longitudinal phenotypic health records data to assess clinical and prognostic significance;
- data from clinical trial and pharmaceutical agents to know if there have been medicines developed against this disease that target this protein.

All of the above requires the availability of curated, structured reference information to be readily available.

3.2 Challenges of data integration

Much of the data required to determine clinical significance is deposited and curated in bespoke biological repositories such as those hosted by the European Bioinformatics Institute (EMBL-EBI) or the NIH National Center for Biotechnology Information (NCBI). However, data integration for genome-wide bioinformatics analyses and the conversion of data to knowledge needs continuous development of analytical pipelines and systems.

One important component required for data integration is the careful curation and mapping of data to controlled vocabularies or ontologies. For example, for genomic data integration with clinical information, data from primary care, hospitals, outcomes, registries and social care records should be first recorded using controlled clinical terminologies, such as SNOMED CT and the Human Phenotype Ontology (HPO). Ontologies in themselves are not ever complete and end-users such as clinicians will need to work with ontology developers to continuously improve the precision and accuracy of terminologies.⁵ The DDD study and now Genomics England also demonstrated that systematic recording of relevant clinical data, curation of a gene–phenotype knowledge base, and development of clinical decision support software was crucial for scalable prioritisation and review of possible diagnostic variants.⁶

3.3 Cumulative knowledge

As the volume of genomic data grows with associated clinical data, it is also useful to note that aggregation and reanalysis of such data will result in new and improved understanding of clinical value over time. For example, a novel variant discovered in a patient today may have little or no information associated with it. However as genomic data grows and this variant is analysed in conjunction with other similar variants, more statistically significant results can result in greater confidence of this variant being associated or not with disease.

The use of genomic data for research purposes can improve existing data and resources that provide reference datasets for clinical research as well as further our understanding of basic human biology. Secondary research can also result in the development of new tools and algorithms such as those used to model the genetic diversity and evolutionary patterns of individual cancers.⁷ Large-scale projects such as the International Cancer Genome Consortium, which aim to generate genomic, transcriptomic and epigenomic changes in 50 different tumour types and/or subtypes, have also shown the value of integrating genomics data from the same patient. These types of analyses will lead to novel targets and disease mechanisms and should in turn drive enhanced diagnostic and therapeutic yields for individual patient benefit.⁸⁻¹⁰

3.4 Data Analysis Environment

Bioinformatics analysis leading to clinical interpretation is an expensive part of the pipeline. For costs to go down, there needs to be simultaneous improvements in data sharing and use of common standards.¹¹ Currently much of the human genomic data generated so far is deposited into public databases for broad research reuse. This cannot scale nor is appropriate for the growing volume of genomic data from national health studies. Managed storage systems which follow national legislation and which allow such data to be accessible for research purposes are essential. The system established by the 100,000 Genomes Project for researcher access to their database of whole genome sequences is exemplary in its creation of a research community that will be connected and contribute directly to the NHS and add to the knowledge base of the genetic basis of disease.

In addition, federated standards such as those developed by the Global Alliance for Global Health (GA4GH)¹² should be adopted by national healthcare systems to drive efficiencies and drive down costs.¹³ The goal of the Alliance is to create data standards and strategies for storage and analysis of medically relevant genomic data, and to catalyse the creation of data sharing standards and methods to ensure worldwide interoperability of medical genomics data. GA4GH includes institutions like EMBL-EBI that play a key role in facilitating the transfer of knowledge and expertise in data management and analysis of big data projects.

New data sharing mechanisms are also needed to minimise the movement of large volumes of data and allow instead for analyses to take place at the point of private data stores. The cloud computing framework allows for data to be stored remotely, analysis scripts to be uploaded to the cloud and analysis performed remotely in virtual machines. This greatly reduces data transfer needs because only the script and analysis results are transferred to and from data that reside permanently in the cloud.¹⁴

4. Clinical interpretation of genomic data

The recent evolution in high throughput sequencing (HTS) has enabled us to generate the full sequence of a patient's exome (20,000 genes) or entire genome as an affordable clinical test, liberating us from the clinical guesswork of serial single gene testing. However, to address our clinical questions, interrogation of a variant set of commensurately larger scale is required: ~20,000 variant alleles for a particular patient's exome and ~4-5 million variants for a genome.¹⁵

4.1 Rare disease diagnosis

In the context of high throughput sequencing analysis for identification of the single mutation (or pair) causative of Mendelian rare disease, bioinformatics algorithms allow prioritisation of genes according to matching by phenotypic terms,¹⁶ complemented by filtering of variants based on family inheritance, variant rarity and variant impact. Whilst emerging clinical decision support software integrates these functionalities to facilitate more efficient exploration of prioritised variants, currently such tools are biologically rudimentary and are limited by both the size of reference variant datasets and the clinical validity of input algorithms. Whilst increasing automation and deepening understanding of genomic data will reduce the currently substantial manual component of genomic interpretation, detailed and expert multidisciplinary clinical review will remain essential to appropriately interrogate and interpret the molecular data for the specific clinical context and to define appropriate family and functional validations. Expansion across medical specialities of the currently limited pool of clinicians with subspecialty expertise in molecular genomic pathology will likely be required, as well as evolution of the skillsets of clinical scientists and expansion of clinical bioinformaticians.

Some variant types enable more ready prediction of their impact, such as those that cause premature truncation and loss of the protein. However, the vast majority of variants are non-truncating (missense, synonymous or in splice regions), or lie in the 98% of the genome outside of the genes. Our approaches are poorly evolved to disentangling the minority of such variants which are of clinical significance from the likely neutral majority.

In silico (computational) prediction tools have emerged in sizeable numbers but remain poorly predictive of clinical pathogenicity.¹⁷ Useful functional (laboratory) assays exist for a few genes but for most the technical reproducibility and clinical validity have not been established to sufficient standards for clinical application.¹⁸ Lack of robust 'truth sets' of pathogenic variants has been a limitation to development in both these areas: commercial and academically-maintained variant databases are improving but remain significantly contaminated by errors and conflicting classifications.¹⁹⁻²¹ Only through two related endeavours in data centralisation and federation will our investment in sequence generation yield commensurate returns in genomic diagnoses. Firstly, there are many cases we could readily solve if we better utilised our existing variant-level 'intelligence': we must centralise from across the UK molecular diagnostics laboratories our collective accrued evidence about specific genetic variants leveraged from three decades of genetic testing and functional enzyme/splicing assays undertaken upon valuable rare disease families.²² Secondly, it will only be possible to make genomic diagnoses in 'unsolved' patients with extremely rare and/or atypical diseases through systematic centralisation of fuller sets of patient-level genomic and phenotypic data: iterative cross-comparison of phenotypically similar cases is required to pull out the causative genes and variants to solve the more obscure cases.

4.2 Clinical molecular oncology

Whilst in Mendelian rare disease we are seeking the single causative mutation (or pair), clinical interpretation of a cancer genome should leverage the multitude of 'driver' genomic changes serially accrued in the tumour tissue which (i) define diagnostic subtype (ii) indicate tumour behaviour (prognostic category) (iii) inform on sensitivity to specific conventional chemotherapeutic agents or (iv) indicate aberration in a specific pathway amenable to exploitation using a 'targeted' drug.²³⁻²⁵

Available molecular oncology knowledge bases are typically rudimentary, often inaccurate and most catalogue only modest numbers of individual 'actionable' variants, non-systematically annotated from the published literature.^{26,27} Current molecular oncology management therefore typically involves manual one-by-one review of detected variants against such knowledge bases, followed by non-standardised clinical decision strategies leveraging only the one or two most prominent variants. For molecular tumour analyses to deliver improvement in outcomes and cost-efficient drug administration, clinical decision support tools should integrate in an unbiased fashion the totality of that patient's informative genomic data and reference these data against validated multi-marker tumour-specific data models.²⁸⁻³⁰ Development and validation of such models has begun, but is a long-term and iterative endeavour. To maintain the statistical tenets of evidence-based medicine in the context of integration of a multitude of low-frequency driver mutations requires very large clinical datasets (knowledge banks), well annotated for both molecular tumour features and longitudinal treatment and outcomes. To date such datasets have been of modest size, having been derived largely from research collections and clinical trials: generating knowledge banks of sufficient scale may only be achievable through systematic capture of accurate molecular and clinical data from the totality of cancer patients passing through routine clinical (NHS) care. This substantial commitment will require new funding models coupling the clinical care of the patients of today with knowledge generation to benefit the patients of tomorrow. Ultimately, (i) systematic and routine broad molecular characterisation of cancer patients using quality assured, standardised platforms (ii) ongoing improvement to the current national systems for capture of cancer diagnoses, outcomes and therapies (including the National Cancer Information Network, held by PHE) and (iii) systematic centralisation of molecular data alongside longitudinal clinical data will be required. Areas for priority focus of this model might include patients involved in late stage trials or post-marketing drug studies: systematic comprehensive molecular profiling of patients in this context as a condition for NICE adoption would contribute towards molecularly-stratified drug indications, enriching for those more likely to respond.

Whilst current clinical molecular oncology practice largely focuses on discrete variants (small mutations, amplifications or gene fusions), there is emerging data supporting therapeutic prediction based on pan-genomic features, such as total mutational load and genomic 'signatures' of base substitutions or copy number change.³¹⁻³⁴ Emerging molecular oncology decision-support systems will need to integrate these more complex pan-genomic features, as well as predictive biomarkers derived from transcriptomic, epigenomic analyses and other emerging molecular approaches for patients to get the most benefits.

As well as knowledge banks and decision-support tools for clinical interpretation which encompasses molecularly-stratified rationalisation of established oncological interventions, there is increasing focus on leveraging genomics to direct experimental repurposing of a broader range of therapeutic molecules.^{35,36} Large-scale computational biology databases have started to integrate banks of cellular and animal data with human molecular and clinical data, to evaluate potential candidate molecules, with renewed focus on opportunities in oncology for drugs licensed or in pre-clinical development for other clinical indications as well as molecules previously abandoned during development.^{37,38} Commercial organisations are already offering services for individualised "molecular-matching" of a patient's tumour genomic features against a broad drug set.³⁹ However, assiduous overview is required to evaluate robustly the clinical effectiveness of agents used in novel contexts: ad hoc administration off-licence outside of a clinical trial framework is unlikely to be informative beyond anecdote, as well as being highly challenging to fund in a rationale and equitable manner.

Establishing a robust infrastructure for systematic data centralisation will be critical to advancing clinical interpretation of genomic data for both rare diseases and cancer. Urgent attention is required to address the existing barriers leading to variant-level 'intelligence' and patient-level data remaining siloed in local laboratories, rather than centralised within the NHS for collective patient benefit, namely (i) legal and regulatory ambiguity regarding consent, de-identification and primary versus secondary usage of patient data (ii) lack of consistency in the formats used for the collection and storage of clinical and genomic data (iii) absence of appropriate centralised national data systems for deposition of variant-level and patient-level data (iv) lack of local resource, recognition, remuneration or mandate for data curation and deposition.

Box 2 Lifelong UK health record data for patient benefit and research in genomic medicine

National structured records

Primary care electronic health records available in nearly all of the UK's 65 million citizens (from three main providers: Vision, via CPRD, EMIS, and TPP).

Secondary care coded data on hospital admissions, procedures, Accident and Emergency attendances, 111 (telephone triage), prescribing and other datasets are curated for England at NHS Digital, Information Services Division Scotland, Patient Episode Database for Wales, and Health Information Branch Northern Ireland.

Public Health England (for example, national screening programme data; cancer registries; cancer treatments, for example, SACTS; infectious disease data).

National registries of disease (for example, for cancers, heart attack and stroke), national registries of procedures (for example, for renal transplants), or drug treatments (for example, "biologics" for treating rheumatoid arthritis).

Local, more detailed ('deeper') hospital data:

Electronic health records used for decision support in hospitals with high level of digital maturity (e.g. Leeds, Birmingham); 12 Global Digital Exemplar hospitals identified by the Wachter report.⁴⁷

EHRs for specific diseases or settings, e.g. from the NIHR Health Informatics Collaborative programme and Clinical Record Interactive Search. For example the critical care theme has built the information governance and informatics pipelines to share very rich (15,000 data points per individual per day) data across five large NHS trusts.

Imaging data

Research using NHS imaging for research at national or regional scale is at an early stage: with early efforts starting with natural language processing of imaging reports for phenotype extraction among 100,000 brain MRIs.

Wearables and mobiles

Puts actions in the hands of the user (patient), with suites of apps being built off "omics" technologies to support or inform nearly any decision relevant to health, from participating in a randomised trial, to advanced medication adherence solutions. For research only: already adding phenotypic information ("always on") in heart rate, ECG etc., and for areas not or poorly covered in health records, for example symptoms and quality of life.

Societal data

The patient outcomes of genomic medicine extend beyond the strictly medical. For example, the UK more intensively "phenotypes" and records educational attainment in the whole population compared to any other country in the world. Charting progress through school by linkage to the National Pupil Database could help to understand outcomes of early life disorders. Already diseases are being linked to Department of Work and Pensions data to evaluate return to work after serious illness – an outcome important for the patient, their family, and society.

5. Lifelong electronic health records for genomic medicine

5.1 Why are lifelong health records vital for genomic medicine?

An ideal health system in which to embed genomic medicine will have, as the NHS does, lifelong ('cradle to grave') structured and high quality electronic health records in order to better understand (and improve) the pathways to diagnosis (sometimes experienced as an odyssey) and subsequent health outcomes. Clinicians follow up their patients observing a sequence of disease events and processes relevant to health; when these are efficiently captured at scale in electronic health records then such 'phenomic medicine' will complement genomic medicine in two ways. Such data will aid the diagnostic process in better understanding prognosis and disease progression, then help to select appropriate treatments to alter prognosis.

5.2 Diagnosis and informing decisions now

If genomic information is embedded in an electronic health record (EHR) there are several potential near term patient benefits. Firstly, the health record may assist in the diagnosis, if structured high quality information is integrated across different points of healthcare contact. Records, if efficiently integrated, should support earlier diagnosis of genetic disorders to help understand and mitigate the diagnostic odyssey. In a research context, manual entry of Human Phenotype Ontology terms may in part be automated through tools such as CogStack⁴⁰ to provide elastic searches across diverse structured and unstructured (e.g. text) records. Second, when genomic information is embedded in the EHR it becomes possible to build in pro-active decision support on drug prescribing in which genetic testing is recommended or considered to avoid adverse effects or help with dosing. This pre-emptive pharmacogenomics approach has been demonstrated at Vanderbilt.⁴¹ Third, where health records can be linked across mother-father-child it becomes possible, where confidentiality allows, to inform approaches to reproductive counselling. Fourth, of wide patient interest, irrespective of any disease genetic or otherwise, is the immediate value of 'knowing about me'; for example people buying 23andMe value finding out about their ancestry, variants relevant to behaviours (e.g. the ability to metabolise coffee), and the ability to metabolise certain drugs. All this information may be linked to a personal health record.

5.3 Prognosis and informing decisions over time

The ability to follow patients up long term for a range of health outcomes through their health records allows a further range of potential benefits for care and research. For care, the clinical management of known or emerging actionable variants is informed by the availability of relevant clinical features in the patient's record. The patient journey of quality of care and treatment over time can be evaluated with longitudinal patient records. For research, long term follow up for a wide range of fatal and non-fatal outcomes is essential to understand the genetic architecture and mechanisms influencing the future course of disease. Large scale record research in 110 million patients has been used to demonstrate how Mendelian genetic disorders are associated with common complex disorders.⁴² Furthermore, disease onset and disease progression may not share the same genetic basis, for example in Crohn's disease none of the variants associated with disease onset are associated with progression to severe complications; and variants associated with progression are not associated with onset.⁴³ Drug repurposing and drug discovery efforts are facilitated by access to large scale disease diagnostic electronic health record collections linked to genomic information.⁴⁴⁻⁴⁶

5.4 Does the UK have a special position?

The richness of longitudinal phenotypic information on an individual which is available for care and research varies widely between countries. Because the UK has a single healthcare provider (the NHS), and because nearly every (>98%) citizen is registered with a GP, whose care is based on a structured true electronic health record, the UK has important opportunities for advancing genomic medicine. Furthermore, with a unique identifier (the NHS number in England and Wales, Community Health Index in Scotland and the Health and Care number in Northern Ireland) there is a rich array of record linkage opportunities: taken together the UK has the potential to add a lifelong, detailed understanding of phenomics to rival or surpass efforts in Scandinavia or in the different health systems within the US.

It is important to distinguish phenotypic data held in a true electronic health record in real time and used for decision support from that which is available ('cold') later after a record linkage or data sharing process. True electronic health records in hospitals which are 'genomic medicine' ready are uncommon.

Diverse sources of phenotypic information are illustrated in Box 2.

5.5 Integrating health record data

But there are major challenges in ‘converging’ the phenomic information from these health record sources with genomic information. Challenges include the governance framework to share data for care and research (see ethics chapter), and the extent to which ‘all’ health records, which are held by diverse national and regional data controllers can be brought together. Initiatives to improve the quality of the electronic health record, for example the Professional Standards Record Body are important; although existing evidence supports the validity of linked primary and secondary care records in a wide range of research uses.⁴⁸ By contrast, with the rapid developments in genome sequencing and the associated bioinformatic pipelines there has been much less attention on the methods and tools to unlock the value of health record data. Efforts to reproducibly define disease phenotypes across multiple sources of records are important for replicable research and standardised care and such efforts are underway.⁴⁹

Building on multiple investments in health informatics and bioinformatics, a new (2017) national institute Health Data Research UK (HDR-UK) funded by MRC-NIHR-Wellcome Trust⁵⁰ is being established. HDR-UK embraces data challenges and is well placed to set standards for the linking of the lifelong phenotypic data from different health record sources with genomic information to improve the care of individual patients and their families and for research. The UK should leverage its genomic-health record linked resources (including Genomics England and UK Biobank) to develop international standards for the collection, recording and storage of clinical and genomic information. One example of this is the definition of disease, currently recognised sub-phenotypes and the discovery of new phenotypes relevant to the development of new therapies. The UK is well placed to launch the Human Phenome project in which the rapidly increasing, lifelong data of health and disease from multiple diverse sources is marshalled in order to improve health.

5.6 Developing the skills and capacity

The data, methods and tools of genomic medicine do not, of themselves, deliver health benefit: it is the people who are trained in the multiple necessary disciplines to gather, analyse, interpret and communicate the results of the data that deliver the benefit. There is a need for new researchers trained at the interfaces of quantitative sciences, clinical medicine and biology and software engineering. The list of research disciplines which might usefully work together is long and extends well beyond genetics, bioinformatics and health informatics, ranging from computer science, mathematics, statistics, and epidemiology to ethics, health economics and other social sciences. The NHS capacity in Genomic Medicine is being expanded, for example through Health Education England. The NHS Digital Academy, announced in the Wachter report, seeks over the next three years to equip 300 Chief Information Officers, Chief Clinical Information Officers and other senior staff with the necessary leadership and informatics skills.

6. Suggestions for policy makers

- National approaches to embedding consent in clinical care as part of the social contract (see Chapter 16 of this report) should be broad and include in principle datasets held nationally (or example, NHS Digital, Public Health England, CPRD) and those unlikely to ever be held centrally (for example, 'deeper' hospital data). Legal and regulatory ambiguity regarding consent, de-identification and primary versus secondary usage of patient data requires clarification and the legislation that may be required when the EU General Data Protection Regulation comes into force in 2018 may be an opportunity
- A national genome informatics network, acting in conjunction with NHS Digital, the UK 100,000 Genomes Project, Health Data Research UK (50), UK BioBank and Scottish SHARE could act as a driver to coordinate health and biomedical informatics research. The challenge for partner organisations in this network will be to work in a seamless, integrated way with NHS Trusts, hospitals, research organisations and other national health initiatives to maximise the utility of genomics and electronic health data.
- The interface between basic and clinical research needs to be strengthened and explicitly funded. This should build upon the impressive work already performed by the National Institute for Health Research (NIHR) where we can continue to build collaborations and research studies and establish standards and guidance.
- Education and skills in the data sciences needs to be dramatically expanded. Programs should be established to ensure the long-term generation of proficient investigators who can undertake the multi-disciplinary nature of genomics and phenomics in clinical practice and research. The 700 person years of Masters level education in genomic medicine from HEE for Genomics England needs to be matched with a comparable effort in capacity developing in health informatics and its inter-relation with bioinformatics in understanding the phenotype through lifelong health records.

7. References

1. Caldicott F. Review of Data Security, Consent and Opt-Outs National Data Guardian. London; 2016.
2. <https://genomicsengland.github.io/GelReportModels/>.
3. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*. 2015;385(9975):1305–14.
4. Wright CF, Middleton A, Burton H, Cunningham F, Humphries SE, Hurst J, et al. Policy challenges of clinical genome sequencing. *BMJ*. 2013;347(nov22 3):f6845–f6845.
5. Bowdin S, Gilbert A, Bedoukian E, Carew C, Adam MP, Belmont J, et al. Recommendations for the integration of genomics into clinical practice. *Genet Med*. 2016;18(11):1075–84.
6. Fitzgerald TW, Gerety SS, Jones WD, van Kogelenberg M, King DA, McRae J, et al. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2014;519(7542):223–8.
7. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*. 2014;346(6206):251–6.
8. Tirode F, Surdez D, Ma X, Parker M, Le Deley MC, Bahrami A, et al. Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov*. 2014;4(11):1342–53.
9. Bender S, Gronych J, Warnatz H-J, Hutter B, Gröbner S, Ryzhova M, et al. Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. *Nat Med*. 2016;22(11):1314–20.
10. Kirby MK, Ramaker RC, Gertz J, Davis NS, Johnston BE, Oliver PG, et al. RNA sequencing of pancreatic adenocarcinoma tumors yields novel expression patterns associated with long-term survival and reveals a role for ANGPTL4. *Mol Oncol*. 2016;10(8):1169–82.
11. Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, et al. The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol*. 2016;17(1):53.
12. Global Alliance for Genomics and Health. [cited 2017 Mar 21]. Available from: <http://genomicsandhealth.org/>
13. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. *Nature*. 2015;526(7573):336–42.
14. A federated ecosystem for sharing genomic, clinical data. *Science* (80-). 2016;352(6291):1278–80.
15. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
16. Smedley D, Schubach M, Jacobsen JOB, Köhler S, Zemojtel T, Spielmann M, et al. A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. *Am J Hum Genet*. 2016;99(3):595–606.
17. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125–37.
18. Sobrido M-J. Are Functional Assays for Pathogenicity Assessment of Genetic Variants Overrated? *Hum Mutat*. 2017;38(1):5–5.
19. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862–8.
20. Stenson PD, Mort M, Ball E V., Shaw K, Phillips AD, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014;133(1):1–9.
21. Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth H V., Bevan AP, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res*. 2014;42(D1):D993–1000.
22. Chatzimichali EA, Brent S, Hutton B, Perrett D, Wright CF, Bevan AP, et al. Facilitating Collaboration in Rare Genetic Disorders Through Effective Matchmaking in DECIPHER. *Hum Mutat*. 2015;36(10):941–9.
23. McDermott U, Downing JR, Stratton MR, Downing JR, Stratton MR. Genomics and the continuum of cancer care. *N Engl J Med*. 2011;364(4):340–50.
24. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med*. 2014;20(6):682–8.
25. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell*. 2013;153(1):17–37.
26. Micheel CM, Lovly CM, Levy MA. My Cancer Genome. *Cancer Genet*. Elsevier; 2014;207(6):289.
27. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D777–83.

28. Gerstung M, Papaemmanuil E, Martincorena I, Bullinger L, Gaidzik V, Paschka P, Heuser M, Thol F, Bolli N, Ganly P, Ganser A, McDermott U, Döhner K, Schlenk R, Döhner H, Campbell P. Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat Genet.* 2017;49(3):332-340. doi: 10.1038/ng.3756. Epub 2017 Jan 16.
29. Yuan Y, Van Allen EM, Omberg L, Wagle N, Amin-Mansour A, Sokolov A, et al. Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat Biotechnol.* 2014;32(7):644–52.
30. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MG Della, Jädersten M, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun.* 2015;6:5901.
31. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med.* Massachusetts Medical Society; 2015;372(26):2509–20.
32. Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature.* 2016;538(7625):378–82.
33. Petljak M, Alexandrov LB. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis.* 2016;37(6):531–40.
34. Alexandrov LB, Nik-Zainal S, Siu HC, Leung SY, Stratton MR. A mutational signature in gastric cancer suggests therapeutic strategies. *Nat Commun.* 2015;6:8683.
35. Rubio-Perez C, Tamborero D, Schroeder MP, Antolín AA, Deu-Pons J, Perez-Llamas C, et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell.* 2015;27(3):382–96.
36. Workman P, Al-Lazikani B. Drugging cancer genomes. *Nat Rev Drug Discov.* 2013;12(12):889–90.
37. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2017;16(1):19–34.
38. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 2017;45(D1):D985–94.
39. Cellworks - About Cellworks [Internet]. [cited 2017 Mar 22]. Available from: <http://cellworksgroup.com/>
40. Jackson R, Agrawal A, Lui K, Folarin A, Wu H, Groza T, et al. CogStack - Experiences Of Deploying Integrated Information Retrieval And Extraction Services In A Large National Health Service Foundation Trust Hospital. *bioRxiv.* 2017;
41. Roden DM, Wilke RA, Kroemer HK, Stein CM. Pharmacogenomics: the genetics of variable drug responses. *Circulation.* 2011;123(15):1661–70.
42. Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabani H, et al. A nondegenerate code of deleterious variants in mendelian loci contributes to complex disease risk. *Cell.* 2013;155:70–80.
43. Lee JC, Biasci D, Roberts R, Gearry RB, Mansfield JC, Ahmad T, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn’s disease. *Nat Genet. Nature Research;* 2017;49(2):262–8.
44. Finan C, Gaulton A, Kruger F, Lumbers T, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *bioRxiv. Cold Spring Harbor Labs Journals;* 2016 Jul.
45. Dewey FE, Michael F. Murray, John D. Overton, Lukas Habegger, Joseph B. Leader, Samantha N. Fetterolf, Colm O’Dushlaine, Cristopher V. Van Hout, Jeffrey Staples, Claudia Gonzaga-Jauregui, Raghu Metpally, Sarah A. Pendergrass, Monica A. Giovanni, H. Lester Kirchner, DJC. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR Study. *Science (80-).* 2016;354(6319).
46. Abul-Husn NS, Manickam K, Jones LK, Wright EA, Hartzel DN, Gonzaga-Jauregui C, et al. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science.* 2016;354(6319).
47. Wachter RM. Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England. 2016.
48. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. *Lancet.* 2014;383(9932):1899–911.
49. Hemingway H, Feder G, Fitzpatrick N, Denaxas S, Shah A, Timmis A. Using nationwide “big data” from linked electronic health records to help improve outcomes in cardiovascular diseases: 33 studies using methods from epidemiology, informatics, economics and social science in the ClinicAI disease research using Linked Besp. *Program Grants Appl Res.* 2017;5(4):1–330.
50. <https://www.mrc.ac.uk/about/institutes-units-centres/uk-institute-for-health-and-biomedical-informatics-research/>

Chapter 14

Economics of sequencing

Chapter lead

Sir John Chisholm¹

Chapter authors

Sir John Chisholm¹ and Dr Sarah Wordsworth²

¹ Genomics England

² Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford

With contributions from Katarzyna Witkowska, Calum Boyd, Mark Bale and Tom Fowler (all of Genomics England)

1. Background

Economics is the study of how individuals and societies choose to allocate scarce resources among competing alternative uses, and how to distribute the products from these resources. In health care, assessing the economics of a 'disruptive' new technology such as genomic sequencing poses numerous challenges, perhaps more than those for new drugs in well-established clinical conditions. For instance, there is the issue of how wide to draw the boundary of the economic analysis. Do we merely assess the immediate and obvious costs and effects for the health care system in relation to the treatment of the patients involved? Or should we broaden our assessment to consider wider implications for the health care system in terms of transformational processes with spill-over benefits? Or should the assessment go even wider to a macro-economic level and consider inward investment and wealth generating economic activity from undertaking sequencing on a large scale?

In addition, there is the challenge of capturing useful data that can be used to provide a meaningful assessment of future costs when the disruptive technology is in its early introduction stage. Historical data can be misleading, yet future projections are inherently uncertain. Moreover, the disruptive nature of the technology can mean that the benefits (and costs) of its introduction are felt way beyond the narrow remit of the existing alternative diagnostics and treatments. Furthermore, there is the challenge of technologies, such as genomic sequencing, which produce a vast amount of information of which only a very small fraction is understood at the point of the technology's introduction. That small fraction will provide the basis for a diagnosis in a minority of patients, with the remaining information being fertile ground for research which could lead to new insights and increased diagnostics yield for future patients. This is a pattern that has been noted as being of particular relevance with whole genome sequencing (WGS).¹ It is important in assessing the economics of sequencing to disentangle these two activities. The diagnosis should be the largely automated process which provides a result based on the knowledge state at the time. The ongoing research aimed at increasing the knowledge base should perhaps count as a separate activity.

In Part A of this chapter, we start from the wider perspective that has driven worldwide investment in genomic technologies. We seek to understand where we are on its development curve, what it is able to deliver now, over what timescale might further developments be introduced, and what the consequences might be for the health system and the wider economy. In PART B, we explore what the current economic evidence base is for the use of genomic sequencing in routine clinical practice in the NHS. In particular whether this evidence is sufficient to make recommendations on the use of genomic sequencing across a range of clinical conditions including rare diseases, cancer, pathology, risk assessment and newborn screening.

2. Part A

Economic opportunities from genomic sequencing

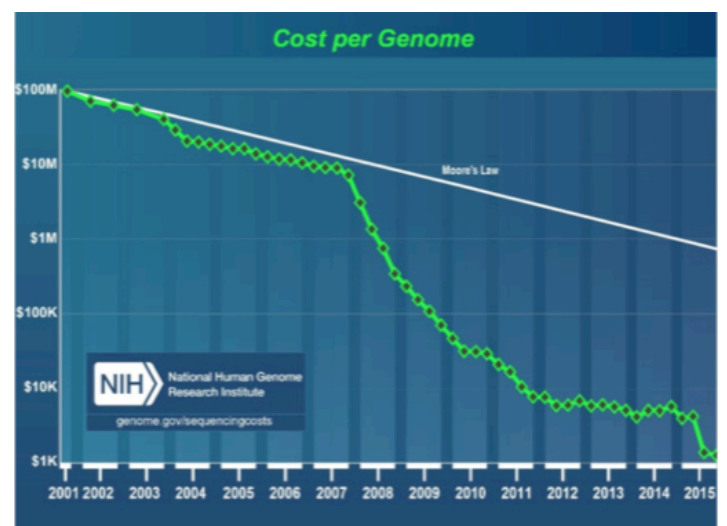
In the century between the rediscovery of Mendel's work on inheritable traits and the multi-national Human Genome Project (HGP), genetic science was driven by the inspiring vision that at least some debilitating human conditions might be ascribable to genetic phenomena and possibly cured. In the wake of the heroic endeavors of the HGP, the heritability of complex diseases susceptibility is now estimated to be 30-60%, depending on the disease.² But that optimism has to a degree become balanced by growing realization that three billion years of evolution had endowed the human genome with an enormous interrelated complex of codes and pathways. Very few conditions had singular genetic causes, and most common diseases were further complicated by the influence of environmental factors. It became clear that to unpick all that colossal complexity would require not just the sequencing of a single standard human genome (which had cost \$3 billion) but the sequencing of very large numbers of human genomes and the correlation of those with data on each individual's clinical conditions. Coupling WGS data with other multiomics large-scale analyses, such as epigenetics, transcriptomics, proteomics and others, is also likely to further our understanding of disease mechanisms and drive diagnostic and therapeutic yields for patient benefit

In the first decade of the 21st century sequencing was impossibly expensive. However, from 2008 with the introduction of Next Generation Sequencing (NGS) there began a significant decline in sequencing costs. The National Institutes for Health (NIH) in the US published a widely used chart in 2013 which illustrates that the reduction in cost had easily outpaced the well-known Moore's Law of microelectronics (1965), which suggests that scaling (doing activities in large numbers) was a key determinant to driving costs down.³

Despite widespread media commentary about the approaching \$1000 genome there was a levelling out after 2012 at just under \$10,000 per genome, with the next significant fall recorded by NIH in July 2015 to approximately \$1,400 per genome.⁵ When Genomics England examined the market in 2013 to prepare for the 100,000 Genomes Project, the commercial price available was around \$5,500 per genome.

Much of the sequencing capability available resided in individual research or clinical laboratories who purchased their equipment, mainly from the market leader Illumina, and ran their own sequencing pipelines buying reagents as required. However, Genomics England challenged suppliers to offer a turnkey service against a volume requirement of 100,000 whole genomes and this resulted in two suppliers developing prototype machines optimised to sequencing at scale. Both suppliers were able to offer prices at around \$1000 per genome including post processing and the production of the variant file. The price was achieved principally by process engineering rather than fundamental changes to technology or chemistry. This illustrates the crucial importance of scale at this juncture in genomic sequencing technology.⁶

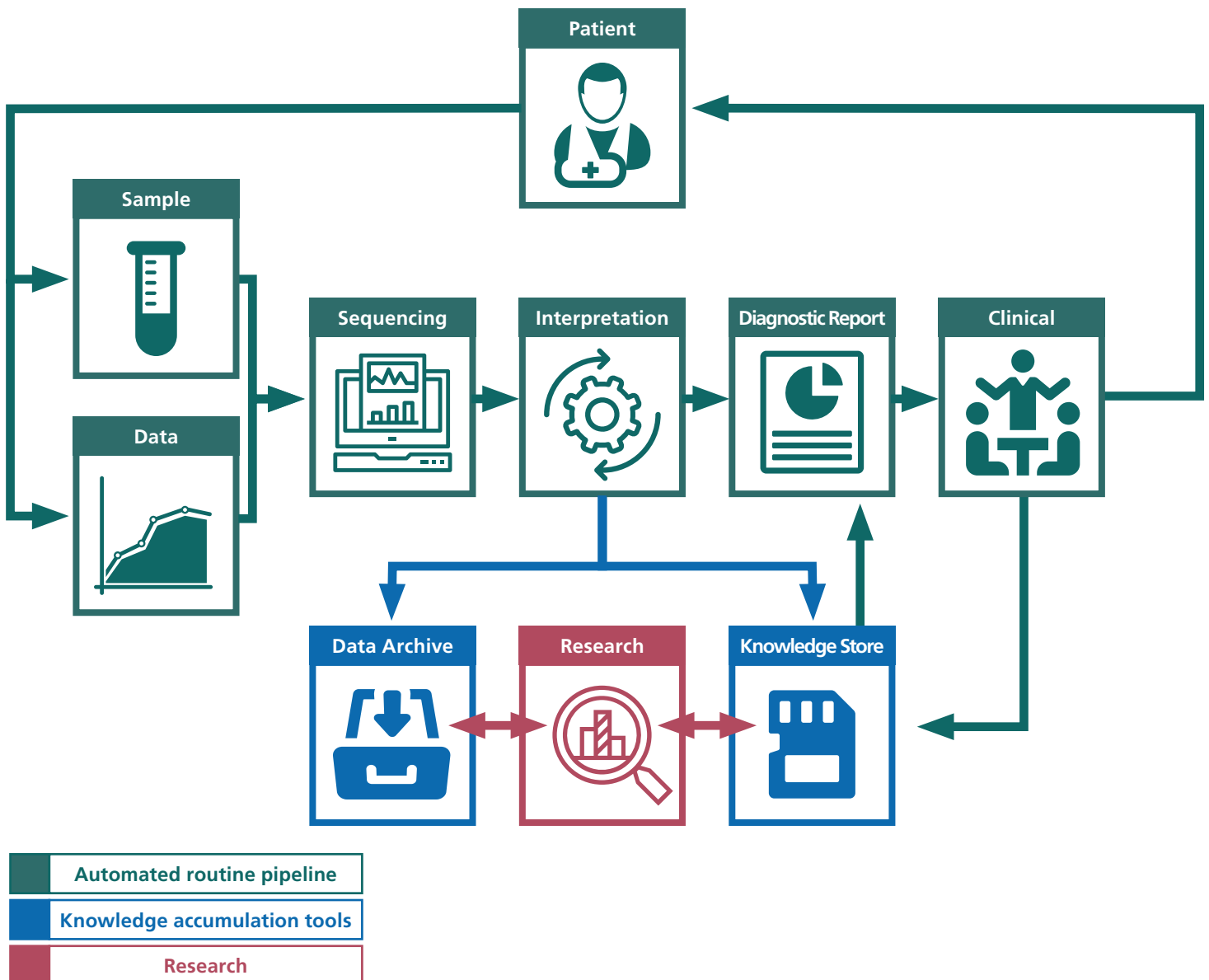
Figure 1 - Cost of sequencing per genome, National Human Genome Research Institute (NHGRI)⁴



Source National Human Genome Research Institute

Despite the substantial reduction in sequencing costs a key challenge has been the bioinformatics analysis, clinical interpretation and reporting still have significant technical and scientific challenges and their costs (unlike sequencing costs) are not experiencing a significant decline. This needs to be considered in two parts. First, the interpretation when running an automated pipeline to produce a routine diagnostic, and second the discovery research needed to gain insight into new conditions or greater insight into existing conditions. Taken together these represent substantial costs often not included when estimates are provided on the costs of sequencing (indeed they are not included in the NIH analysis in Figure 1 above). To clarify this point the simplified schematic below (Figure 2) shows the principal elements of a genomic pipeline.

Figure 2 - Resources required for an automated pipeline for whole genome sequencing

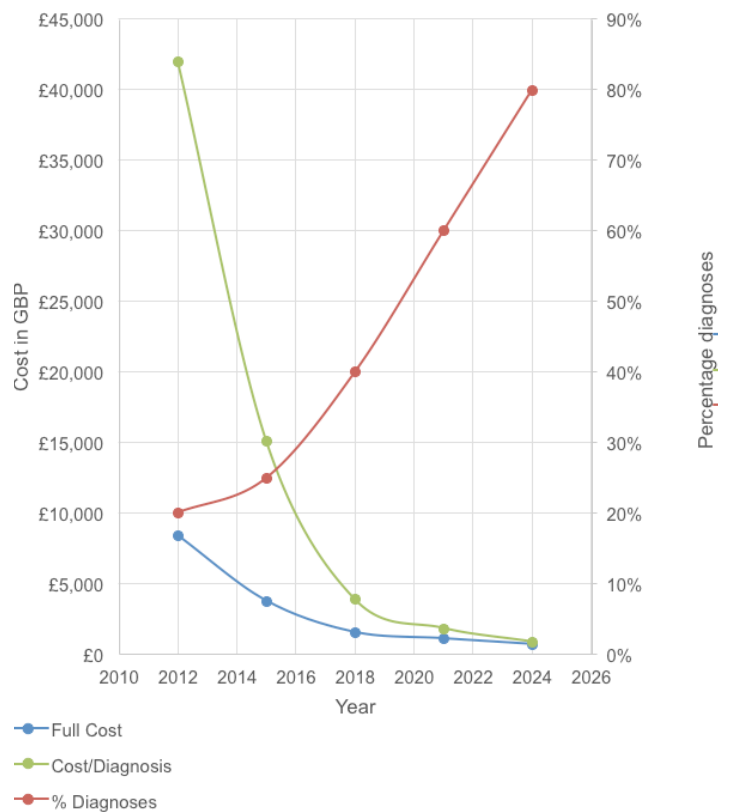


All the stages shown in the red boxes constitute a semi-automated routine pipeline starting from the point of gaining consent from the patient to perform the sequencing on their sample of blood/tissue/tumour etc, the sequencing itself and going through to the patients clinical team to feed back a diagnosis (or report a lack of a definitive diagnosis if one is not made). Those stages in the blue boxes are knowledge accumulation tools which would be expected to be part of a national infrastructure, as in the Australian Genomic Health Alliance’s comparable target to develop a national data repository tied to centres of diagnostic expertise.⁷ Indeed, the cost structure for computational and analysis resource in genomics points towards the efficiency of developing a single, large centre for data analysis and processing, to enable aggregate analysis.⁸ Such a resource would, through virtual access, be combined with a more widely distributed network of expertise, which can be imbedded in the healthcare closer to the patients. The green box is the research stage which takes place for patients who do not receive a diagnosis through the automated tools on the pipeline. Their data are assumed to be diagnosable (hopefully) at some point in the future through the improvement of tools not yet invented and the accumulation of knowledge not yet discovered. It is the role of the research to provide those insights which then accumulate on the pipeline for the benefit of future patients. All these different boxes require the use of health care resources (staff, equipment, consumables etc.), which need to be identified, measured and have costs attached to them.

There is considerable variation in the costs of genomic analysis reported both in academic papers and more widely in the media. The cost information we can obtain with perhaps the most accuracy currently are those actually borne by Genomics England in executing the 100,000 Genomes Project. Genomics England’s full costs for the project are expected to come in at £3,600 per patient after allowing for cancer sequencing of both the germ line (at 30X) and the tumour (at 75X), and rare disease sequencing (at 30X) of the patient and both parents (when available). This cost includes all the set-up and research and development costs of establishing the facility as well as the running costs of undertaking the sequencing and bioinformatics analysis and interpretation. It therefore includes all the costs in the red and blue boxes above, but not the green (the research element). Indeed, at the same time as the substantial reduction in sequencing costs, a key challenge is that the bioinformatics analysis, clinical interpretation and reporting still have significant technical and scientific challenges, and the cost structure for computational and analysis resources is diverging from the cost structure for sequencing itself.⁹

Projecting into the future and trying to make reasonable assumptions about the optimisation opportunities available for a routine service, the cost of sequencing based on 2016 technologies, could perhaps be halved. Figure 3 uses cost data from the 100,000 Genomes Project and tracks the cost per patient and the cost per diagnosis from 2012 to 2024 based on current experience with regard to rare diseases in the 100,000 Genomes Project after removing the cost of researching undiagnosed cases and the cost of long term knowledge curation, but including an estimate of the pipeline costs in NHS hospitals handling the patients.

Figure 3 - Cost per rare disease patient/per diagnosis projection



Note Genomics England, various communications, September 2016. Sequencing cost and cost estimates are taken from current project costs and professional projections from receipt of sample to production of Variant Call File (VCF). The projection of a reduction in cost per diagnosis, and increase in percentage of diagnoses depend on an assumptions that some WGS diagnoses can be achieved without a trio structure, as they are now in some cases, and that the diagnostic rate rises for any assay which becomes a first-line assay of choice, as it no longer handles only the hardest cases (Stavropoulos et al. 2016, and Sagoo GS, Mohammed S, Barton G, et al. Cost-effectiveness of Using Array-CGH for Diagnosing Learning Disability. Appl. Health Econ. Health Policy. 2015;13(4):421-32).

Source Genomics England, 2016

This chart relates to rare diseases for the Project because more rare diseases than cancer samples have been sequenced to date. For cancer the situation is more complex, primarily because Genomics England has found pathways need more re-engineering in order to collect viable tumour samples for sequencing, but also because the hurdle to provide an improved standard of care is higher in cancer. In rare diseases we are concerned with undiagnosed patients who experience a 'diagnostic odyssey' through NHS facilities at this time. Although confirmed diagnoses do not often lead to available therapies at this time, they do provide a relief from uncertainty for the patient family, and a saving of nugatory expenditure for the NHS.

For cancer the hurdle is higher – the objective is to produce a diagnosis which can be used to commission a therapy. For that to work the cancer pipeline has to be able to produce a result in a clinically meaningful timeframe (circa 3 to 4 weeks), and has to reliably identify mutations that are recognised in cancer pathways, which are linked to approved NHS therapies. For WGS to become part of the regular commissioned service it will have to demonstrate superior efficacy (and efficiency) to alternative sequencing regimes. While there are strong indications that all these conditions will be met as the technology develops, more progress will be required in the 100,000 Genomes Project to help provide the evidence.

The assumptions behind these scenarios in Figure 3 assume continuing efficiencies in hospital processes plus the steady effect of the continuing investment in digital capabilities. Sequencing will also continue to develop and the scenario assumes improvements in current technologies. Emerging technologies, such as solid state sequencing, will eventually make further inroads into these costs but the date by which these become viable for routine production of whole human genomes remains uncertain. Similarly, full digitisation of the NHS and the introduction of machine learning technologies into areas such as pathology will also offer significant efficiency and improvement opportunities. However, with safety and security being of prime importance in health care it is probable that the routine application of those technologies will only start to make a significant cost reduction impact towards the end of the timescale addressed in Figure 3, 2024.

2.1 Consequences for the wider health care system

The NHS is already preparing for transformation, for which genomic sequencing is a critical part. It has mandated the establishment of Genomic Medicine Centers which provide regional hubs to concentrate expertise of up to 70 hospital trusts and has launched a Personalised Medicine Strategy. There are plans to modernize the current network of NHS genetic and molecular pathology laboratories to take account of the developments in high throughput clinical sequencing. The plans will include requirements for routine data-sharing to improve clinical interpretation. The importance of data sharing and common standards is highlighted by Muir et al (2016) who highlight that storage and computation costs have not reduced as quickly as sequencing costs. They conclude that “if the sequence data generated by individual labs is not processed uniformly and sequence databases are not made easily accessible and searchable, then analysis of aggregated datasets will be challenging”.

Furthermore, the current Paperless NHS 2020 investment will provide much of the critical infrastructure to make possible the wide exploitation of genomic medicine and thus the facilitation of the spill over benefits that the technology can make available. It has been noted that the rationale for investment in sequencing depends on these spillover effects as well as direct clinical benefit.¹⁰ Paperless 2020 is a £4 billion investment in digital technologies which exactly matches the introduction period of genomic sequencing.¹¹ The challenge of gathering reliable clinical records has been the single biggest problem area for the 100,000 Genomes Project. But those difficulties have to be seen against the context of significant learning having accrued from previous digitization initiatives, the commitment of the participating hospitals to enhance their digital capabilities, the scale of the central commitment to the Paperless 2020 Project, and the success of the parallel programme to digitize primary care. The commitment to greater introduction of genomic medicine thus works very much with the grain of these initiatives because it is a technology which depends on system-wide digitization and provides the opportunity for patients to see that the benefits of genomic medicine could not be provided by any other means. It is therefore well placed to be a vanguard example of the success that can be achieved.

A second spillover effect of a world leading programme in genomic sequencing is the magnet such a programme could provide for clinical trials. As reported by PharmaTimes, the world market for clinical trials amounted to \$33 billion in 2015 and is likely to exceed \$65 billion by 2021.¹² With the globalization of this industry many major pharma companies, such as Pfizer and Novartis, have retrenched from developed economies, so it can be speculated that much of this capital growth will migrate to emerging markets. However, despite this trend, the UK may still provide a competitive niche by demonstrating its track record of conducting ‘some of the biggest and most smoothly-conducted trials’, as noted by The Economist.¹³ Demonstrating expertise in clinical trials design (currently the most costly and most risky stage of the drug development process) backed by well-developed accessible de-identified high quality clinical datasets, and particularly accurate targeting due to availability of high quality genomic data, together with responsive regulatory regimes has the potential to bring many more precisely targeted high performing trials into the UK’s NHS as evidenced by the Matrix trial based around the CRUK SMP2 infrastructure.

Bringing these trials to the UK has advantages on at least two levels. First, by their nature late stage clinical trials provide access by patients to the most advanced medicines and provide the NHS with accelerated access to new innovations. Second, by being a renowned global hub for such trials draws in many of the brightest and most able professionals in the field thus further improving patient results.

A third area of spill over relates to secondary (additional) findings and the opportunity to advance precision medicine. These can be thought to pull in two directions. Genomic data have the potential to reveal patient vulnerabilities beyond those the patient has sought treatment for. This is not new in medicine but it is likely to be on a much greater scale than previous medical advances. As highlighted by a recent report from the US & Canadian Clinical Genetics Think Tank, secondary findings are an important ‘additional’ benefit at a population level, driving knowledge of the genetic determinants of disease to support primary prevention and earlier diagnosis.¹⁴ At the same time the greater understanding of the particular patient’s condition offers the opportunity to avoid wastage of ineffective medicines and use of more targeted remedies. There are of course economic risks in these spill-over consequences from increasing demand to treat previously undetected syndromes, and also more expensive medicines due to the narrowing of the target patient community. But these are issues for the health system to tackle eventually under any scenario, and the progressive awareness of patients of their own risks with support from appropriate health professionals could be key to managing these issues as highlighted by the report from Nesta, People Powered Health.¹⁵

The focus here has been the introduction of WGS because of the rich knowledge base on this that is being accumulated in the UK as a consequence of the 100,000 Genomes Project. In practice WGS is merely the pinnacle of a pyramid of molecular diagnostics which will be increasingly deployed in the NHS over the next decade. More and more medicines are being approved with complementary single gene diagnostics and the use of panels and exome sequencing already has a place in many genetic laboratories for rare diseases and cancers. These alternative molecular diagnostics are significantly smaller compared to WGS in processing burden and therefore potentially less costly, or able to sequence a targeted area more deeply for the same cost. For instance, one study has shown Whole Exome Sequencing (WES) in children's rare disease to have improved the diagnostic rate five-fold compared with standard care, at the same time as reducing costs.¹⁶

They are less likely to compete with the capability of WGS to give a complete picture of the genome and to be more informative about structural variation in rare disease cases. It is possible that because WGS can be industrialised as a single common process it will become more cost effective to draw panels or exomes from a WGS processing factory than invest in many bespoke diagnostics for particular conditions. WGS is a multiplex technology, a single diagnostic test that can facilitate further analysis of a subset of the broader data set, or "virtual panels", without the need for commissioning several separate diagnostics. The analogy with mass manufactured products like smart phones could be a useful guide. Although smart phones are today manufactured with vastly more capability than any single user will require, it is more cost effective for the manufacturer to build a single product line which can cover all users rather than fragment the production capability into many smaller volume offerings.

There are also other complementary applications for molecular diagnostics emerging. For example, Circulating Tumour DNA (ctDNA) is very promising in relation to early cancer detection before tumours have reached the size when they present with physical symptoms, and also in following the course of cancers post treatment. Significant commercial investments have been announced in this field indicating the realistic promise it is seen to have. Similarly, other highly portable devices such as the desktop Minion device from Oxford Nanopore may well play a role in quickly triaging patients into relevant care paths and thus concentrate the high quality sequencing where it is most needed.

As with the introduction of any transformative technology it is extremely difficult to call the best investments upfront. That is why the 100,000 Genomes Project has integrated health economic analysis into its programme of work and presents an ideal opportunity to systematically collect high-quality cost and health outcome data within a large sequencing programme in an NHS setting. The Health Economics Genomics England Clinical Interpretation Partnership (GeCIP) will undertake analysis to include the examination of both appropriate diagnostic points in the care pathway and also the downstream consequences of testing (such as the use of targeted drug therapies). The Australian Cancer 2015 Programme also has integrated the collection of health economic data including quality of life data (using the EQ5D questionnaire) into its programme of research and is collecting information on the downstream costs and consequences of WGS.

2.2 Becoming a global hub: Opportunities for inward investment and economic benefit

When the UK Government announced the 100,000 Genomes Project it explicitly targeted benefits for the NHS in being a world centre for genomic technology and therefore a magnet for private sector investment. This is an industry where the UK starts with important assets. The most important is the NHS as a large and very capable integrated health care system serving a heterogeneous population. This is backed by a world class base of clinical and scientific expertise which is especially strong in the field of genomics. Size, capability and coherence of approach matter in genomics and the UK offers a very rare opportunity in the global market as a where all these factors come together in the same place. The UK is thus in a good position to expand on its 10% share of the £8 billion global genomics market, which will continue to produce increasing clinical utility and investment opportunity as the number of biomarker targeted drugs increases.¹⁷

Evidence that these advantages are recognised can be found in the readiness of the 13 company GENE consortium to come together to participate in the 100,000 Genomes Project and the numerous related investments that have been announced since the project started. Companies such as Congenica and Genomics plc have been formed and found venture backing, Seven Bridges has already built a team in the UK and Alexion and Berg Health plan to do similarly. The Wellcome Trust invested £27 million in the sequencing centre supporting the 100,000 Genomes Project and Illumina are investing £75 million on the back of the project including a significant drive to develop an automated genomic analysis pathway. In April 2016, AstraZeneca, along with its global biologics research and development arm, MedImmune, announced the launch of a global-reaching approach to interlink large genomics projects with basic research opportunities, clinical trial readiness and drug development programmes. With a London-based genomic centre and a bespoke database set up in Cambridge (Centre for Genomic Research) the project aims to tap into DNA data from up to two million study participants. Historically, national investments in genomics have proved very successful in stimulating private investment.¹⁸

Since the original announcement of the 100,000 Genomes Project (in December 2012) numerous other national (or near national) scale genomic sequencing programmes have been announced but the UK remains the only one which is significantly advanced with a quarter of the target already in the processing pipeline. Ultimately, the science of genomics will call for very large datasets so international collaboration is not only desirable but essential. By being first in the field the UK has the opportunity to place itself as an epicenter of global investment in this field that has the power to continue to deliver transformative therapies for the rest of this century.

Whilst all of the above points to considerable opportunities and excitement, we need to recognize that WGS should demonstrate clear economic benefit when compared directly to sequencing options such as whole exome sequencing or panels. If it is to be used in routine care in the NHS, it must demonstrate that it is effective in terms of health outcomes (and non-health outcomes which might be important to patients) and be undertaken at a cost which is considered as affordable for the NHS. Therefore, the next section of this chapter will explore the evidence base so far for genomic sequencing.

Box 1 - Summary of key points (PART A)

1. Genomic sequencing is a transformative technology which is becoming a practical possibility due to the dramatic fall in the cost of WGS sequencing.
2. Population scale sequencing initiatives are being developed to introduce genomic medicine into routine health care. These aim to capture benefits of improved patient care, provision of a data resource for discovery leading to new therapies and diagnostics.
3. Important spillovers from such projects include the stimulation of investment in related industries in the UK which will create further economic benefit.
4. The end-to-end cost of WGS continues to reduce but the balance of benefit for direct care, for the health system at large, and for the country as a whole remains work-in-progress. As with any major transformation the whole picture will not be clear until completion which is why the UK's 100,000 Genomes Project has committed to on-going health economics analysis under the auspices of the Health Economics GeCIP.

3. Part B

Collecting the evidence base

3.1 Health economic evaluation methodology

Economic evaluation compares different interventions/treatment in terms of their costs and consequences.¹⁹ There are several different types of economic evaluation (as shown in Box 2), but the most common type is cost-effectiveness analysis (CEA). In common, all types of economic evaluation approach costing in the same manner and it is the outcomes component which determines which type of economic evaluation a study is.

Box 2 - Economic evaluation methods

Cost-effectiveness analysis – (CEA) uses effects such as life years saved, cases/mutations detected as outcomes measures.

Cost-utility analysis – (CUA, a special form of CEA) assesses both survival and quality of life together, using quality adjusted life-years (QALYs). With CEA and CUA results are expressed as ratios of incremental costs to incremental outcomes. QALY's are the main form of evaluation for most international health technology assessment agencies, including NICE in the UK, and quality of life is measured using the standardised EQ-5D™ questionnaire. The EQ-5D is a generic health status questionnaire (i.e. not disease specific) and consists of a descriptive system and an EQ (EuroQol) visual analogue scale (VAS). Five dimensions are included in the descriptive system: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The latest version includes five levels in each dimension (EQ-5D 5L); from which respondents select the level that most closely matches their health state: no problems, slight problems, moderate problems, severe problems and extreme problems

Cost-benefit analysis – (CBA) has been proposed as possibly more helpful for some genomic tests where there might not be an obvious survival benefit. In CBA health outcomes and non-health outcomes, such as the value of information from a diagnosis (as well as costs) are valued in monetary terms, e.g. using willingness-to-pay questionnaires and discrete choice experiments. With CBA results are commonly expressed as either a ratio of costs to benefits, or a sum representing the net benefit of one intervention compared to another (Buchanan et al, 2015).

Cost-consequence analysis – Costs and consequences are reported separately to each other for comparison.

Decision making criteria – cost-effectiveness analysis is the incremental cost-effectiveness ratio (ICER), where the healthcare budget = quantify QALYs 'acceptable' £ per QALY gained, Cost-benefit analysis is net benefit (£benefit > £cost), cost-consequence analysis- the decision maker examines the disaggregated data and makes a decision on the relative merits of costs and effects.

3.2 Measuring outcomes

Disease specific outcome measures are occasionally informative in genomics but in general they do not capture all relevant dimensions of outcomes, in particular non-health outcomes (e.g. process-related outcomes such as improvements in waiting time). This is especially tricky for genomic interventions such as sequencing which sometimes only provide some diagnostic information, which can reduce anxiety and help patients to make future plans, but does not increase survival or quality of life necessarily. Generic preference-based measures such as the EQ-5D improve comparability by collecting data across a broad range of health related quality of life domains. These methods can work well in many scenarios, for instance using QALYs in cancer care can be informative, as chemotherapy can extend survival but quality of life can be reduced due to treatment side effects. However, the QALY paradigm has been questioned by some health economists,^{20,21} especially in the context of rare diseases, where there might not be a treatment to follow a diagnosis achieved by using genomic sequencing (e.g. microarray testing in learning disability provides a diagnosis but does not lead to a treatment). Also, QALYs do not capture family spillover effects.²⁰ In a decision making context, as some genomic interventions do not improve health or extend life, preference-based measures like QALYs may not pick up differences in outcomes, resulting in high incremental cost-effectiveness ratios ICERs which suggest that these tests are poor value.

Other economic evaluation approaches, such as cost-benefit analysis (CBA) have been proposed as possibly more helpful for some genomic tests where there might not be an obvious survival benefit (or reduction in pain/increase in mobility etc.). In CBA health outcomes and non-health outcomes, such as the value of information from a diagnosis (as well as costs) are valued in monetary terms, e.g. using willingness-to-pay questionnaires and discrete choice experiments. With CBA results are commonly expressed as either a ratio of costs to benefits, or a sum representing the net benefit of one intervention compared to another.²⁰

3.3 Evidence on the economics

To create a picture of the current economic evidence on sequencing for this chapter, we reviewed the health economic evaluation literature on WES and WGS. We were primarily interested in papers which reported information on the costs and outcomes of these sequencing approaches and any other sequencing (and non-sequencing approaches) such as panel tests which they were compared against. We found that the costs of WES ranged from £372 per patient to £3257. For WGS the costs ranged between £6,741 and £16,180 per patient and £40 to £481 for pathogen WGS. The outcomes assessed in the papers were largely diagnostic yield, rather than survival (life years gained) or QALYs. Most of the papers concluded that WES and WGS were superior in economic terms to other testing methods. However, nearly all the papers were reported very small patient samples, which is in total contrast to the data which the 100,000 Genomes Project will provide. Also, with a few exceptions, the papers tended to be of fairly low quality and some were not even full-economic evaluations (some had cost data but little data on effects and vice versa). A further important limitation in many previous and current economic evaluations in sequencing is that long term data on costs and effects are not available. The 100,000 Genomes Project has linked Health Episode Statistics data and CRPD data, so it will be possible to examine patient use of NHS resources before and after WGS.

From this review we concluded that published cost data in the literature probably lags current reality owing to the rapidly evolving technology. However, having taken some account of that effect, it provides a useful starting point to explore some of the potential economic considerations for sequencing in cancer, rare disease and pathogens, further discussed in detail in subsequent sections.

Cancer

In the UK over 300,000 cases of cancer are diagnosed annually. For cancer treatment, sequencing could help reduce or avoid treatment adverse events and reduce time delays in treatment selection. A well-known sequencing treatment example is the identification of the HER2/neu gene and the development of trastuzumab (Herceptin®) drug therapy. This has increased the cost of breast cancer treatment, partly as a result of the additional cost of HER2/neu expression testing, but mainly because of the cost of the associated drug (trastuzumab) for the 25% to 30% of women who test HER2/neu –positive. The total cost of this genomic advance, with substantial clinical benefit, has been estimated at more than \$750 million per year in the United States.²² Conversely, a cost-saving application of sequencing can be found in colorectal cancer, as the cancer is closely related to the epidermal growth factor receptor (EGFR) pathway. KRAS forms a vital part of the EGFR mediated pathway and mutations in this gene have now been established as a mechanism for the development of resistance to anti-EGFR antibodies. KRAS mutation testing in metastatic colorectal cancer is now routinely used to identify patients unlikely to benefit from treatment with expensive anti-EGFR monoclonal antibodies. Using sequencing to reduce the use of interventions in patients who will gain little or no benefit could have important economic implications, especially if a treatment is used frequently or is very expensive.

In cancer diagnostics, sequencing can help identify disease causing mutations. A recently completed UK study at Oxford University (funded by Innovate UK) showed that sequencing using a 46 gene panel more accurately identified disease causing mutations than sequential sequencing single gene testing using Sanger sequencing and did so at a lower cost per patient. To then move to using WGS instead of cancer panels will require overcoming some technical challenges, such as determining the amount of DNA required for sequencing and how best to obtain the DNA. This suggests that in the short-term, cancer panels might still be used widely, although other ongoing work at Oxford University (funded by the Health Innovation Challenge Fund) is assessing the cost-effectiveness of WGS in cancer care (and rare diseases) and will provide a useful comparison to the cancer panel cost-effectiveness study.

Rare diseases

In the Chapter 6 of this report, the sheer scale of rare diseases in the UK is highlighted. Research suggests that one in 17 people may suffer from a rare disease at some point in their lifetime. In the UK, this means that more than 3 million people may have a rare disease. At least 80% of rare diseases have an identified genetic origin and 50% of new cases are in children. Many cases of rare disease are undiagnosed, so these figures are likely to be an underestimate. For some rare diseases such as inherited heart disease, if effective treatments are available following disease diagnosis, sequencing could provide clinical and economic benefits as shown in the cost-effectiveness analysis presented (see Case study 1).

Case study 1 - Inherited heart disease

Hypertrophic cardiomyopathy (HCM) is the most common monogenic cardiac disorder and the most frequent cause of sudden cardiac death in young people and competitive athletes. People with HCM have enlarged hearts and prevalence amongst adults is around 0.2% (1:500). HCM is caused by mutations in over ten genes and the child of an affected parent has a 50% chance of inheriting the disease-causing allele. Most people with HCM are asymptomatic and sudden cardiac death can be the first sign of disease. Traditionally, those at risk of sudden cardiac death from HCM had clinical tests (ECHO and ECG) and family history taken. Treatment for those thought to have a disease causing mutation include life-style changes, drug therapy and ICD implants (Maron, 2003).

Genomic sequencing for HCM has been shown to be more effective and cost-effective than clinical testing to diagnose individuals with HCM and at risk of sudden cardiac death. A UK study using an economic model showed that the incremental cost per life year saved was £13,372 which is highly cost-effective in a NICE type decision making context (where under £20,000 is considered as cost-effective) for cascade genomic sequencing compared with cascade clinical approach. The upfront costs for sequencing were slightly higher than clinical testing, but this was largely because sequencing is more effective and identifies more individuals at risk.

Comments

Sequencing to diagnose and manage HCM was shown to be a cost-effective approach to the primary prevention of sudden cardiac death. This economic evidence was key to the testing being adopted by the NHS. The evidence on cost-effectiveness was considered as pivotal in generating new European clinical guidelines making genomic testing the first line approach to management of families (European Society of Cardiology, 2014).

Source: Wordsworth et al, 2010.

3.4 Solving the diagnostic odyssey

For other rare diseases a diagnosis may not lead to immediate treatment if one has yet to be developed, so there might not be any health improvement which could be measured using survival (as in the HCM example above where life years gained were used) or QALYs. In this case, other approaches to measuring the economic benefit could be used. For example an analysis could be undertaken of the diagnostic odyssey of the medical journey travelled by patients with a rare disease (and their families). Symptom realisation to a final diagnosis may involve many referrals to several specialists and numerous, sometimes invasive tests.²³ This odyssey can span many years and have serious consequences for patient welfare and waste NHS resources, especially if this journey does not provide a diagnosis. A 2015 survey found that more than a third of rare disease patients in the UK had received three or more incorrect diagnoses, and studies have identified an associated range of health-related suffering and misuse of healthcare resources.²⁴

In this context, if QALY information is not easily available, sequencing could be assessed using simpler cost-effectiveness measures such as cases detected (diagnoses made) where the costs and consequences of 'solving' the diagnostic odyssey using sequencing are measured. For example, a US economic analysis examined the cost and diagnostic yield of WES compared to traditional diagnostic trajectory in children with intellectual disability and concluded that WES can be a cost-effective option to diagnose children with a range of genetic conditions, if the diagnostic yield is higher than existing tests, especially if the costs of those other tests are higher than a one-off test using sequencing.²⁵ Evidence from clinical paediatrics in a Canadian context has shown that WGS can also be an effective primary test which increased the diagnostic rate four-fold compared to Chromosome Microarray Analysis (CMA), also potentially reducing the time to diagnosis.¹ Alternatively, cost-benefit analysis could be used instead of cost-effectiveness analysis as mentioned previously using willingness to pay or discrete choice experiments could help to measure the utility gained from finally having a diagnosis.

Pathogens

Genomic sequencing can provide high resolution information to distinguish pathogen strains that differ by as little as one SNP (Single Nucleotide Polymorphism), to help inform disease diagnosis, and predict which antibiotic would work best (i.e. to which antibiotic the bacteria is less likely to be resistant). The case study described below on WGS in tuberculosis (TB) is atypical in a sense because when compared against standard laboratory testing, using the Illumina Miseq was both more effective and less costly than standard testing, as most new interventions (especially new drugs) are often more expensive than existing options, but often provide more benefit.

Case study 2 - Whole Genome Sequencing for tuberculosis

Slow routine laboratory diagnostics for pathogens such as *Mycobacterium tuberculosis* complex (MTBC) risk delayed treatment and poor patient outcomes. In a UK based prospective study, WGS was compared against routine MTBC diagnostic workflows across eight laboratories in Europe and North America. Diagnostic accuracy, processing times and cost were compared for the two diagnostic methods.

Compared with routine results, WGS predicted species with 93% accuracy and drug susceptibility also with 93%. Full WGS diagnostics could be generated in 9 days, compared to routine diagnostic workflows taking 31 days. This came at a cost of £481 per culture-positive specimen, compared to routine diagnosis cost of £518 per patient. This equates to WGS-based diagnosis being 7% cheaper annually than are present diagnostic workflows for this infectious disease.

Comments

WGS for diagnosing and helping treat TB is faster, more accurate and cheaper than current routine diagnostic methods.

Source: Pankhurst et al, 2015.

Newborn screening and neonatal intensive care units (NICUs)

Newborn screening for metabolic inherited disorders such as phenylketonuria and MCADD, a rare genetic condition manifested by reduced conversion of fat into energy (medium-chain acyl-CoA dehydrogenase deficiency) can improve health outcomes and is cost-effective, because screening and disease management costs (simple dietary measures) are less than the costs of treating children if they become ill. Many countries include these and other disorders in their newborn screening programmes, using biochemical tests first and then genomic testing. A study in Sheffield is currently assessing whether next generation sequencing of the newborn could be used to diagnose several neonatal conditions when applied as first line screening. Another area where sequencing could provide economic (and health benefits) is in helping provide a diagnosis for babies in neonatal intensive care units (NICUs), although more health economic assessment needs to be done in this area.

3.5 Primary Disease Prevention

A recent report by Nesta (2016)¹⁵ explored the hypothesis that digitally enhanced diagnostics coupled with insights from behavioural psychology could have significant impacts on both the cost of healthcare and the wellness of the population. Sequencing could possibly increase patient willingness to undertake primary disease preventative measures, including behavioural changes. Type 2 Diabetes is an example of a genetically complex chronic disease, in which knowledge of genetic risk factors could alter individual behaviour and help with disease prevention, thus reducing healthcare costs and driving associated economic benefit. Over a hundred common DNA variants are associated with increased risk for Type 2 Diabetes. Recommended shifts in behaviour, such as diet change and exercise are difficult to implement on a population-wide scale. The relatively scarce research into the clinical utility of genomic testing for diabetes has produced mixed results. Some studies have found evidence that providing results of genomic sequencing for other chronic diseases increases patients' preventive behaviour, which could provide substantial economic benefit, although more research is required in this area.

Box 3 - Summary of key points (PART B)

- The care pathways involved will be novel in most parts of the NHS if WGS is introduced into routine care, therefore the introduction will need to be informed by informed economic evaluations to understand where the relative costs and effects will reside.
- The QALY approach may have limitations for WGS, especially for rare diseases, where changes in survival might not be a key outcome. Therefore, solving the diagnostic odyssey is likely to be an important outcome measure for economic assessment of sequencing in rare diseases. The use of other economic evaluation approaches such as cost-benefit analysis might also be informative.

4. Conclusion

Genomic sequencing shows significant potential to increase the diagnostic yield of diseases and provide valuable information on treatment options. However, it needs to provide value for money for the NHS relative to other uses of funds. In terms of effectiveness, there is some evidence that diagnoses are increasing through the use of next generation sequencing panels, WES and WGS. In terms of costs, sequencing costs are reducing, although we need to create an infrastructure to support sequencing on a larger scale to help push sequencing costs down further and efforts need to be directed at reducing the costs of bioinformatic analysis and validation of results. WGS should ideally replace rather than just add to existing tests, therefore serious discussions need to happen concerning dis-investing in sequencing and non-sequencing tests if WGS is shown to be more cost-effective than these existing tests. The optimum mix of these diagnostic methods will depend on the speed of technology developments both in sequencing hardware and in bioinformatics software. Initiatives such as the Australian Cancer 2015 Project and the 100,000 Genomes Project with integrated health economics data collection and analysis provide a massive step forward in providing clear evidence on the economics of sequencing, especially on the relative costs and effects of alternative sequencing options and their likely impact upon patient health and well-being.

5. Suggestions for policy makers

- Capitalise on the infrastructure and knowledge established by the 100,000 Genomes Project to develop a robust commissioned service that will succeed it.
- Centrally commissioned WGS could be explored as a transformative diagnostic through the recommended Accelerated Access Pathway as outlined in the Accelerated Access Review, final report, October 2016 including exploration of commercial arrangements by a new strategic commercial unit.
- Ensure that maximum value from all molecular diagnostics is obtained in terms of healthcare and spill over benefits by linking genomic, clinical and health care resource use data together on a curated national repository.
- Use the scale of the NHS volume to create efficiencies by increasing clinical utility and driving down costs.
- Continue with an active programme of Clinical Interpretation Partnerships for research to stimulate discovery, and continue with the outreach to industry to stimulate a vibrant concentration of genomics-active companies in the UK.
- Increase the level of health economic assessment of genomic interventions to assist in the optimum allocation of NHS resources.

6. Commentary from Chapter lead

When the 100,000 Genomes Project was announced in December 2012, the notion of a project of this scale and technological challenge being conducted within a national health service was novel and hugely ambitious. By 2016 numerous countries as large as the US and as small as Estonia have announced their ambition to launch similar projects. The UK has a lead which can be used to help shape the future of genomics medicine globally. Important though the dataset from the 100,000 Genomes Project may be, the greater contribution is the demonstration of how a technology which requires data and protocol coherence can be implemented across what is nearly always a heterogeneous health system, and develop into a future candidate for centralised commissioning into routine health care. By being an active and generous participant in the global genomic community the UK, as a leader, can reap not only health benefits for its patients but also spill over benefits in research and in the economy.

7. References

1. Stavropoulos et al., 'Whole genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine' in *Genomic Medicine*, 2016
2. Price et al, 'Progress and promise in understanding the genetic basis of common diseases', *Proc. R. Soc. B*, 282: 20151684, DOI: 10.1098/rspb.2015.1684
3. Wetterstrand, 'DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Programme (GSP) retrieved from [<https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/>] on 18.10.16
4. *ibid.*
5. *ibid.*
6. Monitor Deloitte, *Genomics in the UK: An industry study for the Office for Life Sciences*, 2015, p.7
7. Australian Genomics Health Alliance, National Data Repository, retrieved from [<https://www.australiangenomics.org.au/our-research/national-data-repository/>] on 18.10.16
8. Bowdin et al, 'Recommendations for the integration of genomics into clinical practice', *Genetic Medicine*, 18(11): pp.1075-1084, 2016
9. Muir P, Li S, Lou S et al. The real cost of sequencing: scaling computation to keep pace with data generation. *Genome biology*. 2016 Mar 23;17(1):53.
10. Dubow & Marjanovic, *Population scale sequencing and the future of genomic medicine*, RAND Europe, 2016
11. NHS England, *Annual Report and Accounts 2015-16*, 2016, p.37
12. Mansell, 'Over 50% growth to 2015 seen in global clinical trials market', in *PharmaTimes*, 2011, retrieved from [http://www.pharmatimes.com/news/over_50_growth_to_2015_seen_in_global_clinical_trials_market_980753] on 18.10.16
13. 'Testing, Testing' in *The Economist*, July 2014, retrieved from [<http://www.economist.com/news/britain/21608775-nation-losing-its-grip-valuable-industry-testing-testing>] on 18.10.16
14. Bowdin et al, 2015
15. Horne et al, *People Powered Health: Health for People, By People and with People*, Nesta, April 2013
16. Stark et al, 'A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders', *Genetics in Medicine*, March 2016
17. Monitor Deloitte, 2014, p9-11
18. *ibid.*
19. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes* (Oxford Medical Publications) Paperback – 24 Sep 2015.
20. Buchanan J, Wordsworth S (2015) Welfarism versus extra-welfarism: can the choice of economic evaluation approach impact on the adoption decisions recommended by economic evaluation studies? *Pharmacoeconomics* 33: 571-579.
21. Payne K, McAllister M, Davies LM (2013) Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Econ* 22: 258-271.
22. Hillner et al, 2007
23. Black et al, 2015, All Party Parliamentary group on rare, genetic and undiagnosed conditions, 2016
24. Angelis et al, 'Socio-economic burden of rare diseases: A systematic review of cost of illness evidence' in *Health Policy*, 119(7), 2015, pp. 964-979; Genetic Alliance UK, *The Hidden Costs of Rare Disease: A Feasibility Study*, July 2016
25. Monroe et al, 2016

Chapter 15

Genomic information and insurance

Chapter lead

Dr Mark Bale¹

Chapter authors

Dr Mark Bale¹, Prof Dame Sally C. Davies²

1 Genomics England and Department of Health

2 Department of Health

1. Summary

- Genomic medicine has always highlighted the need to address public concerns about genetic discrimination or the emergence of a 'genetic underclass'. This is particularly problematic for countries without a socially funded healthcare system or where employers fund healthcare costs. In the UK the debate has mostly involved concerns about access to life or protection insurance, often related to the large mortgages that many families have.
- Reflecting on the debate in the UK suggests that our voluntary arrangements have reduced the polarised debate about the concerns of patients, clinicians and insurance professional. The principles described below apply to a small number of late-onset single gene conditions. The 100,000 Genomes Project aims to transform NHS delivery to provide genetic testing for a wide range of NHS patients (and their families). As well as the initial reason for being invited to join the project, secondary findings that are severe and actionable – such as hereditary cancer risks – may (with consent) be fed back to patients and their healthy family members.
- Beyond the project, estimates are that by 2020 there will be 5 million genomes sequenced worldwide, primarily for research but with an increasing uptake of genomic sequencing and testing by healthcare providers or private consumers. Public concerns about the use of genomic data by insurers or employers may limit their willingness to consent to join research studies or to consent to feedback of actionable findings. This could limit the potential for the NHS to exploit the benefits of genomic science both to improve patient outcomes and to improve efficiency.

2. History of the genetic and insurance debate in the UK

The debate around genetic discrimination was prompted by the international Human Genome Project in the late 1990s which raised concerns about the discriminatory use of genetic information. There was an increasingly availability of clinical genetic testing for single gene tests for a highly penetrant, late onset, single gene condition – e.g. Huntington’s Disease (HD) or hereditary breast and ovarian cancer caused by BRCA1/2.

In the UK, the debate has generally focussed on an inability to purchase affordable life insurance. The existence of our NHS means that unlike many developed economies there is no concern about medical insurance. In contrast, other countries, notably the USA, have had detailed debates about healthcare insurance and the strong link to employment status. The protections enshrined in US legislation such as the 1996 Health Insurance Portability and Accountability Act (HIPPA), and the 2008 Genetic Information Nondiscrimination Act (GINA) are not directly relevant to the UK because they primarily relate to specific aspects of the US healthcare insurance system.

In the UK, large mortgages are often covered by life, critical illness and/or income protection insurance policies. At first the insurance industry was slow to engage with the public concerns, treating genetic data like other health data. Under the principle of *uberrima fides* (“utmost good faith”), both parties to an insurance contract should have the same information. The initial response to concerns raised by the Nuffield Council on Bioethics in a 1993 report on Genetic Screening, and repeated in a detailed report from the House of Common’s Science and Technology Committee in 1995 and by the Human Genetics Advisory Commission in 1998 was the establishment of a Government expert committee. (For a summary of the discussions in the 1990s, see Thomas 2017¹)

The Genetics and Insurance Committee (GAIC) was established by Government in 1999 to assess the genetic tests results that were already being used by insurers (who had initiated a temporary moratorium). However, a decision by GAIC in 2001 to approve the use of (HD) genetic test results for life insurance caused a major public and Parliamentary reaction. In response to a highly critical House of Common’s Science and Technology committee report², the Association of British Insurers (ABI) agreed to a more comprehensive voluntary moratorium on the disclosure of genetic test results unless the test had been approved by an independent body. It also agreed that only customers seeking large insurance policies (over £500K for life insurance, with comparable values for other insurance types) would need to disclose approved genetic test results.

The moratorium left just one test (HD) that needs to be disclosed for large policies but it was expected that other tests, such as BRCA1/2 would in due course be submitted for GAIC scrutiny. The insurance moratorium has been reviewed and extended several times.

In 2005 a number of key policy agreements between the Government and insurance were incorporated into a Concordat which was combined with an extended Moratorium. This included agreement as to the types of test and insurance products that are in scope; the process for approving new tests and arrangements for monitoring and seeking arbitration in cases of dispute. There was also agreement that applicants could offset a poor family medical history by voluntarily disclosing favourable (negative) genetic test results. It is assumed that most insurers would look favourably on such results even if they were not assessed by the independent committee.

The Concordat also included two key principles; that no insurer would ask an applicant to have a genetic test as a condition of insurance and that research results were exempted from disclosure. It also accepted the key principle for insurers that family medical history and information on referrals, diagnoses or treatment was potentially disclosable.

3. Keeping the policy under review

In 2007, the Government's disability law review consulted on the possibility of legislating to protect against discrimination on genetic grounds. The Equality Bill consultation response (CM7454³) concluded in July 2008 that, on balance, there was little evidence of discrimination on the grounds of genetic predisposition, that the Moratorium appeared to be working and that it was not clear that discrimination law was the right route to address this problem in the future. It noted that:

"Should legislation appear necessary in the future, it may be more appropriate to strengthen data protection legislation or make specific rules as to what can be done with genetic information in this respect".

The Government response to the House of Lords' Genomic Medicine report (CM7757⁴) in 2009 welcomed the Committee's conclusion that there was no evidence to support specific legislation on genetic discrimination. It also undertook to consider a longer-term agreement at a future review of the agreement between the Government and the Association of British Insurers.

The UK policy position on genetics and insurance has been reviewed regularly and in each case the Concordat and Moratorium has been extended. The Human Genetics Commission considered specific aspects of the policy in relation to the concerns raised by people who may be considering a genetic test but may wish to buy insurance after the end of the moratorium – the "test now, buy later" question. This was also raised in the 2009 House of Lords report on Genomic Medicine.

The review of the agreement in 2011 introduced a more formal arrangement for review and extension so that a review would happen every three years and that there would never be less than 3 years between the review and the end of the moratorium. For example, in 2014 the moratorium was extended until 2019 with a review in 2016. The 3-year period was to enable anyone who has an adverse genetic test result to decide whether to buy insurance before the end of the moratorium. The agreement is due to expire in 2019 and a review is underway which will conclude in late 2017.

4. Research findings and insurance

Since the early debates on genetics and insurance it has been recognised that many patients and families were part of research studies seeking the genetic basis for severe disease. As such tests were in a development phase, the 2005 Concordat clarified that they did not need to be disclosed for insurance purposes. In the 2014 revision of the Concordat it was clarified that findings from 100,000 Genomes Project were primarily research findings and the information fed back to participants will not need to be disclosed.

In 2015, the early experiences from NHS Genomic Medicine Centres seeking consent showed further complexities around the optional feedback of incidental findings. As these were potentially actionable – for example through early screening for cancers or prophylactic interventions – it was possible that the additional medical interventions would be disclosed to insurers and therefore side-step the protections under the Concordat and Moratorium. Following a helpful dialogue between experts from Genomics England, the Genomic Medicine Centres, Genetic Alliance UK and the insurance industry advice has been given that emphasises that in most cases taking part in a research project will not alter a person's insurance arrangements. It outlines the normal expectations of insurers in seeking access to medical information and in particular their desire to look favourably on preventative measures that were proven to be effective.

<https://www.genomicsengland.co.uk/the-100000-genomes-project/insurance/>

5. The UK's international position

The UK experience has been considered with interest by several other countries. In the United States the main policy concern has always been related to health insurance, and the potential challenges for those with a genetic condition have been covered by legislation such as GINA and HIPAA. In other countries, also with insurance-based healthcare and/or concerns about genetic tests (such as France and Germany), there is legislation to prevent insurers using genetic test results.

Comparisons with the UK are complicated by the universal access to healthcare afforded via the NHS and by the fact that UK private insurance contracts are fundamentally different from those in many other parts of Europe. Put simply, UK policies pay a defined benefit whereas many in EU jurisdictions pay out variable sums depending on the level of claims against the policy fund.

This debate is currently active in Canada which provides two interesting elements. The first was a ruling in July 2014 by the Privacy Commissioner that challenged some of the assumptions made about disclosability of medical (and genetic data) to insurers under the principles of proportionality and necessity. The conclusion was that there were strong grounds for not disclosing genetic data for insurance and that the Canadian insurers should continue their voluntary moratorium for reason similar to the UK experience. In parallel, the equivalent of a Private Member's Bill was introduced to the Canadian Parliament by Sen. Cowan. The final Genetic Non-Discrimination Act (Bill S-201) received Royal Assent on 4th May 2017. It prohibits anyone from requiring a genetic test or the disclosure of the results as a condition of any contract or agreement, with exceptions for health care providers and researchers. It also amends the Canadian Labor Code to extend the protections to employment contracts.

Some of these differences in approach have been apparent in the Council of Europe's work to develop a non-binding Recommendation based on the principles in the Convention on Human Rights and Biomedicine. The UK experience has been considered in detail and in October 2016 the Council of Europe adopted the "Recommendation on the processing of personal health-related data for insurance purposes, including data resulting from genetic tests". This is one of the first international instruments on genetics and insurance that provides a framework for national legislation from a human rights perspective (see https://search.coe.int/cm/pages/result_details.aspx?objectId=09000016806b2c5f).

Note

The Council of Europe's Recommendation sets out seven principles based on the Council of Europe conventions on biomedicine and data processing. The experience of the UK debate has informed a number of the detailed requirements. There is a focus on the importance of consent, the justification for requesting medical information and the importance of respecting the confidential nature of such information. It repeats Article 12 of the Convention on Human Rights and Biomedicine such that insurers must not insist on a genetic test as a condition of insurance. However, if national law permits it, then pre-existing genetic information may be taken into account if it meets the specified criteria for accuracy and relevance.

The seven key principles are important to balance the interests of patients and their families with the insurers' statutory duties to accurately assess the risk of an insurance product:

- Insurers should justify the processing of health-related personal data
- Insurers should not process [data] without consent
- Insurers should have adequate safeguards for storage of [data]
- Insurers should not require genetic tests for insurance purposes
- Insurers should take account of new scientific findings
- Member States should facilitate risks coverage that is socially important
- Member States should ensure adequate mediation, consultation and monitoring

The Recommendation is not binding on national governments but will be taken into account in the current review of the UK policy.

Box 1 - International agreements on genetic information and insurance

Council of Europe - Convention on Human Rights and Biomedicine (1997)

Article 11 – Non-discrimination

Any form of discrimination against a person on grounds of his or her genetic heritage is prohibited.

Article 12 – Predictive genetic tests

Tests which are predictive of genetic diseases or which serve either to identify the subject as a carrier of a gene responsible for a disease or to detect a genetic predisposition or susceptibility to a disease may be performed only for health purposes or for scientific research linked to health purposes, and subject to appropriate genetic counselling.

UNESCO International Declaration on Human Genetic Data (2003)

Article 14: (b)

Human genetic data, human proteomic data and biological samples linked to an identifiable person should not be disclosed or made accessible to third parties, in particular, employers, insurance companies, educational institutions and the family, except for an important public interest reason in cases restrictively provided for by domestic law consistent with the international law of human rights or where the prior, free, informed and express consent of the person concerned has been obtained provided that such consent is in accordance with domestic law and the international law of human rights. The privacy of an individual participating in a study using human genetic data, human proteomic data or biological samples should be protected and the data should be treated as confidential.

6. Conclusion

The public and political concerns in 2001 have reduced and the industry has shown itself able to self-regulate and manage the inadvertent or inappropriate disclosure of test results. There have been few complaints and these have almost all been resolved by the existing financial services regulation and arbitration. In view of the low workload, the Genetics and Insurance Committee (GAIC) was disbanded in 2009. The UK's voluntary and soft-law regulation of the use of genetic data by insurers has proved to be flexible and responsive to changes in the genomic technologies. Many elements of the UK position have been reflected in the recent Council of Europe Recommendation. The Government has said it will take careful account of the experience and international consensus in the forthcoming review of the Concordat and Moratorium to ensure that public concerns are addressed beyond 2019.

7. Authors' commentary

We support the long-standing Government policy of maintaining been to maintain a flexible semi-voluntary regulatory structure for this fast moving technology. The concordat and moratorium were developed to prevent individuals from being deterred from obtaining predictive genetic tests due to the fear of potential insurance consequences. Trying to legislate on this basis, however, would raise questions about the use of other non-genetic information that is predictive of ill health.

The Concordat and Moratorium is complex and not widely understood. It is based on a series of positions arrived at through the application of legislation such as the Data Protection Act, the Equality Act and various rules relating to the financial services sector. It also adopts a series of pragmatic measures based on the insurance industry's underwriting principles for different types of insurance. It may also expire as soon as 2019 if not extended.

Insurers have different requirements for underwriting insurance contracts based on the size of the sum insured. One of the key features of the Moratorium that is not widely appreciated is that no-one needs to disclose a genetic test result if the policy is worth less than £500,000 for life assurance or £300,000 for critical illness cover or £30,000 per year for or income protection policies. This means that at current estimates more than 95% of insurance customers would not need to disclose genetic test results.

We suggest that the next review of the Concordat and Moratorium is based on the following:

- The adoption of the key points in the Council of Europe recommendation reflecting international consensus.
- That it is renamed into something more readily understood – such as the Genetics and Insurance Code – and clearly linked to the relevant statutory requirements and to industry best practice.
- That the Code is long-term (preferably open-ended), albeit with regular reviews, to reassure people considering a genetic test or joining a research study.
- That it clearly adopts appropriate financial limits that are index-linked to recognised economic measure, such that patients and the public are reassured about never being required to disclose genetic information for most insurance policies.
- That it recognises the importance of properly conducted medical research and reinforces the principles of the current Concordat, including with the detailed agreements reached for the feedback of additional findings arising from the 100,000 Genomes Project.

One consideration that does need to be addressed here is the question of genetic exceptionalism. Past policy debates have questioned why other predictive health information is not equally protected. Some argue that it is potentially futile to protect the use of genomic findings but to accept that as soon as someone acts on the findings, for example by accessing diagnostic or screening services, then they are required to disclose the information when buying insurance. Also, many people's genetic information is clear from routine laboratory testing e.g. sickle cell haemoglobin. Similar concerns apply to people who are identified at high risk of ill health, for example from diagnostic DNA tests, screening programmes or other technologies such as X-rays or imaging. Furthermore, a large proportion of genetic conditions are diagnosed in childhood and individuals can face difficulties accessing affordable insurance.

These challenges are complex and the response must vary according to the overlap with other equalities legislation. The main issue is that genomic information is sensitive, impacts on relatives and future generations, is of uncertain actuarial relevance and that not protecting against misuse may deter people from taking up the offer of diagnostic tests or research opportunities. That is why it remains a public concern and also why the response must be suitably flexible and responsive to particular situations.

Box 2 - Key UK policy reports

■ 1995

House of Commons Science and Technology committee report "Human Genetics: the science and its consequences" summarised the potential impact of advances in genetic testing and the impact on insurance. It recommended legislation if insurers did not address the concerns.

■ 1997

Human Genetics Advisory Commission report on genetics recommended a two-year moratorium on the use of genetic tests by insurers

■ 1999

Establishment of Genetics and Insurance Committee (GAIC) – an expert committee of scientists, actuaries and lay members to evaluate the medical and actuarial quality of genetic test information.

■ 2000

Application for HD in life insurance, approved by GAIC

■ Early 2001

House of Commons Science and Technology committee report "Genetics and Insurance" recommended a two-year moratorium, enforced by legislation if necessary, to allow further research on the scientific and actuarial relevance of genetic test results for insurance

■ November 2001

Industry offers a moratorium on genetics and insurance which was accepted in the Government response

■ 2005

Extension of the Moratorium and agreement of a Concordat between Government and the Association of British Insurers on the wider principles and procedures for review.

■ 2007/8

Consultation on legislation against discrimination on the grounds of genetic predisposition as part of the review of equalities legislation

■ 2009

House of Lords Science and Technology Committee: "Genomic Medicine" which concluded legislation was not needed at present but the rapid advances should be kept under review by a body such as the Human Genetics Commission.

8. References

1. Thomas GR. 'Genetics and insurance in the United Kingdom 1995–2010: the rise and fall of "scientific" discrimination. Published online. 2012
2. Select Committee on Science and Technology,. Fourth Special report, 2001. Genetics and insurance. London: House of Commons
3. Government Response to The Equality Bill. London. Crown Copyright 2008
4. Government Response to the House of Lords Science and Technology Committee Inquiry into Genomic Medicine. London. Crown Copyright 2009

Chapter 16

Ethics and the social contract for genomics in the NHS

Chapter leads

Prof Anneke Lucassen¹, Prof Jonathan Montgomery²
and Prof Michael Parker³

Authors contributed equally

- 1 Faculty of Medicine, University of Southampton
- 2 Faculty of Laws, University College London
- 3 Ethox, Oxford University

1. Introduction

Previous chapters in this report have illustrated the potential of genomics to lead to important improvements in our understanding of health and in the diagnosis and treatment of disease. They have also suggested that the achievement of these benefits is going to require significant changes in the ways in which healthcare is understood, organised and practised in the NHS. One of the most important of these is the need for a greater degree of integration of, and complementarity between, healthcare and medical research. A second is a growth in the importance of the collection, storage and appropriate sharing of information at scale: in the care of individuals and families, in research, and in the improvement of health systems. A third will be the importance of a faster pace of learning and a consequently greater degree of uncertainty and open-endedness in healthcare practice. Each of these is going to have profound implications for the way the NHS works and for how the obligations and responsibilities of health professionals and institutions – as well as those of patients – are to be understood. Together, however, these changes have the potential to bring important benefits to patients and their families.

* In this chapter we use the term 'social contract'. In similar discussions others have sometimes used the term 'social license'. We prefer the term social contract because of its helpful implication of the location of healthcare and medical research in a broader context of social arrangements, practices and institutions.

The founders of the NHS were committed to two guiding principles.¹ The first of these was to the availability of healthcare on the basis of need and independent of the ability to pay. The second, which is less widely discussed than the first, was that this health care should be of a very high standard of excellence. The value of the first is clearly enhanced by that of the second. Development of genomics and its integration into day-to-day practice of the NHS has the potential to provide important improvements to the quality of care provided equitably by the NHS, but the pace of development and the need for large scale data interpretation mean that research and innovation will need to become more integral to routine NHS practice than we have seen to date.

In this chapter, we reassert the importance of these important founding principles of the NHS and argue that if the potential benefits of genomics are to be realised, there is a need for a rethinking of the wider 'social contract' for medical practice and research in the UK.* After introducing and discussing the social contract and its importance, we go on to identify and discuss four areas to which we believe particular attention needs to be paid. The first of these is consent. The second is the use of information in the care and treatment of both patients and their relatives. The third is the need to rethink the duties and responsibilities of health professionals, and the fourth concerns the responsibilities of health systems in the context of rapid developments outlined in previous chapters.

2. Rethinking the social contract for medical care and research

The NHS Constitution reminds us that our health service is founded on a common set of principles and values that bind together patients, the public and staff in order to ensure that it can be effective and equitable. It recognises that each party has important rights that must be respected, but also that each owes each other responsibilities. Through this combination of reciprocal rights and obligations the NHS aims to operate fairly and effectively for mutual benefit. The NHS Constitution is thus the expression of a form of 'social contract' which aims to bring the highest levels of human knowledge and skill to save lives and improve health. We need to understand and agree how those rights and responsibilities work in genomics if we are to harness its potential to fulfill the promise that the NHS "is there to improve our health and well-being, supporting us to keep mentally and physically well, to get better when we are ill and, when we cannot fully recover, to stay as well as we can to the end of our lives".²

Under this social contract, the health service has important responsibilities. We feel safe in entrusting our bodies and intimate personal information to health professionals in part because of the rights we have to protect our ourselves through the giving or withholding of consent, in part because we have confidence that those staff will act with integrity in our interests – for example in maintaining high standards of confidentiality – and in part because there are systems in place that protect our interests and hold professionals to account. In the era of genomic medicine the basis of this contract needs to be revisited. Most obviously, this is because linking up of large data-systems containing personal identifiable data,[†] on a scale not previously necessary (or possible), is a prerequisite for success, but also because genomics will provide both diagnoses and predictions and will affect patients, families, the general public in different ways over time.

Whilst it is clear from earlier chapters in this report that genomics has the potential to bring great benefits to patients, there has been considerable, and understandable, public concern over the handling of personal data by the NHS,³ coupled with suspicion over the involvement of commercial organisations in the handling of 'big data'.⁴ A recent survey commissioned by the Wellcome Trust on commercial access to data suggests that the public see genomics as both of great potential benefit and as presenting important risks.⁵ The success of genomic medicine will depend on patients having confidence that the way genomic information is generated, held and used will properly protect their interests. This requires re-examining the traditional rules around confidentiality, which focused on secrecy and the keeping of information as separate and private. Such a rigid separation cannot operate in genomics, which requires clinicians to consider the patient's specific genetic situation in comparison with knowledge gleaned from others. The most important structural implications of the move to genomic, big data-driven medicine is the requirement for a greater degree of interdependence between the care and treatment of individual patients on the one hand and the collection and analysis of data relating to the care of very large numbers of other patients, often in real time, on the other. The clinical interpretation of genome findings requires information about clinical features in others with similar findings: the genotype-phenotype association remains an important clinical tool. Genomic medicine will require use of patient level information to support better clinical decisions in the future and for others. For this to be ethical, and acceptable to patients, a stronger focus on information security, data analysis and decision-support will be required as will greater clarity about and broad agreement on the relevant obligations of health professionals and systems.[‡]

[†] In the past, research studies providing the data for health care improvement would either have been collected using specific consent, and or a non-identifiable form. Genomics raises problems with anonymization because the data cannot be completely anonymised and can result in identification of individuals. Attempts to deal with this issue through consent are problematic because of the wide range of possible outcomes, over time, from genomic research.

[‡] In addition to the benefits to individuals of interpreting their information in the context of a larger dataset, there will also be situations in which information arising in the direct care of one person, can prove vital in informing the care of others e.g. where it leads to the identification of an infectious disease, or where it identifies others at high risk of a condition. Provision will also need to be made for dealing with such situations.

As the Nuffield Council on Bioethics has suggested,⁶ the successful and appropriate use of data-driven approaches to healthcare and research will require the NHS to provide the public with a mutually acceptable statement of the expectations that they can have of the use of data. This would need to include a realistic explanation of the ways in which genomic medicine can personalise care, addressing the sometimes excessive assumptions about the predictive or diagnostic power of genetics. It would require an explanation of how the management of data protects privacy, but necessarily uses information that cannot be completely anonymous because of the uniqueness of our genetic identities.⁵ It would need to recognise that one of the consequences of advances in genomic medicine is that initial consents to the use of genomic information cannot be fully informed about future uses or interpretations, and so there is a need for high standards of broad consent to be complemented by continuing oversight of the way genomic data is handled.

This needs to be coupled with assurances of the competence of the NHS to deliver on the information governance requirements of genomic medicine. People need to be able to trust it to hold data securely, make it available to clinicians reliably when needed for patient care, and to patients themselves in an intelligible manner. Current information systems in the NHS are unlikely to be sufficient to earn this level of trust since they often seem to operate on a tacit permission to continue until the public withdraws their support due to mistrust. Instead, people need to be satisfied that genomic medicine operates in their common interests, whilst protecting their individual privacy, and does not exploit some to benefit others. Protection of individual privacy cannot be absolute,** nor can data ever be guaranteed as entirely secure, but there needs to be an understanding of the associated risks and reassurance that breaches are appropriately prosecuted.

The basis for greater trust and confidence created by such a social contract could encourage the growth of “genomic citizenship” or the genetic altruism and solidarity described by the Human Genetics Commission⁷: Genomics offers benefits and responsibilities for the individual, the family, the broader community and globally that cannot be realized by keeping the secrets revealed from one genome separate from others.

This requires a mutuality that is not captured by current systems. The idea of a social contract provides a basis for such an arrangement because it endures over time, brings benefits (and obligations) for both patients and the professionals (and services) who offer care. To achieve this, processes for creating common understanding are required, as well as mechanisms for revising the agreement when necessary.⁸

⁵ Even though we differ in only roughly 0.1% of our genetic codes, this still equates to some 3 million variants.

** We will also need to address how protection of privacy is related to identifiability of genomic information, i.e. just because a sequence is potentially identifiable because of its uniqueness, does not mean that the privacy of a person is more invaded than were the data truly anonymous.

3. Renegotiation of the social contract - reasons

3.1 Overview

We began this chapter by highlighting three key requirements for the achievement of the benefits of genomics in the NHS for patients and families. These were: a greater degree of integration and complementarity between research, innovation, and clinical practice; the collection, storage and more effective use of health data; and, a more central role for learning and open-endedness in day-to-day clinical practice. In what follows, and against this background, we discuss a selection of some of the key areas of medical practice and research in which new ways of thinking about and practising medicine – each an important element of the social contract – are going to need to be considered in the renegotiation of the social contract. They are: (1) consent, (2) confidentiality and caring for families, (3) the obligations of health professionals and researchers, (4) the appropriate uses of data and samples, and a range of governance and system responsibilities.

3.2 Valid consent

The obtaining of valid consent is an important part of good ethical practice in healthcare and research. Whilst consent is an important component of ethical practice, it is not in itself, however, a guarantor of high ethical standards. In genomics as elsewhere, consent needs to be understood as an important component of an ethics ecosystem along with, for example, the duties and obligations of health professionals to treat patients with respect and care, and the requirement for health systems to provide protections to ensure that those who provide their consent are not exploited, discriminated against or unfairly treated.

Consent is nonetheless an important part of good medical practice and high standards of consent are essential. In genomic medicine both the importance and the limits of consent become increasingly apparent. The wider introduction of genomics into medical practice will present significant challenges for the achievement of understanding. Many of the key concepts in genomics are both complex and likely to be new to many patients (as well as the health professionals offering them) and may present problems of explanation and understanding as will many of the features the healthcare system in which genomics plays a central

role: the close relationship between research and clinical practice; the collection, storage and use of health data; and, the open-endedness and uncertainty at the heart of a learning healthcare system. These latter two factors will be important both at the time of consent and at the time of communication of results which may be revised over time as new evidence accumulates. There will often be a degree of uncertainty about findings and their current or future implications as well as uncertainty about the potential future research uses to which data may be put and what additional (or incidental) findings this may produce in the future. This uncertainty may be at the level of evidence available; more big data is needed to ensure that findings are reproducible, and that confidence limits are minimised. It may also be reflected in the fact that even where good evidence exists, the chances that the finding will result in a particular symptom or condition is uncertain because it may be just one factor in amongst several that determine whether the condition manifests. NHS health professionals will also need to improve their acknowledgement of such uncertainty, as too often the language of single highly penetrant gene mutations is used for susceptibility factors that may never manifest as signs or symptoms.

Case study 1

Results of genomic investigations in patients investigated for neuro-developmental delay (whose samples have been collected and stored in a national resource) reveal a mutation in a gene that increases the chance of a brain tumour. The relative risk that this mutation appears to confer is very high, but the absolute risk less than 15%. Records of the consent taken at the time of testing reveal no mention of the possibility of tumour/ cancer risks being found. Health professionals are concerned about disclosing this finding since no specific consent was given to find it. Some argue that the patients have a right to know about their increased risk, but others argue that the lack of clear evidence based interventions and the 85% chance of not developing such a tumour would go against disclosure. However, had consent been explicit about this possibility the health professionals would have disclosed regardless of 'actionability'.

The quality of consent needs to be sufficient to reflect the importance of respecting patients' autonomy, but what kind of understanding of genomics is good enough for a decision to be seen as autonomous? This problem is sometimes presented as a problem about 'broad consent'. To what extent can consent to participation can be thought to be genuine 'consent' where significant implications of the decision are unknown, or unknowable at the time? Can such consent meet the requirements for validity? How broad can consent be and still be valid, or indeed prevent claims of insufficient information, to make a decision?

In this context, a key question is going to be how might the validity of consent be judged in contexts of such complexity and uncertainty? It is clearly not reasonable for the answer to be that consent to genomic testing, storage of the sample and communication of the data is only valid with 'full understanding'^{††}. This would mean that the benefits of genomics could not be realised. It would also mean the imposition of a highly paternalistic approach to consent in which patients were not allowed to come to the conclusion – in the real world, against a background of significant uncertainty, that this is something they would like to pursue.

All of this suggests a need not only for the development of new evidence-based approaches to best practice in consent but also a clearer statement of the complementary roles of consent and of other protections. It is our view that an important question should be what protections and controls need to be in place such that when people do give their valid consent – inevitably on the basis of a degree of uncertainty and open-endedness – they are not exploited, discriminated against or unfairly treated.

It is also going to be important to consider what function consent is required to play, given the familial aspects of some genetic findings. Unlike an operation or procedure where there is a physical intrusion for which the operator requires consent, consent to genomic testing may perhaps be better seen as being explicit about entering a relationship with agreed ground rules about mutual responsibilities and rights. These mutual responsibilities and rights extend to the individual, their relatives who may be unwitting 'stakeholders' in the outcome of genomic testing and to the population as a whole who stand to benefit from largescale geno[me]type/ pheno[me]type correlations.

Finally, it may be that the challenges presented by the uncertainties and open-endedness of genomic medicine require a rethinking of some aspects of the role of non-directedness in the doctor-patient relationship. It may, for example, turn out to be the case that patients are more likely to be content with the decisions they have made where the process of decision making involves a greater degree of clinician involvement/deliberation than is the case elsewhere. So whilst the genome 'sequencing' is a technical step, that can be undertaken with minimal medical intervention (spitting into a pot and sending it through the post) the complexity of the possible outcomes of analysis may require an extended clinical interaction to ensure that different types of outcomes (clear/ uncertain, mild/ severe, current or future) are assimilated in the consent process in a more clinician directed way than would normally be expected. Might this, perhaps, be a place where the evidence reveals joint decision-making comes to be seen as 'better' (by patients) than one that is more 'non-directive'?

^{††} This is of course also true for existing care outside of genomics. The difficulties in promoting 'full understanding' are rarely acknowledged in policy documents, but have been examined by N. Manson and O. O'Neill 'Rethinking Informed Consent in Bioethics' (2007).

3.3 Confidentiality and the availability of the best care for patients and families

High standards of confidentiality and the securing of potentially sensitive health care information are going to be at the heart of good genomic medicine practice. However, there are at least two important ways in which patients and the public are likely to be supportive of new practices in the use of patient information, each of which suggests the need for new thinking on the appropriate uses of health-related information and their limits:

The first of these relates to the potential benefits for individual patients of having at least some of their clinical data analysed together with genetic findings from others in population-scale (secure) data bases so that evidence can be acquired on the nature of the link between genotype and phenotype. This might improve their own care, now, in the future and improve the care of others.

Case study 2

A mutation in the BRCA1 gene was thought for many years to confer a high risk of breast and ovarian cancer. More recently, evidence suggests that it is a benign variant and that the surveillance and interventions offered to those with the variant were therefore wrongly directed. This evidence has only come to light through international efforts and database linkages of family history details and segregation of the variant with disease in families. Although those with the variant have previously been advised they are at high risk, they, their relatives and future individuals can now receive more up to date clinical advice.

Source See for example, E.T. Rosenthal, et al, 'Exceptions to the rule: Case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes', *Clinical Genetics* 2015

The second situation when the sharing of patient information in new ways might be expected to command support is to distinguish individual clinical information about a disease or condition from the inherited mutation(s) that led to the clinical findings. Whilst professional guidelines such as those from the GMC⁹ specifically list genetic information as one possible reason for breaching individual confidence (if doing so would protect people from serious harm), it might in certain cases also be possible to share relevant information without any breach of confidence. In practice, it is not always necessary to disclose to relatives (existing or future ones) that a specific patient has been diagnosed with, say, inherited breast cancer. They can be informed that in a particular family there is an inherited tendency to cancer that could be usefully tested for in family members who are worried about their risk. On some occasions, the second approach might raise concerns that discussing the test would identify a particular family member and constitute a breach of her or his confidentiality but this need not always, or even often, be the case – particularly in large or multi-generational families where others have the familial disease in question. For example, a woman who is concerned about her family history might simply be offered an appropriate genetic test without this raising any confidentiality concerns about the individual in whom the familial cause was first identified. We would argue that where this is the case, a social contract that would allow such information to be available for use by clinicians in the appropriate care of family members (for example, testing for the particular familial mutation to determine if extra surveillance is warranted) would be publicly acceptable.** Although this involves the use of information beyond the individual care of the person, this approach would only use familial information and not disclose any individual details thus maintaining confidentiality.

** See for example, Confidentiality and sharing genetic information with relatives. Lucassen A, Parker M. *Lancet* 2010 May 1;375(9725):1507-9. Genetic information: a joint account? Parker M, Lucassen AM. *BMJ*. 2004 Jul 17;329(7458):165-.

Case study 3

A man with a mutation in a mismatch repair gene resulting in a high lifetime chance of bowel and other cancers steadfastly refuses to inform his siblings, or allow his doctor to do so, of the risk they might be at. The health care professionals know that one sister has had bowel cancer and is therefore likely to harbour the same mutation. This sister is at increased lifetime risk of endometrial cancer and might therefore benefit from a risk reducing hysterectomy. The heritable aspect of the cancer is insufficiently common to justify testing unless there is a family connection. Unless we allow the use of the familial information, we have to choose between either testing everyone, for little clinical utility in most cases, and at a cost to the NHS, or not testing at all.

The health professionals have 3 options (1) to respect the man's wishes (2) to breach his confidence on the basis that it is justified by the opportunity to prevent harm to relatives who might unknowingly have the mutation as per GMC guidelines (3) Use NHS tracing to contact the sister's GP and tell him/her that a referral to a genomic service is recommended because she might be at increased risk. Option (3) does not need to breach the man's confidence because only information that is familial is communicated.

One way forward therefore is for the boundaries of confidentiality in genomics to be seen, at least in some situations, at a familial rather than individual level. Taking a recent court case (*ABC vs St Georges*¹⁰) as an example, it may indeed not have been good practice for clinicians to tell the daughter that her father had the genetic condition, Huntington's disease without his consent, but it might have been perfectly good practice to tell her that the facts of his case, and his family history – both which were in the public domain - indicated a potential familial risk about which she could seek independent advice and treatment. This separation of approaches to confidentiality of individual clinical information from familial genetic information has not yet gained widespread traction in practice, in part because of the limited situations to date where it was required, and in part because in some situations communication about familial information could lead to inference about individual clinical information. Evidence from qualitative research on the topic suggests that although health professionals find thinking in a familial way about genetic information difficult,

many patients assume such an approach is happening and are surprised to hear that sharing and familial use are not standard practice.¹¹ This suggests that there may be patient and public support for an approach adapting the default position on confidentiality such that: Instead of breaching confidence only if one can prove they are preventing serious harm in specific others (current GMC guidance), the default becomes that relevant information that might prevent harm is communicated unless there are good reasons not to.

There will, inevitably be some situations of a different kind in which the question of using of properly confidential patient information in the care of family members will need to be considered. Confidentiality is an extremely important part of good medical practice. The provision of evidence based medical advice and treatment requires patients to undergo tests and to entrust confidential information to professionals. Confidentiality also shields patients from embarrassments and intrusions into their private lives. Protecting patients' confidences is as important in genomics as any other area of medicine. However, since certain genomic information may also be relevant for others, for example, biological relatives, or be dependent on data from them before it becomes information, there is a higher degree of interdependence in the generation of information than most areas of medicine. The principles of confidentiality and data protection therefore require special attention in this context, especially as the scope of genetic and genomic testing increases.

On the one hand genomic findings may convey or predict sensitive, potentially stigmatising facts, on the other hand much genomic information is common to many, and particularly common to that of biological relatives. The information generated in circumstances of confidence to one person, may allow inferences to be drawn about its significance to another, whose views may, or may not be known. Conversely, inferences about a particular genomic output may only be possible if confidential information is first obtained from others. These aspects of genomics can mean that health care professionals do not know whether they are balancing their duties of confidentiality with the rights and freedoms of others appropriately. This needs to be borne in mind when calls for better data sharing between laboratories and countries are made. Data is most useful if linked to some clinical information and those submitting the data need to be clear what is acceptable within the rules of data protection and confidentiality. This is captured in the idea of 'fair processing principles' which can be developed through the elaboration of a new social contract for genomic medicine.

Case study 4

Communication is not good between the different family members of one family at risk of sudden cardiac death through a pathogenic gene mutation that alters cardiac repolarization. Although information letters about the condition, the risks and the surveillance and treatments available have been given to the index patient in whom the mutation was found, it is clear that these have not been passed on. When another member of the family is referred for assessment of his family history of sudden cardiac death, health professionals are unsure whether utilizing the genetic result of the index patient in this assessment would breach his confidence. Whilst it would be inappropriate to reveal the clinical details of this index patient, telling the family member that there is reason to believe he might be at risk of a heritable mutation, is not.

One way in which such situations might be preempted is through the obtaining of consent at an early stage for the use of such information for the caring of family members as well as for submission to (inter)national databases for the benefit of wider family or other families. Through consent, patients can authorize the use of their confidential information for other purposes, including research and the treatment of others, and it may be appropriate to take steps to encourage patients to adopt this form of altruism or 'genetic solidarity' as a routine step in genomic medicine, or a social contract for confidentiality.¹² Whilst much clarity can be achieved by such encouragement and explicit 'up-front' statements, there will be times when consideration needs to be given to whether it might be legitimate to use the information in question without specific consent. GMC guidelines on the limits of confidentiality with respect to genetic information⁹ are helpful in clarifying this possibility, but increasingly genomic testing is creating situations where the wishes of an individual are not known, and not easy to obtain, yet a result - perhaps not anticipatable at the time of testing - is relevant to others. As genomics reaches into many more areas of routine medical practice, consideration will need to be given as to whether conventional notions of a duty of confidentiality are realistic or appropriate and how genomic findings of different kinds ought to be dealt with.

3.4 The obligations of health professionals, laboratory staff and researchers

Genomic medicine will have implications for what it means to be a good and ethical health professional in the NHS. It is likely that careful thought is going to be required on the question of how are we understand the obligations of doctors to their patients in this new world e.g. in the context of greater uncertainty, evolving knowledge and ongoing feedback, and a greater concern for the care of families. The clinical use of genomics is likely to take place in the context of a greater degree of interdependence between clinical practice and research, the collection and use of large datasets, and a much greater emphasis on ways of improving understanding and interpretation through the on-going refreshing of datasets with new data in real time. It will be increasingly difficult to argue that research and clinical ethics involve separate sets of principles, a distinction on which many current professional guidelines are based. Originally, the Declaration of Helsinki denied the acceptability of 'therapeutic' or 'clinical' research by professionals on their patients unless direct benefit to them was expected. Since, 2000 there has been a slight relaxation, with additional safeguards applied in the category of research combined with care, but maintaining a clear separation between the two.¹³ Article 14 now recognises that research may be permissible provided that harm is avoided. It states that 'Physicians who combine medical research with medical care should involve their patients in research only to the extent that this is justified by its potential preventive, diagnostic or therapeutic value and if the physician has good reason to believe that participation in the research study will not adversely affect the health of the patients who serve as research subjects'.¹⁴ However, rapidly developing medical practices, and particularly genomics, will force us to revisit these positions since these areas involve research and care being alloyed together so that each activity is dependent on the other.

This also works the other way around. It means that researchers and data managers may increasingly come to be seen to have responsibilities that cannot easily or completely be divorced from clinical care. As the clinical predictions from genomics become clearer, it will become the norm rather than the exception that research will produce information that has potential clinical significance. It will be important to clarify when researchers are expected to liaise with clinicians, and which clinicians they should contact. Novel ways of linking research laboratories with clinical teams and quality assurance approved laboratories for validation will be important for effective interaction between research findings and clinical practice.

In relation to clinical care, the level of detail of stratified and personalised medicine means that the resources to support decision-making and underpin evidence-based care will be different. Randomised controlled trials the gold standard evidence generating tool in many areas of medicine, will be more difficult to employ. Evidence about a particular genomic finding will require large scale (often across national boundaries) phenotype-genotype correlation that take ancestral genomic background and environmental factors into account. Nuanced yet uncertain diagnoses or predictions will remain the norm in genomics for the bulk of clinical practice over the next few years. Yet this is in the face of a discourse about genomics that often mixes appropriate claims of technological accuracy with claims about their clinical predictions which remain far less deterministic than commonly perceived. This will require judgment to play a greater role, something that is consistent with the origins of the evidence based medicine (EBM) movement but has become less prominent.¹⁵ This recognition of the importance of the subjective views of patients on what is material to their decisions in a 2015 UK Supreme Court decision in *Montgomery vs Lanarkshire* (2015) will need to be examined in the wake of genomic medicine's possibilities including its uncertain predictions.¹⁶

3.5 System responsibilities

These issues cannot be resolved by individual clinicians, but need to be addressed collectively through the appropriate design of health systems.

Case study 5

A large study of whole genome sequences reveals a series of 'pathogenic' mutations in a sample of well individuals who took part in sequencing in order to make the search for a diagnosis in a relative more effective. Questions are raised about whose responsibility it is to communicate this information, how the downstream implications for the NHS in terms of clinical follow up, surveillance and treatment are managed, and how the evolving evidence about the predictive value of these mutations in terms of disease can best be communicated and by whom.

Decisions on whether such contact should be made require complex analysis and awareness of the ethics of risk communication. It seems unreasonable to place that burden on primary care alone. Provision needs to be made by the health service for analytical capacity and ethics support to advise researchers on when contact might be appropriate and clinicians on the significance of the information and how best to communicate it without causing confusion. It will also be important to establish when contact should be expected, and the NHS could be held liable for failing to seek to achieve it. For example, what level of risk, certainty or medical interventions would need to be available for lack of contact to be negligent and who would judge? When might contact be a discretionary matter and how could this respect potential rights not to know? It would be inappropriate to create legal obligations beyond those of fair and non-discriminatory processes. If we are to achieve a consistent service, these issues will need to be tackled at a health system level not on individual clinical or research responsibilities. In any event, NHS clinical services and research studies are not currently resourced to be able to take on this role.

Further, if we recognise both the importance of consent and its limitations as a guarantor of ethical practice, and that genomic medicine challenges conventional approaches to confidentiality in significant ways, then a key question becomes what complementary protections and controls need to be in place such that when people do give their valid (but inevitably imperfect) consent, they are not exploited, discriminated against, unfairly treated and have their privacy unacceptably encroached upon. If, furthermore, we acknowledge that even where health professionals and laboratory staff perform their duties to the best of their ability there may be structural or institutional factors affecting the care of patients and the protection of their interests, this suggests a need to think carefully about the responsibilities of systems. That is, the responsibilities beyond those of individual health professionals, research groups or hospitals.

Such responsibilities are likely to include questions relating to appropriate and accountable governance, oversight, data-security and where required, regulation. The ability of the NHS to show that it can be trusted on these issues will be an important foundation for the reasonableness of the new social contract that we propose. It will need to create systems that ensure widespread sharing of linked genomic data that helps interpret the patient's specific information. This will necessarily originally be derived from individuals, but will need to be (a) available in a way that obscures identities where possible, (b) be subject to information governance safeguards. This is unlikely to be achievable in a fragmented provider system without national co-ordination. It may also require specific legislative authority.

However, it will also be important to establish clear responsibilities for ensuring equitable access to the benefits of genomic medicine in clinical guidelines and national commissioning standards. It will not be reasonable to expect people to accept the new social contract unless the health service accepts responsibilities for ensuring that the benefits will be available to all.

Finally, protections against unfair discrimination will need to be enhanced. The Human Genetics Commission recommended on a number of occasions that specific provision providing protection against discrimination on the basis of genetic characteristics should be introduced. This would play an important role in making the new social contract a reasonable one to propose.

4. Conclusion

In this chapter we have outlined some of the ethical challenges presented by the greater use of genomics in the NHS. We began by noting that the realisation of the important benefits of developments in genomics for patients is going to require significant changes in the ways in which health care is understood, organised and delivered. We picked out three particular aspects as having particular significance: the greater integration and complementarity between research and clinical practice; the central importance of data collection and analysis; and, the increasing role of uncertainty and open-endedness in genomic medicine. Against this backdrop, we have argued that the sustainable achievement of the benefits of genomics requires a broad renegotiation of the social contract for medical research and medical practice in the NHS. We picked out four areas in which this is likely to be particularly important: (1) consent; (2) confidentiality and the care of family members; (3) the duties and obligations of health professionals including laboratory staff; and, (4) system responsibilities and governance. There are a number of other important issues we could have discussed. Perhaps the most important of these concern the use of health data for research, and the potential importance of commercial companies in such research. Beyond the immediate clinical uses of data, the quality of care and the quality of knowledge about disease and treatments will be greatly improved by encouraging research activity on the data. Much progress is going to require the involvement of commercial and technology partners. If it is accepted that such activities are in the public interest and are a necessary condition for the NHS to meet its commitment to improvements in the diagnoses and treatments available to patients, careful thought is going to need to be given to the question of how this can be achieved in a way that commands public trust and contributes to, rather than undermine, higher standards of equitably available health care. Despite their importance, we have not discussed these issues at great length in this chapter because they are already the subject of a great deal of academic and policy debate.

The working out and agreement of the terms of any such contract requires the active involvement of many stakeholders including patients, health professionals, researchers, policy-makers, and wider society. This suggests a key role for public engagement and involvement. Evidence suggests that members of the public are aware that genomics has the potential for great benefit but that its use presents a number of risks and challenges. Whilst the risks cannot be entirely eradicated it is reasonable to expect that given certain safeguards and adequate oversight there will be strong public support for the development of a health service with dynamic genomics and the effective use of health data at its heart.

5. References

1. Klein, R., *New Politics of the NHS* (2nd Edition)
2. NHS constitution
<https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england> Accessed April 2017
3. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. *J Med Ethics* 2015; published online Jan 23. doi:10.1136/medethics-2014-102374].
4. *New Scientist: Technology News* April 2016: Hal Hodson
<https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/>
5. Wellcome Trust/ Ipsos Mori report: The One-Way Mirror: Public attitudes to commercial access to health data
<https://www.ipsos-mori.com/Assets/Docs/Publications/sri-wellcome-trust-commercial-access-to-health-data.pdf>
6. The collection, linking and use of data in biomedical research and health care: ethical issues. February 2015.
http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf (Accessed June 2016)
7. Inside Information: Balancing interests in the use of personal genetic data Human Genetics Commission. 2002. http://webarchive.nationalarchives.gov.uk/20061023110946/http://www.hgc.gov.uk/UploadDocs/DocPub/Document/insideinformation_summary.pdf Accessed June 2016
8. <http://www.acmedsci.ac.uk/more/news/a-new-social-contract-for-medical-innovation/>
9. General Medical Council. Confidentiality: good practice in handling patient information.2017.
http://www.gmc-uk.org/guidance/ethical_guidance/confidentiality.asp
10. ABC versus St George's (2015). Retrieved from www.bailii.org/ew/cases/EWHC/QB/2015/1394.html
11. Dheensa et al 'Is this knowledge mine and nobody else's? I don't feel that.' Patient views about consent, confidentiality and information-sharing in genetic medicine. *J Med Ethics*. 2016 Mar;42(3):174-9.
12. Human genetics commission report
13. The revision of the Declaration of Helsinki: past, present and future' *British Journal of Clinical Pharmacology* (2004) 57:6 695-713 DOI:10.1111/j.1365-2125.2004.02103.x
14. World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. World Medical Association *JAMA*. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053
15. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ* 2014;348:g3725.
16. *Montgomery vs Lanarkshire* [2015] UKSC 11.

Acknowledgements

This is a report of over 100,000 words including references. I cannot thank all contributors by name as you are many; I believe hundreds of people have assisted in some way to make this report the success it is. Know that if you helped out whether by attending a workshop, or contributing or commenting upon text, I thank you.

I will pay particular thanks to my Editor-in-Chief, Claudia Langenberg. Claudia has stayed with this programme with enthusiasm and has worked extremely hard, bringing diligence and creativity to her work. I thank Mark Bale for his invaluable policy input and Orla Murphy for managing the production and editing text. I have produced a report to be proud of and I am proud of the team that helps me to do it.

Authors

I would also like to give sincere thanks to each of the Chapter Leads and to the Authors of chapters in this report. I hope your work will inspire others as it has inspired me.

Significant contributors

The authors of Chapter 2

The authors of Chapter 2 remark;

Genomics England is a wholly owned company of the Department of Health and this programme was made possible by funding and assistance from the National Institute for Health Research, NHS England, Public Health England, Health Education England, Cancer Research UK, the Medical Research Council and the Wellcome Trust. We wish to acknowledge the support and advice of the NHS, Universities, academic researchers worldwide and participants which has been crucial to the 100,000 Genomes Project.

The authors of Chapter 6

The authors of Chapter 6 remark;

We are particularly grateful to Dr Marijcke Veltman and the support team for the NIHR RD-TRC for their contributions to this chapter.

The authors of Chapter 11

The authors of Chapter 11 and I would like to thank Professor Bobby Gaspar.

The editor of Chapter 13

The editor of Chapter 13 would like to thank Giovanna Ceroni for her assistance with referencing.

The authors of Chapter 14

The authors of Chapter 14 would like to remark;

We would like to thank Kate Witkowska and Calum Boyd (Genomics England) for their generous support in pulling this chapter together and Mark Bale and Tom Fowler for their comments. Sarah Wordsworth is partly funded by the Oxford NIHR Biomedical Research Centre (Genomic Medicine Theme), University of Oxford.

Glossary

Claudia Langenberg and I would like to thank Ellen Thomas for her assistance with the Glossary and for her assistance in running stakeholder events run during the report scoping period.

Report production and promotion

I thank Alix McCulloch at APS for her professional management. I thank media colleagues at Department of Health for their work to promote my report and my digital media colleagues for ensuring it is available online. I thank the Xerox team at Department of Health, who always deliver quality products at short notice, with good humour.

My private office staff

Hugo Jones, my Head of Office has provided policy support and brought his drafting skills to this project. My Assistant Private Secretaries (past and present), have worked very hard to ensure this report is successful. My thanks also go to Lee Morris and Hasim Miah, who have continued to provide sterling support, keeping the show on the road.

Glossary

Cell-free DNA

Most DNA is found inside the nuclei of the body's cells. When a person has their genome sequenced, this mainly looks at the genome of the infection-fighting white blood cells, because blood is the easiest source of cells for obtaining DNA. It is also possible to study DNA which is in the blood outside the body's cells. This DNA may come from damaged cells, for example cancer cells, or from a fetus in a pregnant woman. Studying this cell-free DNA can tell us things about the tumour or the fetus that the DNA came from.

Consanguinity

A consanguineous family is one in which related individuals have had children together, e.g. the parents are first or second cousins.

Digenic inheritance

This refers to a disease which is caused by pathogenic variants in two different genes. Some renal and cardiac disorders are thought sometimes to show digenic inheritance. Research analysis and genetic counselling are more complex in the context of digenic inheritance.

Dizygotic twins (fraternal twins)

Twins who are genetically non-identical, but are related as closely as any brother or sister.

DNA

DNA stands for deoxyribonucleic acid. DNA contains the biological instructions that make each species unique. During reproduction, DNA is passed on from the parent generation to offspring. Almost all DNA is located in the nucleus of a cell and within a person nearly every cell has the same DNA. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases and their order (or sequence) determines the information available for building and maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences. More than 99% of those bases are the same in all humans.

Exome sequencing

A test which determines the sequence of the 1% of the genome which codes for proteins (the protein-coding parts of the genes, known as exons). Exome sequencing is cheaper than genome sequencing and produces a more manageable amount of data, but doesn't capture all types of genetic variation, for example larger insertions/deletions/rearrangements of DNA.

Expressivity

Some genetic diseases show very variable features in different people with the same disease, even within the same family. This is known as variable expressivity.

Genetic variant

A place in the genome where an individual's DNA sequence is different from the reference sequence. Some variants are very common (present in half of the population) while some are very rare (only ever seen in one family) – plus everything in between. Some variants are seen in people from every population around the world, while others are specific to a particular population or ethnic group.

Genome

A full set of all 20,000 genes, plus the DNA in between the genes. The genome is made up of 3 billion DNA 'letters', of which 3 to 5 million are variant in any individual (see genetic variant).

Genotyping

Doing a test which looks at multiple places where the genome sequence is known to vary between individuals. Genotyping only looks at sites of common genomic variation, so rare variation won't be picked up – sequencing is needed to find rare variants. Genotyping can be used to study common polygenic diseases (for example in genome-wide association studies), or to check how people are related to each other (for example in paternity testing).

Germline genome

The genome which is passed down from parents to children. Every cell in the body contains a single copy of that person's genome, but when cells divide new genetic changes can happen which affect the daughter cells, but aren't passed down to that person's children. The new genetic changes are known as somatic mutations.

Heterozygous variant

We all carry two copies of most of our 20,000 genes. A heterozygous variant is found in only one copy, while the other copy of the gene has the reference sequence.

Homozygous variant

We all carry two copies of most of our 20,000 genes. A homozygous variant is found in both copies of that gene.

Inheritance pattern

The way in which a disease is inherited in families. Autosomal dominant diseases only need one faulty copy of the gene to cause the disease; these pass down through the generations of a family (eg Huntington's disease). Autosomal recessive diseases need both copies of the gene to be faulty, one inherited from each parent (eg cystic fibrosis); these tend to happen in several siblings in the same generation, and are more common in consanguineous families. X-linked inheritance causes a different pattern in males and females in the family, with females usually being asymptomatic or mildly affected carriers, and males usually being more severely affected. Mitochondrial inheritance passes down the female line and tends to affect different family members with variable severity.

Library preparation

The laboratory process of preparing DNA for next-generation sequencing. A sequencing library contains millions of DNA fragments which have been altered so that they are recognised and processed correctly by the sequencing platform. Different platforms require specific library preparation techniques, which may lead to particular artefacts or omissions in the resulting sequence data.

Linkage

A method of studying how particular sections of DNA have been inherited together through a family. This helps to identify which regions of the genome are likely to contain the genetic variant which is causing the family's disease. Linkage studies in large families was the most successful way of finding disease genes until next-generation sequencing became available.

Modifiers

Genetic variants which change the way a monogenic disease affects an individual. Modifiers are likely to contribute to the penetrance and expressivity of many genetic diseases but at present are poorly understood due to the difficulty of achieving adequately powering studies.

Monogenic disease

A disease which is caused by a single genetic change which has a big impact on the patient's phenotype. Some diseases are purely monogenic, for example Huntington's disease, where there is 100% correlation between the gene change and the disease. Others are monogenic but have variable penetrance or expressivity, presumably caused by other modifying genetic or environmental factors.

Monozygotic twins

Genetically identical twins.

Mosaicism

If a genetic change happens after an embryo is made, but at an early stage of embryonic development, that change will be present in some of that individual's cells or body parts, but not in others. This is known as mosaicism. Mosaic conditions can be passed on to the next generation if the genetic change is present in the gonads (ovaries and testes).

Mutation

A genetic variant which has been shown to cause a particular disease, otherwise known as a pathogenic variant. The majority of genetic variants do not cause any medical problems. Note - the word mutation is avoided by clinicians discussing genetic results with patients; 'DNA change' or 'DNA variant' are preferred terms in this context.

Next generation sequencing

Also known as massively parallel sequencing; the Human Genome Project drove an international race to increase the pace and capacity of genome sequencing, which included many different developments in the chemistry and engineering related to DNA. The results have been highly successful, and the translation of these new technologies from research to clinical use has also occurred in a much shorter timeframe than would be traditionally expected. A new generation of DNA sequencing platforms are under development which will add further speed, power and accuracy to genome sequencing.

Oligogenic disease

A disease which is caused by a small number of genetic variants, probably together with environmental risk factors. These conditions often seem to run in families but no single causative genetic mutation can be found. The genetic basis of most oligogenic diseases is currently poorly understood.

Panel test

A test which determines the sequence of a specific set of genes which are known to cause a particular medical condition. Some panels are fairly small (only a handful of genes); others contain hundreds of genes, for example panels of genes affecting the retina of the eye have several hundred genes on them, because retinal conditions are very heterogeneous (they can be caused by mutations in many different genes) and it is not possible to tell which gene is the cause by looking at the patient's phenotype. A virtual panel refers to an analysis technique in which an exome or genome sequence is examined just for variants in a list of genes known to be relevant to the patient's condition; the gene selection occurs at the stage of data analysis and interpretation, not at the wet lab design of the test, as for the original panel tests.

Penetrance

Probability that a person with the disease-causing genotype or combination of genotypes will show clinical signs of the disease. For a fully penetrant disease, 100% of people with the genotype will have clinical features by a certain age. Diseases with incomplete penetrance can appear to skip generations.

Phenotype

The clinical features of a disease or condition, which result from a combination of genotype (the individual's genetic make-up) and environmental and/or lifestyle factors.

Polygenic or complex disease

A disease which is caused by an accumulation of common genetic variants which each have only a small effect on the patient's phenotype, together with environmental factors. Great progress was made in studying the pathogenesis of polygenic disorders using genome-wide association studies (GWAS), but the predictive power of genetic testing for any individual is still low for these disorders, and they have not been used widely in the clinical context.

Predictive test

Offering a genetic test to a healthy, asymptomatic person to give them information about whether they may be at risk of developing the specific genetic condition identified in their family at some time in the future.

Proband

Variably used to mean the first person in a family to be identified as being affected by a genetic condition, or the youngest or most seriously affected person in the family.

Reference sequence

Every human genome is unique; the reference sequence is a consensus sequence of the human genome which defines what is considered 'normal' at each position in the genome – although no-one with this exact sequence has ever or will ever exist.

Sanger sequencing

Also known as 'capillary sequencing'; the first automated process to determine the sequence of DNA fragments. Most of the Human Genome Project was completed using this technique. Sanger sequencing is very accurate, but it is not suitable for very large-scale sequencing due to cost, DNA input requirements and limitations on scalability. Before being used in the clinical diagnostic context, genetic variants found using next-generation sequencing are still checked using Sanger sequencing to ensure the technical validity of the result.

Segregation

A genetic variant segregates with a disease if the variant is present in all the people who have the disease, and absent in all the people who don't have the disease. This helps to establish whether the variant is causing the disease in that family. Segregation analysis can be confused by non-penetrance (where a family member has the genetic variant but doesn't have the disease) or by phenocopies (where a family member has a disease but doesn't have the variant; this is more common in conditions with a high population frequency e.g. in familial breast cancer families, there are likely to be relatives who have breast cancer by chance and not because of the familial condition).

Sequencing of DNA

Determining the order of the DNA 'letters' which make up the genome, like reading all the way through a very large book.

Simplex case

The only person in a family to have a particular medical condition.

Somatic mutations

See Germline genome.

Whole genome sequencing

A test which determines the order of all 3 billion DNA 'letters' in a particular individual.



Department of Health

Crown Copyright 2017.

Produced by APS Group for the Department of Health.

First published (online only) July 2017.

You may re-use this information (excluding logos and copyrighted material) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/ or email: psi@nationalarchives.gsi.gov.uk.

This publication is available for download at the Department of Health pages of the .gov.uk website: <https://www.gov.uk/government/organisations/department-of-health>

Please cite as:

Davies, S.C. "Annual Report of the Chief Medical Officer 2016, Generation Genome *London: Department of Health (2017)*