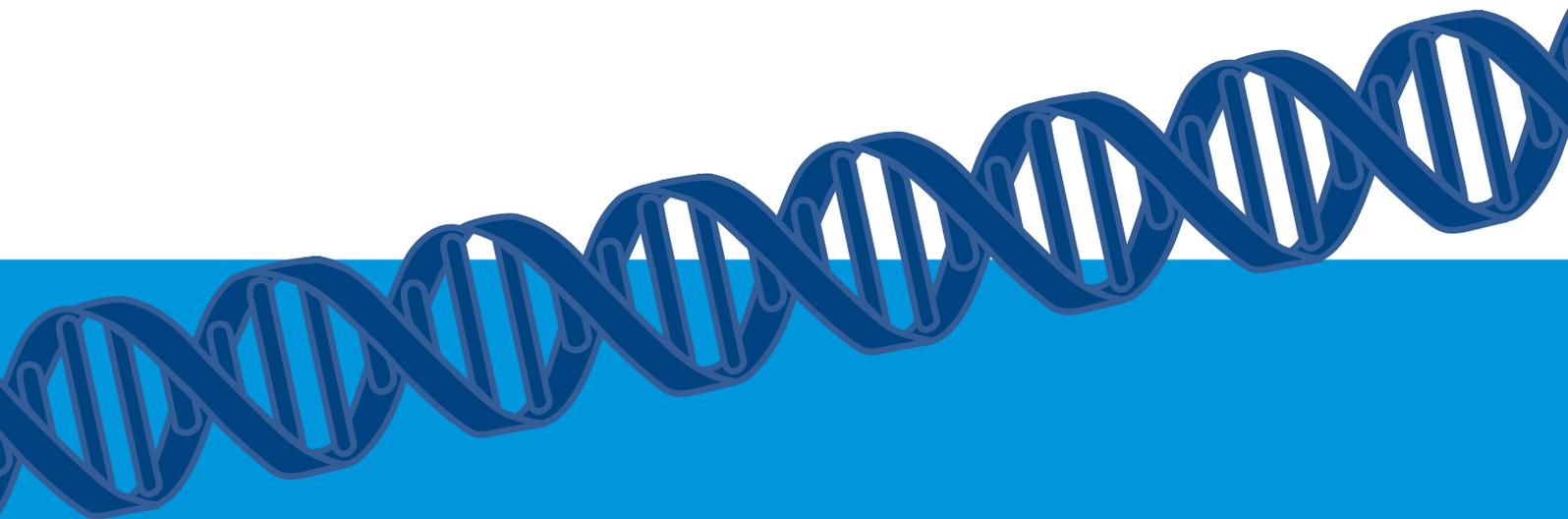


# Clinical genome analysis evidence review

Update 2020



## Authors:

Laura Blackburn, Chantal Babb de Villiers, Joanna Janus, Emma Johnson, Mark Kroese

## Acknowledgements:

The authors would like to thank Karina O'Rourke and Sian Morgan, authors of sections 2.1 and 2.3 respectively.

The PHG Foundation is grateful for the insight provided by the individuals consulted during the course of this evidence synthesis. Full acknowledgements are listed in Appendix 1.

This review was commissioned and funded by the National Specialist and Screening Services Directorate, NHS Services Scotland

NB: URLs in this report were correct as of June 2020

## Published by PHG Foundation

2 Worts Causeway  
Cambridge  
CB1 8RN  
UK

Tel: +44 (0)1223 761 900

Report produced by PHG Foundation on behalf of NHS Scotland.  
Copyright NHS Scotland, under licence from PHG Foundation.

How to reference this report:

**Clinical genome analysis evidence review: update 2020**  
**PHG Foundation (2020)**

This report is the result of PHG Foundation's independent research and analysis and is not linked to a third party in any way. PHG Foundation has provided occasional analytical services to Oxford Nanopore Technologies (ONT) as part of a consultancy agreement.

The PHG Foundation is a health policy think-tank and linked exempt charity of the University of Cambridge. We work to achieve better health through the responsible and evidence based application of biomedical science. We are a registered company, no. 5823194

# Contents

<b>Executive summary</b>	<b>6</b>
Context	6
The UK genomics policy landscape	6
Summary of findings	7
The future	9
Conclusions	9
<b>1 Introduction</b>	<b>11</b>
1.1 Background	11
1.2 Objectives	11
1.3 Structure of the report	12
1.4 Methodology	12
<b>2 The UK genomics policy landscape</b>	<b>13</b>
2.1 Genomics in Scotland	13
2.2 Genomics in England	18
2.3 Genomics in Wales	23
<b>3 Clinical genome analysis</b>	<b>28</b>
3.1 Uses of clinical genome analysis	28
<b>4 Genome analysis methods</b>	<b>31</b>
4.1 The genome analysis spectrum	31
4.2 The evolution of genome analysis methods	33
<b>5 Genome analysis by non-sequencing methods</b>	<b>35</b>
5.1 Karyotyping	35
5.2 Fluorescent <i>in situ</i> hybridisation (FISH)	37
5.3 Microarrays (Array CGH and SNP arrays)	39
5.4 Polymerase chain reaction (PCR)	43

<b>6</b>	<b>Genome analysis by DNA sequencing</b>	<b>48</b>
6.1	The sequencing pipeline	49
6.2	Sample and library preparation	49
6.3	DNA sequencing	51
6.4	Data interpretation	63
6.5	Associated technologies and techniques	69
<b>7</b>	<b>Factors that determine choice of genome analysis methods</b>	<b>72</b>
7.1	Clinical exome sequencing (CES)	72
7.2	Whole exome sequencing (WES)	73
7.3	Whole genome sequencing (WGS)	73
7.4	Comparison of clinical exome, whole exome and whole genome sequencing	74
7.5	Gene panels and targeted sequencing	77
7.6	Trio analysis	78
7.7	Supporting clinicians in the use and understanding of genetic information	81
7.8	Considerations around the identification and interpretation of variants of unknown significance (VUSs)	82
7.9	Retention, resequencing and reinterpretation of data and samples	83
7.10	Additional looked for findings	84
<b>8</b>	<b>Genetic testing through the lifecourse</b>	<b>86</b>
8.1	Genetic testing in the prenatal period	86
8.2	Genetic carrier screening and testing in the prenatal period	87
8.3	Pre-implantation genetic diagnosis	89
8.4	Prenatal genetic screening tests	90
8.5	Prenatal genetic diagnostic tests	93
8.6	Future developments in prenatal screening and diagnosis	96
8.7	Conclusions for prenatal genomic analysis	97
8.8	Genetic testing in the postnatal period	97
8.9	Clinical utility of genome analysis in the postnatal period	98
8.10	Genomic analysis of newborns	99
8.11	Genomic analysis during childhood	101
8.12	Reasons for undertaking genomic analysis in adolescence	102
8.13	Reasons for undertaking genomic analysis in adulthood	103
8.14	Reasons for undertaking genomic analysis postmortem	103
8.15	Future opportunities for genomic analysis not confined to lifecourse stage	104
8.16	Factors determining current approaches to postnatal constitutional genomic analysis	106
8.17	Impact of rapid developments in genomic technology and knowledge on postnatal constitutional genomic analysis	107

<b>9</b>	<b>Genome analysis for cancer management</b>	<b>111</b>
9.1	Current approaches to germline testing for inherited cancer susceptibility	112
9.2	Current approaches to diagnostic and prognostic somatic tumour genome analysis	114
9.3	Technical challenges to effective genomic analysis in cancer	117
9.4	Samples for solid tumours: FFPE, fresh tissue and liquid biopsy	119
9.5	NGS-led approach – use of WGS and panels	120
9.6	The future	122
9.7	Conclusions	123
<b>10</b>	<b>The future clinical genomics landscape</b>	<b>125</b>
10.1	Technological developments in sequencing technologies	125
10.2	Applications of genomic technologies	127
10.3	Wider implications for the health system	131
<b>11</b>	<b>Conclusions</b>	<b>133</b>
<b>12</b>	<b>Appendix</b>	<b>135</b>
12.1	Appendix 1: Acknowledgements	135
12.2	Appendix 2: Glossary of key terms	136
<b>13</b>	<b>References</b>	<b>140</b>

# Executive summary

## Context

In 2016 PHG Foundation published a report for NHS National Services Scotland which contained a literature review in the field of genetics and molecular pathology. This was to inform a review by the National Services Division (NSD) on the provision of clinical genetics laboratory and molecular pathology services in Scotland. In the time period since the 2016 review, there have been significant developments not only in genomics technologies, but also in the clinical genomics infrastructure in the UK home nations. This document is an update on developments in genomics technologies and their applications since the 2016 review, to inform the development of the long-term strategy for clinical genomics in Scotland.

This report contains:

- » A review of developments in the UK clinical genomics policy landscape since 2016
- » A description of current and emerging DNA-based testing technologies and an analysis of their likely contribution to diagnostic testing over the next three years

## The UK genomics policy landscape

### *Scotland*

The four genetic laboratories in Scotland belong to a formal consortium, the Scottish Genetic Laboratory Consortium (SGLC), and work collaboratively to deliver and support decision making on the provision of genetic testing. The Scottish Genomes Partnership – a group of universities, NSD and SGLC – researches the use of whole genome sequencing for clinical benefit. The bridge to a Scottish strategy for genomics currently provides funding to support evidence gathering needed to inform implementation in rare diseases. A pharmacogenomics pilot is also ongoing. The Scottish Genomics Leadership Group (GLG) is overseeing the development of a future strategy to maximise the benefits of genomic medicine to Scottish patients while aligning with the wider UK vision.

### *England*

The 100,000 Genomes Project, launched with the aim of harnessing whole genome sequencing technology to inform improved diagnoses and management for participants with rare disease and cancer, has laid the foundation for the introduction of whole genome sequencing as part of the NHS National Genomic Medicine Service, launched in October 2018. Genomic testing is being delivered through a network of seven genomic laboratory hubs (GLHs), supported by a national directory of genomic tests. Seven genomic medicine service alliances are being developed to support research and collaboration across the geographical area of each GLH.

Genomics strategy is overseen by the national genomics board, which has the goal of realising the implementation of the recommendations in Generation Genome (Chief Medical Officer for England's annual report, 2016) [1] and the Life Sciences Industrial Sector Deal [2].

## Wales

The Welsh Government's genomics for precision medicine strategy sets out plans to utilise genomics for the benefit of Welsh patients, and the Partneriaeth Genomeg Cymru – Genomics Partnership Wales (GPW) was established in 2018 to support a collaborative approach to genomics in Wales across a range of stakeholders. The aim of the Welsh national programme is to ensure that genomics enters mainstream medicine across health and care, and is aligned with wider Welsh and UK genomics strategy. Investment has driven the expansion of testing provision – for example, to include non-invasive pre-natal testing and circulating tumour DNA testing – as well as supportive digital infrastructure. Research and innovation projects with higher educational institutions and other collaborators are supporting efforts to integrate research into healthcare in a timely manner, and also contributing to education, training and engagement around genomics.

## Summary of findings

There are a number of ways in which clinicians can use the analysis of genomic variation in individuals to manage their health:

- » **Diagnosis/classification of existing disease** – including newborn screening and tumour testing
- » **Predictive/pre-symptomatic disease risk testing** – including adult inherited cancer mutation testing and carrier testing
- » **Antenatal testing** – including non-invasive prenatal testing
- » **Pharmacogenomics** – to guide appropriate treatment

There are a wide range of genome analysis methods available and the choice of method will vary by:

- » The type of genomic variation being analysed
- » The resolution of the analysis required
- » Time sensitivity of the application

There are a wide range of methods available, some of which have been in use for decades and are still the 'method of choice' for certain clinical applications and to measure specific genomic variants due to their ease of use, reliability and cost. These include non-sequencing methods utilising cytogenetic-, microarray- and PCR-based techniques that can detect a range of genomic variation from chromosomal rearrangements to single nucleotide alterations and are still in widespread use providing focused analysis of genomic variation for specific clinical indications. In addition, there are methods to analyse other molecules related to the regulation and expression of the genome.

There are situations where sequencing is a more appropriate approach, for example if clinical presentation is complex and does not indicate an obvious choice of targeted assay, or if there are multiple variations/variation spread over large areas of sequence to be investigated. Developments in sequencing technologies since the 2016 review have seen improvements in the reliability of techniques, reduction in sequencing costs and improvements in analytical pipelines. While Sanger sequencing, an older and lower throughput method, is used for clinical confirmation of single genes or small sections of sequence, higher-throughput next generation technologies are used for examining large parts of or whole genomes. The latest long-read sequencing technologies which offer portable and cheaper sequencing are not ready for clinical application but are likely to have an impact in the future.

Despite many advances in sequencing technologies there is not one assay currently that meets all the requirements for clinical genomics testing. While an increase in demand is expected for the sequencing of clinical exomes, whole exomes and whole genomes, many longer established approaches, such as karyotyping and PCR techniques, remain part of the genomic testing toolkit.

### *Application: testing through the life-course*

Genetic testing is used throughout the life course from pre-implantation diagnosis, through diagnosis of genetic disease in children and adults, to post-mortem analysis. Demand for genetic testing is increasing year on year; in particular, there have been significant changes to the use of sequencing technologies to manage rare and inherited diseases due to initiatives such as the 100,000 Genomes Project. As knowledge about the genetic causes of disease increase, further expansion of testing provision is expected that utilises both non-sequencing and sequencing methods; for example, the use of larger gene panels is becoming more common. While whole exome sequencing (WES) and whole genome sequencing (WGS) are demonstrating clinical utility for example in terms of supporting diagnosis of critically ill babies and infants, overall the picture is still one of varied testing provision.

### *Application: testing in cancer*

For cancer applications, testing is used in the contexts of risk management – identification of those with inherited cancer risk – precision diagnosis and prognosis, and personalised treatment management. A wide range of non-sequencing and sequencing approaches are used, including targeted single gene tests and panels. It is anticipated that WES/WGS could become more widely used as the number of potentially actionable targets increases. One major development is the implementation of circulating tumour DNA (ctDNA) testing as a companion diagnostic. Further implementation of this approach is expected for other companion diagnostic tests.

## The future

Further technological advances in short read sequencing will continue; however, the fastest developments are likely to be seen in the improved accuracy of long-read sequencing platforms. With this technology, validation of clinical tests when the technology is evolving rapidly could create challenges.

In terms of testing demand and provision, expansion of current testing services and improved access to meet current unmet needs can be expected. In addition, further demand can be expected through use of non-invasive prenatal testing (NIPT), potential WGS of cell free fetal DNA, expanded post-natal analysis including through existing newborn screening programmes, and reactive pharmacogenomic testing. In cancer, ctDNA technologies continue to develop and further uses such as residual disease monitoring could be expected.

One area where there is currently little activity, but in which considerable increase is likely, is through the testing of healthy individuals. This is most likely to occur through the use of polygenic scores as a tool to predict individuals' risk of disease, and through pre-emptive pharmacogenomic testing, where a person's pharmacogenomic profile is determined while they are still healthy, or when they present to the health system for another reason.

Wider implications for the health system – due to more extensive use of testing (particularly more WES and WGS) are centred around:

- » **Infrastructure and provision of testing** – in terms of availability of laboratory space, equipment, staffing and expertise
- » **Data storage and analysis** – including longer term storage and data formats that remain usable over long periods of time
- » **Re-contacting patients with new information** – as genetic knowledge develops, or when people tested as children come of age
- » **Downstream effects on clinical pathways** – impact on other clinical services of increased diagnostic rates
- » **Links between clinical practice and research** – optimising input of patients and clinicians in research, and realising the benefits of research for the health system.

## Conclusions

A wide range of sequencing and non-sequencing methods will still be used and there is not 'one method' to cover all testing needs – the tipping point towards more universal WES/WGS is still unclear, and has not yet been reached. Therefore, retaining capacity for existing methods is necessary.

Laboratory capacity is increasing, particularly as panels get larger and more WES/WGS is used, and new methods are brought online e.g. ctDNA testing. However, the scale of genomic data being produced brings new challenges, such as issues around longer-term data storage and analysis.

Technological developments in genomics, along with our evolving knowledge of the relationship between genes and disease, will continue to facilitate more accurate, sophisticated and cost-effective clinical genome analysis in the next three years. Sequencing technologies have already been increasingly harnessed by the NHS to improve healthcare in a range of disease areas, in particular in the diagnosis of rare disease and the management of cancer. In order to support this ongoing progress, it is vital that health systems are able to rapidly exploit advances in genomic analysis and research as they emerge, both from the biomedical research sector and the collaborative healthcare research ecosystem. This will ensure that they can implement the most appropriate genomic laboratory approaches in a timely manner.

If the health system is able to harness current and future genomic technologies to their full potential, it will be able to meet expected increased testing volume and demand, through the provision of equitable services for previously unmet clinical needs and improved quality of care to NHS patients and their families.

# 1 Introduction

## 1.1 Background

In 2015 NHS National Services Scotland (NSS) commissioned the PHG Foundation (PHGF) to undertake a literature review in the field of genetics and molecular pathology. This report, published in 2016, was to inform a review by the National Services Division (NSD) on the provision of clinical genetics laboratory and molecular pathology services in Scotland, as part of a five-year plan for the service.

The 2016 PHGF report contained:

- » A description of current and emerging DNA-based testing technologies and analysis of their likely contribution to diagnostic testing over the next five years
- » Identification and description of methodologies available to establish the need for genetic testing and for predicting future diagnostic genetic test activity
- » Description of available evidence on the cost-effectiveness of diagnostic genetic testing in the areas of cancer and inherited diseases

In the time period since the 2016 review there have been significant developments, not only in genomic technologies and policy, but also in the clinical genomics infrastructure in the UK home nations, for example through the 100,000 Genomes Project and the launch of the NHS Genomic Medicine Service in England.

Following the publication of the NSD report and its recommendations, a number of reviews and initiatives have been undertaken in Scotland, including a bridging strategy to maintain momentum in the delivery of genomic medicine and allowing the health system to gain experience and evidence to support the development of a long-term Scottish strategy for genomics.

## 1.2 Objectives

The key objectives of this review are to update the PHGF literature review of genetic technologies published in 2016 by:

- » Describing current and emerging DNA-based testing technologies and analyse their likely contribution to diagnostic testing over the next three years
- » Reviewing developments in the UK clinical genomics policy landscape since 2016

Cost-effectiveness evidence review of genomic testing is not included in this review since it is being covered by extensive research elsewhere.

### 1.3 Structure of the report

We begin the report by describing advances in the UK genomics landscape since 2016, with updates from three of the four UK home nations. Next we describe the technologies currently being used for clinical genome analysis and their relative strengths and weaknesses. This review provides context to the considerations and decision making around the future configuration of genomic laboratory service provision and what the impact might be on clinical services of the emergence of new technologies for clinical genome analysis.

We then review how these technologies are used for the purposes of clinical genome analysis across the patient life course of inherited diseases and cancer. This includes a description of the purpose of clinical testing, which technologies and assays are employed, and how this use is likely to change over the next three years. We also include descriptions of genomic assays that are currently used mainly in research settings but which are sufficiently well developed to warrant consideration of their validity and utility for clinical application within the next three years. All the technologies and applications in this report have been reviewed from a clinical grade testing perspective.

The final section of the report briefly explores how the clinical genomics landscape might change in the next five to ten years, including a short description of developments in genomics technologies and what the implications of these developments might be for clinical genomics services.

### 1.4 Methodology

This review and analysis are informed by our assessment of peer-reviewed literature and other public sources of information. We have also drawn on consultations with a number of NHS clinical scientists who currently deliver various types of clinical genome analysis. These contributors are acknowledged in Appendix 1.

## 2 The UK genomics policy landscape

Since the publication of the 2016 Clinical genome analysis evidence review, there have been significant developments in the landscape of genomics policy in the UK, and subsequently in infrastructure to support the delivery of genomic medicine.

### 2.1 Genomics in Scotland

**Author: Karina O'Rourke, National Services Division, NHS Scotland**

#### *Background*

NHS National Services Scotland (NSS) National Services Division (NSD) commissioned the four genetic laboratories in Scotland which are based within the Regional Genetic Centres in Aberdeen, Dundee, Edinburgh and Glasgow. The Laboratories work as a formal consortium, the Scottish Genetic Laboratory Consortium (SGLC), enabling collaborative discussions and decision making on the provision of genetic testing in Scotland [\[3\]](#).

Molecular genetic testing was nationally designated in 1985 and cytogenetics in 2009. Molecular pathology testing was nationally commissioned as a single designated multi-site national specialist service from 1 April 2013.

Following the 2016 review of the Nationally Designated Genetic Laboratory Testing Services, the National Specialist Services Committee (NSSC) confirmed that the current Consortium service model for the provision of genomic testing would retain its national designation. The review recommended a revised governance structure including the establishment of a National Genetics Laboratory Management Committee (GLMC) and the appointment of three designated national leads for Clinical Genetics, Laboratory Science and Molecular Pathology to drive forward the strategic development of all genetic testing in NHS Scotland. Members of the GLMC and national leads are committed to leading on the strategic planning and decision making from an all Scotland perspective in line with the NHS National Services Scotland (NSS) Strategy 2019-2024 to 'provide national solutions to improve the health and wellbeing of the people of Scotland' [\[4\]](#).

A key strength of the Consortium has been providing a forum for clinicians, scientists and commissioners to identify, evaluate and implement the most clinically and cost-effective approach to genetic testing consistently across Scotland. Another benefit of consortium working is that it provides accessible genetic, scientific and clinical expertise on a national level directly benefiting patient care through this interaction. Operating over four sites, the Consortium also offers in-built contingency and resilience to protect the delivery of genetic and molecular pathology services.

## *The Scottish Genomes Partnership (SGP)*

SGP was established in 2016 and is a major Scotland-wide research programme between the Universities of Edinburgh, Glasgow, Aberdeen and Dundee, with NSD and SGLC. It was funded by the Scottish Government's Chief Scientist Office and the UK's Medical Research Council to undertake research on the use of genome sequencing for medical benefit. Research by SGP included a collaboration with Genomics England on the 100,000 Genomes Project [\[5\]](#), which aimed to sequence 100,000 genomes from UK families, including those affected by rare conditions, in order to provide a diagnosis for some patients with rare diseases and develop future NHS healthcare services.

One of the aims of the SGP research project was to provide evidence to inform the future delivery of NHSS genetics services and to ensure that learning from genome-based science can be applied within NHS clinical services in Scotland.

Recruitment to the 100,000 Genomes Project in Scotland (and the rest of the UK) concluded in February 2020 with the clinical results of Whole Genome Sequencing (WGS) being returned to NHS Scotland. Health economics data is being gathered by SGP to support this evidence base and help inform future health service decision-making. The SGP launched a project which will evaluate the effectiveness of WGS as an intervention within NHS Scotland clinical genetic services. Working in collaboration with the Health Economics Research Unit (HERU) at the University of Aberdeen, the analysis will assess the costs and benefits of WGS for the diagnosis of rare disorders, in comparison to other potential health care interventions. The evaluation will form part of the basis of recommendations about the use of WGS within the Scottish NHS.

## *Genomics for Rare Disease in NHS Scotland: The bridge to a Scottish Strategy for Genomics*

The Scottish Government included a commitment in the Programme for Government 2018/19 to continue the development of genomic medicine through the enhancement of NHS Scotland genetic capabilities for the diagnosis of rare diseases. This has allowed the SGLC to build on practice and experience gained in genomic technologies; including the continued development of genomic tests, ongoing collaboration with the SGP as well as improved data analysis, sharing and storage.

The allocation of £4.2 million has supported the continued momentum of genomic work in Scotland and has acted as a 'bridge' to provide sufficient evidence to inform a strategy for the implementation of genomic testing in Scotland. This bridge funding only provides support for inherited rare disease (germline) testing and does not currently cover testing for non-inherited (somatic) cancer which is a major and rapidly increasing workload for the Consortium.

The bridge funding made available by Scottish Government allowed time for the robust assessment of genomic testing approaches in an NHSS setting. It was anticipated that the experience and evidence gained through the application of the two-year bridge funding will be considered in the broader context of the accumulating evidence in the field of genomics and will clearly provide evidence for the value of genomic testing for Scottish patients.

A stepwise strategy has been adopted for the implementation of genomic testing into mainstream medicine for Scotland. This strategy links several initiatives whose progress is being regularly evaluated by the Consortium to inform further strategic planning and ensure cost-effective delivery of the most appropriate testing methodology within routine service for optimum patient care.

The stepwise approach under the bridge to a Scottish strategy for genomics scheme of work includes the following work-streams:

- » Testing using clinical exome sequencing (CES) and targeted gene panels has been rolled out across the SGLC effectively creating specialist testing centres according to the disease type (see table 1: Rare Disease Allocation by Centre)

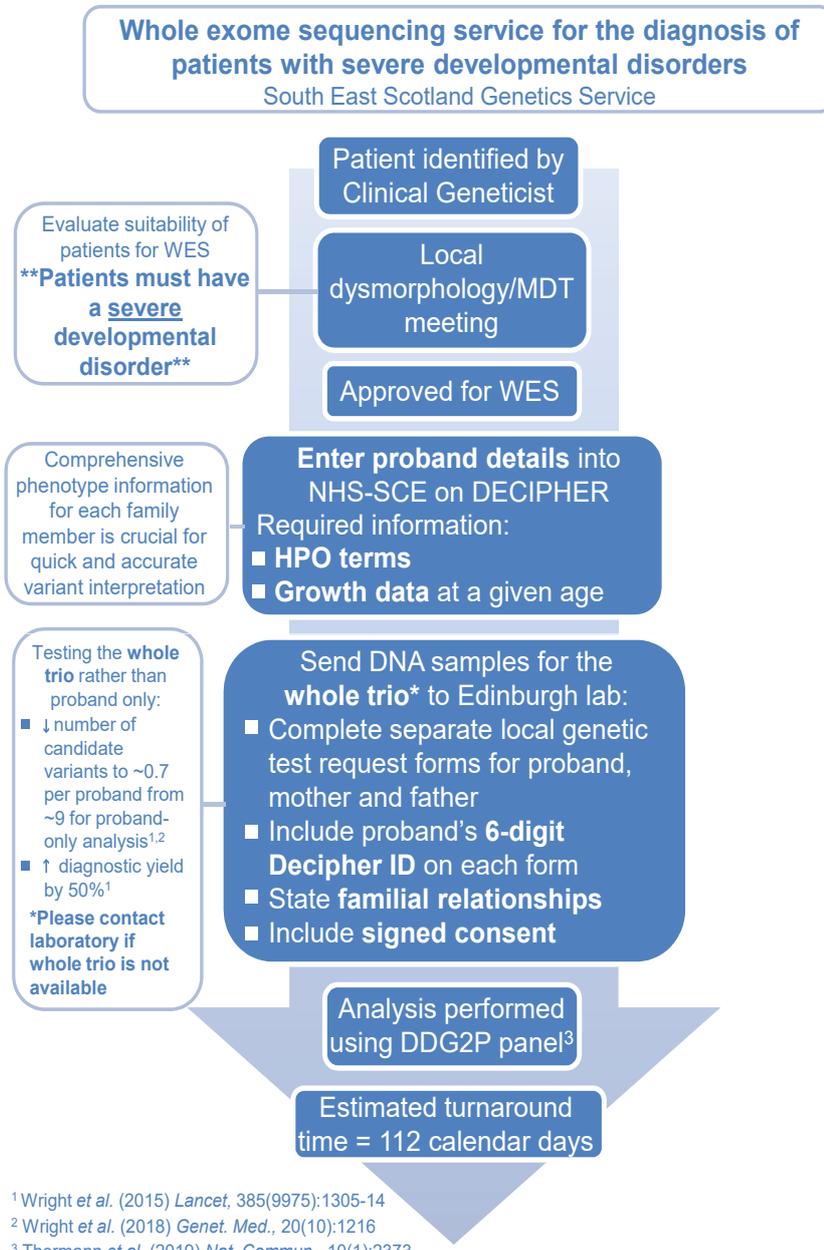
**Table 1: Rare disease allocation by centre**

<b>NHS GG&amp;C</b>	<b>NHS Grampian</b>	<b>NHS Lothian</b>	<b>NHS Tayside</b>
Inherited cancer	Cardiology	Connective tissue	Endocrinology
Musculoskeletal	Chromosome Breakage	Haematology	Hearing Loss
Neurology	Eye Disorders	Musculoskeletal	Neurology
Respiratory	Gastrohepatology	Neurology	Renal
	Haematology	Dev Delay	
	Metabolic		

- » Trio-based whole exome sequencing (WES) and analysis for patients and their families across Scotland has been provided by NHS Lothian Molecular Genetics Laboratory since September 2019. This builds on the findings and processes developed through the highly successful UK-wide Deciphering Developmental Disorders (DDD) project (see figure 1: WES Referrals Workflow).
- » All processes developed for SGP have continued to be used as we extend into the second phase of work within the bridge funding ('SGP2'), which includes sequencing and analysis of a further 500 whole genome sequences and using the cutting-edge technique of long-range sequencing.
- » The SGLC and NSD are working with NES colleagues, tasked to implement key elements of the 2018 Digital Health & Care Strategy on Genomics opportunities within the National Digital Platform (NDP) to establish an expandable Cross-Scotland secure genomic data repository within the NHS, to enable data sharing and shared analysis.

As well as successfully integrating CES into routine service delivery, the SGLC is also keeping pace with the emergence of other new areas. For example; colleagues at the NHS Tayside laboratory have been working in collaboration with NSS Pregnancy & New Born Screening Programmes with a view to implement complementary testing strategies future proofing the delivery of Non-Invasive Prenatal Testing (NIPT) testing services in Scotland utilising a next generation sequencing platform which is expected to be available from summer 2020.

Figure 1: WES referrals workflow



### Cancer genomics

Next generation sequencing (NGS) panels for cancer are already being delivered in the NHS in Scotland for some cancers such as melanoma, lung and colorectal cancer. These panels are targeted and enable the testing of approximately 20 different genetic biomarkers or genes using a single assay and a unified workflow.

A number notable of areas in cancer testing are being pursued collaboratively across the consortium, these include:

- » Mainstreaming testing, e.g. colorectal cancer / Lynch Syndrome and BRCA testing in the germline and somatic setting
- » Ongoing development of more comprehensive NGS panels, e.g. lymphoid NGS panel and fusion panel/s
- » Glasgow Precision Oncology Laboratory (GPOL) UK Cancer Panel Pilot
- » Companion diagnostics for tumour-agnostic indications / treatments
- » Increased application of circulating tumour (ct) DNA testing for disease monitoring

### *NHSS National Planning Board*

A new NHSS National Planning Board was established in 2018 to provide oversight, governance and decision making in relation to national planning of NHSS services and identify priority areas over the coming 5-10 years.

A short life planning group pulling together expertise from across cancer specialties was established in 2019 to undertake a horizon scan and consider the planning needs in detail. The group reported to the NHSS National Planning Board in February 2020, and highlighted that there is a rapidly evolving diagnostic and treatment environment with major advances in precision medicine, including genomics and companion diagnostics to inform treatment selection for patients. It was agreed by the Board that, in addition to work around ATMPs, workforce, surgery (including robotic surgery) and multi-disciplinary/rapid diagnostic centres, the National Planning Team should lead on the development of a plan for precision medicine for NHSS by creation of a NHSS Precision Medicine Steering Group, which will link to the SGLC and newly established Scottish Pharmacogenomics (PGx) Working Group.

### *Pharmacogenomics*

In the area of pharmacogenomics, all four consortium laboratories have undertaken a pilot that tests for common clinically important variants in the DPYD gene that encodes for the liver enzyme dihydropyrimidine dehydrogenase (DPD). A deficiency in the DPD enzyme could make the side effects of certain chemotherapy drugs (fluoropyrimidines), commonly used in the treatment of colorectal and breast tumours, much worse. For patients who have one of the common DPYD gene variants, these side effects can be particularly severe or even be life threatening therefore testing is of great importance to inform treatment decisions.

DPYD testing is just one example of a growing number of applications for pharmacogenomic testing. NSD is working towards integration of a pharmacogenomics (PGx) workstream under the Consortium governance structure to ensure that the same level of scrutiny is applied to appraise the evidence based on clinical validity, analytical validity and clinical utility of testing before being approved to be provided on a national basis.

## *Scottish Genomics Leadership Group*

The Scottish Scientific Advisory Council (SSAC), tasked by Scottish Government to review the planning and investment required to implement genomic medicine in Scotland, has produced a report (Informing the Future of Genomic Medicine in Scotland [\[6\]](#)) that contains a number of key recommendations, the first being the establishment of a Scottish Genomics Leadership Group. This group is tasked with developing a strategy that will deliver on the SSAC recommendations to maximise the benefit of genomic medicine to Scottish patients whilst aligning with the wider UK vision.

Upon the recommendation of the Scottish Science Advisory Council (SSAC) report entitled Informing the Future of Genomic Medicine in Scotland, the Scottish Government has convened the Scottish Genomics Leadership Group (SGLG), chaired by Chief Scientist Prof David Crossman.

As well as establishing the Scottish Genomics Leadership Group (SGLG), the report set out five other broad recommendations:

1. Clinical implementation - Scottish Government should support NHSS to expedite evaluation and adoption of genomic testing where there is good evidence of improved patient outcomes.
2. Workforce - NHSS working with NHS Education for Scotland, ScotGen, Skills Development and academic institutions should lead in co-ordinating the development and delivery of the training courses and educational resources required to develop essential expertise and drive genomic medicine capabilities.
3. Digital health - Scottish Government should take account of the digital infrastructure needed to enable genomics within clinical pathways and to support the use of genomic data.
4. Research and innovation - Scottish Government, working with research funders, industry, enterprise agencies and higher education providers, should consider how best to support genomic research and innovation.
5. Industry-facing activity - Scottish Government, working with enterprise agencies, should build on the strengths of Scotland's 'triple helix partnership' between academia, the NHS and industry to accelerate genomics development as an integral part of precision medicine.

The role of the SGLG is to provide leadership for the further development of medical genomics in Scotland with the aim of producing a draft Scottish Government strategy and action plan for consideration by the Scottish Government Health and Social Care Management Board and the Cabinet Secretary for Health and Sport. The strategy and action plan will take forward the recommendations of the SSAC medical genomics report, defining the opportunities for Scotland.

Members of the SGLG have been working with colleagues in the Office for Life Science (OLS), who are compiling a UK National Genomic Healthcare Strategy for presentation to the National Genomics Board. The strategy is intended to facilitate collaborative working across the UK to help make the whole greater than the sum of the parts. The vision of the

Strategy is to make the UK the global leader in genomics and home to the most advanced genomic healthcare in the world, with the NHS, researchers and industry collaborating to benefit patients, the public and the UK economy.

## 2.2 Genomics in England

### *Background*

In 2012, the then-Prime Minister David Cameron launched the 100,000 Genomes Project. The goal of the Project was to harness whole genome sequencing (WGS) technology to uncover new diagnoses and improved treatments for participants with rare inherited diseases and cancer, with a view to making the UK a world leader in genomics within five years. Genomics England was established in 2013 by the Department of Health and Social care to deliver the Project.

To support the implementation of the Project, NHS England created 13 NHS Genomic Medicine Centres (GMCs), and Genomics England procured the use of a state-of-the-art sequencing centre run by Illumina, Inc. and developed an automated analytics platform to return whole genome analyses to the NHS.

The first participants were recruited in 2014 – with the first participant diagnoses returned in 2015. From 2017, the first users accessed their data and by December 2018 100,000 genomes had been sequenced; currently, all results are in the process of being returned to clinicians. This Project has laid the foundations for the introduction of whole genome sequencing as part of the NHS Genomic Medicine Service, which will provide equitable access to genomic testing to patients across the NHS.

The systematic application of cutting-edge genomic technologies has the potential to transform patient's lives by:

- » Enabling a quicker diagnosis for patients with a rare disease, rather than years of uncertainty, often referred to in rare disease as the 'diagnostic odyssey'
- » Matching people to the most effective medications and interventions, reducing the likelihood of an adverse drug reaction
- » Increasing the number of people surviving cancer each year because of more accurate and early diagnosis and more effective use of therapies

Genomics England also launched the Clinical Interpretation Partnerships (GeCIPs) in June 2014. Researchers and clinicians, from both academia and the NHS, are collaborating via the GeCIPs to analyse data from the 100,000 Genomes Project, to inform particular disease-focused or other topic areas, also known as 'domains'. There are currently 42 active GeCIP domains.

### *The NHS Genomic Medicine Service in England*

Plans to commission and embed genomic medicine into routine care pathways in England began in 2016. In March 2017, the NHS England Board set out its strategic approach to

create a National Genomic Medicine Service (GMS) with the commitment outlined by NHS England in Next Steps on the NHS Five Year Forward View [\[7\]](#). The NHS GMS officially launched in October 2018, and comprises five key elements:

### **1. A national genomic laboratory service provided through a network of Genomic Laboratory Hubs**

Since launch, genomic testing in the NHS has been provided through a single national testing network, consolidating and enhancing the existing laboratory provision. An NHS England procurement exercise was begun in November 2017 to centralise genomic testing throughout England into a network of seven Genomic Laboratory Hubs (GLHs). Each GLH is responsible for coordinating services for their region and for delivering a standard set of genomic tests which are outlined in the National Genomic Test Directory.

The seven GLHs are:

- » East Midlands and East of England GLH led by Cambridge University Hospitals NHS Foundation Trust
- » London North GLH led by Great Ormond Street Hospital for Children NHS Foundation Trust
- » London South GLH led by Guy's and St Thomas' NHS Foundation Trust
- » North West GLH led by Manchester University NHS Foundation Trust
- » South West GLH led by North Bristol NHS Trust
- » Wessex, Oxford and West Midlands GLH led by Birmingham Women's and Children's NHS Foundation Trust
- » Yorkshire and North East GLH led by Newcastle upon Tyne Hospitals NHS Foundation Trust

### **2. A new National Genomic Test Directory to underpin the genomic laboratory network**

The National Genomic Test Directory has replaced the NHS Directory of Genetic Disorders/Genes for Diagnostic Testing developed and maintained by the UK Genetic Testing Network (UKGTN). The UKGTN was established in 2001/2 as a national advisory organisation for the commissioning of NHS clinical genetic testing services. UKGTN evaluated and recommended genetic tests to be added to the Directory for rare and inherited disorders for NHS provision across the UK. The new Genomic Test Directory was initially generated by combining the UKGTN and a review of the cancer genomic testing that was already being performed across the country. The Directory specifies which genomic tests are to be commissioned by the NHS in England, the technology by which they are available, and the patients who will be eligible for a test. There are currently two parts to the Directory: one for rare and inherited disorders, and the other for cancer. Pharmacogenomic tests are expected to be included in the Directory in the near future; evidence gathering is currently underway to inform which gene-drug pairs are suitable for clinical implementation. Over time, as evidence develops, the Test Directory is also expected to include other functional genomic tests, for example RNA based technologies and proteomics.

The Genomic Test Directory will be updated on an annual basis by NHS England, supported by input from clinical and scientific expert panels covering rare disease, cancer and pharmacogenomics and with input from experts in the devolved nations. The 2019/2020 version was released in August 2018 and updated in March 2019.

### **3. A national whole genomic sequencing provision and supporting informatics infrastructure, developed in partnership with Genomics England**

In January 2019 the NHS Long Term plan was published; it included the commitment that from 2019, all seriously ill children in whom a rare genetic disorder is suspected, all children with cancer, and adults suffering from certain rare conditions or specific cancers, will be offered whole genome sequencing (WGS) by the new NHS Genomic Medicine Service [\[8\]](#). Under current plans there are 22 clinical indications in rare disease that will be eligible for WGS launching in April 2020. Sequencing will be provided by Illumina, Inc.

In addition, rapid whole exome sequencing to support the diagnosis of critically ill babies and children is being provided by the South West GLH from January 2020 [\[9\]](#). It is expected that up to 700 babies and children will benefit from this testing each year.

### **4. A Clinical Genomics Medicine Service and NHS Genomic Medicine Service Alliances**

Clinical genetics services will run in parallel with genomic laboratory hubs and will incorporate clinical genetic consultants and genomic counsellors.

NHS Genomic Medicine Centres (GMCs) played a central role in the success of the 100,000 Genomes Project. It was clear from this work that co-ordination and strong NHS partnerships across large geographical areas were key success criteria in supporting both the clinical utility of whole genome sequencing and creating new pathways of care. To support the systematic implementation of genomic medicine into the NHS and building on the strengths of the NHS GMC model, seven Genomic Medicine Service (GMS) Alliances are being developed across the country. Key to the model is that each GMS Alliance is able to demonstrate strong partnership and collaboration, working with all NHS organisations across a geography.

### **5. A national co-ordinating and oversight function within NHS England and NHS Improvement (Genomics Unit)**

The NHS England and NHS Improvement Genomics Unit sits within a governance network that collaborates to support the delivery of genomic medicine in the NHS. The GLH Partnership Board reports to the NHS England Genomics Programme Board Chair (SRO Genomics, Chief Scientific Officer for NHS England, Prof Dame Sue Hill).

There are a number of elements of the NHS GMS that Genomics England are delivering in partnership with or on behalf of the NHS. Therefore the joint elements of the programme are reported to the NHS England and Genomics England partnership board, which in turn informs the Department of Health and Social Care Ministerial National Genomics Board [\[10\]](#). The goal of the National Genomics Board is to oversee the implementation of Generation Genome (CMO 2016 annual report [\[11\]](#)) and the genomic elements of the Life Sciences Industrial Sector Deal [\[2\]](#).

## *Wider policy landscape*

In the past few years, genomics has been a topic under consideration in the broader policy landscape in the life sciences as well as in health. Various policy documents and initiatives have included genomics while also considering the broader implications of this technology for the health system and for the country as a whole.

In 2014, Health Education England (HEE) launched the Genomics Education Programme (GEP) to ensure that the 1.2 million-strong NHS workforce has the knowledge, skills and experience to keep the UK at the heart of the genomics revolution in healthcare.

The Chief Medical Officer for England's 2016 annual report, *Generation Genome*, presented a thorough examination of how genomics can improve health and prevent ill health. The recommendations within it have helped to inform the development of the GMS, including the establishment of the National Genomics Board.

In parallel, the Life Sciences Industrial Strategy (Aug 2017), recognised the opportunities of genomic medicine and the broader impact of building a genomics industry on the life sciences industrial sector [\[11\]](#). In particular, the Strategy outlined proposals for the Health Advanced Research Programme (HARP), large scale and multi-stakeholder programmes of research and development in target areas. One of these, genomics in medicine, outlined ambitions to: i) whole genome sequence 500,000 people in UK Biobank including supporting informatics infrastructure; ii) capture data generated by NHS England-commissioned WGS services; iii) complete the pathway in the NHS for routine WGS of cancer, including sample and analysis pathways; iv) support the UK's commercial capabilities in DNA sequencing.

The House of Lords Science and Technology Committee published a report in April 2018 on its inquiry into the Strategy, which raised concerns about the government's delivery of the Strategy, particularly around engagement of the NHS [\[12\]](#). Another inquiry report from the House of Commons Science and Technology Committee in April 2018, *Genomics and genome editing in the NHS*, made a number of recommendations, including around infrastructure and workforce planning. The government's response (July 2018) outlined its commitment to launch the Genomic Medicine Service and support the initiatives outlined in the Life Sciences Industrial Strategy [\[13\]](#).

The Life Sciences Sector Deal 2, published in December 2018, contained a range of programmes to further boost genomics in the UK, including the commitment to sequence one million genomes within five years (half from the NHS and half from UK Biobank), as well as wider support for uptake and adoption of innovation in the NHS [\[2\]](#). As part of the wider government Industrial Strategy, the Accelerating Detection of Disease (ADD) programme was announced in July 2019, with up to £240 million investment to recruit up to five million healthy people and use their data to explore new ways to detect and prevent the development of diseases [\[14\]](#).

The NHS Long Term Plan (January 2019) included a commitment to increase sequencing and genetic testing provision in health care, such as for children diagnosed with cancer, increasing the diagnosis of familial hypocholesterolaemia, further access to cancer testing and access to genomics research.

The UK Strategy for Rare Diseases (February 2019) summarises an implementation plan to support patients with rare diseases, including ongoing commitment to exploring the use of genomics as a diagnostic tool [\[15\]](#).

The National Genomic Healthcare Strategy, announced in February 2019 and expected in 2020, will include plans to build on the ambition to sequence five million genomes in the UK by 2023-4, an increase from the one million announced in the Life Sciences Sector Deal 2 [\[16\]](#). There will also be proposed improvements in services for people with rare diseases, including each patient having a dedicated health care professional responsible for coordinating their care.

The Topol Review (February 2019), while not directly focused on genomics, provided advice on how digital technologies, including genomics, will change workforce roles and as such measures need to be taken to support the workforce to keep their skills up to date [\[17\]](#).

The 2019 Green Paper Advancing our health: prevention in the 2020s presented an opportunity to provide a mid-term look at proposed prevention strategies, of which genomics forms one of many strands [\[18\]](#). The response to this consultation is expected in 2020. Along with the Independent Review of Adult Screening Programmes, it included discussion on polygenic scores, with a view to identifying individuals most at risk of common diseases such as cancer or heart disease, or to stratify individuals according to risk prior to screening efforts [\[19\]](#).

There are also further discussions ongoing in the area of WGS of healthy newborns – in November 2019 the Secretary of State for Health and Social Care announced an NHS-based pilot project whereby 20,000 babies would have their genomes sequenced at birth [\[20\]](#). No further details are available at present (May 2020).

Most recently, the Life Sciences Industrial Strategy Update (January 2020) outlines progress made in delivering the strategy, including the launch of the project to sequence 500,000 UK Biobank participants, launch of the ADD programme (which will include up to five million polygenic score assessments) and investment in digital infrastructure [\[21\]](#).

## 2.3 Genomics in Wales

**Author: Sian Morgan, All Wales Medical Genetics Service**

### *Genomics for precision medicine strategy*

The Welsh Government's Genomics for Precision Medicine Strategy, published in July 2017, sets out an ambitious plan "to create a sustainable, internationally-competitive environment for genetics and genomics to improve health and healthcare provision for the people of Wales" [\[22\]](#).

'Partneriaeth Genomeg Cymru – Genomics Partnership Wales (GPW)' was established in 2018 to support a collaborative approach to genomics in Wales. It represents several disciplines coming together to deliver a transformational programme of work. Partners include: patients and the public, NHS Wales' All Wales Medical Genomics Service, Public

Health Wales' Pathogen Genomics Unit, Wales Gene Park, Higher Education Institutions in Wales, Welsh Health Specialised Services Committee, NHS Wales Informatics Service, Health Education and Improvement Wales, NHS Wales and Welsh Government.

The national programme follows the Prudent Healthcare philosophy and through its five core themes, described below, represents the Welsh approach to ensure that genomics for precision medicine becomes mainstreamed across health and care in Wales [23]. The programme is in line with 'A Healthier Wales' - the long term plan for Health and Social Care [24] – as well as the UK Life Sciences Industrial Strategy [11], in recognising the importance of earlier and more personalised diagnostics to prevent illness, intervene appropriately and prolong independent living. The Genomics for Precision Medicine Strategy also aligns with the principles of the Quadruple Aim; 1) Improved population health and wellbeing; 2) Better quality and more accessible health and social care services; 3) Higher value health and social care; and 4) A motivated and sustainable health and social care workforce. Progress to date is noted below.

### *Co-production*

GPW are committed to working in an open and transparent manner with patients and the public, using their collective experiences to shape and add value to the work of the Genomics Partnership and future precision medicine services in Wales. A Patient and Public Sounding Board has been established to provide expert guidance on framing discussions such as ethics and consent, and to ensure that the programme remains focused on delivering against quality of life metrics as part of the wider Value Based Healthcare approach. Representatives from the Sounding Board are also part of the GPW Board membership, ensuring co-production at the highest level within the Strategy.

### *Clinical and laboratory Services*

The commitment to transform the diagnostics services in Wales is well described in parallel and connected national programmes that include genomics, pathology [25] and imaging [26]. Combined with the plan for emerging advanced therapies [27] and a national digital strategy [28] there is a clear direction for precision medicine in Wales [29]. Advances in genomic tests for the diagnosis and targeted treatment of diseases such as cancer, point of care testing for infection control and the use of artificial intelligence to support clinical decision making, together with digitally connected health records, are just a few examples of how technology will support the provision of personalised care [30].

The Welsh plans include ambitions to develop integrated precision medicine services, to establish a hub of excellence for Wales that will not only expedite the development of local expertise, but also attract collaborations from beyond its borders. The intention is to fully explore opportunities emerging from the UK Government through the UK Industrial Strategy and Life Sciences Sector Deal to benefit Welsh citizens as well as businesses in the growing life sciences sector in Wales.

Development of both human and pathogen genomics services is a critical dependency in enabling the full potential held by personalised medicine to refine the diagnosis, treatment and prevention of disease in Wales. Major reconfiguration to human genomic NHS services in England has driven significant expansion of testing provision in Wales to ensure equity

of services is offered to Welsh citizens. Additional investment in Wales has provided additional sequencing capacity and integrated IT infrastructure between academic, NHS and PHW institutions.

Recent national NHS human genomic services implemented include (but not exhaustive) Non-Invasive Pre-natal Testing (NIPT), Circulating Tumour DNA (ctDNA) service, cancer and rare disease next generation sequencing (NGS) panel testing and whole genome sequencing, an All Wales Lynch Syndrome Service for all patients with colorectal cancers, and an NGS panel service for Familial Hypercholesterolaemia (FH). National drivers such as the Single Cancer Pathway [\[31\]](#) and the Rare Disease Implementation Plan [\[32\]](#) ensure that a whole system approach is taken to the potential held by genomics.

New national services implemented by Public Health Wales include the mycobacterial service providing identification and outbreak support for SW England and Wales using whole genome sequencing and the HIV service. Whole genome sequencing (WGS) for influenza has attracted international recognition as Wales is the first nation to report its WGS data in real time to Europe. The data is an integral part of ongoing surveillance and vaccine development.

It is recognised that for progress to be made at pace, genomics cannot act alone. Alignment with other Welsh Government priority policies in the areas of digital innovation, such as the National Data Resource and the commitment to work transparently with the public on the handling of personal data, is central to progress and builds on national assets such as the Secure Anonymised Information Linkage (SAIL) data bank hosted by Swansea University [\[33\]](#).

### *Research and innovation*

To advance the genetics and genomics research agenda, Wales has developed a strong academic network of Higher Education Institutions working seamlessly with the NHS and external collaborators, and supported by the infrastructure of the Wales Gene Park. An emphasis on translational research in Wales remains key to driving advances in areas such as genetic and genomic determinants of psychiatric disorders, cancer, rare diseases and microbial resistance. Local strengths in clinical trials, health services research and health economics enable Wales to be an early adopter of new findings to transform the health and wellbeing of its population.

There is full commitment to increasing training opportunities for future genomic scientists and clinicians, encouraging and supporting young people to pursue degrees in the life sciences, for them to continue with post-graduate education and to develop exciting and rewarding careers within genomics in Wales.

In 2018, Wales successfully recruited rare disease patients to the 100K Genomes Project. Wales' involvement in initiatives such as this and the UK Biobank demonstrate the ambition to continue to work at a UK level to strengthen the collective research portfolio and outcomes for the benefit of citizens UK-wide.

Partners within the programme are developing models to manage consent to genomic data and samples for research.

## *Workforce*

Health Education and Improvement Wales leads on education, training, development, and shaping of the healthcare workforce in Wales, supporting high-quality care for its citizens.

The Genomics Strategy emphasises the importance of the work to:

- » Inspire the public and future workforce
- » Enthuse the trainee workforce
- » Strengthen the genomics specialists
- » Upskill the wider workforce
- » Embed genomics into everyday practice

Raising the profile of careers in genomics to graduates and early career professionals in Wales will ensure that opportunities to develop and contribute to genomics in Wales are accessible to all and are seen as attractive opportunities. Developing a dynamic scheme for recruiting to scientific training and higher scientific training posts e.g. Genetic Counselling, Bioinformatics, Genomic Science, Microbiology and Molecular Pathology in Wales is also key in enabling Wales to fulfil its commitment to mainstreaming Genomics across healthcare disciplines [\[30\]](#).

The commitment to grow and develop the workforce is evidenced by the successful ongoing investment in new staff to join the clinical and academic genomics services since 2018. Started in 2019, the genomics clinical workforce is being reviewed and redesigned to increase service capacity to meet patient need. The 'One Health' approach that Genomics Partnership Wales is taking allows for the development of joint teams across partners, to share expertise and meet the needs of the patients in Wales. The initial steps were taken at the beginning of 2020.

In 2018, Swansea University launched the first MSc in Genomic Medicine course in Wales. HEIW provide funded places to support mainstreaming of genomics across healthcare disciplines.

A number of engagement and education initiatives have been implemented in order to ensure that the wider NHS and academic workforce are aware and ready for the improvements to healthcare that genomics can make. A survey for nurses and midwives was launched in January 2020 to inform the genomics training needs for these disciplines. Genomics Champions are being recruited to raise awareness of genomics across all relevant healthcare disciplines.

Since 2018, GPW Genomics Roadshows have been delivered in all hospitals across Wales to engage with and educate the wider clinical workforce and demonstrate how they can use genomics to improve the care that they provide to their patients.

In June 2019, the Genomics Café initiative was launched. These free networking events held at different locations across Wales are for people affected by a rare or genetic condition; they provide a relaxed and informal opportunity to find out more about new advances in genomic medicine, and provide a chance for people to tell Genomics Partnership Wales how this group of people can be better supported.

### *Strategic partnerships*

New partnerships and further development of existing relationships between clinical services, academia, industry and patients and the public are being progressed to realise the benefits of genomics for precision medicine in Wales. Wales benefits from having a strong relationship between the health system and government. There is a single partnership model for genomic services, where decision making is agile and can facilitate rapid adoption of new technology in a consistent manner across NHS Wales. GPW are building on a strong track record of working with industry partners in areas of strength for service provision, clinical trials, technology adoption (Health Technology Wales) [34], and clinical education and a vibrant SME sector linked to MediWales [35] and the Life Sciences Hub [36].

In 2019, engagement with Healthcare UK and Department for International Trade was established to increase overseas opportunities for GPW partners, who recently hosted trade delegations from the USA and Bahrain. Attendance at national and international trade conferences will continue.

Wales is working closely with the other devolved nations to develop collaborative working for sharing services and knowledge for both human and pathogen services.

UK-wide collaborations include the UK Genomics Board, UK Genomics Working Group, the UK Genomics Strategy (developed with the Office for Life Sciences) and the upcoming NHSE Test Evaluative Working Groups; whilst each home nation has an individual approach to the delivery of genomics services, patient benefit is maximised by sharing best practice, innovation, evidence and expertise where opportunities can be found.

**Information was requested from Northern Ireland, however, it was not received by the time of completion of this report.**

## 3 Clinical genome analysis

Our understanding of the genetic basis of disease has become a powerful tool in providing strategies for the treatment and prevention of disease. Human genome sequence varies between individuals and while much of this variation has no impact on our health, an important subset of human genomic variation contributes significantly to our risk of disease. For example, the average human genome contains approximately 3.5 million single nucleotide polymorphisms and 2,100 to 2,500 structural variants [37]. Over the past 25 years there have been major advances in genome analysis technologies and in subsequent understanding of the human genome and its role in health.

More than half of the 7,000 known rare (monogenic) diseases have now been found to have a specific gene associated with the disease [38]. Single (or small numbers of) genetic variants are the dominant causal factor in the development of these rare diseases, which cumulatively affect an estimated 1 in 17 of the population. Genetic testing for symptomatic individuals and at-risk relatives occurs routinely in many medical specialties.

In cancer, a proportion of people will develop cancer because they have inherited a genetic fault that substantially increases their risk. Cancer arises when a series of errors in multiple genetic variants, normally arising from the cumulative effect of DNA damage over time, create abnormal cell growth. Certain inherited genetic faults can accelerate this process. Significant environmental factors such as ultra-violet light exposure, smoking and obesity are also known to drive the development of cancer.

Most common diseases such as diabetes, osteoporosis, and cardiovascular conditions can have genetic influences, as has been shown in many twin studies. For common diseases, it is possible that the dominant contribution to their development is environmental and behavioural, with genetic variants playing a smaller but potentially significant role in modifying the overall risk of developing the condition. These smaller contributions can be amalgamated into what is known as a polygenic risk score to assess their combined effect, though the clinical utility of this approach is still being investigated.

### 3.1 Uses of clinical genome analysis

There are a number of ways in which clinicians can use analysis of genomic variation in individuals to manage their health. Broadly, these can be classified as follows:

1. **Diagnosis/classification of existing disease** – Where a patient is identified as having clinically significant symptoms suspected to have arisen from a genetic abnormality, genetic testing can be used to provide a definitive diagnosis. This genetic diagnosis often provides utility for the patient through identifying the cause of their condition and assisting reproductive decision making. It may enable more accurate prognosis, direct more effective therapeutic interventions, or unlock access to disease-specific social and peer-to-peer support services. Genetic testing can also be used to diagnose rare diseases, distinguish between subtypes of rare genetic disease, or to distinguish cases of apparently common diseases (such as chronic kidney disease) that have genetic origin – and potentially heritable by other family members – from those that

do not. This information has benefits not only for individual patients, but also for family members, as it allows them to be tested and to understand their risk and if necessary, take actions to reduce it. In particular this is relevant for rare diseases, where there is often a protracted 'diagnostic odyssey'; genomics can enable a faster diagnosis. Approximately 80% of rare diseases are of genetic origin.

- a. **Tumour testing** – Examination of genetic markers in a tumour can aid diagnosis of cancer, and determine which genetic alterations are present to inform which targeted therapy would be most effective. Knowing which subtype of cancer is present can also assist in determining prognosis of the patient.
2. **Predictive / pre-symptomatic disease risk testing** – Analysis of genomic variation can be used to identify individuals who may be at increased risk of disease. Examples include:
  - a. **Adult inherited cancer mutation testing** – where healthy individuals with a strong family history of particular types of cancer can be tested for the presence of genetic variants that indicate they may be at much higher risk of developing these diseases than the general population. This type of testing is not able to predict definitively that the patient will develop the disease, but can be used to guide decisions about potential risk reducing measures such as enhanced screening, chemoprophylaxis, or preventive surgery.
  - b. **Carrier testing** – including cascade testing to family members (for example for familial hypercholesterolemia), where relatives of a patient who has been diagnosed with the condition and has a positive genetic test can also be tested, to determine whether they are affected or carriers of the same disorder.
3. **Antenatal testing** – To establish the diagnosis in high risk pregnancies, genomic testing of amniocentesis or chorionic villus samples is carried out. Non-invasive prenatal testing (NIPT) for aneuploidies such as Down's syndrome is a relatively recent development, whereby genetic testing can be used to very accurately determine whether or not a fetus is at risk of being affected with a disorder. Where a high risk is indicated by testing, a further invasive test is required to confirm diagnosis. For those pregnancies at risk of specific rare inherited disorders, non-invasive prenatal diagnostic testing is also available.
4. **Newborn screening** – Newborn blood spot screening involves taking a blood sample from a heel prick for biochemical testing to find out if a newborn baby has any of nine rare but serious health conditions. A positive result (for example, for cystic fibrosis) is generally followed with a genetic test to confirm the diagnosis. There was a call from Genetic Alliance UK in 2019 for genetic screening to be part of the standard newborn screening programme [\[39\]](#).

5. **Pharmacogenomics to guide appropriate treatment** – The effectiveness of a number of therapeutic interventions, and in particular a number of drugs, is dependent on the genomic variation of the patient to whom they are prescribed. Over 100 drugs have pharmacogenomic information or actionable labels [40]. Genetic variation can determine both the likely effectiveness of a drug and also the risks of unwanted toxicity. Genetic testing can be used to guide drug selection in a number of scenarios including:
- a. **Prescription of targeted therapies supported by companion diagnostics** – primarily in cancer treatment. More than thirty cancer drugs that have a companion diagnostic have been cleared or approved by the FDA and EMA, each having a specific genetic test associated with their use [40, 41]. It is anticipated there will be a large increase in these companion diagnostics. The genetic test is used to identify whether the patient has a type of tumour that is more or less likely to respond to the therapy, and so enables the clinician to select and monitor for the most appropriate treatment for each patient.
  - b. **Drug selection and adverse drug reaction risk reduction** – for example, the anti-HIV drug abacavir can only be prescribed following genetic testing of the patient to ensure they do not carry a rare genetic variant associated with a severe and potentially fatal adverse drug reaction.
  - c. **Apply more accurate dosing strategies** – in certain situations, precision dosing can alleviate adverse drug reactions and can also improve efficacy of various medications.

## 4 Genome analysis methods

There are a wide range of genome analysis methods available suited to different clinical scenarios, which may affect the type of genomic variation for analysis, the resolution of analysis required, and the time sensitivity of the application. The number of indications for genome analysis in clinical practice continues to expand. In response to this demand, conventional analysis methods are being continually refined and improved.

### 4.1 The genome analysis spectrum

The human genome is the complete set of DNA molecules in humans, including gene-coding regions and non-coding regions. Many variations in the DNA sequence are found between different individuals, and additional variants are also acquired within an individual's own genome during their lifetime. In practical terms, genome variation can be defined as a change in one or more nucleotides in comparison to a reference genome. The exact definition of a reference genome varies on the context used. In this context of describing broad genome variation, it is used to refer to the digital representative of the reference human genome that is generated from multiple donor genomes and considered 'standard'. However, it is worth noting that in a clinical context it can also be used to refer to any other genome against which genome variation in a DNA sample is being compared to. For example, in trio analyses, the genome of the offspring (the proband) is compared to the parental genomes, which are considered the reference genomes.

The human reference genome is a publicly-available digital representative human genome that can be used for comparison with sample (or experimental) genomes for the identification of variants. Reference genomes exist for different species and do not represent the DNA sequence of a single individual from that species, but rather are generated from multiple donor genomes. The current human reference, Genome Reference Consortium human 38 (*GRCh38*), was released in 2013 and has received many smaller updates since then. Whilst referred to as 'the human reference genome', this is not the only human reference genome in existence. Areas of the genome that vary notably between sample populations are annotated separately and are assembled alongside the reference genome – these are referred to as scaffolds.

Although used as a representative for human DNA sequences, the human reference genome is primarily constructed using European-derived DNA samples, leading to some populations being under-represented. The reference genome has been built upon and improved since the completion of the human genome project in 2003, but it still contains some sequence gaps, owing to the difficulty in sequencing some areas of the genome. Release of *GRCh39* is anticipated in the near future.

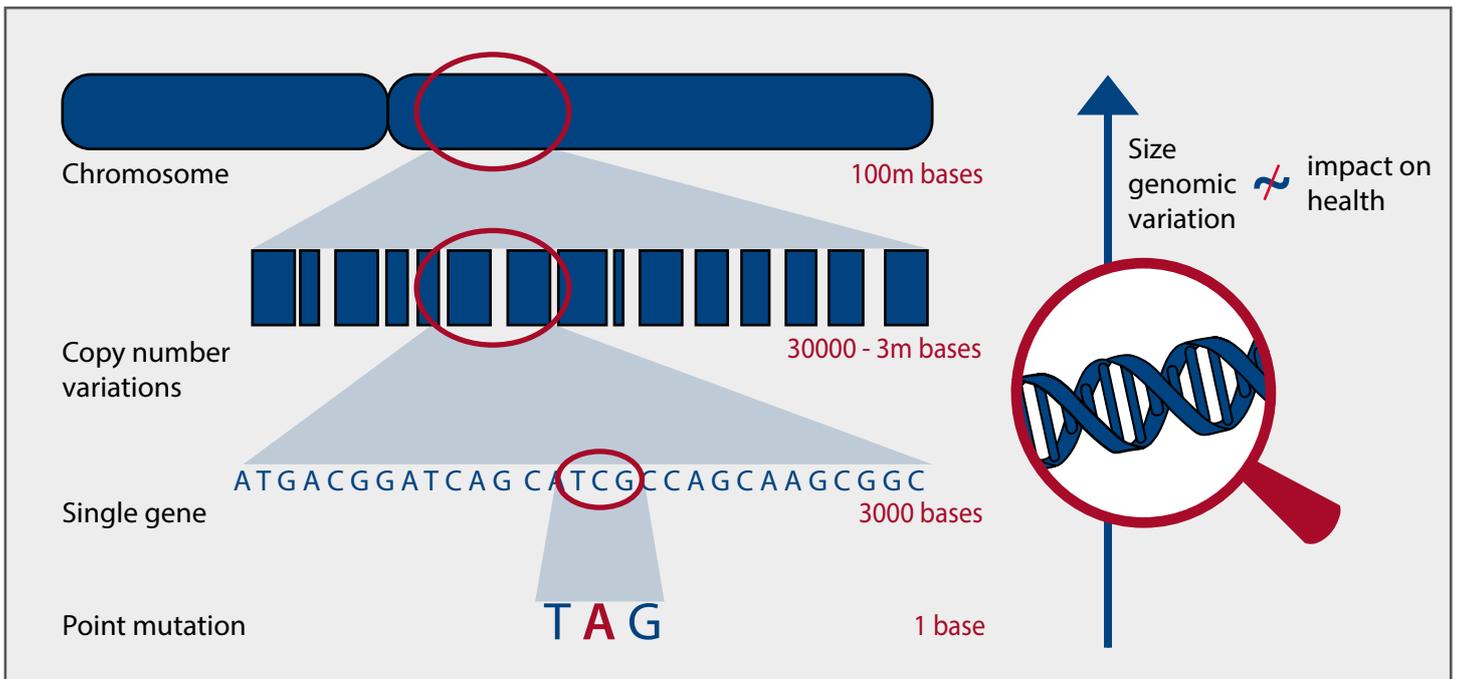
All sequencing within clinical contexts utilises a reference to which the patient DNA sequence can be compared to identify variation. Without a reference, *de novo* genome construction would be needed, a prohibitively costly and laborious process. Patient genomes are 'mapped' to reference genomes for comparison – making the process significantly cheaper and easier.

Genome in a bottle (GIAB) provides standards, methods and data around reference genomes for use in a clinical context, with the aim of facilitating translation of genome sequencing into clinical practise. GIAB makes available both complete and partial references for clinical use, including a range of references from genetically diverse groups e.g. European, East Asian, Ashkenazi Jewish.

There are millions of such variants in each human genome, which range in size from single nucleotide variations to deletions or duplications of single genes or even entire chromosomes. Importantly, the size of the genomic variation does not always correlate to the impact it has on health. The effects of a single nucleotide variation can range from having no effect on health at all, to health consequences as severe as those from deletions that span many thousands of nucleotides. Methods for clinical analysis of genomic variation need to be able to detect genomic variation across this spectrum (figure 2).

In addition to analysis of DNA directly, there is growing use of methods that analyse other molecules related to regulation and expression of the genome. Genes can be switched on and off through epigenetic mechanisms that modify the structure of DNA molecules. One such mechanism investigated in clinical genome analysis is gene methylation. This occurs when a methyl group is added to a DNA molecule, typically repressing gene activity. Other methods measure the level of gene activity by analysing the amount of messenger RNA (mRNA) that is transcribed when a gene is switched on. As the mRNA sequence is based on the DNA sequence, the RNA can also be used to detect the same variations found in the genomic DNA.

Figure 2. Spectrum of genomic variation



## 4.2 The evolution of genome analysis methods

The methods for analysing genomes have evolved significantly over the past 100 years, and it is this technological evolution that has driven an exponential expansion in our understanding of the function and dysfunction of our genome. Early genome analysis methods involved looking for genomic variation using optical microscopy. As optical microscopes have limited resolution, only the largest deletions, duplications or changes in chromosome number can be assessed with these methods. Smaller variations, at the level of one to a few nucleotides, remained undetectable until the development of a range of 'molecular' biological techniques which exploited the increased understanding of the structure of DNA and how it was replicated. The two key advances made in molecular biology in the 1970s-80s were:

**Amplification** – The ability to increase the quantity of DNA obtained from a biological sample by making millions of copies of it either in a test tube or inside cells

**Hybridisation** – The ability to identify and generate fragments of DNA that would strongly and selectively bind to one another

Together, these step changes in DNA analysis formed the bedrock on which modern clinical genome analysis is built. Amplification of DNA in bacterial cells enabled the development of a DNA sequencing methodology by Fred Sanger, which in various modernised forms remains in use to this day, and was used to sequence the first whole human genome [\[42\]](#).

A combination of hybridisation and *in vitro* amplification are also the basis of PCR (polymerase chain reaction), a method to selectively amplify fragments of DNA that can produce sufficient quantities of genomic material for higher resolution biochemical and optical analysis of genomic variations, and which has become a key component of many types of clinical genome analysis used today.

Over the past 30 years the number of genome analysis methods has continued to grow, exploiting varying combinations and adaptations of amplification and hybridisation to enable the sensitive and specific detection of large- and small-scale genomic variation from small initial quantities of DNA. Each of these methods have their own strengths and weaknesses and so many have found 'niches' of clinical utility for analysing particular genes, genomic regions or types of genomic variant. Collectively, however, these methods have revolutionised clinical genome analysis, allowing an ever-wider variety of types of genome variation to be detected in an ever-increasing proportion of the genome. New methods that have been developed to analyse RNA and methylation also make use of amplification and hybridisation technologies. RNA analysis requires an additional process to convert RNA into DNA, known as reverse transcription, after which the DNA copy (cDNA) is analysed using the same methods as conventional genome analysis [\[43\]](#). DNA methylation analysis typically requires use of a methods to convert the methylated DNA into a nucleic base called uracil, normally found in RNA. A DNA sequencing method called pyrosequencing is then often used to identify the methylated site [\[44\]](#).

This historical perspective on the evolution of genome analysis methods is important because this wide variety of 'older' lower resolution methods remain in extensive use throughout the NHS, despite the move towards using ever more high-resolution genome analysis methods for detecting clinically significant genome variations of all sizes. The historical development of the field provides the context to understanding why this is the case.

Other technologies and methods used both prior to and post DNA testing are just as important to the genome analysis process as the methods directly used to detect and sequence DNA. DNA must be of sufficient quality and quantity for DNA analysis to be successful. The type of sample used for DNA extraction and the methods used for sample extraction, processing and preparation have a vital role to play in this. For example, whether a patient sample is fresh, frozen or embedded in paraffin will affect the quality of the DNA extracted. For DNA analysis in cancer, the extraction of DNA from a sample where it is hard to distinguish tumour from normal tissue may affect results, whilst blood samples require different processing from solid tumour samples. For DNA sequencing, the bioinformatics methods required to make use of the information and interpret results will also influence the capabilities of DNA sequencing methods.

In the following sections of the report we review the full breadth of assays currently being used for clinical genome analysis. We then review how genomic analysis is currently being used by clinicians across the patient life course and for cancer management – including which assays and technologies are employed, the purpose of the clinical testing they provide – and how this use is likely to change over the next five years. We also include descriptions of genomic assays that are currently used mainly in research settings, but which are sufficiently well developed to warrant consideration of their validity and utility for clinical application within the next five years.

## 5 Genome analysis by non-sequencing methods

This chapter provides an overview of genome analysis methods which are not sequencing based. These methods range from relatively low resolution techniques suitable for analysing large scale variations in the genome such as chromosomal rearrangements, to high resolution techniques able to detect single nucleotide alterations. Each of these techniques has different strengths and limitations, making them suited to different diagnostic workflows. They fall broadly into three main categories:

### *Cytogenetic techniques*

Cytogenetics refers to the study of chromosomes. Methods include:

- » Karyotyping: a classical low resolution cytogenetic technique using microscopy to analyse large scale changes in the number or structure of chromosomes, such as whole chromosomal deletions or duplications
- » Fluorescent *in situ* Hybridisation (FISH): A molecular karyotyping technique that uses fluorescent probes designed to detect regions of interest in a patient's DNA, such as copy number variations.

### *Microarray based molecular cytogenetics techniques*

Techniques based on hybridisation of fluorescently labelled patient DNA to a microarray, a solid platform (such as a microscope slide) containing thousands of DNA probes that bind specific sequences. Methods include:

- » Array Comparative Genome Hybridisation (Array CGH): Compares the quantity of patient DNA bound to an array to that of a reference DNA sample, to detect genetic imbalances such as copy number variations and deletions.
- » Single nucleotide polymorphism arrays (SNP array): Uses an array containing DNA probes that bind patient DNA to reveal sites of single nucleotide variation.

### *PCR based techniques*

Molecular analysis techniques based on the polymerase chain reaction (PCR) where selected short DNA sequences are amplified to allow regions of interest to be detected and measured, for example to detect the presence of a specific mutation.

## 5.1 Karyotyping

An individual's 'karyotype' is their complete set of chromosomes in a particular cell. Normally human cells have 46 chromosomes grouped into 23 pairs, with one chromosome in each pair inherited from each parent. These chromosomes have recognisable shapes and structures that allow them to be identified under a microscope. Therefore, biological errors that result in large scale changes to these chromosomes, such as a loss or gain of

an entire chromosome or large structural changes, can be visually detected. The process of 'karyotyping' refers to methods for visually inspecting the number and/or structure of chromosomes in a sample of cells, to detect any abnormalities present.

Tissue samples (such as amniotic fluid, blood or a tumour biopsy) are collected from the patient, and the cells extracted and cultured *in vitro* (to expand the number of cells available for analysis). Cells are harvested at a particular point in their cell division cycle where the chromosomes are condensed, to allow them to be visualised more easily, and spread onto a microscope slide. The chromosomes are stained using dyes which produce characteristic banding patterns on the chromosomes, a technique called G-banding. The chromosomes are imaged using microscopy and photographed. The image is examined for any changes in the number, morphology or banding patterns, which could indicate different genetic syndromes.

### Types of genetic variants identified by karyotyping

**Aneuploidy** – Karyotyping can identify changes in the number of chromosomes and is commonly used as a first or second line test to look for aneuploidy.

**Large copy number variants** – The highest resolution karyotyping procedures can detect genetic changes of around 5-10Mb (5-10 million bases) allowing for detection of some large microdeletions.

**Chromosome breakage** – Some genetic disorders are caused by repeated breakage of chromosomes. Where these are suspected the sample is cultured with a DNA cross-linking agent and the metaphase chromosomes visualised. Samples from affected individuals exhibit highly fragmented chromosomes resulting from defective DNA repair mechanisms.

**Mosaicism** – The term 'mosaicism' is used to describe the presence of two genetically distinct cell populations within an organism. This happens when one cell in a tissue acquires a genetic alteration and proliferates to produce more identical cells containing that alteration (clones). Karyotyping can be used to analyse cells for mosaicism that occurs as a result of large-scale chromosomal abnormalities, such as a change in chromosome number or a large structural variation. This can be used to detect germline mosaicism, which arises in germline cells and means a child may inherit the genetic alteration. It can also be used to detect somatic mosaicism, which arises in somatic cells during malignancies. Tracking different clones can allow disease progression and transplant success to be monitored.

**Chromosomal rearrangements** – Karyotyping can detect where there have been 'balanced' or 'unbalanced' rearrangements of DNA within or between chromosomes. Balanced rearrangements do not result in any change in the total amount of DNA, and so are not detected by array CGH (see next section), but can disrupt genetic function and cause disease.

### Clinical uses of karyotyping

- » Prenatal screening for genetic disorders known to result from chromosomal alterations, e.g. Turner Syndrome (a single X chromosome in females) and Down's syndrome (extra copy of chromosome 21).
- » First line testing for clinical conditions that may be caused by genetic abnormalities, such as short stature, indeterminate gender, delayed puberty, infertility and chromosome breakage syndromes [45].
- » Diagnosing cancer by identifying known chromosomal alterations, e.g. chronic myelogenous leukaemia, where a part of chromosome 9 breaks off and attaches itself to chromosome 22 (a chromosomal translocation).

### Strengths and limitations of karyotyping

Strengths:

- » Karyotyping provides a snapshot of the entire genome, useful for searching for all different potential chromosomal alterations simultaneously, such as in prenatal screening for genetic abnormalities.
- » While other methods offer higher resolution analysis of structural genomic variation, only karyotyping can currently reveal changes in three-dimensional chromosome structure.

Limitations:

- » Compared to other methods (discussed below) for detecting large scale genomic variation, karyotyping is very low resolution, being limited to changes large enough to observe by conventional microscopy. This means that for most rearrangements smaller than 5 Mb it is insufficiently sensitive to be of use.
- » It requires specialist skills, is time consuming, and as such is costly to perform.
- » Detailed karyotyping is reliant on successful cell culture, which means it is of limited use where tissue is scarce, or is difficult to transport and culture in a timely fashion.

## 5.2 Fluorescent *in situ* hybridisation (FISH)

FISH is a molecular cytogenetic technique that can be used to confirm the presence and location of specific DNA sequences. The technique uses DNA probes designed to bind to a target sequence of interest, for example a specific region of duplicated DNA. The probes are fluorescently labelled and added to patient cells taken from a tissue sample, where they hybridise to the target sequence if it is present in the patient DNA. A fluorescent microscope is used to detect the presence/absence and the location of any bound probes.

### Types of genetic variants identified by FISH

**Copy number variations** – FISH can be used to detect deletions or duplications of regions of chromosomes too small to identify by karyotyping, such as microdeletion syndromes (e.g. Prader-Willi, DiGeorge). Copy number variations are also referred to as Copy Number Alterations (CNAs).

**Gene fusions** – FISH is commonly used to identify gene fusion products, where chromosome breakage and reattachment leads to parts of two different genes becoming attached to one another, creating a new and abnormal gene such as the BCR-ABL fusion gene involved in chronic myeloid leukaemia (CML).

**Mosaicism** – FISH may be used to establish constitutional chromosomal mosaicism and can also be used to detect the development of tumours with abnormal genomes, or the successful growth of transplanted cells (which may have different genomes from their host) e.g. following bone marrow transplant for treatment of haematological malignancies such as CML.

### Clinical uses of FISH

- » Confirming the diagnosis of a suspected well-defined clinical syndrome, for example Edward's Syndrome (presence of an extra copy of chromosome 13).
- » Can be used to detect chromosomal abnormalities in cancer cells, for example extra copies of HER2 in breast cancer.
- » Can be used alongside karyotyping to help better identify chromosomal abnormalities.

### Strengths and limitations of FISH

Strengths:

- » FISH is higher resolution than karyotyping (it can detect regions about 100kb-1Mb in size) [45], and so can detect smaller changes in DNA that may not be visible with karyotyping.
- » Cells can be used at any stage of the cell cycle, so cell culture is not required, and archived tissue samples fixed in formalin (known as formalin fixed paraffin embedded-FFPE) can also be used, as is often necessary with tumour samples.
- » Any number of novel rearrangements can be examined, as probes can be developed for any specific DNA sequence required.
- » Results can be obtained from individual cells, requiring much less sample than needed for karyotyping and allowing different cells to be compared.

Limitations:

- » The specific design of the probes targeting pre-defined regions of DNA means FISH is limited to use in clinical cases where the type of genomic abnormality is suspected. It is not an efficient method for genome-wide interrogation for genomic alterations.
- » The detection of microduplications is more difficult than microdeletions, owing to the difficulty of discerning duplicated probe signals.
- » The time taken to design and deliver probes for novel alterations can be problematic in situations where there are time constraints, such as in preimplantation genetic diagnosis. As it is unlikely probes designed for a specific individual will be re-used for other individuals, it is hard to overcome this problem.

### 5.3 Microarrays (Array CGH and SNP arrays)

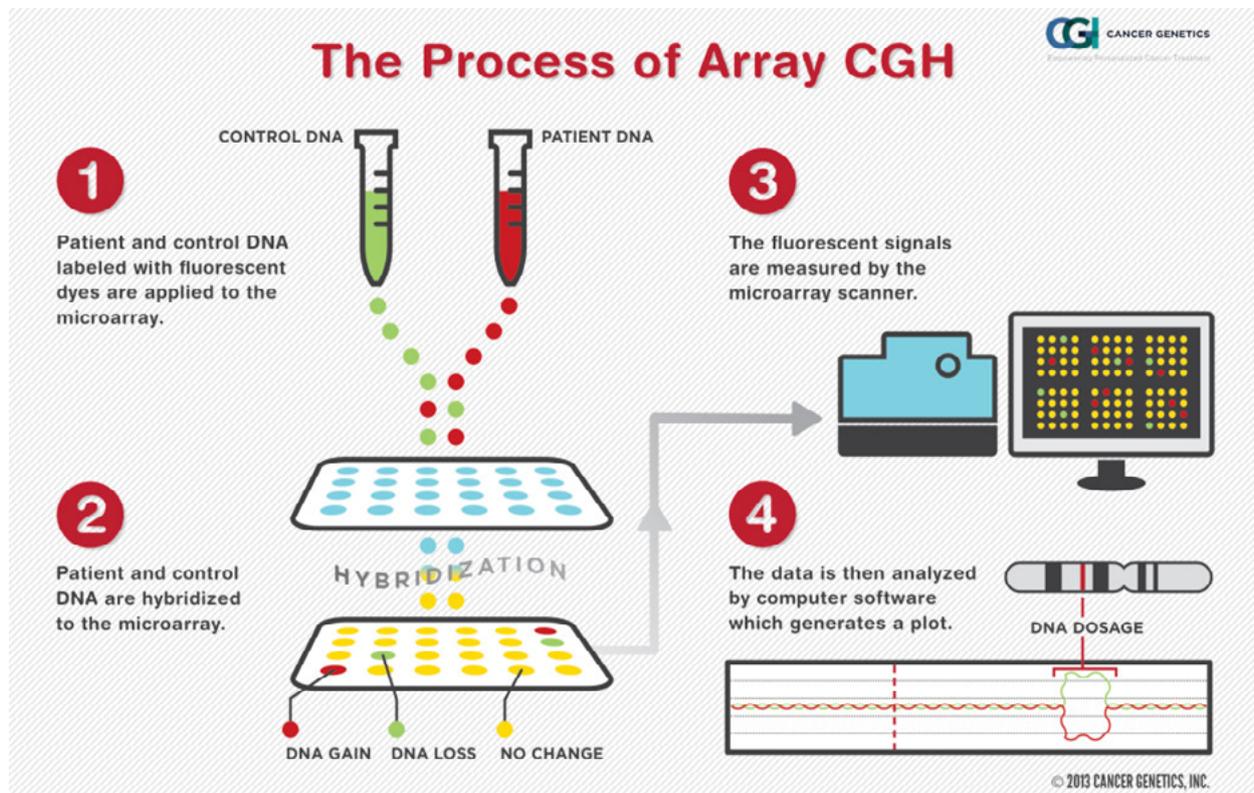
Array Comparative Genome Hybridisation (Array CGH) and Single Nucleotide Polymorphism arrays (SNP arrays) are molecular cytogenetic techniques that use microarrays to identify specific sequences of patient DNA present in a sample. A microarray is a solid platform such as a microscope slide, containing thousands of short, single stranded DNA fragments, arranged in precise microscopic locations. The DNA sequences act as probes which can hybridise with complementary DNA sequences present in patient DNA. They can be used to interrogate thousands of potential DNA sites at the same time. Microarrays can be designed for broad low coverage scans of sites across the entire genome, by containing DNA probes that are evenly distributed across the genome, or for more in-depth coverage of pre-determined genomic regions. The resolution depends on the microarray platform used and the size of the intervening genomic regions between probes; more closely spaced probes increase coverage and therefore produce higher resolution.

#### *Array CGH*

Array CGH is used to compare the amount of specific regions of DNA present in a patient's genome to a reference (normal) genome [46]. This allows genomic imbalances in a patient's DNA compared to the reference DNA to be detected, i.e. if a patient has more or less of a particular region of DNA as a result of copy number alterations or duplications.

An overview of the process is shown in figure 3. Patient DNA is heated to make it single stranded, and fragmented into many small regions of known length. The reference DNA sample is made single stranded and fragmented in an identical way, then each set of DNA is labelled with different coloured fluorescent probes (e.g. green for control and red for patient). The fluorescently labelled DNA are added to a microarray, where they hybridise to any complementary DNA probes on the microarray. The microarray is then scanned and software is used to detect the fluorescent colour and intensity information at each location. If a sequence of DNA is present in the same quantity in both the reference and the patient samples, equal proportions of each will hybridise to that location on the array, producing an orange colour as the fluorescence from the probes blend. If the patient DNA contains duplications of a particular sequence, more patient DNA will bind to the array than the reference sequence, producing more red colour. Conversely if the patient DNA contains deletions of a sequence, less of that sequence will bind to the DNA on the array than the reference, producing more green colour.

Figure 3. The process of array CGH



### SNP microarrays

Single nucleotide polymorphism (SNP) arrays are similar in principle to Array CGH, but are used to detect SNPs, sites within DNA that vary by a single nucleotide from the same site in a typical reference genome. The term SNP is often used to refer to inherited variants that occur commonly within the human genome i.e. at a frequency greater than 1%. There are an estimated 4-5 million SNPs in every individual's genome. Less frequent single nucleotide changes can simply be described as single nucleotide variants (SNVs). SNVs can also be acquired in somatic cells during an individual's lifetime, in which case they are often called 'point' mutations.

SNP microarrays can contain up to several million DNA probes designed to cover different SNP sites, which are fixed on to a solid platform (such as a microscope slide) in a known order. For each SNP a pair of probes is used, to ensure that both potential alleles can be detected. Similar to Array CGH, single stranded patient DNA is fragmented and fluorescently labelled. The patient DNA is applied to the microarray, and the level of fluorescence at each site reflects the degree of hybridisation, and therefore whether the SNP is present.

## Types of genetic variants identified by microarrays

**Copy number variation and aneuploidy** – Array CGH can detect small copy number variants: deletions, duplications and amplifications, as well as larger genomic imbalances such as aneuploidies. The size of variant detected is dependent on the design of the specific platform and is therefore influenced by probe type, size and spacing.

**Single base changes** – SNP arrays can provide information at the single base level of systematic changes e.g. loss of heterozygosity occurring across multiple pre-defined genomic locations.

## Clinical uses of microarrays

- » Array CGH is recommended as a first line test for behavioural, developmental and cognitive disorders arising from genomic imbalances, which can be caused by a large number of different variants.
- » SNP arrays can be used to reveal point mutations, including those that may arise following on from an initial mutation. These can result in loss of heterozygosity, an important mechanism in tumour development.
- » SNP arrays can reveal a lack of polymorphisms in a longer stretch of DNA, contrasting with normal inheritance from unrelated parents, where a more random distribution of alleles would be expected. This can be indicative of:
  - a. **Consanguinity** – Two people descended from the same ancestor (usually second cousins or closer) are more likely to have similar DNA variants. Children arising from consanguineous relationships have a greater likelihood of inheriting identical recessive disease-causing variants.
  - b. **Uniparental disomy (UPD)** – UPD occurs when both copies of a chromosome, or parts of a chromosome, are inherited from one parent rather than one from the father and one from the mother. In most cases this is not a problem clinically, but if it occurs on chromosomes containing certain genes it may cause disease. This is due to a phenomenon called genomic imprinting – when copies of certain genes are turned on or off depending on whether they are inherited from the mother or father. In UPD this can lead to both copies of an essential gene being turned off, resulting in a number of diseases, including Prader-Willi and Angelman's syndromes.
- » SNP arrays are currently used in clinical research to conduct Genome Wide Association Studies (GWAS), where associations between specific SNPs and a certain trait such as a disease are identified. Single SNPs may have a very minor or no discernible effect on a trait, but when effects of multiple SNPs are added together, they may have a significant impact.

## Strengths and limitations of Array CGH

### Strengths:

- » Depending on how many probes are used, resolution of the method can range from more than 1MB to less than 10kb for targeted arrays, higher than karyotyping and FISH.
- » Can be used to interrogate regions over the entire genome, or to more closely analyse selected regions in more detail.
- » Does not require any living cells, just a DNA sample. This has advantages in terms of resource: cell culture is not required and results are obtained more quickly as a result.
- » Can be used to interrogate hundreds of thousands of genomic loci simultaneously, significantly faster than FISH which is limited by the number of probes that can be used at the same time.

### Limitations:

- » The number of genomic sites explored is still limited by the number of predefined DNA probes that can be fixed to the array. Practical limitations on the number and density of probes means that they cannot detect very small deletions and duplications e.g. of single exons within genes.
- » Can only be used to detect changes in copy number, it is not useful for detecting alterations such as chromosomal inversions, which do not affect the amount of DNA in a sample.
- » Many copy number imbalances occur naturally that are not pathogenic, care must be taken to avoid misinterpreting results.

## Strengths and limitations of SNP arrays

### Strengths:

- » SNP arrays may allow identification of changes at the single base pair level e.g. loss of heterozygosity and important non-numerical genetic changes e.g. uniparental disomy.
- » Does not rely on cell culture as karyotyping does, so has advantages in terms of laboratory personnel resources and the time to generate a result.
- » Arrays have an element of flexibility as they can be customised to offer genome wide or targeted analysis, and can be designed to vary in the resolution with which they can detect copy number variation at different genomic locations.
- » Unlike array CGH, SNP arrays can be used to detect copy neutral loss of heterozygosity, as they do not rely on comparing relative amounts of DNA.

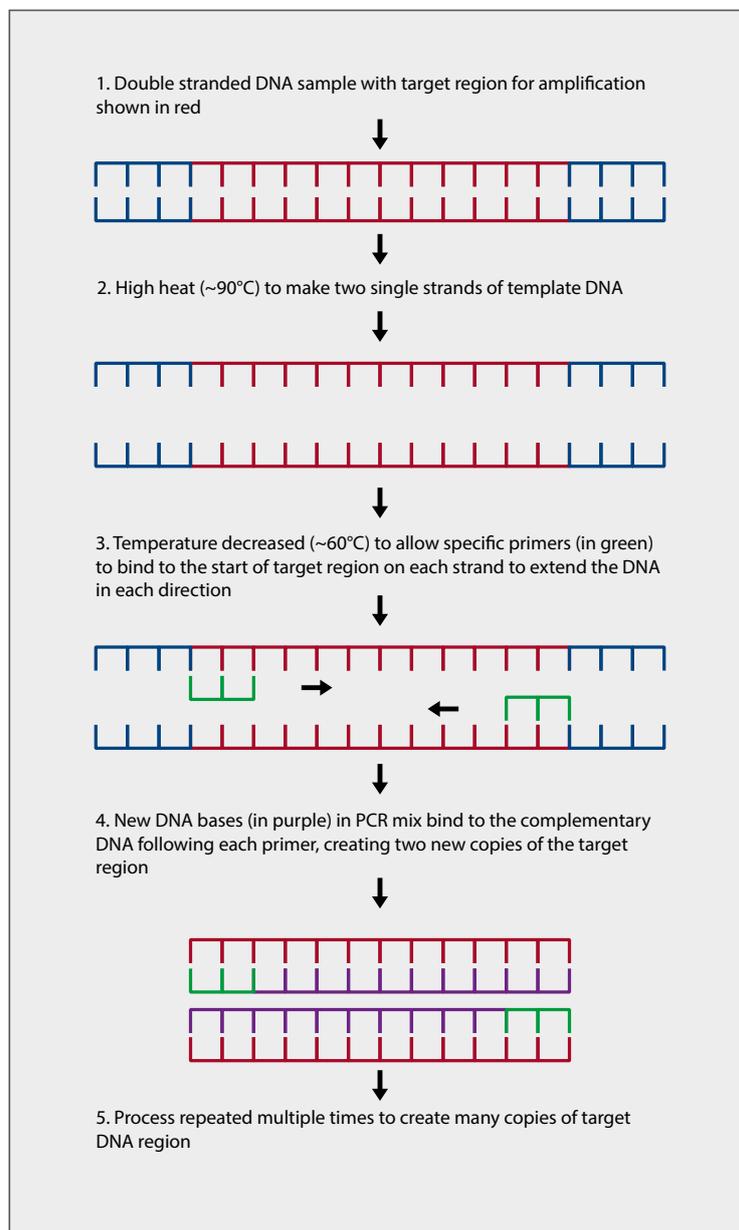
### Limitations:

- » As with array CGH, SNP arrays rely on pre-designed probes for exploring specific regions of the genome, and there is a limit to the number of probes that can be used on an array.
- » Arrays are not typically used to detect single base changes in individual genes, as there are other more cost-effective methods to achieve this.
- » Arrays are not suitable for detection of small insertions or deletions within genes as these are usually too small to be reliably detected.

## 5.4 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is a method used to selectively amplify one or more small segments of DNA for analysis. The method uses short sequences of DNA called primers – which can be designed, customised and synthesised within 1-2 days – to bind to very specific sequences in a sample of the patient’s genomic DNA. The primers act as templates to which DNA building blocks called deoxynucleotide triphosphates (dNTPs) are added, to allow a copy of the target sequence of DNA to be made. The process is repeated many times to allow multiple copies of the target sequence to be generated. This targeting and amplification process can be used to detect and quantify with great sensitivity and specificity the presence of a range of different genomic variations from small initial quantities of patient DNA.

Figure 4: Principles of PCR



## *Types of PCR methods*

Since it was developed in the 1980s many refinements have been made to the PCR method that have broadened its use in clinical genome analysis. Some of the main ones in clinical use are:

### **Quantitative PCR (qPCR)**

Often PCR is qualitative – it is sufficient to understand whether a target DNA sequence is present or absent in a patient sample. However in other circumstances it is useful to determine the absolute or relative quantity of the target DNA sequence, for example when determining if a therapy targeting a specific gene is working, investigating genetic duplications or deletions, or characterising the level of gene expression i.e. gene expression profiling (see RT-PCR below).

Quantitative PCR (qPCR) is often performed in real time by using fluorescent primers or fluorescent dyes, which become incorporated into the amplified DNA during the PCR process. The intensity of the fluorescence will be proportional to the quantity of amplified DNA and can be monitored in real time during the amplification process. This is sometimes also called RT-PCR which should not be confused with reverse transcriptase PCR (see RT-PCR below) which is often referred to using the same abbreviations.

Another type of quantitative PCR that uses fluorescence is called quantitative fluorescence PCR (QF-PCR). This technique uses fluorescently labelled primers which are designed to amplify specific short tandem repeat markers (regions of repetitive DNA) found on chromosomes 13, 18, 21, X and Y. PCR products are then separated by their molecular mass using gel electrophoresis, and the amount of fluorescence used to quantify the amount of each PCR product present. This technique is specifically used to quantify aneuploidy and abnormal numbers of chromosomes.

### **Reverse Transcriptase PCR (RT-PCR)**

PCR requires a sample of DNA, however in some situations it is more useful to analyse the RNA copy of a particular gene:

- » Quantifying RNA can determine how many copies of a gene are being produced by a cell (the level of gene expression). This can be useful for diagnostic or prognostic purposes
- » In some cases, an RNA transcript is more similar across different patients than the DNA copy of the gene, making it easier to design primers based on the RNA copy for use in multiple patient samples. This is often required for detection of oncogenic fusion products such as the BCR-ABL1 fusion gene in leukaemia [\[47\]](#)

In these cases the RNA is isolated from a patient sample, then a type of enzyme called a reverse transcriptase is used to produce a DNA copy of the RNA sample, known as cDNA. The cDNA can then be used in the desired type of PCR reaction. This technique is called reverse transcriptase-PCR (RT-PCR). If RT-PCR is performed using a quantitative PCR method it is commonly called (RT-qPCR). Care must be taken not to confuse the acronym RT-PCR with real time PCR, which is also commonly called RT-PCR.

## Multiplexing

Multiple genomic variants can be investigated in a single assay from a single sample of patient DNA by mixing together multiple sets of PCR primers, each of which selectively amplifies a different variant or region of a gene. The multiple amplified DNA fragments are then separated according to their size by capillary electrophoresis so that each can be individually analysed. It can be more complicated to develop multiplexing methods, and the technique is often less sensitive than when using single primers. However, PCR multiplexing is time and cost-saving, and can include the use of primers as an internal control to prevent false positive and false negative results [\[48\]](#).

## Multiplex Ligation Dependent Probe Amplification

Multiplex Ligation dependent Probe Amplification (MLPA) is a multiplexing based technique which can be used to detect copy number variation in specific genomic regions, such as a specific gene. This technique is different from other PCR based methods, as it is based on amplification of DNA probes that target specific sequences of DNA in the region of interest, rather than amplification of the patient's DNA itself. Up to 60 probes can be used simultaneously, with the resolution of the technique dependent on the number of probes used and the amount of coverage they provide [\[49\]](#). One set of primers is used that can fluorescently label and amplify all probes bound to patient DNA. Probes are designed to be different lengths, allowing them to be identified by their size. The level of fluorescence at the end of the experiment correlates with the quantity of an amplified probe present, and the results from the patient DNA is compared to that of a control DNA. If the patient is missing copies of a particular sequence, there will be less DNA for the probe for that sequence to bind to. Therefore, the probe will undergo less amplification compared to the control and the resulting fluorescence will be lower than the control signal, indicating a deletion. Conversely if the patient has more copies of a DNA sequence, more probes will bind and the fluorescence will be higher. MLPA techniques are also available for analysis of methylation (see Methylation Specific PCR below) [\[50\]](#).

## Methylation-specific PCR (MSP)

Methylation-specific PCR (MSP) is a qualitative PCR technique that is used to detect high levels of promoter methylation (hypermethylation) at CpG islands. These are regions of DNA with a high frequency of cytosine nucleotides adjacent to guanine nucleotides, where methylation of cytosine residues frequently occurs. MSP relies on use of bisulphite conversion, where unmethylated cytosine residues in a patient DNA sample are converted to uracil, whilst methylated cytosines are not converted. Two sets of primers are then designed which can detect both the methylated and unmethylated DNA in the region of interest [\[51\]](#).

## Droplet Digital PCR (ddPCR)

Several digital PCR methods have been developed and commercialised which have higher levels of sensitivity than conventional PCR methods, the most widely used of which is a technique called droplet digital PCR (ddPCR) [\[52\]](#). This technique is based on the partitioning of a fluorescently labelled PCR mixture into thousands of nanoscale droplets, with the PCR reaction for a sample then taking place independently and in parallel within each droplet. At the end of the reaction droplets are classified as positive or negative for the presence of the target sequence, based on a set threshold of fluorescence.

As there are very few DNA copies in each droplet, there is more chance that very low frequency DNA variants will be detected, when otherwise their signal might have been masked. The multiple parallel reactions also reduce the chance of error. The high sensitivity of ddPCR is important for detecting DNA variants present at very low frequencies. This method is often used in cancer companion diagnostic testing, for detecting very small quantities of circulating tumour DNA (ctDNA) in a blood sample, which can have variant allele frequencies as low as 0.01% (i.e. one copy of tumour DNA in 10,000 copies of non-tumour DNA) [53].

### Types of genomic variant detected by PCR

- » **Single base change detection** – PCR primers can be designed that can distinguish between two segments of DNA that vary at only a single nucleotide position (a so-called ‘point mutation’). This type of analysis is known as allele specific amplification. By selectively amplifying only DNA that contains the point mutation (the mutant allele), PCR can be used to sensitively and specifically detect the presence of disease-causing mutations in patient samples.
- » **Changes in repeat length** – Many genes have sequences in which the same pattern of nucleotides is repeated over and over again. Expansions in the length of these repeating sequences in some genes can cause disease. Customised PCR strategies can be designed to detect these expanded repetitive sequences and to quantify the number of additional repeats.
- » **Copy number variation** – Quantitative and multiplexed PCR approaches enable the accurate detection of copy number variants i.e. duplications or deletions of all or part of a gene. Deletions are detected as a reduction (in the case of heterozygotes) or complete absence (in the case of homozygotes) of the amplified DNA fragment. Duplications are detected as an increase in the quantity of amplified DNA fragment, when compared to a reference standard sample known to have a ‘normal’ number of copies of the DNA sequence being investigated.
- » **Aneuploidy** – A combination of multiplexing and quantitative PCR approaches can also be used to detect abnormal numbers of whole chromosomes, known as aneuploidies. Multiple sequences unique to each of the chromosomes of interest can be amplified from a single sample and quantified by QF-PCR. Abnormally high or low quantities of amplified DNA fragments from a particular chromosome are indicative of an abnormal number of chromosomes being present.
- » **Methylation** – Methylation specific PCR and MLPA can be used to detect altered levels of methylation. This can be used to diagnose both inherited disease, including Prader-Willi and Angelman syndromes, as well as certain types of cancer including neurological tumours and colorectal cancer [54-56].

### Strengths and limitations of PCR

Strengths:

- » **Amplification** – PCR amplifies the initial amount of DNA in the sample millions of times within a few hours. This enables detection of genomic variants from very small initial quantities of DNA, which can be crucial where only a limited amount of patient DNA is available e.g. some tumour biopsies.

- » **Multiplexing** – Multiple target regions of the genome can be investigated in a single assay from a single patient sample.
- » **Quantitative** – PCR can be used to quantify the amount of DNA present, which is important in detecting variations in the number of copies of chromosomes or smaller genomic regions present in a patient’s sample.
- » **High sensitivity and specificity** – PCR relies on the inherent stringency of pairing between complementary nucleotides (A with T and C with G) in DNA molecules to enable highly sensitive and specific detection of genomic variants. The amplification process itself has an extremely low error rate (less than one error in every 10 million nucleotides incorporated) meaning that the likelihood of generating false positives or negatives from copying errors in the amplification process itself is extremely low.
- » **Ease of use: assay design and deployment** – The PCR technique itself is generic and relies on basic equipment that is standard in all molecular diagnostic laboratories. Assay design is simple, with numerous online tools available to automate the design of primers for targeting novel genomic loci and for optimising the specific reaction conditions for each assay. Most PCR reagents are cheap and generic, and the specific primers required for each different genomic target can be easily purchased from one of many commercial primer synthesis companies at minimal cost (a few pounds) in only a few days.

Limitations:

- » **Not suitable for genome wide screening** – PCR relies on design of primers for amplification and detection of known regions of interest. Therefore, it is only suitable for targeted analysis of specific suspected genetic alterations. It cannot be used to detect unknown mutations, or screen whole genomes to detect abnormalities.
- » **Indirect sequence inference** – The presence or absence of particular DNA sequence assayed by PCR is determined indirectly, by quantifying the fluorescence of an amplified fragment. It reports whether or not the assay primer was able to bind to the patient’s DNA, and while the presence/absence of an expected variant can be inferred from this event, it can be confounded by other unexpected variations giving rise to false positive or negative results. For example, primers designed to detect an exon deletion might fail to bind, despite the exon being present and intact, because of an unexpected single point mutation in the primer binding site. This would result in the absence of an amplified DNA fragment, from which exon deletion could be incorrectly inferred.
- » **Limitations to multiplexing** – The number of different genomic loci that can be simultaneously assayed by multiplexing are constrained by the fragment size resolution limits of the instruments used to detect amplified DNA, the number of different fluorescent probes available to distinguish between different fragments and the challenge in finding amplification conditions that are optimal for multiple different chemical reactions between the patient’s DNA and the different primer sets being used. These limitations cannot always be easily predicted and may have to be determined empirically, adding time and cost to assay development.
- » **Assay validation** – While the machines and reagents used for PCR are mostly generic, each assay targeting a new genomic locus and/or using a new set of primers must be separately optimised and validated for diagnostic use.

## 6 Genome analysis by DNA sequencing

Sequencing is distinguished from other genomic analysis technologies in that it allows the order of nucleotide bases (the sequence) of the DNA to be read. Previously described techniques and technologies (chapter 5) can be used to determine whether a specific variant, nucleotide, or short sequence exists at a particular point in DNA, but not to determine the sequence of bases from scratch. DNA sequencing involves the conversion of physical information in the form of DNA into written or more often digital data, which can then be further investigated to infer clinically significant (or not) information.

In total, the genome is approximately 3.2 billion base pairs (bp) in length, whilst chromosomes average around 130 million bp each – currently too long to be sequenced as one continuous string. In preparation for sequencing, DNA is broken into shorter fragments that can be sequenced to produce ‘reads’; these are strings of nucleotides representing sections of DNA that have been read as one continuous piece by the sequencer. Depending upon the technology used, the length of retrieved reads varies from around 30 bp to hundreds of thousands of bp – with no theoretical limit except that imposed by the DNA itself.

Non-sequencing methods are useful where small numbers of known variants in pre-specified genomic loci are being investigated in a patient. However, there are many situations in which this type of focused analysis is either not possible, or becomes impractical to undertake because of the number or complexity of assays required to provide an adequate answer. Such situations include:

- » Where clinical presentation is complex and/or unusual and does not indicate an obvious choice of targeted assay
- » Where variations in multiple genes that might be causative or related to the patient’s condition need to be investigated
- » Where there are multiple point or small disease causing mutations within a single gene
- » Where variation in a single gene is likely to be the cause of the disease, but that gene is very large (many thousands of base pairs) and variants are widely distributed within it

In these cases it becomes more efficient and effective to ‘scan’ the entire sequence of one or more genes of interest, rather than to target analysis to particular known variants and loci. The process of nucleotide by nucleotide DNA sequencing potentially enables all types of clinically relevant variants to be identified across any part of the genome without any prior knowledge about the suspected type or location of the clinically relevant variation being required to conduct the assay.

DNA sequencing can provide a more comprehensive and detailed view of genetic variation than other non-sequencing techniques. Sequencing techniques do not require detailed prior knowledge of the sequence under investigation and allow for the discovery of novel variation at scale.

## 6.1 The sequencing pipeline

Retrieval of clinically relevant information from DNA samples by sequencing includes several stages which all impact upon the information retrieved and the subsequent interpretation of that information. Whilst DNA sequencing is the process by which DNA in a sample is converted into data that can then be analysed, the steps towards interpretation and prior to the actual retrieval of DNA for analysis are crucially important and substantially variable – there are many options for how this can be conducted.

These stages can be summarised as follows:

**Sample preparation:** encompasses the steps from extraction of a sample from a patient to the entry of a prepared sample library into a sequencing instrument for DNA sequencing.

**DNA sequencing:** covers the steps from entry of a prepared sample library into a sequencing system to retrieval of raw sequence information following the sequencing run.

**Bioinformatics analysis:** includes the processing of the raw sequence data into a format which is suitable for clinical interpretation.

**Clinical interpretation:** Interpretation of processed sequencing data, which may include further computational analysis alongside clinical professional insight, to provide an answer to the clinical question under investigation.

## 6.2 Sample and library preparation

Patient samples can be collected in one of several forms appropriate for the indication and circumstance, such as blood, saliva, or tissue/tumour biopsy. Samples must be appropriately processed before they can be run on DNA sequencing equipment.

A few sequencing systems now provide integrated sample and library preparation steps with sequencing, however the majority do not. Dependent upon the sequencing equipment to be used, different sample preparation methods may be recommended for each sample type. In general, sample preparation includes DNA extraction and library preparation. Following sample collection, these include the following stages:

### *DNA extraction*

DNA extraction is the isolation of DNA from other cell contents and debris. DNA extraction and purification are important as failure to remove contaminants can compromise results.

1. **Lysis** – the breaking open of cells and their nuclei in order to release DNA. This can be done through mechanical, chemical or enzymatic disruption.
2. **Precipitation** – steps involve the removal of unwanted cellular debris (e.g. proteins) and precipitation of DNA from water in which it is soluble. This can be achieved using phenol, chloroform, salts, ethanol and/or spin columns that bind DNA. It includes neutralising the charged DNA (making it more stable) in preparation for suspension in an ethanol solution.

3. **Washing and resuspension** – this step involves the washing of the extracted DNA using ethanol to reduce the presence of non-DNA impurities such as salts and phenol, and final suspension in ultra-clean water or buffer solution.
4. **DNA quality checks** – testing DNA quality is important, as it will affect the performance of subsequent steps and results retrieved. Quality checks can be performed on a gel or using specialised equipment such as the to check for excessive degradation or contamination.

### *Library preparation*

Steps in library preparation vary depending upon the sequencing system being used. Here, an example overview is given of library preparation for next generation sequencing (NGS).

During library preparation, DNA is prepared for sequencing by adding identifiers to strands of DNA and adding adapters that allow the DNA to adhere to the sequencing flow cell. Library preparation is an essential step prior to sequencing using most systems.

1. **Fragmentation or target selection** – DNA fragments may be required to fall within a specified length range for preparation. DNA is fragmented using one of several mechanical methods e.g. enzyme treatment, sonication or shearing with a needle. If a specific section of the DNA is required for analysis then this region may be amplified using PCR (see chapter 5.4) to generate what are known as ‘amplicons’ [\[57\]](#).
2. **Adapter ligation** – After fragmentation, the ends of the DNA are repaired and an adapter – a known sequence which allows for the attachment of strands to a surface for sequencing and sample identification – is ligated to the ends of the DNA.
3. **Size selection** – Adapter ligated DNA strands can be size selected using gels or beads. More similar fragment sizes lead to more efficient sequencing runs. This may not be required if using amplicons.
4. **Amplification** – If required, ligated DNA strands can be amplified by PCR to ensure enough sample is available for sequencing.
5. **Library quantification and quality control** – after library preparation, the sample library is quantified to ensure enough DNA is available for sequencing and to ensure the correct amount of sample is loaded onto the sequencer. The quality of the library is assessed i.e. examined for the presence of contaminants which may impede sequencing. Several of these steps can be done using, for example, qPCR or assessment on a Bioanalyzer.

Sample and library preparation kits are often sold by the companies that produce the associated sequencing equipment; however, a range of kits are also available from other suppliers and some of the preparation may be performed without commercial kits. The extent and type of sample and library preparation required depends upon several factors, including:

- » **Sequencing platform and type of sequencing** – Each sequencing platform provides recommendations on preparatory techniques. Platform will influence, for example: the amount of input DNA required, reagents and/or kits utilised, the type of adapters provided, whether PCR is required, and more.

- » **Amount and quality of starting material** – The amount of DNA retrieved following extraction will influence whether amplification (by PCR) is required. Without enough DNA, adequate coverage of genomic regions may not be achieved. If DNA is highly degraded e.g. formalin-fixed paraffin-embedded (FFPE) samples, a different method of preparation may be used such as FFPE-specific library preparation methods.
- » **Required read length** – Required read length will impact fragmentation and size selection steps. These are in turn impacted by the type of sequencing to be conducted and the target sequence type. For example, exome sequencing requires reads around 200bp in length, whereas micro RNAs (miRNA) are around 16-35bp in length.
- » **Time taken** – ‘Fast’ library preparation techniques have been devised for situations in which preparation time is restricted. Kits and reagents are available from several suppliers and can allow for sample preparation in under two hours. These can incur some trade-offs in quality and cost.
- » **Molecule being investigated** – If sequencing of RNA or investigation of DNA methylation is required, additional or different steps will need to be taken during sample preparation to ensure that these molecules are preserved and can be sequenced. Specific commercial sample preparation kits exist for extraction and preparation of particular types of DNA/RNA.

Selection and execution of appropriate sample and library preparation can subsequently impact upon: overall sequence quality, genomic coverage and uniformity, error rate, selection of bioinformatic pipelines, and even interpretation.

### 6.3 DNA sequencing

DNA sequencing is the reading of the order of nucleotides (bases) that make up sections of DNA – translating a physical molecule into digital information. From the emergence of few techniques developed in the 1970s, there are now hundreds of different sequencing systems available for this purpose. Many of these platforms use similar foundation techniques – Illumina is the largest (by revenue) of the sequencing technology companies globally, and utilises ‘sequencing by synthesis’ technology; these rely on DNA synthesis to generate sequence data [58]. Whilst many other methodologies exist, clinical sequencing in the UK relies on this and a handful of other methods. Manufacturers of sequencing systems frequently produce a range of platforms that differ in cost, infrastructure required, throughput and other capabilities; of which one or more may be most suitable to the desired application. Factors such as required coverage, throughput and paired/single end sequencing can be adjusted for on many systems. Sequencing platforms can be used to provide sequencing in several formats, including whole genomes sequencing (WGS), whole exome sequencing (WES) and targeted or panel sequencing. The desired format will also influence the system and associated tools of choice. These formats are discussed further in chapter 7.

In this section we provide details of key sequencing systems, covering:

- » Sanger sequencing
- » Next generation sequencing
- » Single molecule and long read sequencing
- » Complementary technologies

Sanger sequencing was developed by Fred Sanger and colleagues in the 1970s and has been built upon and modified from its original conception. It has been extensively used to read DNA in a wide range of contexts. The Sanger sequencing method was used between 1990 and 2003 in an international effort to retrieve the first 'full' human genome [59]. The Sanger method, also referred to as the 'chain-termination method' of sequencing, can be performed with relatively limited equipment and analysis to accurately read short sections of DNA in a laboratory setting. Sanger sequencing involves the synthesis of identical amplified DNA molecules which terminate at different read lengths and emit nucleotide-specific fluorescence at the termination point. Examining these reads on a size-separation gel allows for the sequence to be read. This process can be broken down as follows:

1. First a single selected section of DNA must be replicated using the polymerase chain reaction.
2. The amplified DNA is then degraded using heat to separate the two strands that make up each section of DNA, and a short primer is attached to allow for the subsequent addition of free nucleotides.
3. Free nucleotides of all four varieties (dNTPs) and DNA polymerase are added to four aliquots, alongside fluorescently-tagged 'terminator' versions of nucleotides (ddNTPs) - one of each type of ddNTP (A, C, G, and T) is added to each aliquot.
4. DNA synthesis then proceeds, generating a new DNA strand. The tagged nucleotides are incorporated into the forming strands randomly and infrequently, preventing the polymerase from incorporating any more nucleotides into that chain therefore terminating the synthesis of the strand where they are incorporated.
5. As the reaction proceeds, reads of different lengths are produced, which can then be separated by size on a gel and the fluorescent signal detected. Manual reading on gels or automated capillary-based electrophoresis can be used to read the sequence produced. This creates a map showing where specific nucleotides are present in the many copies of the sequence.

This method has evolved from the original form devised by Sanger, in which radioactively labelled fragments were separated on long polyacrylamide gels that had to be visually inspected to 'read' the sequence. Modern Sanger sequencing machines automate this process using fluorescent labels and capillary electrophoresis. Accompanying software is able to automate the determination of the sequence. Automated systems – sequencers – available for Sanger sequencing include Thermo Fisher Scientific (Applied Biosystems) SeqStudio, 3500 Series, and 3730 Series Genetic Analyzers. The older 3130 Genetic Analyzer is no longer available for purchase from the manufacturer, but is still used in some clinical genetics laboratories in England [38, 60, 61].

Sanger sequencing is highly accurate and can produce reads of up to around 900bp in length [62]. Sanger sequencing is particularly useful when there is the need to examine a 'gene of interest', single and small number targets can be examined cost-effectively using this technology. Sanger sequencing remains useful in the clinic and many clinical investigations still rely on its availability. In many circumstances, Sanger sequencing is considered the 'gold standard' of sequencing and is also frequently used to confirm results retrieved from other sequencing techniques, though the need for this is reducing as advances are made in the accuracy and reliability of other sequencing techniques.

In theory, Sanger sequencing can detect all forms of genomic variation from single base changes through to deletions or duplications of entire genes or regions of the genome. However, its use in diagnostic molecular genetics in practice is limited to detecting single base changes and small (1-100 base pair) insertions or deletions in the sequence being analysed. This is because it does not reliably identify all variations, and there are now much more cost-effective alternatives that can be used to detect large scale structural variants in the genome. It is also difficult to elucidate heterogeneous samples such as tumour samples using Sanger sequencing. Because it is low-throughput, Sanger sequencing is not the most appropriate technology now available for examining whole, large parts of, or multiple genomes.

### Next generation sequencing

Next generation sequencing (NGS) is the commonly used term for the high throughput sequencing methods first commercialised in 2006/2007. Many of these methods are based on DNA synthesis, as used in Sanger sequencing, but allow for 'massive parallelisation' of the process, meaning thousands or millions of DNA fragments can be sequenced and reads generated at the same time using a single machine. These technologies now supersede Sanger sequencing for many diagnostic applications.

NGS technologies are referred to as high throughput, meaning they are capable of producing massively increased output over older techniques, with some modern-day sequencers producing in excess of 3000GB of data in around two days [63, 64]; this compares to Sanger sequencing using automated capillary electrophoresis, which can produce 2.76Mb per day [62].

Several different technologies are included within the class of 'next generation sequencing technologies', each of which uses a slightly different approach. Whilst a select few of these technologies are in common use in research and to some extent in the clinic, many have fallen out of favour, have been superseded, or have never passed into use in UK healthcare. In the UK, Illumina produces the major NGS technologies being used for large-scale clinical and research applications (such as the 100,000 Genomes Project), but other systems are used within clinical settings. In this section, an overview of the following sequencing technologies is given:

- » Illumina sequencing (sequencing by synthesis)
- » Ion Torrent sequencing (semi-conductor sequencing)

### Illumina sequencing by synthesis

Sequencing using Illumina platforms relies on a technique called 'sequencing by synthesis'; it is the most commonly used system in clinical genomics for sequencing of multiple genomic regions. The main principle is the detection of nucleotides as they are incorporated in a complementary manner to single strand versions of the target DNA i.e. as DNA is synthesised. This is similar to Sanger sequencing in that modified (fluorescently tagged) bases are incorporated and subsequently detected. The major differences between these techniques is that during Illumina sequencing, the incorporation of a fluorescently tagged base does not result in termination of DNA synthesis due to subsequent removal of tags. The fluorescent tags attached to modified nucleotides can be washed away, allowing for more modified bases to be added subsequent to this point.

Illumina sequencing by synthesis occurs as follows:

- » Adapter-ligated DNA fragments (see 6.2) are then hybridised to a flow cell, a glass surface containing short nucleotide sequence probes (oligonucleotides or oligos)
- » Bridge amplification occurs where the DNA is replicated and, as it is reproduced, binds to local oligos, creating a cluster of amplified strands of DNA. Only one strand of each DNA fragment is retained. The term 'bridge' comes from the shape formed by the single stranded DNA as it binds to an embedded oligo at either end of the strand.
- » This occurs across the flow cell, creating clusters of identical sequences that, in combination, represent all DNA included in the original sample.
- » Sequencing begins as complementary nucleotides are incorporated into the hybridised DNA strand. Each type of nucleotide (A, T, C or G) has a distinct fluorescent tag. These nucleotides are incorporated into the forming strand one by one. As each nucleotide is incorporated, a fluorescent signal is emitted, which can then be detected and the type of nucleotide (and therefore its complementary base) inferred. This happens simultaneously across millions of DNA strands across the flow cell.
- » The fluorescent tags are removed by 'washing' in between each incorporation step to allow synthesis to continue. The number of cycles of synthesis and washing determines the length of the read returned.
- » Bridge amplification is repeated, this time retaining only the reverse strands, and the process of sequencing by synthesis is repeated to obtain information about the same stretch of DNA.

Example platforms available for Illumina sequencing include the HiSeq, MiSeq, NextSeq, and NovaSeq, which differ in size, capacity, and cost. Whilst the high throughput NovaSeq is better suited for sequencing multiple whole human genomes, the MiSeq in particular is suited for panel-based targeted sequencing (see chapter 7.5).

Illumina NGS sequencing has several advantages over conventional Sanger sequencing, the major advantage being throughput. NGS is capable of producing a significantly higher number of reads than Sanger technologies in a certain time period. The higher throughput Illumina sequencing systems (e.g. NovaSeq) provide the ability to sequence multiple whole human genomes in a single sequencing run. These systems are costly, very large and require specialised infrastructure e.g. cooling systems and reinforced floors. Whilst still highly accurate, Sanger sequencing is considered the more accurate technology across short sequences. Illumina sequencing by synthesis can cope better with short homopolymeric regions – regions that contain many copies of a single type of nucleotide in sequence - than some other NGS technologies.

Illumina have built partnerships with Genomics England and the Scottish Genome Partnership, and provide sequencing services for these and other large sequencing projects internationally. Illumina sequencing platforms are used by the majority of clinical genomics services across the UK.

### Ion Torrent sequencing

Also referred to as semi-conductor sequencing, Ion Torrent sequencing by Thermo Fisher uses a semi-conductor chip and a bead-based system to allow for the sequencing of DNA fragments. This technology also works by DNA synthesis, but instead of detecting light, the sequencer detects hydrogen ions that are released as nucleotides are incorporated.

DNA is fragmented and each fragment is bound to an individual bead. The fragment is copied until the copies cover the surface of the bead. The bead is then washed into one of thousands of wells on the surface of the semi-conductor plate, so that each well contains only one bead. Free nucleotides are washed over the chip and wells. Nucleotides of one type only are added each time i.e. A, C, G or T. If a nucleotide is incorporated into the strands anchored in the beads (i.e. if there is a base-pair match), then a hydrogen ion is released and a voltage change detected by the sequencer. Where two or more bases of the same type are present in succession on the target strand, then multiple bases are incorporated - the voltage change is relative to the number of bases incorporated, so multiple additions can be detected. This process happens simultaneously across up to millions of wells and different DNA strands.

Sequencing platforms include the Ion GeneStudio S5, Ion Torrent Genexus, Ion Personal Genome Machine (PGM) and Ion Proton.

### Other NGS systems

Until recently, Thermo Fisher Scientific SOLiD and Roche 454 sequencers were also used in clinical laboratories. Production of 454 sequencers ceased in 2013, and of SOLiD sequencers around 2016. Although no longer used in NHS clinical genetics laboratories in Scotland, pyrosequencing using 454 platforms and small-scale DNA sequencing using SOLiD platforms continues in some laboratories in England for purposes such as methylation analysis or variant detection in small genomic regions (see section 6.5 for more on methylation sequencing).

### Strengths and weaknesses of NGS technologies compared with Sanger sequencing

Next generation sequencing (NGS) has some broad advantages and disadvantages compared with Sanger sequencing, in addition to platform specific considerations. NGS technologies provide the following opportunities and challenges for clinical genome analysis:

#### Advantages

- » **High throughput** – The throughput of NGS machines varies by manufacturer and model, but at the highest end there are now sequencers able to deliver 3000GB of data, equivalent to around fifteen human genomes at 30x coverage in around 2 days in a single sequencing run [63, 65] (estimate 4-48 from Illumina dependent upon flow cell technology [64]). Unlike Sanger sequencing, NGS is a 'massively parallel' method, enabling billions of template DNA fragments to be sequenced simultaneously.
- » **Batching** – NGS allows for multiple independent patient samples to be sequenced in a single sequencing run (multiplexing). This allows for batching of samples, which can significantly reduce the per-patient cost of analysis.

- » **Flexibility** – Depending on the platform, the range of different tests that can be performed is very wide, from targeted single gene analysis through to ‘panels’ of genes implicated in particular disorders and ultimately whole genome analysis. Laboratories can either use pre-fabricated kits from manufacturers that enable targeted sequencing of single genes or gene panels, or can custom build their own assays using primers similar to those used for PCR. The easy customisability of NGS assays is significant, as it allows laboratories to add new genes into their analyses in response to new knowledge of gene-disease associations.
- » **Lower cost** – The cost of sequencing a single whole genome have reduced significantly since the introduction of NGS in 2007. Whilst sequencing alone now costs less than \$1000 [66] (~£820 at May 2020 exchange rates), this does not reflect the full cost of generating a clinically actionable report from whole genome sequenced, which often requires more than one sample. When all costs are taken into account, including sample collection and preparation, sequencing, bioinformatics and reporting, the costs are £6841 per cancer case (tumour and germline sample – two genomes) and £7050 per rare disease case (trio – three genomes), therefore £2350-£3420 per genome [67]. By comparison, it would still cost around £6 million to sequence a single genome using Sanger sequencing.

### Disadvantages

- » **Errors** – Individual sequence reads generated by NGS are more likely to contain errors than Sanger sequencing. This is usually not a problem, because a sufficient number of parallel reads covering the same region of DNA can be generated to distinguish true variants from sequencing artefacts. However, NGS platforms are vulnerable to some systematic errors, such as difficulty sequencing repetitive regions e.g. homopolymeric or duplicated regions of DNA sequence. This is an important limitation to note in the context of clinical genome analysis service provision, as a number of genetic diseases, such as spinocerebellar ataxias and Huntington’s disease, are caused by increases in the length of repetitive DNA sequences, and the precise determination of the length of the repeat is vital for accurate diagnosis.
- » **Informatics burden** – NGS requires a far greater level of bioinformatic expertise, computational infrastructure and time to analyse and interpret data than Sanger sequencing. This is partly due to the need to reassemble the enormous quantity (billions) of short reads from a single run accurately into longer contiguous stretches of DNA sequence, and to accurately map and interpret vast numbers of potentially clinically significant variants. The applications for which NGS is more suited i.e. whole genome or larger scale analysis also require greater informatics input than examination of small genomic variation does. The increased availability of purpose-built bioinformatic tools and pipelines designed to aid NGS analysis have helped to address the informatics challenge. The increasing availability and use of high performance computing clusters (often linked to academic centres) or commercial cloud computing is helping to provide the computational capacity required for analysing the large amounts of DNA sequence data generated by NGS technologies.
- » **Turnaround times** – As a consequence of relatively long run times for some NGS instruments, the need to batch multiple samples to ensure full utilisation of the instrument, and time taken (days to months) to analyse large volumes of NGS data, turnaround speed for diagnostic clinical genome analysis using NGS can be slow. NGS approaches may be less appropriate where rapid access is required to a limited quantity of sequence information e.g. urgent prenatal tests looking for a single known variant in a gene.

## Single molecule and long read sequencing

Sometimes referred to as the 'third generation sequencing technologies' after Sanger (first) and NGS (second), long read or single molecule sequencers use different technologies to read substantially longer contiguous strands of DNA than other sequencing platforms. These techniques allow for molecules of more than 100,000bp to be read contiguously, although more commonly these are 10,000 – 100,000 base pairs in length [68]. There are two main providers of long read technologies: Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio). These are UK and US based, respectively. These technologies are relatively new – PacBio's RS sequencer was released commercially in 2011 and ONT's MinION sequencer became commercially available in 2015. Although quite widely used in research, their use in clinical applications is currently very limited.

Long read sequencing technologies are advancing rapidly and are likely to become much more widely used. There are inherent advantages in the generation of long reads compared to short read techniques, and specific advantages and disadvantages associated with the available platforms used to conduct long read sequencing.

**Benefits of producing long reads:** Longer sequences of nucleotides produced by sequencing (reads) are more likely to appear distinct from one another compared with shorter reads. This allows them to be computationally reassembled with less ambiguity. This may be particularly useful for: detection of larger mutations and complex structural features or variants, gene fusion events and chromosomal rearrangements, phasing of alleles, highly homologous regions, and highly polymorphic regions such as the HLA region.

In addition to the advantages of longer reads, the flexible sequencing depth and increased speed of these sequencers may enable same-day analysis and reduce the need for batch sampling in some circumstances. The DNA is read in real-time, meaning results are generated as the DNA is being sequenced, and sequencing can be terminated at any point in the process. These systems also offer amplification-free sequencing, potentially removing some bias often introduced at this stage. Both PacBio and ONT systems use single-molecule sequencing, allowing direct examination of the molecule in question. This enables the examination of epigenetic base modifications simultaneously to nucleotide sequencing.

### Oxford Nanopore Technologies

Oxford Nanopore Technologies' (ONT) technology works by threading single stranded DNA/RNA molecules through a membrane-bound biological protein pore – a 'nanopore'. ONT's systems are designed to be relatively mobile, generate ultra-long reads and be more accessible to those with less expertise.

A sample of DNA is collected and prepared by ligation of a motor protein and adapter sequence at the ends of the DNA strands. The sample is then pipetted into the sequencing platform, which contains a flow cell with embedded nanopores. During sequencing, the tagged DNA is fed automatically through one of the pores; as the molecule passes through the pore, it causes a disruption in the current across the pore – each type of base (A, C, G, T) will cause a slightly different disruption. This pattern of disruption can then be read to determine the base sequence of the molecule. The molecule is read at approximately 400 bases per second, per pore.

In the past, the technology has faced challenges around accuracy, limitations due to the need for high quality input material, and difficulties distinguishing between bases in homopolymeric regions. However, some of these concerns are being overcome as the technology advances and associated bioinformatics pipelines are improved.

ONT offer sequencing on several systems. Listed by increasing throughput and size, these are the portable MinION, and benchtop high-throughput systems GridION and PromethION. The sequencing systems vary in cost from £820 for the lowest capacity system (MinION) to £132,100 for the highest capacity system (PromethION) (correct as of March 2020 [69]).

### Pacific Biosciences

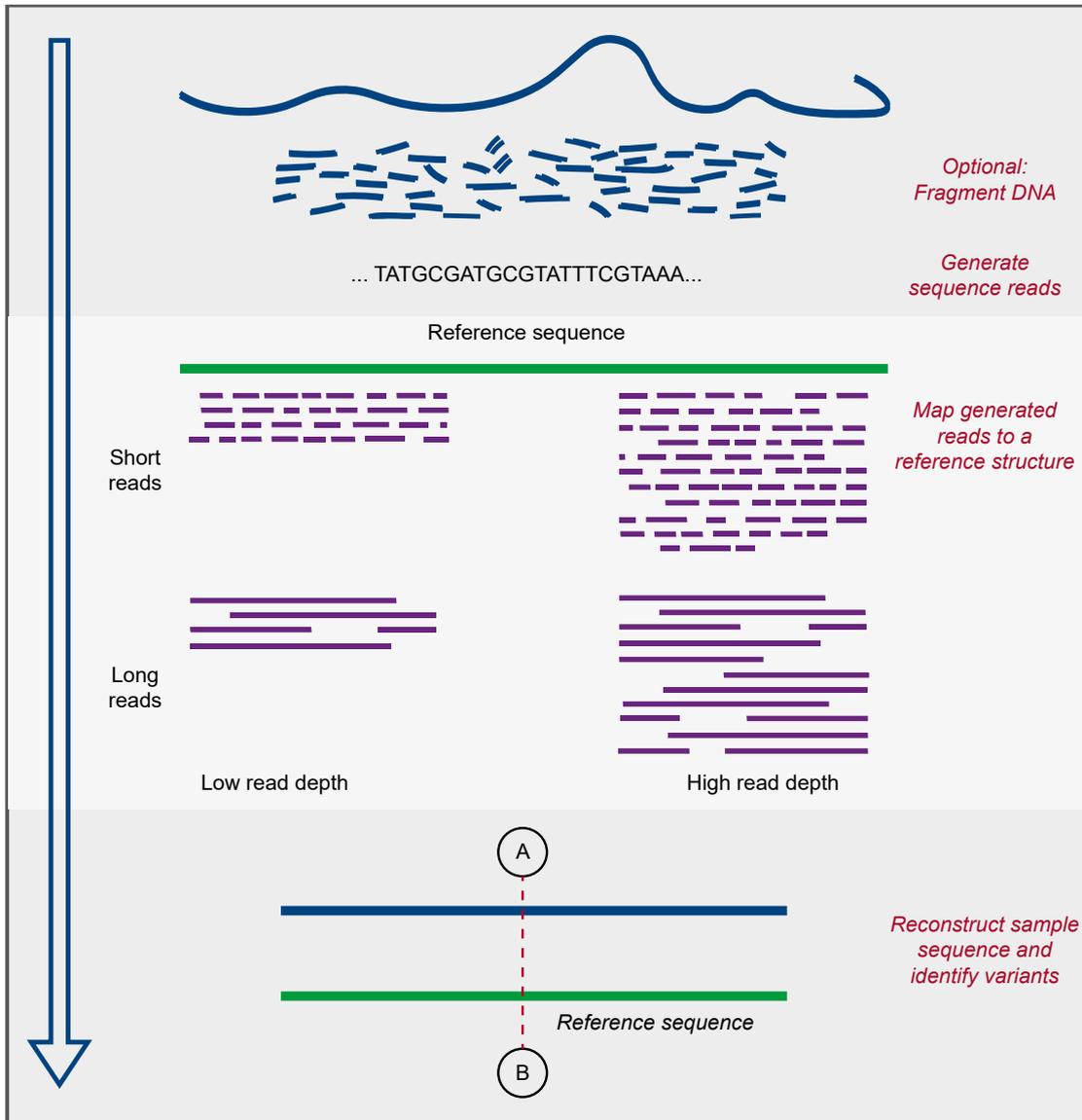
Pacific Biosciences' (PacBio) sequencing technologies utilise the process of DNA replication, single-molecule real-time (SMRT) sequencing and precision light detection to generate high-accuracy long read sequence data.

Sample DNA is isolated and converted into a loop, known as a 'SMRTbell'. These DNA loops are then immobilised in microscopic pits embedded in the sequencer's 'SMRT cells'. These pits are zero-mode wave guides (ZMWs), which contain a tethered DNA polymerase and are open to a pool of free phospho-linked nucleotides. As complementary nucleotides are incorporated into the sample DNA, light is emitted which can be measured by the sequencer. Each of the four nucleotides emits a specific signal due to the different labels, allowing the nucleotide present to be determined and recorded. This process repeats as the DNA polymerase goes along the DNA strand and the sequence is read via the light signals. DNA polymerase can synthesise more than 10 bases per second, and up to around a million ZMWs can be multiplexed to sequence large quantities of DNA in parallel on a single Sequel system. PacBio offer single molecule real time (SMRT) DNA sequencing on their newer Sequel system (Sequel II) and older PacBio RS II machine.

Like many sequencing platforms, PacBio systems require notable investment, with the newest Sequel II system priced at \$495,000 for its launch in 2019 [70]. The older Sequel system cost \$350,000 on release in October 2015 [71]. The release of the Sequel II system saw a marked increase in throughput over previous systems – at the end of 2019, PacBio reported their early customers were achieving around 160GB per SMRT cell [72]. The PacBio system offers extremely high accuracy under certain conditions, which further limits throughput but can achieve accuracy similar to that gained through Sanger sequencing.

PacBio provides some tools for bioinformatic analysis, and additional tools are being developed by researchers. Sequencing with PacBio platforms is not used in clinical settings in the UK, but has been used in research internationally, predominantly to aid in the assembly of whole genome sequences.

Figure 5. Summary of short- and long-read sequencing



### Complementary technologies

Technologies that do not directly utilise sequencing but are complementary to sequencing technologies may be employed to increase accuracy or provide context to sequencing outputs. These include: optical mapping systems, such as BioNano Technologies' Saphyr system; cell isolation systems such as those produced by 10x Genomics; and 'synthetic long read' systems such as those produced by Illumina.

These systems can expand the functionality of existing equipment and it may be possible to integrate them into existing workflows.

- » Optical Mapping platforms use tags applied across long strands of DNA to allow visualisation of various genomic reference points which can then be used to construct consensus genome maps.
- » 10x Genomics platforms utilise microfluidics to isolate single cells and nucleic acid molecules. These molecules can then be barcoded in such a manner as to provide high-throughput single cell sequencing data or linked-read data ('synthetic' long reads) when combined with conventional NGS short-read sequencing.
- » 'Synthetic' long read sequencing (or linked-read) technologies are typically based on high throughput short-read sequencing technologies but use additional computational or molecular methods to provide information akin to that retrieved from longer contiguous reads.

These technologies are currently limited to research use.

### BioNano Technologies

In 2017, BioNano Technologies released the Saphyr Genome Imaging Instrument. The Saphyr system is not a sequencing system, but works in complement with sequencing systems to facilitate the generation of longer-range genomic information from samples which may also be used for short read sequencing. The primary purpose of the optical mapping platform is for the identification of structural variation and other large genomic changes. As previously discussed, short-read NGS technologies have many advantages, but these technologies are limited including their ability to correctly identify repetitive regions, structural variations and chromosome rearrangements. Rather than offer an alternative to the sequencing technologies listed above, the Saphyr system represents an alternative to cytogenetic techniques such as karyotyping.

The Saphyr system utilises ultra-long DNA molecules, extracted using a proprietary extraction technique, which are then tagged at specific points across the genome. The system linearises the ultra-long DNA molecules and images the labelled DNA, scanning across the whole genome. The accompanying informatics pipeline converts the images into information that can be further analysed, and is then used to align the sample genome to consensus genome maps. Any variation detected can be investigated further.

It is suggested that the BioNano Saphyr system can produce 100x coverage of 6 human genomes per day. However, it is worth noting that although the term 'coverage' is the same as is used in above sections to describe the extent of sequencing conducted on a sample, this term describes retrieved information that is less detailed than that which can be obtained from sequencing the sample.

Studies have taken place internationally examining the suitability of the Saphyr system for use in the diagnosis of haematological cancers, comparing its functionality against conventional cytogenetic techniques such as karyotyping, FISH and southern blotting (see Chapter 5 for more information on these techniques) [\[73–76\]](#). The studies generally suggest concordance of results with those gained through conventional techniques and

some advantages in terms of expanded functionality and ease of use, with optical mapping being culture-free and requiring only a single platform for detection of different structural variants. However, the studies are few in number and many have not been published in peer-reviewed journals.

## 10x Genomics

10x Genomics' systems provide the ability to perform high-throughput single cell sequencing, produce linked reads and examine chromatin state to identify cell types and states (Assay for Transposase Accessible Chromatin Sequencing - ATAC-seq) by combining 10x systems with conventional NGS performed on a separate short-read sequencing system. 10x systems enable the user to identify sequence reads from single cells by isolating individual cells or strands of nucleic acids in droplets and adding a unique barcode to the DNA or RNA. This index allows reads originating from the same cells to be identified post-sequencing.

The 10x systems also facilitate production of linked reads; these are produced by tagging (barcoding) individual single molecules of DNA prior to short-read sequencing, providing additional information regarding their position in the context of the genome. This means the short reads can be identified as originating either from the same or different long single molecules of DNA. This provides information similar to that gained through long-read sequencing and can be referred to as producing 'synthetic' long reads. Much like true long reads, linked reads can provide haplotype phasing information and enable more accurate examination of heterozygous regions without the need for trio sequencing.

10x Genomics currently produces two systems - the Chromium Controller, and the larger Chromium Connect. 10x systems are compatible with several types of NGS sequencer including those produced by Illumina and BGI.

## Synthetic long reads

Several companies are now developing systems for the production of 'synthetic' long reads. These involve short-read sequencing of nucleic acids, followed by computational construction of 'long reads' from short-read data, often facilitated by barcodes unique to longer sections of DNA applied before the molecules are sequenced.

Illumina's synthetic long read sequencing service utilises tagging and short read sequencing of nucleic acids to produce synthetic long reads. Illumina's system utilises long read fragments of approximately 1.5-10kb in length which are separately amplified before being sheared, tagged, and sequenced. This is currently provided either as a service by Illumina, or can be purchased as a substitute library preparation kit to run with an Illumina HiSeq sequencing machine. A workflow and informatics app specifically designed for the synthetic long reads, called 'TruSeq Long-Read Assembly App' has been produced by Illumina.

Other companies such as Loop Genomics also produce library preparation kits for synthetic long read sequencing [\[77\]](#).

**Table 2: Summary of the strengths and limitations of the different sequencing technologies for clinical genome analysis**

<b>Technology</b>	Sanger Sequencing	
<b>Generation</b>	1st Generation	
<b>Strengths</b>	<b>Limitations</b>	<b>Especially useful for</b>
<ul style="list-style-type: none"> <li>• Extremely high accuracy</li> <li>• Simple workflow and interpretation</li> <li>• Long reads produced (up to 1000bp) Sequence multiple samples without barcoding</li> </ul>	<ul style="list-style-type: none"> <li>• Low throughput</li> <li>• Low discovery power</li> <li>• Not cost-effective for larger target numbers or WGS/WES</li> </ul>	<ul style="list-style-type: none"> <li>• Single target 'gene of interest'</li> <li>• Confirmation of NGS results</li> <li>• Examining broad range of variants in single or small number of targets</li> </ul>
<b>Technology</b>	Illumina Sequencing	
<b>Generation</b>	Sequencing by synthesis, 2nd Generation (NGS)	
<b>Strengths</b>	<b>Limitations</b>	<b>Especially useful for</b>
<ul style="list-style-type: none"> <li>• High throughput</li> <li>• Low cost per base</li> <li>• Higher coverage enables higher accuracy</li> <li>• Can detect several types of variation</li> <li>• Low sample input required for high quality samples</li> </ul>	<ul style="list-style-type: none"> <li>• Short reads produced (75-300bp)</li> <li>• Relatively long run times</li> <li>• Platforms are expensive</li> <li>• Expensive when sequencing a low number of targets</li> <li>• Quality of read deteriorates as length increases</li> </ul>	<ul style="list-style-type: none"> <li>• Panel sequencing (miSeq platform)</li> <li>• Simultaneous analysis of multiple samples</li> <li>• <i>De novo</i> genome assembly</li> <li>• WGS and WES</li> </ul>
<b>Technology</b>	Ion Sequencing	
<b>Generation</b>	Semi-conductor sequencing, 2nd Generation (NGS)	
<b>Strengths</b>	<b>Limitations</b>	<b>Especially useful for</b>
<ul style="list-style-type: none"> <li>• Relatively fast and inexpensive (some platforms)</li> <li>• Can provide smaller outputs within a few hours</li> <li>• Can detect several types of variation</li> </ul>	<ul style="list-style-type: none"> <li>• Relatively low output - large-scale sequencing projects</li> <li>• Long run times</li> <li>• Relatively short reads produced (&lt;400bp, normally c.200bp)</li> </ul>	<ul style="list-style-type: none"> <li>• Quick analysis of short sections of DNA</li> <li>• Panel sequencing</li> </ul>

<b>Technology</b>	Oxford Nanopore Technologies (ONT)	
<b>Generation</b>	Nanopore sequencing, 3rd Generation	
<b>Strengths</b>	<b>Limitations</b>	<b>Especially useful for</b>
<ul style="list-style-type: none"> <li>• Rapid, flexible run time</li> <li>• Long and ultra-long reads possible (up to ~800K bp)</li> <li>• Low cost sequencer</li> <li>• Some platforms are highly mobile</li> <li>• Can detect a wide range of variation</li> </ul>	<ul style="list-style-type: none"> <li>• Higher error rate</li> <li>• Systematic errors in homopolymeric regions</li> </ul>	<ul style="list-style-type: none"> <li>• Complex genomic rearrangements</li> <li>• Repetitive regions</li> <li>• Sequencing in the field</li> <li>• Simultaneous Extraction of epigenetic information</li> </ul>
<b>Technology</b>	Pacific Biosciences (PacBio)	
<b>Generation</b>	Single molecule realtime sequencing, 3rd Generation	
<b>Strengths</b>	<b>Limitations</b>	<b>Especially useful for</b>
<ul style="list-style-type: none"> <li>• Rapid, flexible run time</li> <li>• Long reads</li> <li>• Very high accuracy at lower throughput setting (consensus sequencing)</li> <li>• Can detect a wide range of variation</li> </ul>	<ul style="list-style-type: none"> <li>• Higher accuracy = shorter reads</li> <li>• Platforms are expensive</li> </ul>	<ul style="list-style-type: none"> <li>• Complex genomic rearrangements</li> <li>• Repetitive regions</li> <li>• Extracting epigenetic information</li> </ul>

## 6.4 Data interpretation

Interpretation of genomic data involves a number of steps that can be divided into two key sub-categories: bioinformatics analysis to interpret the outputs from DNA sequencing, and variant interpretation to further analyse the clinical relevance of findings. Variant interpretation is expanded on further in chapter 7, sections 7.7-7.10.

### *Bioinformatics*

Bioinformatics is the use of informatics processes and data in biological contexts, and is a field of science and study in its own right [78]. Data analysis that uses bioinformatics techniques is required to interpret the outputs from DNA sequencing. The type and extent of analysis performed is dependent upon the sequencing platform used, the amount of genomic information being examined, and the context in which the sample is being considered.

The purpose of effective bioinformatics analysis is to detect all actual/true/real genomic variation in the examined sequence, accurately identify and discount false/unreal variation, and best allow for subsequent accurate interpretation of such true variants and distribution of information by a trained professional.

Sanger sequencing requires relatively simple examination of retrieved bases using base-calling software and simplified pipelines. As Sanger sequencing is normally applied to shorter regions of the genome, analysis is relatively simple, involving the identification of discordant bases between short sections of sequence.

NGS technologies are used to generate whole genome or whole exome datasets and require significantly greater bioinformatics input. The data produced by NGS platforms is of little use to clinicians in its raw format. NGS sequencing technologies, such as those produced by Illumina, produce image files of the fluorescent signals given out during the sequencing process. These are automatically 'base called' and thus adapted into binary base call (.bcl) files which are the output of the sequencer.

Bioinformatics analysis is performed in pipelines. Pipelines can be created or altered to suit a particular application e.g. the examination of suspected causal variants in the case of rare disease or differential expression analysis of RNA sequencing data. Once generated, pipelines can be used repeatedly, and can be made accessible to non-experts, although generation and alteration of pipelines is performed primarily by expert bioinformaticians. Commercial companies, academic researchers and clinical research laboratories generate bioinformatics pipelines suitable for analysis of sequencing data. In the NHS, bioinformatics pipelines can be from commercial providers or produced by in-house bioinformaticians; pipelines will vary between different trusts and hospitals. Although pipelines may differ, appropriate standards of sensitivity and accuracy must be met (as per Association for Clinical Genomic Science (ACGS) guidelines) in order for the pipelines to be used in a clinical context, with even minor changes to pipelines requiring validation before use.

In both research and clinical contexts, the choice and customisation of analysis pipeline(s) can have a profound effect on the interpretation of genomic information, for example resulting in miscalling of bases or misalignment. Each stage of the pipeline involves consideration of the context of testing and the circumstances surrounding sample preparation and sequencing; in addition, each stage has associated challenges.

Key stages in bioinformatics are:

1. Quality control – trim adaptors and remove low quality base calls
2. Alignment – map reads to a reference genome
3. Variant calling – determine which regions/bases differ from the reference
4. Variant filtering and interpretation – identify and prioritise functional variants with potential association to phenotype(s) of interest
5. Variant confirmation through additional techniques e.g. qPCR, Sanger sequencing

Bcl files produced by Illumina sequencers require conversion to FastQ prior to downstream analysis. Quality control is then performed and any remaining adapter bases are removed from the reads. These cleaned reads can then be mapped to a reference genome to determine the genomic origin of the reads. If performing gene expression (RNA) analysis, the number of reads at each location can then be counted and, following normalisations steps, compared between samples and transcripts.

Interpretation of genomic data relies on a series of statistical inferences. Each base in a gene or genome is likely to be represented in multiple overlapping reads, and the sequencing process itself can introduce errors into these reads. Knowing the correct identity of the nucleotide (A, T, C or G) at any position in the gene or genome therefore depends both on the use of probabilistic methods to determine whether there are enough reads covering that position, and sufficient agreement between them, to be sure that the identity of the base that is 'called' is correct. Differences in bioinformatic pipelines can impact upon interpretation and potentially diagnostic accuracy in several ways, and through several means.

### Construction and maintenance of bioinformatic pipelines

The diagnostic pipelines used for bioinformatics analysis will often contain a mix of software that has been developed internally, and that developed externally. External software can be open source, developed collaboratively, or commercial [\[79\]](#).

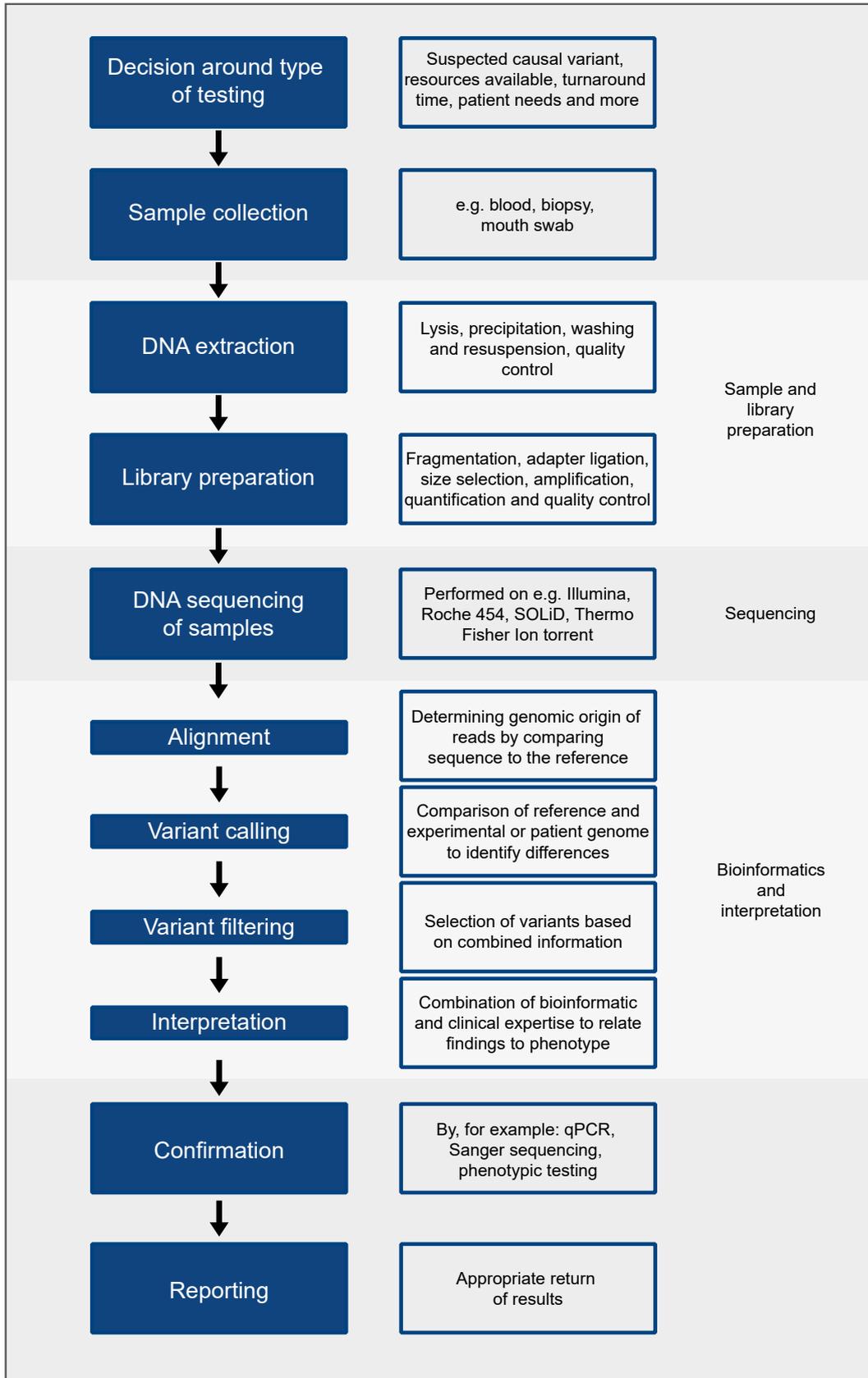
In-house software can be developed either for a specific task or to complement existing external software, but bioinformatics expertise is required in-house to maintain, upgrade, and validate these programs as and when required. The ACGS guidelines state that it is essential that all software or pipeline updates are verified to maintain equivalent or improved analysis standards [\[80\]](#).

The advantages of using externally available software within pipelines are that the development costs in terms of time and infrastructure are outsourced, but the laboratory will still need to validate the software in house and re-validate every time it is upgraded. Externally-produced accredited pipelines will also have been developed with reference to, and make use of, larger and more varied datasets than those available to laboratories in-house.

Cloud technology – which delivers services via the internet with varying degrees of automation – is becoming more widely used, in terms of access to external analysis tools. The advantages of using cloud services are that they reduce software and hardware demands on the users' side and give them the flexibility to increase capacity without requiring additional infrastructure or personnel. However, pipelines have to be developed to work in the cloud environment and there can be lack of flexibility in terms of the form and function of programs in the cloud [\[81\]](#). End-users will need to consider the trade-off between the relative advantages and disadvantages of using cloud services, and act according to clinical need and resource availability.

The ACGS guidelines outline guidance and recommendations in terms of ensuring ongoing accessibility, compatibility and version control of internally- and externally-produced software and/or pipelines [\[79\]](#).

Figure 6. Summary of sequencing and bioinformatics workflow



For internally developed tools, the guidelines recommend that where appropriate, software should be published in online repositories such as GitHub or bioarxiv.org. In addition, a system of version control, such as Git, that records all changes and accompanying notes and data should be used to support any future efforts to trace results obtained using the software/pipeline, and any future audits.

In terms of considerations around software version control and compatibility for external software, the guidelines state that developers are required to consider preferably using heavily documented software with an active support network for users, having a system in place for regularly checking and documenting software updates and bug fixes. Whenever a laboratory uses external software for diagnostic purposes, key details should be recorded such as the details and version of the software used, as well as the reference genome version, annotation sources and other reference data.

## Considerations and challenges in bioinformatics analysis

### Read depth

Read depth is the number of times in a sequencing run a particular position in DNA under investigation is read. The greater the number of reads covering any position, the greater the confidence that the identity of the nucleotide at that position can be called accurately. This is particularly important for detecting heterozygosity - where the two copies of a gene differ at a given position. Many laboratories will aim for a read depth of 30x (less is required for Sanger, 454 sequencing and some long-read technologies) in order to obtain sufficient genome coverage and confidence in calling a heterozygous variant [\[82\]](#).

Read depth is rarely even across all regions sequenced. Inherent chemical properties of different DNA sequences mean some areas of the genome are more amenable to sequencing than others. Easy to sequence regions will have more reads. In practice this means that some genes and regions of the genome cannot be sequenced with sufficient depth to achieve diagnostic levels of accuracy. Alternative methods may be needed to investigate these regions.

Sequencers have a maximum number of reads that can be achieved per run; these can be spread thinly over a wide range of different DNA sequences or samples e.g. a whole genome or multiple patient samples, or can be focused on a low diversity of sequences e.g. a few genes from a single patient, to achieve much higher coverage. Molecular genetics laboratories can optimise these variables for the different uses of NGS and calculate the number of samples that can be run at the same time.

Flexible read depth can be a strength for certain diagnostic applications, particularly where the goal is to detect heterogeneity in a sample. For example, tumours are highly heterogeneous, with different cells potentially carrying different genetic variants. If a clinically significant mutation in a gene (that may drive drug resistance or predict metastasis) is present initially in only a small proportion of cells, then sequencing by the Sanger method is unlikely to detect this. Using NGS, the gene can be sequenced many times over; 100x or even 1000x coverage is possible. These reads can be compared using specialised software that reports the presence of low levels of potentially clinically significant mutations.

## Coverage

The term 'coverage' can be used to refer to one of a few subtly different concepts in the context of sequencing, and can easily be conflated with read depth. There is little consistency in the use and distinction between the terms read depth and coverage; however, useful definitions of the two do exist.

Read depth and coverage are intrinsically linked, but whilst read depth refers to the number of reads aligned to a particular single point in the genome, coverage describes the average number of reads that align to a known reference and does not provide a description of the distribution of reads. Read depth can vary significantly between different parts of the genome. Coverage describes a comparison between sequence reads and a reference, whereas read depth provides an absolute number of reads at a given position.

Coverage is sometimes used to refer to:

- » The mean average number of reads that align to (cover) a reference base or region of the genome (mean coverage)
- » The number of aligned reads that overlap a single coordinate following sequence alignment
- » The proportion of the whole genome or predetermined reference that is read (covered) during sequencing
- » The proportion of the whole genome or predetermined reference that is covered by a specified depth of reads e.g. 90% genome coverage at 10x read depth

In the context of clinical sequencing, coverage is defined as being either vertical or horizontal. Vertical coverage is a measure of how many times a particular base or region has been sequenced – average or absolute read depth for that region or base. Horizontal coverage refers to how much of the genome is covered by reads (a required depth may or may not be specified) [\[83\]](#).

Vertical coverage can help healthcare professionals determine what level of confidence can be applied to retrieved sequence information. Horizontal coverage gives an indication of the equality of coverage across a genomic target, indicating that results may not be equally reliable across the whole target. Coverage requirements differ depending on the sequencing approach applied (see chapter 7) and the type of sample being sequenced. Minimum coverage of cancer samples is greater than that of germline samples owing to greater heterogeneity [\[67\]](#). The ACGS practice guidelines state that it is essential to examine robustness and reproducibility of findings particularly through the examination of coverage. Minimum values of vertical coverage for diagnosis are determined based on 'the required sensitivity of the assay, the targeting/sequencing method and the type of mutation detected'. The guidelines also state that horizontal coverage should, at a minimum, include 'the coding regions of the gene and the invariant acceptor and donor splice sites'. If these minimum requirements are not met, further testing is recommended [\[83\]](#).

Comparative examination of vertical coverage within and between samples can also aid in the identification of copy number variants. Several open access and commercial tools exist for the analysis of comparative vertical coverage.

### **Quality of reads**

Quality scores are generated from raw sequence data, based on coverage and alignment confidence. It reflects the confidence that the base call is correct. The threshold for quality can be modified, and it is possible to retain base calls with low quality scores. The higher the quality score, the more likely the base call is to be true and real.

Errors incurred during sample and library preparation (steps prior to sequencing) will not be detected or accounted for during bioinformatics analysis. Alterations to pipelines can be made to ameliorate some of these issues, but these will need to be identified and considered in addition to errors induced during sequencing and analysis.

### **Assembly and mapping**

Massively parallel sequencing produces large quantities of 'raw' DNA sequence at low cost and high speed. Turning this raw data into accurate and interpretable information presents a computational challenge. This involves reassembly of short reads in the correct order (akin to reassembling a jigsaw puzzle using a picture as a guide) with sufficient accuracy to be able to reconstruct the sequence of the original gene in its entirety and detect any clinically significant variation with diagnostic accuracy.

*De novo* sequencing, where a genome is sequenced and reconstructed from scratch without the use of a reference genome is much more taxing than genome analysis using a reference. *De novo* sequencing is not performed for clinical purposes.

### **Pseudogenes**

Pseudogenes are imperfect copies of functional genes that can exist anywhere in the genome. The interpretation of pseudogenes in any context is complex, as not much is understood about the function of these sections of DNA. What implications these have in a clinical context is often unknown, and their similarity to functional genes can cause issues in interpretation and analysis.

### **Confirmation of findings**

To ameliorate some of the potential for incorrect calling of variants, confirmatory analysis using alternative methods (e.g. qPCR, Sanger sequencing for low numbers of genes) can be employed prior to clinical reporting.

## **6.5 Associated technologies and techniques**

DNA sequencing is helping to advance understanding of disease and provide answers for patients and clinicians across a range of circumstances. As sequencing technology continues to advance and become more robust, additional sequencing methodologies, including the sequencing of different nucleic acids and molecules, is becoming more common place. Although some of these techniques are not yet extensively used in clinical laboratories, they are likely to become more prominent in the next few years. The sequencing element of these techniques can be performed on one or more of the above listed sequencing platforms; sample preparation and bioinformatics steps differ, but are not completely unrelated, from those employed for DNA sequencing.

### *RNA-sequencing (RNA-seq)*

RNA sequencing does not involve directly examining the genome itself, but a product of it: RNA. There are many different forms and functions of RNA, but the most frequently referred to is messenger RNA (mRNA). mRNA is the intermediary molecule between DNA and proteins - DNA is first transcribed into RNA, where it is spliced and further processed, before being translated into proteins. Examination of RNA can allow investigators to assess when and to what extent genes are expressed in specific tissues at specific time points. The collection of all RNA in a sample is referred to as the transcriptome, the examination of which is transcriptomics; this offers a snapshot of gene expression at a particular time point in a particular tissue.

RNA-sequencing requires the same platforms and many of the same processes as DNA sequencing. Differences exist in sample and library preparation, storage and analysis. RNA is less stable than DNA and considered harder to work with. Before RNA is sequenced, it is converted to cDNA (complementary DNA) by reverse transcription. Some newer technologies, such as some long read sequencers, do not require RNA to be converted prior to sequencing.

Examination of RNA can also be carried out using other (non-sequencing) methods such as RT-PCR or microarray, although these are only capable of examining substantially smaller portions of the transcriptome. RNA sequencing is being investigated for the diagnosis of rare disease (complementary to DNA sequencing) and as an adjunct in the prognosis or informing treatment decisions in cancer. Although examination of RNA using methods such as RT-PCR and microarray is common in the diagnosis and/or prognosis of several diseases (chapter 5.3), currently there is no use of whole transcriptome analysis in clinical care beyond research for rare disease diagnosis.

### *Bisulphite sequencing (methyl-seq)*

Bisulphite sequencing is used to determine the presence or absence of methylation marks on the genomic DNA sequence. DNA can accumulate, or can be inherited with, methylation marks. These are chemical modifications to the nucleotide bases and involve addition of a methyl group on the hydrogen of cytosine or adenine. DNA methylation is one of several forms of epigenetic modification which can affect if or how parts of DNA are expressed without changing the DNA sequence itself. Cytosine modification to methylated cytosine (methyl-C) is the most commonly studied epigenetic modification in humans and other mammals. Bisulphite sequencing is able to capture information about the presence of absence of methyl groups on cytosine bases.

As with RNA-sequencing, bisulphite sequencing utilises the same equipment as is used for DNA sequencing, with differences in sample and library preparation, storage and analysis. In preparation for bisulphite sequencing, bisulphite conversion is performed; this involves the conversion of unmethylated cytosine into uracil - an analogue of thymine which is encountered in RNA - whilst methylated cytosine remains unconverted.

Hypermethylation of genomic regions can result in transcriptional repression (gene silencing), whilst hypomethylation can result in genomic instability. Methylation has implications for disease, including imprinting disorders such as Prader-Willi syndrome and Angelman syndrome, for which detection of methylation marks by bisulphite sequencing, methylation-sensitive pyrosequencing or non-sequencing techniques are included in diagnosis.

While bisulphite sequencing is the most common method used, alternative methods for measurement of methylation (and other epigenetic marks) exist. These include methylation specific PCR (msPCR), methylation specific multiplex ligation-dependent probe amplification (MS-MLPA), methylation specific FISH, and, more recently, long-read sequencing technologies.

## 7 Factors that determine choice of genome analysis methods

All sequencing technologies can be applied in different ways in a clinical setting to obtain information about the DNA that makes up the genome. This includes examining only subsets of the genome, particular functional regions of the genome and examining RNA or DNA modifications (e.g. methylation) rather than DNA itself. The main clinical and research techniques are discussed below.

### 7.1 Clinical exome sequencing (CES)

Clinical exome sequencing (CES) is the sequencing of all protein-coding regions of the genome which are known to have a clinical (disease) association. Genes are made up of regions referred to as exonic (exons) and intronic (introns). Exons are the regions of a gene that can be translated to produce proteins, introns are transcribed but removed as RNA matures prior to translation into protein. CES refers to the sequencing of the exons of genes in the clinical exome, which includes around 5,000 genes. If a variant is found in a gene already known to be associated with disease, it can be easier to establish a causal link between the patient's condition and any genetic variants.

There is not one 'fixed' clinical exome. Organisations may produce their own list of genes to be included in a clinical exome, for example: the clinical exome offered by Great Ormond Street hospital, referred to as the 'GOSHome', contains around 5,000 disease-associated genes [84]; the TruSight One sequencing panels (Illumina clinical exome kit) include between 4,800 and 6,700 genes [85]; and the four laboratories within the NHS Scotland Laboratory Genetic Services consortium each have clinical exome panels and sub-panels relevant to particular disease areas [86].

In some circumstances, clinical exome sequences can offer notable advantages over whole exome or genome sequencing. Due to sequencing systems having limited capacity and producing a limited number of reads per run, examining larger parts of the genome is both more expensive and, all else being equal, each region of the genome is covered by fewer reads (reduced coverage). Sufficient and balanced coverage is important for producing reliable sequencing results. CES has both advantages and disadvantages as compared to WES and WGS in the context of clinical genome analysis.

#### Advantages

- » Cheaper and faster to conduct
- » Less likely to encounter variants of unknown significance when examining the clinical exome
- » Improved coverage of examined genes or increase in multiplexing: reads can be distributed at greater depth across one sample or more samples can be examined

## Disadvantages

- » Less comprehensive than either WES or WGS, sequencing around 0.5% of the whole human genome, although it contains regions most pertinent to disease, disease causing variants can exist outside of these genes and outside of exons
- » Limited utility for discovery of disease-gene associations

## 7.2 Whole exome sequencing (WES)

Whole exome sequencing involves the sequencing of all known protein-coding regions of genes in a genome only. The human genome contains around 21,000 genes, although the exact number is still in dispute. Exons only represent about 1-2% of the human genome sequence, but include around 85% of all described disease-causing variants. Whole exome sequencing involves sequencing the exons of all known genes.

As per whole genome sequencing, DNA is first amplified and then fragmented prior to sequencing. An additional step selects exonic regions from the whole DNA sample - this involves the selective hybridisation of these regions to biotinylated probes, which bind to streptavidin beads; anything left unbound can be washed away and will not progress to sequencing.

## Advantages

- » Cheaper and faster to conduct than WGS, but more expensive than CES
- » Less likely to encounter variants of unknown significance than when performing WGS

## Disadvantages

- » Less comprehensive than WGS - although it contains many exonic regions pertinent to disease, there are many regions of the genome that may further impact disease development which are not included in the exome
- » Not useful for the examination of large chromosomal rearrangements or deletions, any variation that exists outside of exons will not be detected
- » As with WGS, coverage of specific genes may be poor. Targeting can increase coverage.
- » More challenging interpretation compared to clinical exome, greater likelihood of variants of unknown significance.

## 7.3 Whole genome sequencing (WGS)

Whole genome sequencing describes the most comprehensive sequencing of the genome that is currently possible. The genome is composed of both protein coding and non-protein coding regions, all of which are included in an assessment using WGS. WGS includes the examination of DNA across all chromosomes, and may or may not include mitochondrial DNA (mtDNA).

WGS is sometimes used to describe the whole process of sequencing and analysis; several applications of sequencing in research and clinical laboratories may be described as WGS, whilst the subsequent analysis may not include all parts of the genome. As genome analysis is composed of several parts, the act of sequencing the whole genome does not automatically mean that the whole genome is analysed. WGS is the assay that captures the raw sequence data across the whole genome, subsequent analysis determines what happens next with the raw data. In many instances, especially within clinical applications, specific interpretation approaches will be used to retrieve from WGS data clinically-relevant information from protein-coding regions or specific sets of genes only, guided by the clinical phenotype of the patient.

### Advantages

- » Producing the most comprehensive sequence from genomic samples may allow for future advances in knowledge to be taken advantage of more easily
- » Can be used to interrogate areas of the genome that other methods do not assess
- » In the context of rare disease, there may be clinically relevant variants outside of regions that are examined by less comprehensive methods
- » The analysed genome can be reassessed in the event of a change in phenotype

### Disadvantages

- » Knowledge is lacking about the clinical relevance of the majority of the genome
- » Whole genome sequencing and analysis is more expensive and time consuming than examining select parts of the genome, such as through CES or WES
- » Significant numbers of variants of unknown significance are more likely to be identified, which creates additional interpretation challenges in terms of generating a clinical report (this could be seen as positive or negative dependent upon circumstance)
- » Coverage of specific genes may be poor. Targeting of specific genes can allow for more comprehensive coverage.
- » Examination of all regions of the genome is more likely to result in incidental findings

Currently, whole genome sequencing is being undertaken for individuals enrolled in the Scottish Genomes project run by Scottish Genomes Partnership and the 100K Genomes project. Following the announcement of the establishment of genomic laboratory hubs across England, WGS will be used in a number of applications in cancer and rare disease.

## 7.4 Comparison of clinical exome, whole exome and whole genome sequencing

The relative benefits and disadvantages of whole exome sequencing and whole genome sequencing are summarised in table 3 below.

**Table 3. Relative advantages and disadvantages of clinical exome, whole exome and whole genome sequencing**

Technique	Advantages	Disadvantages
<b>Clinical exome sequencing</b>	<ul style="list-style-type: none"> <li>• Greater coverage of examined genes than WES and WGS</li> <li>• Cheaper to conduct</li> <li>• Reduced data storage demand</li> <li>• May be easier to identify and understand disease causing variants from those detected</li> </ul>	<ul style="list-style-type: none"> <li>• Low proportion of genome examined &lt;0.5%</li> <li>• Does not provide insight into non-coding regions</li> <li>• Reliant on exome capture efficiency – selection steps can result in bias and non-uniform coverage</li> </ul>
<b>Whole exome sequencing</b>	<p>Compared to CES:</p> <ul style="list-style-type: none"> <li>• More protein coding regions examined – more potential sources of disease examined</li> </ul> <p>Compared to WGS:</p> <ul style="list-style-type: none"> <li>• Cheaper and faster</li> <li>• Fewer variants of unknown significance</li> <li>• Facilitates greater read depth</li> <li>• Simpler bioinformatic analysis</li> <li>• Reduced requirement for data storage ~6GB for one exome at 100x read depth</li> </ul>	<ul style="list-style-type: none"> <li>• Low proportion of genome examined ~1.5%</li> <li>• Does not provide insight into non-coding regions</li> <li>• Exome selection steps can lead to bias - non-uniform coverage</li> </ul>
<b>Whole genome sequencing</b>	<ul style="list-style-type: none"> <li>• Greater proportion of genome examined ~97%</li> <li>• Allows for investigation of non-coding regions</li> <li>• Future proofing - future knowledge and re-examination may facilitate disease insight from VUSs</li> </ul>	<ul style="list-style-type: none"> <li>• Greater computational burden</li> <li>• Lots of redundancy - knowledge about significance of many parts of the genome is lacking</li> <li>• Greater likelihood of uncovering VUSs</li> <li>• More expensive</li> <li>• Greater data storage demands ~90GB for one genome at 30x read depth</li> </ul>

CES, WES and WGS have different qualities and there are clinical situations in which one technique may be more suitable than another. The primary differences between the approaches are of focus in terms of read depth, coverage and the number of VUS generated.

### *Why use CES?*

CES can be used when a diagnosis is strongly suspected given a patient's phenotype and a quicker diagnosis might be required – as a diagnostic it is both quicker and cheaper than WES and WGS. As the genes included in the clinical exome have proven gene-disease associations, it is easier to establish causality of any variants found, and variants of unknown significance are less likely to cause interpretation challenges.

### *Why use WES?*

WES provides high read depth and low coverage whilst WGS provides much higher coverage but lower read depth. Where a selection of genes and/or variants are suspected of causing disease, WES can provide a faster, cheaper diagnostic. Substantially more is understood about protein-coding regions than about non-coding regions and the added value of sequencing these regions is currently debated. The reduced data storage required for whole exomes is also significant – roughly 2% of that required for a whole genome. Whole exome sequencing is currently widely used in rare disease diagnosis and, as exome library preparation becomes cheaper and more efficient, it is replacing some of the use of targeted or panel sequencing (see section 7.4) in many circumstances.

### *Why use WGS?*

WGS is the technique of choice where investigations extend into the non-coding regions of the genome. Some disease has been linked to variation outside exons and, without sequencing these regions, causal disease variants in these regions would be missed, resulting in potential missed diagnostic opportunity. As the whole genome is sequenced, WGS also avoids the need for exon-baiting – targeted enrichment of protein-coding regions – which can result in bias and greater inequality in coverage across the genome; in these circumstances, areas such as those enriched for SNPs may suffer reduced coverage. The selection of exons also means that shorter reads are often used, the majority of human exons are shorter than 200bp, meaning reads above this length will be wasted [87]. WGS can utilise longer reads, which has benefits for genome reconstruction and in the investigation of both structural changes and repetitive regions.

### *Diagnostic yield of CES, WES and WGS*

A meta-analysis of diagnostic and clinical utility of WES, WGS and microarray for children with rare disease reported a non-significant difference between utility in WGS and WES, with diagnostic utility (yield) of 0.41 (41%) and 0.36 (36%) respectively, but that both provided significant benefit over microarray. The use of trio analysis was shown to have a greater impact on likelihood of diagnosis than the choice between WES vs. WGS (section 7.5) [88]. An analysis of around 200 seriously ill infants randomised to receive either rapid WGS or rapid WES found that diagnostic rates in these circumstances were near identical, at 19% and 20% respectively [89]. For the clinical exome, one study carried out in two French genetic centres used the Illumina TruSight One panel (covering 4,813 genes) to determine if there was a genetic basis to neurodevelopmental delay observed in 216 patients. The overall diagnostic yield of the panel was 25.9% [90]. Another study on 35 Korean patients with unexplained developmental delay or intellectual disability, also using the TruSight One panel, gave a diagnostic yield of 29% [91].

## *The future of WES vs. WGS*

The 100,000 Genomes Project utilised WGS, however subsequent clinical analysis focused on protein-coding regions. The use of WGS in this instance helps to future-proof the data – as more information is gained about the health and disease implications of non-coding regions, the data can be re-analysed without the need for additional sample collection and sequencing.

Circumstances in which WGS might be particularly advantageous include clinical settings where research and increasing disease understanding is both a primary goal of sequencing and can be facilitated by expertise and bioinformatics infrastructure, or is anticipated in the near future.

As sequencing costs continue to fall, WGS becomes a more realistic prospect as the actual (not necessarily proportional) difference in cost between the two approaches decreases. In addition, the increasing visibility and use of long-read sequencing platforms provides different dimensions to the WGS vs. WES debate, as they demand different technical considerations when comparing the two approaches. Continued improvement to bioinformatic pipelines and an increasing database of information about non-coding regions will likely mean that the added value of sequencing non-coding regions will increase, though whether these changes will justify the use of WGS will remain a matter of context and purpose.

## **7.5 Gene panels and targeted sequencing**

Targeted sequencing involves selectively sequencing a pre-determined set of genes, which can include non-coding and intronic regions of genes. The human genome contains 5-10 million genetic variants on average, the vast majority of which will not be clinically significant to the condition in question. This means that targeting sequencing to examine only a subset of genes most likely to harbour the/a pathological variant associated with the suspected disease in question can reduce time, cost, and analytical load.

Targeting can occur to different degrees ranging from a few genes, to a near complete collection of a type or portion of genes e.g. exome sequencing (see table 3). Genes selected for targeted sequencing form what is referred to as a 'gene panel', in clinical settings these may be referred to using disease or associated biological system names e.g. cardiac gene panels for suspected genetic cardiac conditions.

Targeted sequencing is performed by first performing capture or target enrichment (also referred to as 'pull down'). There are a number of ways of doing this including commercially developed kits (such as Agilent Sure Select), different types of PCR and hybridisation capture, which may involve the use of microarrays or other formats of target probes e.g. in solution. Targeting can also be bioinformatic – the whole or a large part of the genome may be sequenced by WGS or WES, whilst only a subset of this is analysed. This is also referred to as a 'virtual panel'.

The Genomics England PanelApp, a publicly-available gene panel resource, lists different types of panel that can be generated and used within healthcare in England. These include 'real' and 'virtual' panels, and both research and clinical panels [\[92\]](#).

As read capacity is shared across all parts of the sample, targeted sequencing (excluding bioinformatic-only targeting) allows a smaller number of genes or genomic regions to be sequenced at greater depth and coverage. Targeted sequencing can be especially useful in cases of mosaicism or cancer, where more than one version of the genome exists within one individual and some gene variants may be present at low frequencies [93] and where the patient phenotype is clear and candidate causal genes can be easily identified. Targeted sequencing can also reduce the analytical and sequencing costs associated with DNA sequencing that includes more of the genome. Design and use of a gene panel should be carefully considered – some genes, which may be key for diagnosis of specific conditions, may not be suitable for examination using a NGS panel/targeted approach e.g. if the likely causal variant is a large deletion.

### Advantages

- » Greater coverage of targeted areas
- » Reduced analytical and/or sequencing cost compared to WES and WGS
- » Reduced chance of incidental findings
- » Can allow for an increase in multiplexing per run

### Disadvantages

- » More comprehensive genomic investigation required where panel genes are not found to be the cause of disease
- » Suitable where a suspicion as to the cause of genetic disease is already present

In addition, virtual panels can allow for easy redesign where new genes are discovered to be related/found to be unrelated to the condition in question, and can allow for (provided the appropriate processes are in place) reanalysis to include additional genes if a causative gene is not found using the initial set of genes on the panel. However, some of the advantages of targeted sequencing are lost when using virtual panels – more comprehensive sequencing is required which increases cost, lower coverage available per genomic region, and analysis will be more time-consuming and costly than real true panels.

Gene panels using NGS technologies are still subject to the technologies' technical biases, therefore findings in individual genes may require confirmation using single gene analysis by, for example, Sanger sequencing. In the UK, there is some inconsistency in the targets included in a targeted panel. In England, the newly developed National Genomic Testing Directory (NGTD) will address some of this inconsistency.

## 7.6 Trio analysis

Trio analysis is the sequencing of three related individuals – the parents and the proband – for Mendelian gene discovery or identification. Trio analysis can be applied in a whole genome, exome or gene panel format. The comparison of parental and proband DNA allows for the identification of *de novo* disease causing variants versus inherited variants. In many rare disease cases, proband-only sequencing may be sufficient for diagnosis without additional parental sequencing.

The sequencing of three individuals compared to one results in increased costs and analytical burden, but appropriate use of trio analysis can increase diagnostic yield. It can also allow for haplotype phasing – determining which chromosome (the maternal or paternal) a particular variant is present on. This has implications for the diagnosis of some diseases.

Trio analysis is primarily used in the detection of causal disease variants for young children with rare genetic disorders and it is anticipated that its use will expand in the coming years, particularly with application in neonatal and paediatric intensive care units (NICU/ PICU). Initiatives such as the Deciphering Developmental Disorders (DDD) project have utilised trio sequencing extensively in the identification of causal variants and diagnosis of developmental disorders [94]. A number of studies have shown an increased diagnostic yield when using trios compared to single proband sequencing. For example, a recent meta-analysis examining the diagnostic utility of various sequencing approaches for the diagnosis of rare disorders in children showed a far greater increase in diagnostic yield when using trios vs. singles than between WGS and WES. The increased yields from trios were variable, but equated to a roughly 2x diagnostic rate across all studies (ranging from 1.63 - 2.68), compared with that gained from singleton sequencing [88].

A recent UK study utilising rapid sequencing of the genomes of children with rare disease demonstrated the utility of rapid trio WGS for clinical management. The small scale Rapid Paediatric Sequencing (RaPS) project achieved a 42% diagnostic rate for patients in an average of 8.5 days, resulting in an immediate change of clinical management in three out of 10 critically ill children receiving a diagnosis [95].

The additional data storage and analytical burden resulting from the sequencing of three whole or partial genomes versus one genome is substantial. Consideration also needs to be given to broader issues associated with trios such as incidental findings in parents, additional consenting and counselling, and how to manage cases of false paternity. These considerations are applicable to other areas of genetic testing; however, they are amplified in the case of trio sequencing. Determining whether the application of trio analysis is justified depends on the clinical situation, as the relative diagnostic gain will differ between potential applications.

**Table 4. Comparison of techniques used for clinical genome analysis. The genome coverage (C) and resolution (R) possible is indicated in the first column.**

Technology	Variants Detected
<b>Karyotyping</b> C: high R: low	Aneuploidy Large CNVs (5-10Mb) Mosaicism Chromosomal damage/ rearrangements
Strengths	Limitations
Can visualise changes in chromosome structure	Low resolution, slower turnaround times Dependent on cell culture. Analysis is time-consuming, requires specialist skills and viable tissue for cell culture

Technology	Variants Detected
<b>FISH</b> C: low R: low	CNVs (5-10mb) Gene fusions Mosaicism
Strengths	Limitations
Can detect sub-microscopic and cryptic changes Can be used in screening large numbers of cells for mosaicism or clonal analysis	Limited scope of analysis, not genome-wide, therefore most useful for confirming clinical diagnosis of syndromic conditions with distinct phenotypic features Micro duplications may be difficult to accurately discern
Technology	Variants Detected
<b>Array CGH</b> C: medium R: medium	Aneuploidy CNVs (5-10kb) SNPs
Strengths	Limitations
Genome-wide analysis, useful in the absence of clear, syndromic features Cell culture free and allows flexible approach through customised arrays	Cannot detect balanced rearrangements, or low levels of mosaicism Cannot detect very small INDELS (range)
Technology	Variants Detected
<b>PCR</b> C: low R: medium	Aneuploidy CNVs Changes in repeat length SNPs
Strengths	Limitations
Can detect a variety of mutations High sensitivity and specificity Ease of use Multiplexing and targeting	Limited scope of analysis Not effective for gene or genome wide scanning for variants
Technology	Variants Detected
<b>Sanger sequencing</b> C: medium R: high	Indels (1-100bp) SNPs
Strengths	Limitations
Less error prone than NGS Relatively straightforward interpretation	Labour intensive and expensive for higher coverage

Technology	Variants Detected
<b>NGS</b> C: high R: high	Many variants including Indels (1-100bp) SNPs
Strengths	Limitations
Flexible, high throughput and low per-base cost	Not accurate at identification of repeat sequences Analysis and interpretation challenges Turnaround times can be longer than other techniques

Technology	Variants Detected
<b>Single molecule sequencing</b> C: high R: high	As for NGS + large structural variants
Strengths	Limitations
Rapid run time Long reads	High error rate Low throughput

## 7.7 Supporting clinicians in the use and understanding of genetic information

There are two steps of interpretation and reporting: a gene level or molecular report followed by the phenotypic contextualisation of the variant that can influence disease management decisions. Sometimes, where the variants and phenotypic link are well established, the interpretation is straightforward. In other situations, multiple potentially pathogenic variants or VUSs are found, none of which entirely fit the clinical picture. This is particularly relevant when discussing large panels, WES or WGS. Decisions need to be taken concerning how to classify these variants and whether or not to report them. Additional phenotypic information could bring clarity to potentially plausible explanations for the clinical phenotype, or whether further investigations are required.

The multi-disciplinary team (MDT) has a role to play in supporting clinical decision making that utilizes genomic information. The type of MDT will depend on the purpose; a genomic cross-discipline MDT would be more laboratory driven, whereas a diagnostic or disease management MDT would be driven more around the specialty needs and the clinical decisions to be taken such as in oncology, cardiology or neurology. The MDT is likely to include a range of professionals – laboratory scientists, clinical geneticists and the relevant clinical specialists. The referring clinicians may also be involved, although this is not always possible or practical. Involvement in MDTs will vary between services and healthcare professionals' level of expertise.

The contribution of clinical genomics experts to MDTs can facilitate discussions around VUSs, supporting decision making around whether further investigations, phenotyping or molecular analysis may be required. They will also contribute to the improved quality of clinical reporting, patient care and knowledge gathering on and understanding of rare genetic variants.

## 7.8 Considerations around the identification and interpretation of variants of unknown significance (VUSs)

Variants of unknown significance (VUSs) are genetic variants identified through genetic testing for which the clinical significance is currently unknown. As sequencing has become more ubiquitous in clinical practice, the prevalence of detailed, extensive DNA sequence data increases. Although understanding of the genome is increasing greater use of sequencing is leading to an increase in the number of known VUSs. The more comprehensive the method of genetic analysis, the greater the likelihood of uncovering VUSs in the patient genome.

Genetic testing may be carried out for one or more reasons, this can be for diagnosis (as in rare disease), prognosis or treatment selection (as in cases of cancer), or for understanding risk e.g. in breast cancer. All tests are provided on the basis that actionable findings can be derived, whether that be determining treatment or informing reproductive choices. As the importance of any VUSs identified through testing is uncertain, they can pose challenges for clinicians and researchers in determining next steps. Each human genome contains around 4-5 million variants [96] – much of this variation is benign, however the huge diversity in possible variation between individuals and the limited knowledge about the function of genes and the genome itself, make it difficult determine the significance of many discovered variants in a clinical setting.

Evidence for the significance of variants comes from multiple sources:

- » The association of variants with disease
- » Through functional testing, involving for example the development of animal knock-out models
- » *In silico* modelling of potential effects on proteins or other molecules produced from the altered region

Variants of unknown significance present challenges in clinical management, genetic counselling and in the impact on patients [97]. Several issues need to be considered following the discovery of VUSs:

- » **Disclosure to patients and appropriate counselling** – Consideration will be given as to whether the patient is likely to benefit or be harmed from knowledge of this variant. Where a course of action is not immediately apparent, consideration as to how this might be best explained to patients is important. The presence of variants with unknown impacts may contribute to patient anxiety or uncertainty.

- » **Follow-up investigations** – Following the discovery of the VUS, more could potentially be done to understand the significance of the VUS or to continue patient diagnosis. Questions will be raised as to whether the presence of the variant should affect these decisions. Additional time may also be spent in attempting to determine pathogenicity status of these variants, which could lead to delays in patient management and increased analytical or testing costs [96]. Where the patient's condition in question is rare (or ultra-rare) international efforts may be required to inform understanding of variant significance.
- » **Potential future reclassification** – An awareness of the presence of this variant could later benefit a patient if information about its potential impacts becomes available in the future

Other forms of uncertainty from VUSs may not result from a lack of evidence of association, but from current understanding of the individual contribution of variants to the phenotype. Where variant effects are mediated by the presence of other variants or environmental factors, risk and significance are harder to model and understand.

Beyond rare disease, one example in which the difficulties of VUSs are apparent is in BRCA testing. A great number of variants exist for within this gene for which the risk contribution to breast and ovarian cancer is poorly (or not) understood.

Whilst the meaning of VUSs at the time of initial investigation is, by definition, unknown, the identified variant itself will remain constant and the significance of the variant may become apparent over time, following re-analysis or re-interpretation.

Alongside general efforts in research and clinical practice, several larger endeavours have been established which aims to improve the understanding of variants either across the board or within specific diseases, these include: BRCA exchange [98], CanGene CanVar [99], and the Human Functional Genomics Project [100].

## 7.9 Retention, resequencing and reinterpretation of data and samples

Once obtained, genetic samples and data must either be appropriately stored or disposed of. Samples can include raw samples such as blood or tumour tissue, extracted DNA/RNA samples, or prepared DNA/RNA libraries ready for sequencing. Data can range from raw sequence data e.g. bcl files, through to variant-called files, with all the iterations in between also presenting opportunities for storing the data.

There are several advantages to storing genetic data or samples. These include:

- » Potential for re-analysis to gain further insight for the patient or researcher - the sample or data may be used in the future to understand the significance of additional variants
- » Opportunity to apply newer (and probably improved) methodologies to hard-to-obtain or unique samples - where sample is the limiting factor, the ability to apply more accurate or comprehensive technologies could increase information retrieved from the sample
- » Ability to revisit information for the purposes of informing wider endeavours such as research linking variation to cases of disease

However, there are several potential disadvantages and important considerations to sample/data retention:

Storing of samples requires physical space, appropriate labelling and documentation to ensure samples are not confused or lost, and facilities which can provide the right storage conditions e.g. -20°C freezer for DNA, -80°C freezer for RNA. Considerations should also be given to the medium in which samples are stored, for example, formalin-fixation and paraffin embedding (FFPE) is a common method for the storage of tumour tissue, however it can result in lower nucleic acid quality due to degradation.

Retention of sequenced data is both costly and has security and privacy implications. There is also consideration around which state of data should be retained. Whilst fully-processed variant called data is the most overtly useful to clinicians, it does not permit a great deal of reanalysis, unlike raw base-call data, which can be processed through new (and potentially improved) bioinformatic pipelines. There are differences not only in the functionality of the data, but in the size of the associated files too; variant called data following WGS of a human genome (at approx. 30x read depth) uses approximately 125MB of space, whereas raw sequencing data requires around 200GB. Additional meta-data may also need to be stored for accurate interpretation, such as phenotypic information about the patient, details regarding sample collection, and the format of the sequencing run used to produce the data.

The decision to retain either samples or data needs to be carefully considered including the relative benefits, harms, risks and costs, in order to determine the right approach for each clinical service.

## 7.10 Additional looked for findings

Additional looked for findings are variants in the patient genome that are known to be associated with serious or life-threatening disease and are purposefully examined during genetic testing, but that are not related to the condition for which the patient is undergoing genetic testing.

These variants may be associated with a notable predisposition towards future common disease such as breast cancer and mutations in *BRCA1* and *BRCA2* genes, or *LDLR* and familial hypercholesterolaemia. Knowledge of the presence of these variants can allow patients to undergo preventative or otherwise beneficial treatments to avoid the full consequences of the identified genetic condition. Carrier status for serious conditions may also be examined where the patient in question is not affected but has the potential to pass the condition onto future offspring.

Although potentially highly beneficial, the decision to include additional looked for findings should be well-considered. Lists of additional looked for variants change over time as knowledge about genetic disease evolves. In 2017, the American College of Genetic Medicine updated its guidelines to recommend that genes present on a specified list be examined when clinical genetic testing is undertaken, regardless of the original purpose of the test [101]. The genes included are determined to be both serious and clinically actionable. A similar approach has been adopted by the 100,000 Genomes

Project [\[96\]](#). Participants in the 100,000 Genomes Project were able to opt-in to retrieve information about the presence of potentially detrimental variants in genes associated with: predisposition to bowel cancer, breast and ovarian cancer, and select other cancers; the presence of familial hypercholesterolaemia; and carrier status for cystic fibrosis. This is applied with the condition that only early-onset conditions are looked for in child participants [\[102, 103\]](#).

There have been many discussions around the reporting (mandatory or otherwise) of such additional looked for findings; alongside ethical considerations [\[104\]](#), there are costs both in terms of effort and capital to both patients and health care providers of investigating, reporting and acting on these findings [\[105\]](#). In addition to conducting testing to confirm additional looked for findings, patients may require ongoing screening tests or preventative care. The patient may require additional ongoing care from clinical genetics services as information about disease-associated variants changes; recontact may be required.

Considerations around additional looked for findings include:

- » Strength of evidence supporting the disease association
- » Potential for altering clinical care on the basis of variant identification
- » Potential benefits and harms for patient of receiving information about additional variants
- » The need for confirmatory testing of additional findings
- » Decision to recontact patients as information evolves
- » Implications for patient's genetic relatives

## 8 Genetic testing through the lifecourse

In the previous sections of this report we have set out the wide variety of clinically relevant genome variation that can occur, and the multiplicity of techniques with which to analyse this variation. In this chapter we describe and analyse current practice in constitutional clinical genome analysis, i.e. the analysis of inherited DNA rather than the somatic alterations that are acquired throughout an individual's lifetime. This type of genetic analysis is important for helping to diagnose rare and inherited disease. We take a life course approach, detailing the different clinical indications for genomic analysis from pre-implantation genetic diagnosis through to diagnosis of genetic disease in adults. In recent years there have been significant changes to the way in which genomic technologies are used to help manage rare and inherited diseases. This is due to significant milestones being reached, such as the completion of the sequencing phase of the 100,000 Genomes Project in December 2018. In addition demand on services has continued to grow, with the 2015-2016 Association for Clinical Genomic Science (ACGS) audit of genetic test activity showing that in comparison with the previous year, testing for inherited (constitutional) disorders has increased by 9.5% as determined by report number and 17% as determined by GenU (the Laboratory Genetic Units system used by the ACGS) [106].

For each of the major clinical indications for genomic analysis we will describe its clinical utility, the range of methods used and, where data are available, the approximate level of testing currently undertaken in the UK. Where changes in the utility, methods or level of demand can be anticipated to occur in the next five years, we provide an analysis of these changes and the impact they may have on provision of genomic analysis services.

### 8.1 Genetic testing in the prenatal period

Results from data collected from UK Genetic Testing Network (UKGTN) laboratories found that 2016-2017 the prenatal genetic test report rate for residents of Scotland was 2,709.6 per 100,000 conceptions (95% confidence interval 2,586.8 – 2,836.8). This was higher than that of England (2,602.1; 2,576.8 – 2,636.8) [107]. Two main types of genetic tests are carried out in the prenatal (or antenatal) period: screening tests and diagnostic tests. Screening tests can be used to inform parents if their unborn child is at higher risk of certain genetic abnormalities, but are typically not diagnostic. Diagnostic tests are used to definitively diagnose a condition that is indicated either through screening, clinical symptoms, or in some cases family history. There are several scenarios in which these tests may be used:

#### *Genetic carrier screening tests*

Even if they have no symptoms themselves, potential parents may be a carrier for a genetic disorder, especially if they have a family history of a genetic disease or belong to a high-risk ethnic group. Genetic carrier screening tests can be used to test the mother and father for the presence of these genes either prior to or during pregnancy, to determine the chance that they will pass a condition on to their children.

### *Pre-implantation genetic testing*

If couples have been identified as carriers of a genetic abnormality prior to pregnancy, they can opt for *in vitro* fertilisation (IVF) and pre-implantation genetic diagnosis (PGD). PGD can be used to select 'healthy' embryos that do not carry a known genetic abnormality.

### *Prenatal genetic screening tests*

Prenatal screening tests are used to identify if a fetus is at increased risk of a genetic condition. Current non-invasive prenatal screening tests (NIPT) based on genetic analysis do not typically have high enough specificity and sensitivity to be used to diagnose a condition, therefore if a genetic condition is indicated then a confirmatory diagnostic test will be offered. These tests are widely available privately but are not provided by the NHS in Scotland and England, which currently use non-genetic phenotypic tests for screening. However this is planned to change in the near future [\[108-110\]](#).

### *Prenatal genetic diagnostic tests*

If an embryo has a suspected genetic abnormality identified through genetic carrier screening, prenatal screening tests, or clinical symptoms, a genetic test can be used to confirm the diagnosis, either via non-invasive prenatal diagnosis (NIPD) or invasive testing which requires amniocentesis or chorionic villus sampling.

Prenatal screening and diagnostic tests are always optional, as not all parents may want to know if their child is at risk of a condition. There are also potential disadvantages to prenatal testing; screening tests that have lower risk results may not warrant a diagnostic test, but could cause anxiety. As with all screening there is also a chance of false positive and false negative results. Confirmatory diagnostic tests are often invasive and could potentially be harmful, carrying a small risk of causing infection and miscarriage. The use of testing in twin pregnancies can result in additional difficult decision making, as there is the possibility that only one baby may be affected. Therefore, the decision to undergo one of these tests is not straightforward, and the benefits and harms have to be considered carefully for each individual.

## **8.2 Genetic carrier screening and testing in the prenatal period**

### *Current use*

Population level carrier screening for some genetic conditions in the prenatal period, either for all mothers (as in  $\beta$ -thalassaemia screening) or for selected at-risk populations (such as in sickle cell screening) is currently carried out in the UK using non-genetic biochemical tests. These tests can also be carried out on request before conception if a genetic condition is suspected. If results are positive, then genetic tests can be used to confirm the diagnosis and determine carrier status. In other cases, diagnostic DNA tests may be used to screen potential parents for a specific condition where there is a known family history or other reason to suspect they may be carriers for a recessive disorder, such as the birth of a child with a known genetic disorder. An example of this is cystic fibrosis screening. Screening prior to pregnancy can help in planning for future pregnancies,

whereas screening during pregnancy can be used to estimate the risk of the fetus carrying a mutation, determine whether invasive testing of the fetus is necessary, and allow parents time to make decisions based on the result.

### Current methods

As these tests are analysing parental DNA, a blood or saliva sample is requested from the parent, from which their DNA can be extracted. Where the precise mutation causing a disease is known, targeted mutation testing is used to determine the mutation status, such as PCR or targeted gene sequencing. For genes where a variety of variants may be responsible for an inherited disease, the entire gene may be sequenced. Where duplications or deletions are expected, such as in Spinal Muscular Atrophy (SMA), MLPA testing is most appropriate [50]. Methods for different tests currently listed on the National Genomic Test Directory for England are shown in table 5.

**Table 5. Genetic carrier tests listed on the National Genomic Test Directory for England**

Clinical indication	Target/genes	Test method
Carrier testing for known familial mutation(s)	Specific target	Targeted mutation testing
Cystic fibrosis carrier testing	CFTR	Targeted mutation testing
SMA carrier testing at population risk for partners of known carriers	SMN1	MLPA or equivalent
Haemoglobinopathy trait or carrier testing	HBA1;HBB	Single gene sequencing <10 amplicons or MLPA or equivalent
Carrier testing for sickle cell disease	HbS variant	Targeted mutation testing
Carrier testing at population risk for partners of known carriers of nationally agreed autosomal recessive disorders (tbc)	Relevant single gene	Single gene sequencing >=10 amplicons
Congenital adrenal hyperplasia carrier testing	CYP21A2	Targeted mutation testing or MLPA or equivalent

### Future directions

- » The increased use of NGS based technologies means that it may be possible to perform an expanded carrier screening programme, where several genes are screened for at once using panels. For example, in Victoria, Australia, carrier screening is provided for seven conditions to students in Jewish high schools, showing that this approach may be useful and cost effective [111].
- » The question has been raised over whether to continue to provide the majority of carrier screening based on family history, where carriers may be missed, or expand

to the wider population. For example, it has been estimated that 94% of children with cystic fibrosis are from families with no family history. This may be particularly important as populations and ethnicities become more mixed, making screening based on ethnicity harder to carry out [111].

### 8.3 Pre-implantation genetic diagnosis

#### *Current use*

Where couples have been identified as carriers of a genetic mutation or chromosomal rearrangement and a successful assay has been developed, they may be offered pre-implantation genetic diagnosis (PGD). In conjunction with *in vitro* fertilisation (IVF), PGD can be used to analyse one or two cells from the developing blastocyst, and therefore select only embryos unaffected by genetic abnormalities, or with the desired HLA type in the case of saviour siblings. Nearly 600 individual conditions are currently licensed for PGD by the Human Fertilisation and Embryology Authority (HFEA) [112]. In Scotland, eligible couples can undergo up to three NHS-funded cycles of PGD. The current cost of a private PGD cycle at the Royal Infirmary of Edinburgh for UK residents is £7,875 including fertility drugs [113]. PGD testing for single gene disorders is carried out by NHS Lothian in Edinburgh, and testing for chromosomal disorders is performed by NHS Greater Glasgow and Clyde in Glasgow [114-116].

#### *Current methods*

Where one parent is a carrier of a balanced chromosomal rearrangement, FISH is used to exclude embryos which have inherited the unbalanced form of the rearrangement. FISH may also be used for embryo sexing in the case of X-linked conditions, but increasingly pre-implantation genetic haplotyping (PGH) - a form of PCR where multiple markers of the presence of the specific genomic abnormality are measured simultaneously - is being used, as it provides a more sensitive and specific test. PGH is also used where parents are carriers of autosomal recessive or autosomal dominant conditions to identify embryos which have inherited the unaffected version of the gene.

In Scotland, PCR haplotyping is currently performed on whole genome amplified DNA from day five biopsies in Edinburgh, and in Edinburgh and Glasgow FISH is performed on single cell day three biopsies.

#### *Future directions*

**Changes in HFEA policy on PGD** – Whilst PGD represents a small proportion of prenatal testing, a change in the types of conditions licensed by the HFEA would clearly influence the number of tests carried out. As the number of inherited conditions for which a molecular genetic cause is identified increases through programmes such as Deciphering Developmental Disorders and the 100,000 Genomes Project, demand for PGD for these conditions may also increase. This is illustrated by the fact that since the publication of the PHG Foundation Clinical genome analysis evidence review in 2016, the number of conditions licensed for PGD testing has risen from over 400 to nearly 600.

**Use of SNP arrays** – SNP arrays can be used to combine copy number and genotype information to infer haplotype when compared with a reference genome. Two algorithms are currently used for haplotyping: haplarithmisis and the karyomap gene chip. It has been shown to also be important to screen for aneuploidy in embryos undergoing PGD for single gene disorders. In a study of 186 blastocysts undergoing PGD, it was found that approximately a quarter of the 66.8% of embryos identified as normal were in fact aneuploid [117].

**Use of NGS** – Use of NGS for PGD can provide high throughput analysis of both single gene disorders, aneuploidies and translocations using a single biopsy sample and a single process [118]. NGS can be performed after whole genome amplification, or potentially using single cell sequencing methods. It has been shown that in principle, low coverage genome sequencing can be used to predict haplotypes of embryos in a format designed to be suitable for clinical implementation [119]. In a recent study reduced representation genome sequencing using Agilent's OnePGT was able to automate haplotyping and copy number assessment in human single blastomere and trophectoderm samples [120].

## 8.4 Prenatal genetic screening tests

### *Current use*

Genetic testing is not currently used for routine prenatal screening for genetic conditions, and is only used to diagnose conditions indicated through screening. However, this is set to change in the near future, with the introduction of non-invasive prenatal testing (NIPT) for common trisomies.

### *Current screening pathway*

In England and Scotland, all women are offered fetal screening for Down's syndrome (trisomy 21 or T21) between 11 weeks +2 day and 14 weeks +1 day of pregnancy, known as the combined test. In England, this test is also used to detect Edwards' syndrome (trisomy 18 or T18) and Patau's syndrome (trisomy 13 or T13). A business case produced by NHS Scotland in 2017 recommended extending the current screening programme to include T18/T13 testing, but this has not been implemented in Scotland [121]. The combined test requires a blood sample, which is analysed for specific hormonal proteins that could indicate a developmental disorder. In addition, women are also given an ultrasound scan to detect physical abnormalities.

If women present between 14 weeks +1 day and 20 weeks of pregnancy, they are offered screening for Down's syndrome only, known as the quad test. This is also a blood test that measures the levels of three hormones and one protein produced by the placenta and the baby. Currently this is the only test available for pregnant women in Scotland.

If the fetus is found to be at higher risk of a genetic anomaly through a screening test, a confirmatory diagnostic test is offered. If the diagnostic test is positive, the women are then able to discuss whether to continue or terminate the pregnancy. If the diagnostic test is found to be negative, or the women declines the test, they are offered the 18–20 week scan and also followed up at birth. All screening at any stage is optional, and if women decline testing they are still offered a fetal anomaly scan at 18-20 weeks [122].

## *Introduction of NIPT for prenatal screening*

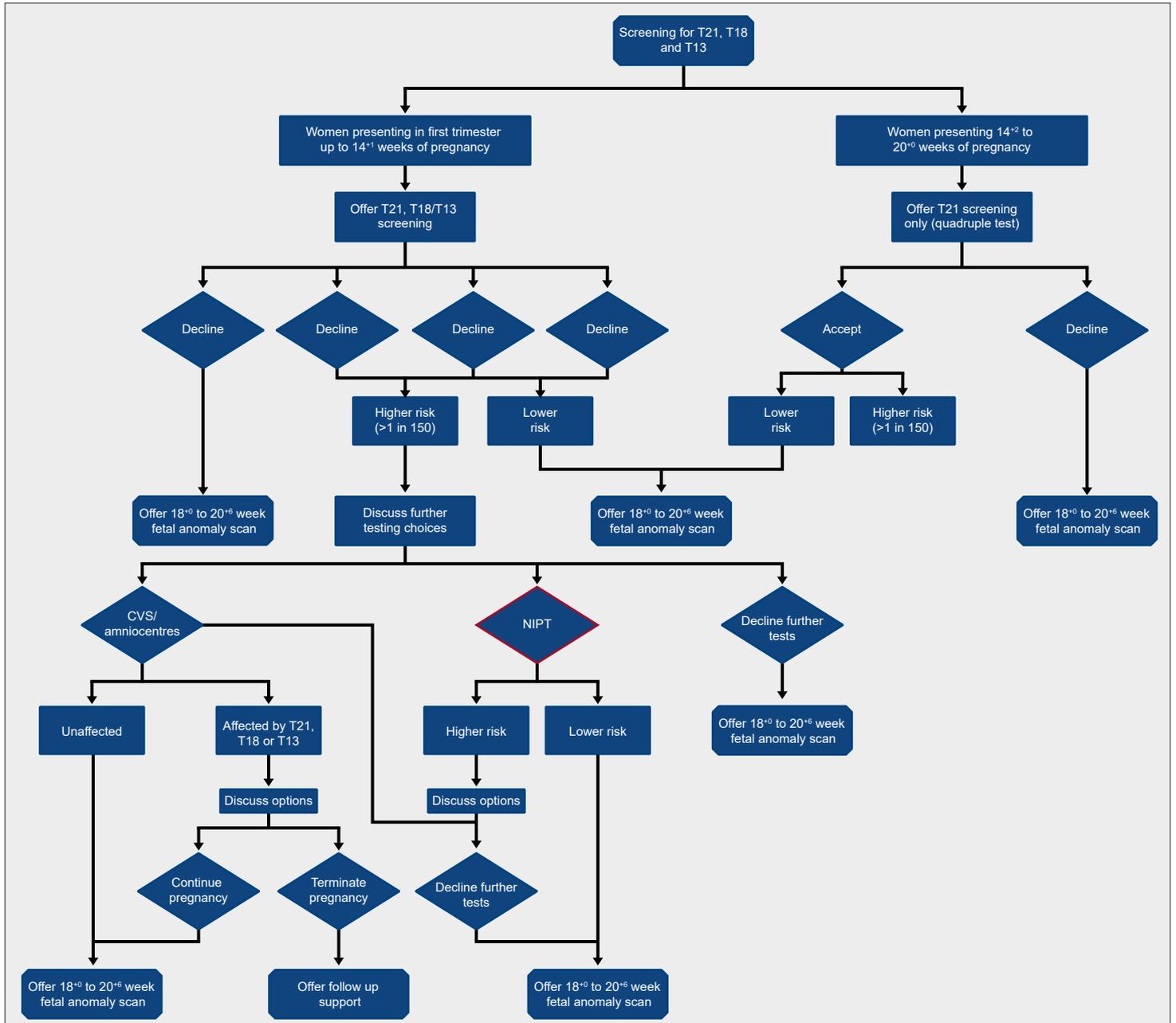
There have been plans to implement non-invasive prenatal testing (NIPT) as a second line test in the prenatal screening programme for common trisomies (T21, T18 and T13) since 2016, after the UK National Screening Committee (NSC) recommended an evaluative pilot. NIPT for detection of common trisomies as well as other chromosomal alterations is already offered privately by several commercial companies, despite uncertain evidence for its performance in detecting other genomic alterations. The technique is based on the analysis of cell free fetal DNA (cffDNA), which is DNA fragments that are released from placental cells and circulate in the maternal blood. This means that to detect genetic conditions, a blood sample can be taken and analysis performed upon the cffDNA, rather than having to take an invasive sample directly from the fetus or placenta. cffDNA is detectable from five weeks' gestation with levels rising throughout pregnancy, followed by rapid clearance after delivery.

NIPT is more sensitive and accurate than current protein and hormone-based tests used in screening, and can potentially also be used for a number of additional genetic testing applications in the prenatal period. Despite its very high specificity, confirmatory invasive testing by chorionic villus sampling (CVS) or amniocentesis is still recommended due to the small number of false positive results, and the technique is therefore referred to as non-invasive prenatal testing rather than non-invasive prenatal diagnosis.

The NSC's recommendation was informed by a review of the literature and the UK RAPID trial, which showed that implementation of NIPT for the common trisomies, as a second line test in high risk pregnancies, would significantly reduce the number of invasive prenatal tests [123, 124]. It has been estimated that use of NIPT for Down's syndrome alone in the 2.3% of pregnant women with a risk of greater than 1:150, out of an annual UK screening population of 698,500, would increase detection by 195 (95% uncertainty interval –34 to 480) cases with 3,368 (2,279 to 4,027) fewer invasive tests and 17 (7 to 30) fewer procedure related miscarriages, for a non-significant difference in total costs (£-46,000, £-1,802,000 to £2,661,000) [125]. NIPT for the common trisomies is currently recommended for use in all pregnant women whose chance of having a baby with one of these conditions is greater than 1 in 150 from an initial combined or quadruple screening test. If the NIPT result is positive, women could then choose to have an invasive confirmatory diagnostic test.

Currently only Wales is offering NIPT, introducing the test across all health boards in April 2018 with evaluation ongoing [126]. Scottish ministers approved an evaluative roll out of NIPT into the Scottish pregnancy programme in July 2018, implementation is currently underway and testing is planned from August 2020 [108]. In England there is an evaluative roll-out of NIPT as part of the Down's syndrome, Edwards' syndrome and Patau's syndrome screening pathway; this follows a recent reanalysis of the original evidence, which resulted in new recommendations on testing technologies and to include twin pregnancies in the programme [109, 127].

Figure 7. Summary of proposed NIPT pathway, adapted from [121].



### Current methods for analysing aneuploidy in NIPT

The cfDNA required for analysis is found in a blood sample alongside cell-free maternal DNA, with the maternal component constituting a greater proportion. Therefore, any analysis of the genetic makeup of the fetus must be conducted against a background of high levels of maternal DNA. Extraction and enrichment methods may be used to increase the relative proportion of fetal DNA in the sample prior to analysis.

There are several methods available for analysis of aneuploidy, including massively parallel sequencing of randomly selected DNA fragments, targeted sequencing of selected chromosomes, and SNP array analysis [128]. For massively parallel sequencing, all the cell free DNA fragments in the sample (fetal and maternal) are sequenced by NGS, with the fragments aligned to a reference genome, and the relative quantities of DNA compared to a normal (non-aneuploid) sample. Bioinformatic analysis then 'counts' the quantity of reads covering the areas of interest in the genome, e.g. chromosome 21 for a Down's syndrome test, in the sample and compares them to the number in the 'normal' standard reference genome known not to contain any chromosomal or large-scale genomic variants.

Stringent thresholds are used to classify samples where there is an abnormally high or low number of reads covering the target region to minimise the risk of false positives. Importantly, the bioinformatic algorithms used to undertake the targeted analysis are focused in such a way that they only report on findings (usually the major aneuploidies) for which the clinical validity and utility are established. As the sequencing itself in NIPT is not targeted, this bioinformatic approach minimises the chances of identifying a wide range of smaller genetic imbalances (as occurs for array CGH), many of which may be difficult to interpret in the clinical context. In comparison, targeted sequencing of specific chromosomes and use of specific microarrays avoids this problem.

Recently the UK National Screening Committee also agreed that DNA microarrays could be used as well as next generation sequencing for evaluation of NIPT [127]. Based on this, NHS England are updating their existing procurement plans for NIPT laboratory services [109].

## 8.5 Prenatal genetic diagnostic tests

### *Current use*

In order to confirm or rule out a condition detected during screening or suspected due to family history; genome analysis must be performed on a sample of fetal DNA. This is most commonly achieved using an invasive test by collecting fetal cells from either amniotic fluid, chorionic villus or in very rare circumstances cord blood. However, it is now also well established that cell free fetal DNA present in maternal blood can be detected and analysed as an alternative to invasive methods of fetal DNA sampling, known as non-invasive prenatal diagnosis (NIPD).

The pathways and choice of method for testing of prenatal samples are complex and depend on:

- » The reason for referral e.g. raised risk of aneuploidy, structural abnormality on ultrasound or known family history of a specific genomic abnormality
- » Variation in clinical and laboratory practice and commissioning guidelines
- » Initial test results e.g. karyotyping to confirm a rapid QF-PCR positive test result for trisomy 21
- » Quality of the available sample e.g. where cells cannot be cultured, full karyotype analysis is not possible

## *Prenatal diagnostic testing: current methods*

### **Chromosomal or large structural genomic abnormalities**

Best practice guidelines, setting out minimum standards for the conduct of prenatal diagnosis have been developed by the Association for Clinical Genetic Science (ACGS) and indicate that for referrals where a chromosomal (e.g. trisomy 21) or large structural genomic abnormality is suspected as a result of antenatal screening, QF-PCR is used to provide rapid (turnaround time of 3 days) testing for the most common trisomies. A positive QF-PCR test may be followed up by confirmatory analysis by karyotyping, which has a longer (two week) turnaround time than the three day rapid QF-PCR test.

Until 2015, karyotyping was recommended as a second line test where QF-PCR did not indicate the presence of a common trisomy but other clinical evidence (such as abnormal ultrasound findings) indicated that another genetic condition might be present. However, in June 2015 the Joint Committee on Genomics in Medicine (JCGM) issued recommendations that array CGH testing should now be used in place of karyotyping following a negative QF-PCR result [\[129\]](#). If, however, there is a known family history of a specific chromosomal or genetic abnormality, a more targeted assay such as FISH or a more targeted array approach might be used as a first line test.

In 2016-2017 the pre-natal microarray test report rate for residents of Scotland was 275.9 cases per 100,000 (95% confidence interval 237.6 – 318.5). This was significantly lower than in England (656.8 per 100,000; 95% confidence interval 639.6-674.3) [\[107\]](#). According to the most recent 2015-2016 ACGS audit of UK laboratory activity, around 17,000 reports were issued for prenatal QF-PCR and cytogenetic tests (FISH, karyotype and array CGH) compared to 17,500 in 2014-2015 and around 20,000 in 2013-2014 [\[106\]](#). Although these types of tests still make up the vast majority of testing (in 2015-2016 approximately 80%) it appears that the numbers requested have slightly declined over the last three years. As data has only been collected for these tests from 2012-2013 and the changes seen are small, it may still be too early to identify any definitive trends. Of the cytogenetic tests reported in 2015-2016, targeted QF-PCR/MLPA accounted for the majority (40%), closely followed by karyotyping (~26% of tests).

### **Single gene disorders**

Where a specific single gene disorder is being tested for, either because of family history or abnormal ultrasound/antenatal screening results, the method used is more likely to be allele specific PCR (that detects a known mutation) or sequencing using Sanger or NGS methods from either amniotic fluid, chorionic villus sampling or increasingly NGS-based NIPD from a maternal blood sample (see below).

According to the most recent 2015-2016 ACGS audit of UK laboratory activity, around 4,000 reports were issued for monogenic tests compared to around 2500 in 2014-2015 and just over 2,000 in 2013-2014 [\[106\]](#). Collectively, monogenic tests constituted approximately 20% of tests in 2015-2016, compared to less than 10% of total reported genomic test activity in the UK. Therefore, it appears that these tests are slowly but steadily increasing to make up a greater proportion of total reports.

## NIPD for fetal sex and single gene disorders

The principles of testing and methods used in NIPD are identical to those of NIPT. The only difference is that NIPD is used to describe applications of cell-free fetal DNA testing that are sensitive and specific enough to be considered diagnostic, whereas NIPT requires follow up with a diagnostic test.

Unlike NIPT, NIPD is already available on the NHS for certain specific applications, reducing the need for invasive diagnostic tests. NIPD can be used for fetal sexing in X-linked conditions. For example, for Duchenne muscular dystrophy, NIPD can guide decision making on whether an invasive prenatal test is warranted (where the fetus is found to be male and therefore at risk). For congenital adrenal hyperplasia (CAH), NIPD can confirm fetal sex and avoid unnecessary treatment being administered in male pregnancies.

It can also be used for some single gene disorders; examples include skeletal dysplasias and cystic fibrosis caused by single point mutations (using PCR amplification followed by targeted NGS). Current NIPD tests listed on the National Genomic Test Directory for England are shown in table 6. Tests for other single gene disorders are currently in development, and assay workup will depend on the specific mutation and technical factors which influence assay validity. The range of single gene disorders that could potentially be tested for prenatally is, in theory at least, very large and will expand as knowledge of the genetic basis of rare disease increases.

**Table 6. NIPD tests listed on the National Genomic Test Directory for England**

Clinical indication	Target/ genes	Test method
Non-invasive prenatal sexing	Other	NIPD
NIPD using paternal exclusion testing for very rare conditions where familial mutation is known	As per tested relative	NIPD
NIPD for congenital adrenal hyperplasia - <i>CYP21A2</i> haplotype testing	<i>CYP21A2</i>	NIPD
NIPD for cystic fibrosis - haplotype testing	<i>CFTR</i>	NIPD
NIPD for cystic fibrosis - mutation testing	<i>CFTR</i>	NIPD
NIPD for Apert syndrome - mutation testing	<i>FGFR2</i>	NIPD
NIPD for Crouzon syndrome with acanthosis nigricans - mutation testing	<i>FGFR3</i>	NIPD
NIPD for <i>FGFR2</i> -related craniosynostosis syndromes - mutation testing	<i>FGFR2</i>	NIPD
NIPD for <i>FGFR3</i> -related skeletal dysplasias - mutation testing	<i>FGFR3</i>	NIPD
NIPD for Duchenne and Becker muscular dystrophy - haplotype testing	Dystrophin	NIPD
NIPD for spinal muscular atrophy - mutation testing	<i>SMN1</i>	NIPD

A significant limitation in the use of NIPD for single gene disorders is that it is currently confined to cases with a paternally inherited or *de novo* mutation. In maternally inherited cases, high background levels of the mutation in the mother's DNA rule out accurate diagnosis of the mutation in the fetus. This also impacts on diagnosis of recessive disorders, where a confident diagnosis can currently only be made (and so the test is only offered) where the paternal mutation is one of a specific set of variants AND differs from the maternal variants.

## 8.6 Future developments in prenatal screening and diagnosis

### *Potential use of NIPT as a first line test in screening*

Health technology assessments carried out in other countries suggest that compared with second line use of NIPT for trisomies 21, 18 and 13, use as a first line test detects more chromosomal abnormalities but increases costs [130]. NIPT screening as a first line test is currently used in Belgium [131] and in the Netherlands as part of the nationwide TRIDENT-2 implementation study [132]. In future the NHS may wish to consider offering NIPT as a first line test. The Royal College of Obstetricians and Gynaecologists is in the process of producing a guideline on NIPT jointly with the British Maternal and Fetal Medicine Society, expected to be published in early 2021.

### *Forthcoming changes in national antenatal screening policy*

QF-PCR, karyotype or array CGH would only be required for the small proportion of women who return a positive NIPT test and require a confirmatory invasive diagnosis, and where other structural abnormalities on ultrasound or family history indicate that such testing might be useful regardless of NIPT status. Demand for cytogenetic methods such as QF-PCR and karyotyping would therefore be expected to decline further in the prenatal sector, and capacity to deliver NGS-based NIPT would need to increase.

### *Demographics*

Population size, age distribution and pregnancy rates impact on demand for testing, in particular the shift towards increased age of the pregnant population will influence demand for testing subsequent to aneuploidy screening, as the risks of trisomy increase with maternal age.

### *Use of whole exome trio analysis in prenatal diagnosis*

The most recent results from the prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE) study show that whole exome sequencing (WES) of fetuses improves genetic characterisation of fetuses with structural abnormalities. WES performed on DNA samples obtained during invasive testing of 610 fetuses, including 596 fetus-parental trios, identified a diagnostic genetic variant in 52 (8.5%) of the 610 fetuses assessed. An additional 24 (3.9%) fetuses had a variant of uncertain significance that had potential clinical usefulness [133]. The study found that WES enabled more accurate predictions of fetal prognosis and risk of recurrence in future pregnancies.

However, the overall rate of diagnosis was lower than previously predicted, and some structural abnormalities were much more strongly associated with a genetic variant than others. Therefore, although there is now evidence that WES improves prenatal diagnosis, it is advised that before clinical implementation it is carefully considered in which cases of structural abnormalities it is best used.

A recent economic analysis based on the results of the PAGE study described above found that it was more cost effective to provide WES alongside chromosomal microarray testing than to use WES alone. When WES is priced at £2,100, performing WES alone prenatally would cost £31,410 per additional genetic diagnosis. A stepwise approach of chromosomal microarray followed by WES would cost £24,657 per additional genetic diagnosis [\[134\]](#).

WES for fetal anomalies with a likely genetic cause is now available as a test on the National Genomic Test Directory in England. Ongoing developments to reduce interpretive complexity, improve turnaround times and demonstrate clinical utility may lead to increased use of these methods in a clinical laboratory setting, initially for clearly defined clinical indications. In addition to making WES available, NHS England are considering use of WGS for prenatal diagnosis of suspected genetic abnormalities in future.

## 8.7 Conclusions for prenatal genomic analysis

Prenatal genomic analysis services are complex and highly varied with respect to the reasons for referral and testing methodologies used. To some extent, this variation is a consequence of the gradual evolution in testing technologies available and the variable speed of adoption of best practice across the NHS. It is also a consequence of the wide range of circumstances under which prenatal testing is undertaken. Family history, clinical question, sample availability, stage of gestation and prior test results can all vary from case to case, each necessitating a potentially different testing strategy. Thus, while the implementation and increased use of non-invasive testing and diagnosis should reduce overall the need for QF-PCR, array CGH and karyotype testing on fetal tissue, capacity for all of these types of testing will need to be maintained as there will remain a significant (although difficult to quantify or predict) number of referrals for which they are the most effective test. Similarly, a shift to use of WES for prenatal diagnostic testing will increase the need for sequencing technologies, however evidence suggests the most cost-effective approach is to maintain use of microarray testing alongside WES. Therefore the capacity to deliver testing via both of these techniques will need to be maintained. If health systems choose to implement WES for fetal anomalies, there will also be additional demands on the bioinformatics and data storage infrastructure needed to deliver testing, especially if trio analysis is used.

## 8.8 Genetic testing in the postnatal period

Postnatal constitutional genomic analysis refers to the genetic testing that takes place from birth throughout the rest of an individual's life. According to data collected by the UKGTN member laboratories, in 2016/2017 the genetic test report rate for Scottish residents was 387.6 per 100,000 population; (95% confidence interval 382.4 – 393.0). This was significantly higher than that of England (293.2; 291.8 – 294.6). Whilst postnatal constitutional genetic testing can occur across the life course, and for a broad spectrum of

conditions, a large amount of testing is concentrated in two peaks in the life course where phenotypic features become apparent: (i) at birth, and (ii) in early childhood, when physical and intellectual developmental signs emerge.

In the following sections we provide examples of clinical applications of genomic analysis in each of the following periods:

- » Newborn
- » Childhood
- » Adolescence
- » Adult
- » Post-mortem

This chapter illustrates some of the more common reasons for referral at specific points in the life course, however genetic investigations for many conditions are not confined to specific life stages. The methods used for postnatal genetic analysis typically depend on the type of genetic variant to be investigated rather than the life stage, as such choice of genome analysis method is discussed separately from life stage in which the test could be used. Where methods have been or are anticipated to be specifically designated for certain life stages this will be specifically mentioned in the relevant section, for example regarding the new initiative to genome sequence all critically ill babies and children.

Whilst genetic testing has been used to establish the basis for rare and inherited disease, recently there has been increased interest in implementing genetic analysis to understand how otherwise healthy people respond to drugs based on the presence of specific genetic variants (pharmacogenomics). There are also initiatives underway to understand the genetic basis of common disease, in order to identify at-risk individuals for these diseases in order to help provide interventions, as well as to help improve disease diagnoses. These types of tests would not necessarily be confined to any stage of the postnatal period, and as they enter mainstream use would be expected to increase the demand on genetic testing services.

## 8.9 Clinical utility of genome analysis in the postnatal period

Genomic analysis in the postnatal period may be requested for a wide range of clinical conditions and the genetic diagnosis provided offers utility for a number of reasons, most commonly to:

In children:

- » Help parents care for a child with a disability and to access educational and social support. A diagnosis allows clearer prognosis, and parents/families can seek support from others with the same condition
- » Help parents make choices about future pregnancies
- » Aid clinicians in the management and treatment of the individual

In adults:

- » Provide an understanding of symptoms and allow individuals to seek specific sources of support e.g. disease support groups, educational or living assistance
- » Help guide reproductive options in the case of infertility or recurrent pregnancy loss
- » Guide current management and/or plan for future management for late onset conditions
- » Enable predictive testing in at-risk relatives: relatives found to have inherited a disease causing variant may be able to have appropriate clinical follow up, whilst those found not inherited a disease causing variant can be discharged from follow up

In both adults and children, understanding the genetic basis of some conditions enable access to therapies that target their genetic alteration. For example, in future those with sickle cell disease may be able to access gene therapies that could potentially cure the disease. As more gene therapies are developed for genetic based diseases, the utility of clinical genome analysis for some conditions may increase further.

## 8.10 Genomic analysis of newborns

### *Current approaches*

In many cases, the earlier a genetic condition is diagnosed, the better the outcome for the individual and their family, both due to the opportunity to begin treatment and/or better management of the condition sooner, and to help provide answers to concerns a family may have. The postnatal period in the weeks following birth is therefore particularly important for the diagnosis of genetic conditions.

There are currently two key opportunities for using genomic analysis to identify and/or confirm suspected conditions.

### *The newborn physical examination*

In the postnatal period, genetic testing may be recommended when phenotypic features, often ascertained through the newborn physical examination, are suggestive of an underlying genetic disorder.

### *The newborn bloodspot screening programme*

In addition to the newborn physical examination, the parents of all babies in the UK are offered screening through the newborn bloodspot screening programme. Using biochemical screening, the programme in Scotland, England and Wales screens neonates for nine conditions when they are five days old:

- » phenylketonuria (PKU)
- » congenital hypothyroidism (CHT)
- » cystic fibrosis (CF)
- » medium chain acyl-CoA dehydrogenase deficiency (MCADD)

- » sickle cell disorder (SCD)
- » maple syrup urine disease (MSUD)
- » isovaleric acidaemia (IVA)
- » glutaric aciduria type 1 (GA1)
- » homocystinuria (HCU)

Suspected cases are offered confirmatory genetic testing. Diagnosis of a genetic condition can inform clinical management of the case, and provide information for the parents to consider in terms of future reproductive choices. Such referrals are considered urgent and prioritised by the laboratory, as prompt treatment for many of the conditions can prevent serious and irreversible adverse health outcomes, in addition to consideration of the emotional impact on parents who are awaiting a potential diagnosis.

Although none of these conditions are currently tested for using genetic analysis, there is potential for genetic tests to replace these phenotypic tests, as is already the case in some countries. Use of genetic testing could improve rates of diagnoses for some conditions such as cystic fibrosis, and also provide additional useful information for potentially affected relatives. Opportunities for use of genetic testing in the newborn blood spot programme were recently highlighted in the UK National Screening Committee's report *Generation genome and the opportunities for screening programmes* [135].

### *Ongoing and future developments*

#### **Addition of genetic test for SCID to newborn blood spot screening**

The UK National Screening Committee (UK NSC) has recently recommended that screening all babies for severe combined immunodeficiency (SCID) should be evaluated in the NHS. Based on this evaluation a recommendation will be made on whether to incorporate SCID into the newborn blood spot programme. The current approach to diagnosing SCID is based on family history which only identifies 30% of babies. Interestingly, the PerkinElmer SCID test that was recently selected by PHE is the first blood spot screening test to make use of genetic analysis, though not in the conventional manner of identifying specific mutations underlying SCID. The test will detect T-cell receptor excision circles (TRECs) — small DNA circles that are created as T cells pass through the thymus [136]. The test is semi-quantitative and uses a combination of PCR and the non-genetic based time-resolved fluorescence resonance energy transfer (TR-FRET) technology.

The screening test could be based on phenotypic markers, but as with other screening tests would then require a genetic test to confirm the diagnosis. The evaluation is due to start in September 2020, with 6 hospital trusts in NHS England taking part. The evaluation is planned to last 2 years, with over 800,000 newborns being tested [137, 138].

## Use of NGS in blood spot test analysis

Use of NGS for cystic fibrosis diagnosis from dried blood spots is currently being piloted in the NHS in England [135]. NGS enables a broader range of pathogenic mutations to be tested for than the alternative targeted mutation testing and MLPA tests used for cystic fibrosis diagnosis, and use of dried blood spots would reduce the need for additional samples, potentially saving time and resources [135]. It has been shown that targeted NGS performed on dried blood spots is feasible and has the potential to improve diagnosis of cystic fibrosis and the conditions in the blood spot tests that currently require genetic diagnosis (MCADD, IVA, GA1 and some suspected haemoglobinopathy cases) [135]. However, this study concluded that it seemed unlikely a targeted NGS panel would be cost effective as a first line test as part of the current blood spot programme, and would likely be used as a second line test after the initial phenotypic tests.

## Considerations over whole genome sequencing of healthy newborns at birth

A recent report by Genetic Alliance UK recommended establishing a pilot scheme to trial genome sequencing of healthy newborns alongside the current screening programme [39]. The aim would be to increase the number of rare conditions tested for and diagnosed early in life. The Secretary of State for Health announced in November 2019 that the NHS planned to sequence the DNA of all babies born in the UK alongside the blood spot test, starting with a pilot of 20,000 children. As of April 2020, no further details are available regarding this pilot.

## 8.11 Genomic analysis during childhood

### *Current approaches*

Individuals may be referred during childhood for genetic testing if their physical and/or intellectual development gives cause for concern. Some critically ill babies and children may be referred for genetic testing if their illness cannot be diagnosed based on clinical symptoms and routine tests. Developmental delay or intellectual disability are common reasons for referral for genetic testing later in childhood. Phenotypic features that may not be evident in younger children can become apparent later in childhood and prompt testing, for example for Fragile X syndrome, microdeletion syndromes such as Williams syndrome, or monogenic disorders such as neurofibromatosis.

As with most genetic tests performed in the postnatal period, the type of test used currently depends on if a specific genetic variant is suspected and if so on the type of variant, or if the likely genetic cause is entirely unknown and a larger scale genome wide investigation is required. These considerations are discussed further in Section 8.16.

Establishing a genetic cause for a critically ill baby or child can provide answers faster than phenotypic clinical investigations, which as well as being time consuming, can also be invasive. Receiving a diagnosis through genetic analysis can confirm the best course of clinical management, and in some cases lead to changes in management and subsequent improvement in the condition. Similarly, understanding a child's developmental issues can improve their clinical management through increased monitoring for physical manifestations of the identified disorder, e.g. monitoring cardiac status due to aortic

narrowing in Williams syndrome. In all cases, having a diagnosis for a condition can provide relief to families, who may then be able to understand the needs of their child better and access support from relevant agencies and groups.

### *Ongoing and future developments*

#### **Use of WES and WGS for rapid diagnosis of rare disease in critically ill babies and children**

NHS England are providing WES to help rapidly diagnose rare disease in critically ill babies and children, with 700 babies and children expected to benefit each year [\[9\]](#).

It has been shown by the Deciphering Developmental Disorders (DDD) / DECIPHER study that trio analysis, where the exomes of both parents as well as the child are sequenced and compared, increases the number of diagnoses returned. Since the service became available through the Genomic Medicine Service in October 2019, there have been over 140 referrals with almost 40% of patients receiving a diagnosis that was previously unavailable [\[112\]](#). In Scotland, NHS Lothian started trio based WES in September 2019, with families already receiving diagnoses for their children. The service aims to provide diagnoses for 300 patients per year (900 samples for trio analyses) [\[86\]](#). The service is currently continuing to use the DECIPHER study database to record patients and facilitating multidisciplinary team (MDT) meetings, as there is no viable alternative available to deliver a service of equivalent quality. The service is currently working towards a turnaround time of 112 days [\[86\]](#).

In future, NHS England plans to move forward to using WGS, once it is available and it is verified that clinical turnaround times are acceptable. WGS performed on 195 parent child trios (567 samples) was able to provide a genetic diagnosis for 21% of critically ill children recruited from neonatal (NICU) and paediatric (PICU) intensive care units in Cambridge University Hospitals NHS Foundation Trust [\[139\]](#). Importantly, in this study it was found that over 65% of the diagnoses resulted in a modification of treatment or care pathways and that the diagnostic timescale needed to impact decision-making took from just 2-3 weeks in some cases. NHS Scotland are continuing work to establish the utility of WGS tests, including plans to sequence and analyse a further 500 genomes following the first 1000 genomes analysed as part of the 100,000 Genomes Project [\[86\]](#).

Both WES and WGS are able to provide more diagnoses for rare disease where the genetic cause is unknown than genomic methods based on panel tests and arrays. In addition the phenotypic symptoms used to drive choice of single gene and panel tests can be unreliable in neonates and infants, therefore using a WGS method that does not rely on physical symptoms can be advantageous [\[139\]](#). These methods have also been shown to be fast enough to be used in the clinic, and can prevent children having to undergo lengthy and intrusive clinical tests to try and determine their condition.

## **8.12 Reasons for undertaking genomic analysis in adolescence**

Issues relating to sexual development such as delayed puberty or amenorrhea may prompt referral for testing e.g. for sex chromosome aneuploidies such as Turner syndrome in females and Klinefelter syndrome in males. Again, diagnosis may help with clinical management, in the case of Turner syndrome; individuals may also be monitored for

cardiac symptoms. The diagnosis may help to inform decisions regarding future assisted conception, for example for males with Klinefelter syndrome.

### 8.13 Reasons for undertaking genomic analysis in adulthood

Adults who are already affected by a disease may be tested in order to confirm a diagnosis, to provide information of prognostic relevance and/or guide therapeutic options. In unaffected adults, predictive testing may be carried out if they have been identified as being at risk of a late onset condition with a known genetic cause where a family member is known or suspected to carry such a variant. Examples include testing for Huntington's disease. Testing may also be offered to individuals who have a high likelihood of carrying variants which confer a high risk of developing disease, such as *BRCA1/2* alleles associated with an increased risk of breast cancer. A genetic diagnosis may allow individuals to prepare for the future, and where available, allow access to prophylactic options. Testing for late onset disorders such as these relies upon careful genetic counselling because of the potential impact on the individual and their family members.

Testing in adulthood may also be initiated for reproductive reasons, for example to establish a genetic cause for infertility e.g. testing for Y chromosome deletions as a cause of male infertility. Genetic testing may be used to establish a cause for recurrent miscarriage, for example karyotyping to identify balanced translocations. A genetic diagnosis may therefore guide reproductive decision making and, in some cases, with the use of techniques such as PGD, invasive prenatal diagnosis or NIPD, allow couples to select healthy embryos for implantation and make choices regarding the continuation of pregnancy based on a prenatal genetic diagnosis.

Testing may also be carried out in parents prior to conception (where there is a known family history or other reason to suspect they may be carriers for a recessive disorder), or after the conception or birth of a child with a known genetic disorder to establish whether the parents are carriers of the genetic abnormality. See the earlier section on prenatal testing for more information.

### 8.14 Reasons for undertaking genomic analysis postmortem

Genetic testing may be conducted on tissue postmortem where a genetic disorder is suspected as the underlying cause of death, for example in cases of sudden cardiac death and expected cardiomyopathy, where there may be significant risk implications for living relatives. Obtaining genetic information alongside personal and family history can help refine a disease diagnosis and aid in designing prevention strategies for other family members, whilst also contributing to research efforts and public health strategies. Additionally genome analysis may occasionally be conducted on fetal tissue following miscarriage or termination of pregnancy, to determine the genetic cause of any observed abnormality or recurrent miscarriage [140]. The utility of testing under these circumstances is to establish the cause of any observed major abnormality or of recurrent miscarriage, with the aim of informing future reproductive choice. Identifying the genetic abnormality can also help inform future approaches to clinical management of pregnancies in women who are deemed at high risk of having children with the same abnormality.

As with all types of genetic testing, the methods used will depend on the type of genetic variation suspected. In the case of fetal miscarriage, the cause is likely to be due to larger scale chromosomal aberrations, therefore QF-PCR can first be used to detect common trisomies, triploidy and sex chromosome aneuploidy. If no results are yielded, array CGH may then be used for genome wide detection of copy number variation [141]. In the case of sudden death, it is more likely that a genetic cause will be due to variation within a gene, therefore higher resolution testing strategies will be required. This could take the form of a panel test to screen for likely mutations, with NGS being used to provide higher coverage to detect rare mutations in suspected genes and increase diagnostic yield.

### *Testing in sudden cardiac death cases*

The use of genetic testing in sudden cardiac death is an area of particular importance. Sudden death is estimated to account for 10-20% of total mortality worldwide, with the majority being sudden cardiac deaths [142]. In those under the age of 35 the incidence of sudden death is much lower, but the deaths that do occur often have no causative pathology and are more likely to have a genetic component, so that testing in this age group may be particularly important [142]. Recent European recommendations endorsed by a number of professional bodies suggest that during an autopsy blood and tissue samples suitable for genetic analysis should be taken and stored alongside detailed phenotypic information, to allow for future genetic analysis if required, as is already recommended by the Royal College of Pathologists in the UK [142, 143]. There are potential issues surrounding consent over the storage and future use of post-mortem genetic material, and family members should be informed and asked to provide consent.

The utility of postmortem genetic testing in different clinical scenarios is still uncertain. A recent study found that pathogenic or likely pathogenic variants were identified in nine of 57 (32%) of cases of sudden cardiac death where the autopsy identified structural abnormalities diagnostic of cardiomyopathy, similar to the diagnostic yield in living patients [144]. This is compared to pathogenic or likely pathogenic variants identified in one of 57 (3%) of cases with autopsy findings of uncertain significance [144]. Therefore, it is possible that genetic testing may be less useful in cases where the autopsy result is uncertain, however more work in larger studies is needed to determine the clinical and cost effectiveness of different testing approaches. In addition, if panels are expanded to include a larger number of genes than the 77 screened by NGS in this study, the diagnostic yield may increase.

## **8.15 Future opportunities for genomic analysis not confined to lifecourse stage**

In addition to the conventional uses of genome analysis described so far regarding rare and inherited disease, there have recently been renewed efforts to make use of genome analysis in the broader population for other clinical purposes. These include pharmacogenomic (or pharmacogenetic) testing to facilitate more targeted prescribing of drugs based on an individual's specific DNA variants, and the use of polygenic risk scores which can be used to estimate risk of common diseases based on the analysis of multiple common DNA variants. As use of polygenic risk scores is still uncertain and no tests have

yet been developed for mainstream clinical use, these shall not be described further here but will be discussed in chapter 10. Pharmacogenomic testing is much closer to clinical application and is described below.

### *Pharmacogenomic testing*

When a drug is being considered for prescription, knowledge of a patient's genotype may be used to aid in determining a therapeutic strategy, determining an appropriate dosage, or assessing the likelihood of benefit or toxicity. There are various time points throughout an individual's life course when they could have their pharmacogenomic status determined, either reactively or pre-emptively.

A reactive test would be done in response to a specific clinical situation, such as:

- » Select targeted therapies appropriately, for example cancer drugs effective against tumours harbouring specific mutations, or in other diseases e.g. PCSK9 inhibitors to treat those with familial hypocholesterolaemia caused by gain of function mutations in the *PCSK9* gene.
- » Alleviate or avoid adverse drug reactions, which have been estimated to cost the NHS around £1 billion annually [145].
- » Apply more accurate dosing strategies, where genetic make-up can affect a patient's metabolism of a drug e.g. variation in the *CYP2D6* gene plays a direct role in the metabolism of many commonly prescribed medications
- » Support the monitoring of drug responses,
- » Determining reasons for treatment failure.

In some clinical circumstances this decision might need to be made quickly – as such, there is research ongoing into rapid point of care testing to offer quick turnaround times. For reactive tests it is likely that detection of specific targets is required, therefore requiring genetic tests that can test for mutations in certain genes rather than broader genome wide testing.

Pre-emptive testing would be carried out in patients before the identification of any specific clinical need, either in healthy individuals or potentially an individual's first contact with the healthcare service, so that results would be available if and/or when necessary. It would likely require a comprehensive pharmacogenomic test through use of either WGS, a panel or genotyping microarray for all known gene-drug pairs. The resulting pharmacogenomics genotype information of an individual would need to be available in their electronic health records (EHR), or as a report that could be reviewed for drug suitability when required. An alternative strategy would be for patients to carry their own pharmacogenomic information in the form of a 'safety-code card' that can be scanned to retrieve pharmacogenomic -based dosing recommendations.

### *Current and upcoming uses of pharmacogenomic testing*

Pharmacogenomic testing is currently not widely used in the NHS, but is anticipated to have major impact on clinical genetics in the future (see chapter 10). The main use of pharmacogenomic testing at present is in cancer, in the form of companion diagnostic

tests to determine the mutation status of a patient's tumour, to understand if they are eligible for specific targeted therapies. This is a rather specialised use of reactive pharmacogenomic testing as it relies on the examination of the cancer genome rather than the 'normal' human genome, and is discussed further in chapter 9.2. Similarly, companion diagnostic genetic tests also exist for non-cancer therapies which target specific rare inherited variants. Examples of drugs requiring genetic tests that are licenced for use by NICE include ivacaftor for cystic fibrosis [146] and PCSK9 inhibitors for familial hypercholesterolaemia [147].

Testing for genes involved in the metabolism of chemotherapy drugs is being trialled for implementation at some cancer centres within NHS hospitals. Dihydropyrimidine dehydrogenase (DPYD) gene variants are implicated in toxicity arising from chemotherapy drugs [148]. Clinical implementation of *DPYD* genotyping for metastatic breast cancer is being successfully trialled in routine clinical practice in a large cancer centre in London, with the aim of reducing the risk of severe side effects of these drugs [149].

The majority of pharmacogenomic tests being considered for clinical use are for the detection of common genetic variants present in the healthy genome, which can alter the biological response to drugs. However, this is the area where there is least routine implementation within the NHS, despite many approved medications containing pharmacogenetic information in their prescription recommendations [150]. For example pre-treatment genetic testing to avoid adverse drug responses is advised in British prescribing guidelines for the HIV drug abacavir (testing for the *HLA-B\*57:01* variant), and for the epilepsy drug Carbamazepine in individuals of Han Chinese or Thai origin (testing for the *HLA-B\*1502* variant) [151].

## 8.16 Factors determining current approaches to postnatal constitutional genomic analysis

The method used to undertake genomic analysis at different stages across the postnatal life course depends primarily on what is known about the genomic variation underlying the clinical condition being investigated.

**Number of genes/genomic regions potentially implicated in pathogenesis** – The number of genes or genomic regions in which pathogenic variation might also explain the clinical condition must also be taken into account. Thus, if all cases of a condition are known to be caused by variation in a single gene, a targeted PCR or sequencing assay for that gene (or parts thereof) would be warranted. However, if a condition such as developmental delay, with varying and complex phenotypes and over a thousand genes and genomic regions known to be implicated in its pathogenesis, is being investigated then a genome wide (either array CGH or WES/WGS) approach would be warranted.

The importance of genome wide assays, in particular WES and WGS is evidenced by the recent decision to perform WES on all critically ill infants and children in England, with a future move to WGS anticipated. Scotland is also in the process of rolling out WES for developmental disorders in babies and children. Importantly, trio analysis will be used whenever possible for whole exome sequencing, to improve the rate of successful diagnoses. In addition, rare disease patients in England will be able to access WGS based testing via the Genomic Medicine Service. This decision follows results from a pilot study of

the 100,000 Genomes Project, which found that whilst the rate of diagnosis varies widely depending on disease type, overall a conclusive molecular diagnosis was returned for 16.1% of patient records reviewed and the discoveries of new variants were able to inform treatment decisions [152]. The same study found use of WGS was able to improve clinical reporting compared to WES.

NHS Scotland is awaiting further evidence of the clinical utility of WGS, but is in the process of implementing clinical exome sequencing throughout the Scottish Genetics Laboratory Consortium. Exome testing has been rolled out in four laboratories and demand is expected to increase as the service develops further [86]. It is planned to formally review the exome panels offered by each of the four centres every three months.

Clinical services may prefer to use phenotype specific panel tests, rather than truly genome-wide approaches, where multiple genes need to be simultaneously investigated. However, the costs and complexity associated with developing, validating and updating these assays as new genes are discovered is such that they are rapidly being replaced by virtual panel tests. In these cases a single genome-wide sequencing or array based assay can be used to test for any combination of genomic variation for any clinical condition by the use of selective bioinformatic analysis pipelines that focus analysis and interpretation on the subset of genes/genomic regions known to be related to the symptoms of the patient being investigated.

**Size of genomic variants being investigated** – If the most likely causal pathogenic genomic variations are large scale, this would indicate that karyotype, FISH or array CGH should be used, whereas if only small insertions or deletions or single nucleotide variations are most likely to be causal, then PCR or sequencing based approaches are more likely to be appropriate.

**Knowledge of familial variants** – Where the genetic variation underlying a condition within a family is already known, then a targeted PCR, sequencing or FISH approach can be used to test only for that specific variation in relatives who may be at risk.

## 8.17 Impact of rapid developments in genomic technology and knowledge on postnatal constitutional genomic analysis

### *Shift to the use of genome-wide assays as first line tests for postnatal constitutional referrals*

The most recent audit of genetic test activity in the UK undertaken by the ACGS shows that around over 264,000 postnatal constitutional referrals were made in 2015-2016, an increase of 9.5% from 2014-15, as determined from report number [106]. This is split about 60%-40% between testing of all or part of individual genes or panels of genes, and testing of larger genomic regions using mainly array-CGH and karyotyping. However, it is expected that there will be continued increased use of genome wide assays that will result in reduced use of individual gene and panel testing. Collectively, the genome-wide NGS based assays CES, WGS, WES and array CGH can now, in principle, provide test coverage of almost all conceivable clinically significant genomic variation.

Importantly, a clinician can now order a 'bespoke' test for their patient, by selecting the virtual panel of genes most relevant to their condition, without the laboratory having to develop a correspondingly bespoke assay. Instead the laboratory is able to use the same generic sequencing and informatics analysis for most patients, and apply tailoring by selective interpretation of genomic variants only in the subset of genes of clinical interest, saving time and a great deal of cost. This approach is already being applied in a number of NHS molecular genetics laboratories. The availability of these approaches is obviating the need to develop new, bespoke assays for new or modified 'panel' tests.

The capacity of laboratories, therefore, need no longer be limited by the speed and cost-effectiveness with which they can develop and validate new single gene or gene panel tests. Instead it will be limited by their ability to filter and interpret which of the millions of genomic variants revealed by genome-wide assays are relevant to the disease of the patient being investigated, and the cost of the sequencing assays and informatics required to achieve this. The challenges associated with interpreting variants of uncertain significance (VUS) that arise from genome wide sequencing methods, and the subsequent impact on the healthcare system are discussed in more detail in chapter 7.

Meeting the interpretative challenge in turn depends on the speed at which research yields new knowledge of genotype-phenotype associations, and the speed with which these can be clinically validated and implemented into bioinformatic analysis algorithms. It also depends on the availability of data storage, but most of all on availability of interpretative 'power' either in human or machine form, to undertake the increasing number of genome analyses that will be requested. The decision to implement trio analysis when performing whole exome sequencing to determine the genetic cause of disease in critically ill children will increase the demand on the system further, by effectively trebling the number of samples requiring analysis for each case (see chapter 7).

### *Increased demand for genetic services outside of conventional clinical genetics*

There is a growing move towards implementing widespread pharmacogenomic testing in the UK, with work currently ongoing in England to evaluate which pharmacogenomic tests should be offered on the National Genomic Test Directory. In addition, there is currently a large amount of interest in generating polygenic scores to help manage common disease. Both these applications, if implemented, may substantially increase the demand on current genetic testing services. In the case of pharmacogenomic testing, the tests currently used will likely be panel tests for specific mutations, whilst the type of tests required for polygenic scores will likely involve the use of genome wide microarrays for specific SNPs. Future considerations for more widespread use of genetic testing are discussed in chapter 10.

### *Convergence towards whole genome sequencing as the genome wide analysis method of choice for most postnatal constitutional referrals*

Previously, genome-wide analysis undertaken in a clinical context was for the most part done using either array CGH (to detect large scale copy number variation) or (more rarely) whole exome sequencing (WES) (to detect single nucleotide variants or small indels within genes and some adjacent non-coding regions). With the use of trio analysis in particular improving its performance, the use of CES and WES is now increasing, for example with

its use as first line test in critically ill children. However, neither method can be used as a comprehensive genome-wide assay for all types of genomic variation as each is limited either by the scale of variation that they can detect, or by the completeness with which they cover the genome.

It is now possible to assay simultaneously structural and copy number variants, single nucleotide variants and indels using whole genome sequencing. As WGS requires no pre-selection of the target regions to be analysed (by probe hybridisation or selective amplification) it provides more even coverage across the genome than WES and greater resolution than array CGH, and has been shown to be equal or superior in diagnostic sensitivity and specificity to both. The comprehensive genomic information provided by WGS makes a compelling argument for its implementation as the clinical genome wide analysis assay of choice. However, the practical challenges of achieving the transformation from the current diverse landscape of testing activity remain significant, as evidenced by the experiences of those engaged in the 100,000 Genomes Project.

Despite the allure of whole genome sequencing as a solution for the provision of clinical genome analysis, it remains the case that during this transitional phase there will continue to be a requirement for other methods of genetic testing. There are some genes and genomic regions that cannot be assayed with sensitivity and specificity by current NGS technologies, such as the triplet repeat region of the Huntington and spinocerebellar ataxia genes. There are also diseases for which only a single gene, or region of a gene needs to be analysed and for which WGS, is an unnecessarily cumbersome and expensive assay. In these cases (and also where only confirmation of a known familial variant is required), Sanger sequencing, PCR, FISH and even in some cases karyotyping will remain important tests for which some, provision must be maintained until technological developments and cost reductions in NGS (or single molecule sequencing) make them redundant.

With the completion of the 100,000 Genomes Project and the subsequent availability of whole genome sequencing on the National Genomic Test Directory in England, it is clear that despite its challenges WGS will become a part of mainstream healthcare. By continuing to evaluate the utility of WGS and planning to continue to contribute samples towards the next UK phase of sequencing another 300,000-500,000 genomes, NHS Scotland will generate evidence for future testing service development. Mainstream implementation of routine WGS differs from previous incremental innovations, in that it is a multiplex technology that allows a single methodology for testing to be applied to a number of disorders. Consequently, it offers a unique opportunity to simultaneously consolidate, simplify and also broaden the coverage of the existing repertoire of tests. Whether or not this occurs within the three year time frame considered by this review depends largely on whether or not investment is made in developing the infrastructure, informatics and scientific and clinical expertise needed to deliver such a radical shift. This includes accumulation of sufficient clinically validated data from research to facilitate this process.

As noted above, whole genome sequencing cannot replace all postnatal genomic analyses, and so retaining capacity for existing techniques (most of which also have utility in prenatal and oncological settings) will be required, at least until NGS or other sequencing methods (such as single molecule long read sequencing) develop to fill the capability gaps left by current short read NGS platforms. The timescale for such developments is uncertain,

but given the pace of sequencing technology and analytical developments in the recent past it is possible that single molecule sequencing, for example, will reach sufficient accuracy and throughput to have clinical utility within the next five years.

## 9 Genome analysis for cancer management

All cancers arise from the accumulation of pathological alterations to the genomes of somatic cells; most of these mutations are not present in the germline (or constitutional DNA), and consequently are not passed on to the next generation. However, there are some families in which several members have been affected with the same cancer type, often at an early age, and the disease shows a Mendelian autosomal dominant pattern of inheritance, suggesting the existence of a single germline mutation conferring high risk of disease. Examples of this include breast-ovarian cancer, familial adenomatous polyposis (FAP) and retinoblastoma.

Given the pivotal role played by somatic genomic variation in the pathogenesis of all cancers, the potential clinical impact of genomic analysis in oncology is vast. However, deficits in technology, knowledge and the current configuration of pathology and wider care pathways mean that the current clinical utility of genome analysis in oncology practice is more limited than might be expected given the fundamental genetic aetiology of this disease.

### *Genomic analysis has established clinical utility for:*

**Risk management** – Identifying currently healthy individuals carrying germline genomic variations (usually inherited) that place them at significantly higher risk of cancer than the general population enables the development of personalised risk management plans, involving options such as more regular screening and surgical or chemo-prophylaxis. The classical example of this use of germline genome analysis is the testing of women with an affected relative and/or significant family history of breast and ovarian cancer for *BRCA1* and *BRCA2* mutations.

**Precision diagnosis and prognosis** – Characterisation of the changes in the somatic genome that underlie tumour development in each individual patient can enable clinicians to identify molecular subtypes of disease with varying prognoses, and in turn direct patients to the most appropriate care pathways. For example, breast cancer has been classified into five molecular subtypes that are strongly associated with clinical outcome and acute myeloid leukaemia is now defined by mutations in genes such as *FTL3* and *KIT*, along with other cytogenetic abnormalities, which heavily influence prognosis.

**Personalised treatment management** – Targeted therapies are agents that block the growth and spread of cancers by interfering with the function of proteins whose abnormal expression or activity is particularly associated with cancer cells. Immunotherapies stimulate the patient's own immune system to recognise cancer cells as foreign bodies and attack these cancer cells. Consequently, the utility of such therapies in individual patients is dependent on the presence of targets in their cancer cells. A number of such therapies have been approved for use by NICE. Some examples where somatic genetic information is being used to guide decisions on the use of these therapies include mutations in the *HER2*, *ABL*, *KIT*, *KRAS*, and *EGFR* genes.

## 9.1 Current approaches to germline testing for inherited cancer susceptibility

Inherited germline mutations within a number of genes are known to increase susceptibility to one or more types of cancer (table 7). For example, inherited genetic variants in the *BRCA1* and *BRCA2* genes can increase risk of breast and ovarian cancer and variants in *TP53*, a 'master regulator' of cell proliferation, causes Li-Fraumeni syndrome and an increased susceptibility to a wide range of childhood and adult cancers.

These inherited forms of cancer have genetic tests available for genes implicated in their development (table 7). Testing for hereditary cancer predisposition is generally offered to affected individuals suspected of having an inherited form of the cancer, or to an unaffected individual who has been classified as having an increased risk based on family history and having one or more affected first-degree relatives. Usually testing of the affected relative is done first in order to try and identify the familial genetic cause, which then allows targeted cascade testing of subsequent family members. Testing for the cancer predisposition in high risk individuals aims to provide information and tailored prevention management for individuals and families identified. These services are primarily offered through Clinical Genetics Departments with the guidance of genetic counsellors and consultants in clinical genetics.

Genetic testing methodologies are primarily sequencing based methods (chapter 6) and include direct sequencing, using either Sanger (simple targeted mutation testing or single gene sequencing) or NGS methods, such as panels. For some clinical indications non-sequencing methods (chapter 5) are used, such as methylation status (relevant in ovarian and colorectal cancer) and MLPA to investigate copy number variations. For some cancers, predisposition is due to pathogenic variants in a single gene, for example retinoblastoma and the *RB1* gene. In other cancers there is emerging evidence that more than one gene could be associated with the cancer. In breast cancer, for example, *BRCA1* and *BRCA2* genes are well established predisposition genes, but there is growing evidence that other genes such as *PALB2* play a role. However, there is some debate about the clinical utility of testing for these genes and as evidence has been building, different panels have had different genes incorporated within them, resulting in variety in panels offered across clinical laboratories in the UK. Initiatives such as PanelApp [\[153\]](#) or specialist groups [\[154\]](#), are underway to try and achieve consensus of which genes are included in panels.

In England, to provide consistent and equitable access to genomic testing the National Genomic Test Directory for rare and inherited disease provides a guide for which tests and methodologies should be used, and for which clinical indication. The National Genomic Test Directory for rare and inherited diseases contains over 25 unique clinical indications for various inherited cancers and tumour disorders (table 7) [\[155\]](#). For these there are over 60 tests available, covered by eight different test methods, namely; MLPA or equivalent, single gene sequencing, small panel, methylation testing, targeted mutation testing, WES or large panel, microsatellite instability (panel or PCR based testing), and DNA repair defect testing. Technical details for these methodologies can be found in chapters 5 and 6.

**Table 7. List of inherited cancers for which genetic testing is listed on the NHS England 2019/2020 National Genomic Test Directory for rare and inherited disease (March 2019 release) [155]**

<b>Clinical indication</b>	<b>Target/genes</b>
<i>CDH1</i> -related cancer syndrome (incl Gastric cancer)	<i>CDH1</i> *
<i>DICER1</i> -related cancer predisposition	<i>DICER1</i> *
Familial melanoma	Small panel <i>BAP1, CDK4, CDKN2A</i>
Fumarate hydratase-related tumour syndromes	<i>FH</i>
Inherited breast cancer and ovarian cancer	<i>BRCA1</i> *, <i>BRCA2</i> *
Inherited colorectal cancer (with or without polyposis)	<i>MLH1,MSH2,MSH6,PMS2</i>
Inherited MMR deficiency (Lynch syndrome)	<i>BRAF V600E</i>
Inherited ovarian cancer (without breast cancer)	<i>BRCA1, BRCA2, MLH1, MSH2, MSH6, PMS2</i> and <i>MLH1</i> hypermethylation
Inherited pancreatic cancer	Small panel, <i>BRCA2, CDKN2A, PALB2</i>
Inherited parathyroid cancer	<i>CDC73</i> *
Inherited polyposis - germline test, Inherited colorectal cancer (with or without polyposis)	<i>APC</i>
Inherited predisposition to GIST	<i>SDHA, SDHC, SDHD</i>
Inherited renal cancer (incl kidney cancer)	<i>FLCN, VHL</i>
Inherited susceptibility to acute lymphoblastoid leukaemia (ALL)	<i>PAX5</i> *, <i>ETV6</i> *
Li Fraumeni Syndrome	<i>TP53</i> *
Multiple endocrine neoplasia type 2	<i>RET</i> *
Multiple monogenic benign skin tumours	<i>FLCN</i>
Neurofibromatosis type 1	<i>NF1</i> *
Neurofibromatosis type 2	<i>NF2</i> *
Nevoid Basal Cell Carcinoma Syndrome or Gorlin syndrome	<i>PTCH1</i> *, <i>SUFU</i> *
Peutz Jegher Syndrome	<i>STK11</i> *
PTEN Hamartoma Tumour Syndrome	<i>PTEN</i> *
Retinoblastoma	<i>RB1</i> *
Schwannomatosis	<i>SMARCB1</i> *, <i>LZTFL1</i> *
Von Hippel Lindau syndrome	<i>VHL</i> *
Wilms tumour with features suggestive of predisposition	<i>WT1</i> *, 11p15 imprinted growth regulatory region

\*currently planned to remain as single gene sequencing

## 9.2 Current approaches to diagnostic and prognostic somatic tumour genome analysis

In those with cancer, analysis of the tumour genomes presents a unique opportunity to better understand cancer biology, achieve diagnosis, improve management and prognosis for individual patients. Genomic analysis is used to varying degrees across the numerous cancer diagnosis and management pathways that exist within the NHS. Utilisation is dependent on:

- » Type of cancer
- » Perceived clinical utility of the test
- » Timely availability of tumour DNA of sufficient quality to undertake accurate analysis
- » Timely availability of clinically meaningful test results that will significantly improve or alter patient management

Current technologies used in diagnostic settings are FISH, karyotyping, PCR, panels, targeted sequencing, MLPA, methylation and more recently WGS (table 4 – strengths and limitations of genome analysis techniques). The National Genomic Test Directory for cancer lists over 950 tests for key groups: haematological tumours, neurological tumours, sarcomas, and solid tumours in adults and paediatrics. These cover over 190 unique clinical indications [156]. The technologies most commonly used are FISH, simple targeted mutation testing, panels, WGS (dual germline and tumour), single gene sequencing, and karyotyping. Simple targeted mutation testing and the use of panels account for over 70% of all the testing technologies listed. Less frequently used are:

- » QF-PCR or equivalent for use in haematological malignancies
- » Complex targeted mutation testing for use in haematological malignancies
- » Targeted mutation testing for methylation analysis used in methylation of MGMT in glioblastomas and methylation of MLH1 in colorectal or breast cancer
- » MLPA or equivalent for use in uveal melanoma
- » Microsatellite instability for use in colorectal cancer
- » Measurable (“minimal”) residual disease (MRD) detection in leukaemia cases using a combination of PCR first, the sequencing and work up using QF-PCR
- » **Gene expression** – using RNA
  - OncotypeDx, Endopredict and Prosignia for use in breast cancer
  - RT-PCR for use in leukaemias
  - PD-1 expression in NSCLC for immunotherapy selection

The directory encourages that when testing for a translocation, wherever possible, DNA/RNA technologies such as panel testing or RT-PCR should be used. However, currently they may not all have an assay available and therefore FISH/targeted mutation testing can also be used when it is the only option available [156].

## Haematological cancers

For haematological cancers, genomic analysis is central to diagnosis and prognosis and is therefore used routinely. Currently, the majority of haemato-oncological analyses are based on karyotyping or FISH assays that detect specific large-scale chromosomal rearrangements, along with less commonly required PCR analysis or NGS panel tests for single nucleotide variants in relevant genes. Additional testing methods used for haematological cancers are RT-PCR to check for translocations and QF-PCR to determine copy number changes. When higher sensitivity is required, such as testing for MRD, alternative methods such as RQ-PCR are required. The use of SNP array karyotyping can be used for the identification of genomic imbalances at a higher resolution than conventional karyotyping and can also detect acquired copy-neutral loss of heterozygosity.

Treatment of patients with a cancer is to achieve complete remission, a response most likely to lead to a cure. However, it is possible for a small cell population to survive treatment that can give rise to relapse of disease and the detection of these cells is critical for prognosis and therapy selection. Several methods now exist that are sensitive enough to permit the identification of small numbers of abnormal cells in bone marrow or blood – accordingly, the topic of MRD is receiving increasing attention. MRD can be detected through phenotypic marker patterns or differential gene patterns by flow cytometry (FCM) (for immunophenotyping), PCR, RQ-PCR, RT-PCR or NGS. Clinically, RQ-PCR is most frequently used for detection of MRD but the development of new diagnostic platforms including NGS are becoming available. Most importantly, the sensitivity and turnaround time needs to be high enough for MRD testing to be clinically useful [\[157, 158\]](#).

## Solid tumours

For solid tumours, genome analysis of the tumour can be undertaken for the purposes of diagnosis and prognosis, but is more commonly carried out to guide treatment decisions, in particular targeted therapies. The majority of tests to guide targeted treatment aim to detect single nucleotide or small indel variants in a single gene, and so can be achieved using a number of different PCR or NGS methods. The type of assay chosen will depend on the particular genetic alteration that is being examined, sample throughput in the laboratory, turnaround time and cost. For example, cost-effective use of NGS is dependent on having sufficient sample quantity and quality to maximise utilisation of expensive sequencing technologies, thus batching to achieve this where sample throughput is low can adversely affect turnaround times and consequently clinical utility. By contrast, PCR based methods rely on much cheaper technology where maximising machine utilisation is less important and consequently more rapid and clinically useful turnaround times. For example, *BCR-ABL1* fusion gene monitoring by qPCR is used in patients with chronic myeloid leukaemia patients undergoing treatment with tyrosine kinase inhibitors.

Genome analysis may also be employed at different points in the care pathway. For example, in the case of breast cancer, biopsy or tumour samples are tested to determine their *HER2* and oestrogen receptor status following first presentation. This provides information relevant to both prognosis – *HER2* positive breast cancers grow faster – and treatment since herceptin (trastuzumab) is a targeted therapy for *HER2*-positive cancers. In the case of non-small-cell lung cancer (NSCLC), *EGFR*-TK testing is also recommended early in the care pathway, usually following initial diagnosis, because *EGFR* inhibitors, which are only effective in *EGFR*-TK positive tumours, are indicated as possible first-line treatments for

NSCLC. By contrast in other cancers, where first line treatment is either surgical or uses non-targeted chemotherapy, molecular testing may only occur if required later in the pathway, usually contingent on the outcome of first line treatment strategies: i.e. if chemotherapy and surgery are ineffective in tackling metastatic disease, targeted therapies that require genomic analysis may be considered as second or third line approaches.

Gene expression via examination of RNA profiles is used in treatment selection, and in breast cancer three such tests are already used in UK clinical practice, namely OncotypeDx, Endopredict and Prosignia. These tests can aid decisions when a clinician may be met with equivocal test results that require these additional tests to assist in determining if a patient would benefit from adjuvant chemotherapy – for example, in this situation luminal, *HER2*-negative, early stage breast carcinomas with up to three lymph node metastases. More recently, expression analysis of PD-L1 in NSCLC tumours has started to play an important role for the selection of the immunotherapies that are PD-1 monoclonal antibody treatments [159].

Some cancers may require multiple methods of testing, for example there are three main genetic pathways in sporadic colorectal cancer: the chromosomal instability pathway, the microsatellite instability pathway and CpG island methylator phenotype pathway. Identification of which pathway is present requires different tests namely immunophenotyping, methylation analysis and molecular testing. There is overlap between these complex molecular pathways and testing is advised because of prognostic and therapeutic implications based on which is present [138]. Additionally, if familial colorectal cancer is suspected patients may be referred to clinical genetics services for inherited cancer risk assessment and possibly germline testing for Lynch syndrome.

Currently, the majority of testing for solid tumours is done on solid tumour biopsy samples. An alternative to solid tumour biopsies for genetic analysis is to use circulating tumour DNA (ctDNA). ctDNA refers to the fragments of cell free DNA released by tumour cells into bodily fluids such as blood and urine, which typically occurs during cell death but can also result from active cellular secretion. In a process commonly referred to as a liquid biopsy, a sample of a bodily fluid such as blood can be taken from a patient and tested for the presence of ctDNA, using similar techniques as those for non-invasive prenatal screening (NIPT). Both the amount of ctDNA present in the sample, and the genetic alterations it contains, can provide diagnostic and prognostic information about a patient's solid cancer [160]. ctDNA testing can potentially be used for a variety of clinical applications, including as a companion diagnostic test for targeted therapies or for broader treatment decisions when no solid biopsy sample is available or hard to obtain, as a monitoring tool for treatment response and tumour progression, and as a general prognostic indicator to guide clinical decision making.

The only current use of ctDNA testing that is widely available on the NHS throughout the UK is as a companion diagnostic for *EGFR* mutation status in advanced NSCLC, to permit prescription of tyrosine kinase inhibitors (TKIs). For initial treatment with TKIs, ctDNA testing is currently used only as a second line test, used if solid biopsy samples do not yield enough material for a genetic test. For patients who develop resistance to initial use of TKIs, ctDNA testing can be then be used as a first line test for detection of the acquired *EGFR* resistance mutation *pT790M*, for which the drug osimertinib is available. In Wales ctDNA testing is also used for detection of specific *RAS* mutations in advanced colorectal cancer to support therapy selection [161, 162].

A key consideration surrounding planning for ctDNA testing includes deciding on the best type of test method to use for an application [163]. Current clinical testing for *EGFR* mutations in NSCLC uses PCR for detection of specific mutations, with digital droplet PCR being the most sensitive test in use to detect the very small quantities of ctDNA in patient samples. The original Roche cobas® *EGFR* Mutation Test v2 is also still frequently used in NSCLC. Potential future uses and requirements of ctDNA testing are described in 9.6 and chapter 10.

### 9.3 Technical challenges to effective genomic analysis in cancer

Unlike germline genomic analysis for heritable conditions, there is a lack of clear consensus around best practice for somatic genome analysis for cancer management. A wide range of techniques currently in clinical use, for detecting the different types of clinically relevant genome variation that occurs within tumours (table 8). Currently, laboratories select one or more of these methods primarily on the grounds of cost and turnaround time, availability of the relevant technology and expertise in their own laboratory. Given the emphasis on rapid turnaround times to meet the need to make treatment decisions within days of sample collection, and the current low throughput of samples in most laboratories, rapid PCR based techniques are generally favoured over NGS techniques. Effective genomic analysis of cancer samples – tissue, cells or fluids - presents a number of specific challenges that are not encountered during genomic analysis of germline DNA. These include:

**Variation in sample size** - Surgical resection specimens tend to be large, but diagnostic biopsy specimens, especially from patients with disseminated disease, may be very small, limiting the amount of DNA available for analysis. With haematological cancers a suitable volume of bone marrow sample may need to be provided, but the volume collected maybe low.

**Tissue/tumour heterogeneity** - Samples taken from cancer patients are highly heterogeneous. The main sources of somatic genomic heterogeneity within a cancer patient are:

- » Admixture between malignant and non-malignant cells (such as normal tissue, fibroblasts, and infiltrating lymphocytes) within a sample
- » Intra-clonal diversity between spatially distinct 'zones' within tumours or between distinct metastases across different organs
- » Inter-clonal diversity between each tumour may have undergone its own clonal evolution and will therefore have different genomic profiles

As a consequence of this heterogeneity, a single sample from a patient may not be representative of the cancer as a whole. Furthermore, over time the cancer will continue to evolve after the sample is taken, meaning that the utility of genomic analysis undertaken at one point in a patient's care is limited in terms of decision making at a later time point when, for example, drug resistance mutations not present at detectable levels in the initial sample have subsequently spread more widely. Given the continuously evolving nature of cancer, numerous samples could be taken through the disease course – at diagnosis, monitoring of the tumour or tumours, treatment failure, or the development of a new growth.

**Variation in sample DNA quality** - Sample handling can impact the quality of DNA extracted from a sample. Appropriate collection methods and storage should be practised for each sample type. Currently the most clinically accessible solid tumour samples are routine histopathology formalin fixed-paraffin embedded (FFPE) samples. Formalin fixation causes alterations to the nucleic acids, low tumour content and biological factors such as extensive tissue necrosis in tumour samples result in poor quality DNA which can decrease test sensitivity or lead to false-positive results [164].

These three challenges – size, DNA quality and heterogeneity – can, to some extent, be addressed through the use of appropriate analytical models able to take into account different types of heterogeneity. Through clinical and laboratory practices including use of ctDNA, microdissection (to analyse distinct tumour regions) as well as the selection of techniques that are more effective when sample volumes are limited (e.g. PCR), or more able to detect genome diversity within a sample (e.g. NGS) can help. Nevertheless, they currently impose significant limitations on the use of genome analysis in cancer management.

**Table 8. The advantages and limitations of the various genomic technologies used in genomic analysis of tumours. Adapted from [165]**

Technology	Feature						Coverage
	Point variation	Insertion-deletions	Fusions*	Copy number alterations	Splice variants	Methylation status	
Whole genome sequencing	Yes	Yes	Yes	Yes	No	No	Complete Genome
Whole exome sequencing	Yes	Yes	Yes	Yes	No	No	Coding genome only
(Shallow) Whole genome sequencing	No	No	No	Yes	No	No	Complete genome for copy number alterations
Targeted sequencing	Yes	Yes	Yes	Possible	No	No	Only target region(s)
Digital PCR	Yes	Possible	Yes	Possible	No	No	Only target region(s)
RNA sequencing	Yes	Possible	Yes	No	Yes	No	Complete transcriptome
Bisulfite sequencing (methylation)	No	No	No	No	No	Yes	Complete genome (methylation)

\* can be translocation, interstitial deletion, or chromosomal inversion

Key – Yes: Good for identifying the feature of interest; Possible: Identify feature of interest but with important limitations; No: Unable to detect the feature of interest

## 9.4 Samples for solid tumours: FFPE, fresh tissue and liquid biopsy

The current paradigm for the analysis of tumour biopsy or resection samples prioritises the sample processing needs of histopathology. Thus samples of solid tumours are usually FFPE prepared immediately after collection and sectioned for microscopic analysis. FFPE samples are used for histological diagnosis where the morphological features of the tumour are investigated microscopically by a pathologist. Histological assessment include classification of the cancer as well as determining the tumour percentage present, determining areas of necrosis and normal tissue within the sample and if the surgical margin is cleared – i.e. that the whole tumour has been removed. These samples are also used for immunohistochemistry (IHC) analysis.

Formalin used in FFPE sample preparation irreversibly modifies and degrades the DNA causing a major issue for molecular analysis. The formalin crosslinks cellular material together, solidifying the tissue and stopping enzymatic breakdown. Cross-linked DNA is difficult to extract without causing damage and makes distinction between damaged DNA and DNA variations caused by the cancer at analysis difficult. This results in high analysis failure rates and if DNA analysis is achieved the quality is generally poor.

Careful preparation of FFPE samples by following a strict protocol, as well as an improved DNA extraction protocol can result in optimised DNA samples, which have been proven to provide adequate quality DNA for NGS and WGS in a clinical setting [166]. But, failure rates at pre-sequencing quality control are substantially higher for FFPE extracted samples (in the region of 20-30%) when compared to fresh tissue and even optimised FFPE extracted DNA has a high false positive variant calling rate and false negative structural variant rate. Formalin fixed tumour tissue samples are therefore generally considered unsuitable for WGS and a shift to fresh or fresh frozen samples is occurring for WGS of tumour samples. Work by clinical scientists as part of the 100,000 Genomes Project demonstrated effective NGS analysis of solid tumours, particularly WGS improvements with fresh, unfixed tissue to achieve a clinically acceptable standard [166, 167]. Pathology services are adapting to these changes ensuring the histopathology diagnostic samples are still available for histological diagnosis [141, 168].

The collection of the solid tumour sample is generally done through invasive procedures such as a biopsy, or surgical sample where the method used is reliant on the location of the growth. With the use of liquid biopsies to obtain ctDNA it is now possible and proven to be effective for genomic analysis of solid tumour cancers. ctDNA analysis is done with the collection of a blood sample and analysis for tumour-specific mutations. Unlike haematological cancer analyses, which require a blood sample collected in EDTA collection tubes, ctDNA requires the use of cfDNA collection tubes for the preservation of the cfDNA. Liquid biopsies have several advantages over solid tumour biopsies including:

- » Easily accessible through minimally invasive procedures
- » The process of obtaining a sample is faster
- » Can be regularly repeated to provide a dynamic and longitudinal assessment of tumour evolution
- » ctDNA is released from the entire cancer, so it represents the genetic heterogeneity of tumours better

The issues discussed above impact on the quality, quantity, and availability of tumour tissue and pose challenges to clinical implementation of testing. From a theoretical standpoint, where the limitations on sample availability and quality do not apply, NGS would be the ideal technique with which to undertake cancer genome analysis. NGS and associated bioinformatics analysis can, in principle, identify the full repertoire of cancer-associated genome alterations as well as the clonal and sub clonal diversity within cancer-cell populations. Together these analyses should, ultimately, enable more efficient and accurate determination of prognosis and treatment suitability than can be currently achieved with the existing repertoire of mainly PCR and FISH based targeted tests.

## 9.5 NGS-led approach – use of WGS and panels

Given the decreasing costs and turnaround time of NGS, the improvement in bioinformatics analyses, and the harmonisation of databases and other sources of knowledge to facilitate the clinical interpretation of genomic results, the move to comprehensive genomic profiling by NGS in clinical services is likely [164]. For the majority of cancers and tumour types, it is anticipated that genetic testing will utilise panels, moving towards larger panels and WGS as the number of potentially actionable targets increases. Significant challenges remain with respect to sample handling, analysis and interpretation, all of which impact adversely on cost and turnaround time, and necessitate significant changes in current pathology practice. Some additional issues to overcome include:

**Analytical complexity** – NGS identifies a very large number of genomic variations in somatic tumour genomes. Separating causal alterations from non-causal alterations in an unstable and evolving genome is problematic and frequently will involve the need to compare germline and somatic genomes, and the need to compare samples over time, both of which increase the sequencing and analytical workload further. This has significant implications for both the cost and turnaround time of testing.

Whilst doing tumour only sequencing there is the potential to identify pathogenic sequence variants in cancer susceptibility genes and it can be unclear whether they are of somatic or germline origin. Recommendations regarding germline-focused analyses of tumour-only sequencing data – as per ACMG guidance regarding return of secondary findings [169] – have been produced, as well as indications for germline follow-up testing and guidance on patient information giving and consent [170].

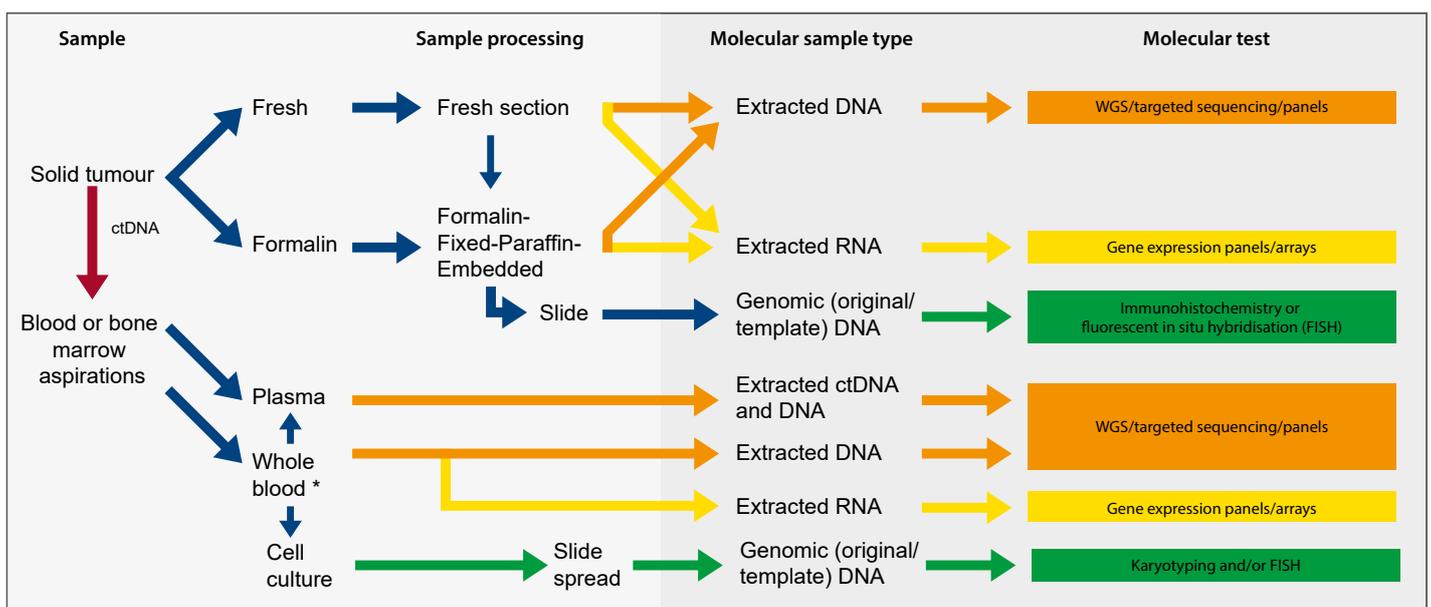
**Limited knowledge for interpretation** – Currently the number of validated associations between somatic tumour genome variation and clinically relevant variables such as prognosis and treatment response are limited. A great deal of research is underway to expand knowledge in this area.

**Need for high sample throughput** – Access to genetic testing is becoming more mainstream which will in the future enable other clinical professionals to request tests. NGS machines are most cost-effective to operate when utilised to maximum capacity. Thus efficient use of NGS-based genome analysis depends on having sufficient sample throughput to justify running the machines with a frequency sufficient to deliver clinically acceptable turnaround times.

Targeted gene panels have been used in the clinical setting because they provide greater depth of coverage in selected areas of interest (e.g., hotspot regions with known actionable mutations), faster turnaround, and more clinically relevant data when compared to broader genomic profiling by WES or WGS approaches. The number of genes included in these panels can vary, ranging from 5, or 20 to over 500 genes. Although the clinical utility of assessing all of the genes that are included in large panels is currently uncertain, the benefit of simultaneous multi-gene testing using NGS and the low incremental cost of including additional genes are reasons for using more comprehensive genomic profiling in the clinic [164]. This approach will be supplemented by targeted testing in situations where rapid turn-around or high sensitivity testing is required, for example in the scenario of MRD monitoring.

The move to WGS is dependent on adequate validation to demonstrate satisfactory performance of the assay for disease-specific requirements. Due to the heterogeneity of tumour samples the read depth required when sequencing a cancer genome is usually higher, ~100, when compared to a healthy tissue genome, ~30. Differentiating an alteration due to the cancer from sequencing errors (for example because of DNA damage from formalin) is difficult at low read depths so a greater read depth of 100 is required to give those interpreting results the confidence to correctly call a base and to then establish which mutations are clinically important. In cancer, to further support clinical validation of WGS, it has been recommended that clinical indications considered eligible for WGS would have WGS done in parallel to the current standard of care testing [156]. For WGS, testing from the tumour and the germline is essential, so appropriate DNA samples are required. However, with cancer there is a lot of information that may need to be cautiously interrogated and identification of the cancer associated changes requires careful examination, in particular if novel and not previously reported. Recent micro costing analysis of WGS for cancer indicated that high throughput, corresponding with a national-scale facility, combined with bulk discounts on consumable costs will likely have the greatest impact on the overall cost of sequencing going forward [67].

**Figure 8. Summary of cancer samples and processes**



\* Different collection tubes required depending on use of sample

## 9.6 The future

### *Liquid biopsies using ctDNA*

Use of ctDNA testing as a companion diagnostic test for targeted therapies in other cancers is approaching readiness for clinical implementation, including in ovarian and breast cancers as well as further use for other genomic alterations in NSCLC and colorectal cancers [171,172]. Companion diagnostic testing is the most clinically advanced application of ctDNA testing, for several reasons:

- » There are clear clinical pathways already in place for genetic testing of solid tumours, requiring relatively minor adjustments to accommodate ctDNA testing
- » There is unmet clinical need in terms of solid biopsy unavailability or failure (particularly in NSCLC)
- » The implementation of ctDNA testing in NSCLC has helped develop the protocols and logistics surrounding sample collection and preparation

In addition to companion diagnostic testing, ctDNA approaches are potentially suitable for a number of established cancer management applications [173]:

- » Before and after treatment to identify patients most at risk of recurrence
- » As a prognostic indicator monitoring for signs of MRD and cancer recurrence after surgery
- » As a prognostic indicator monitoring response to treatment, including emergence of resistance
- » Facilitating stratification of patients into molecularly targeted clinical trials, and investigating drug mechanism during trials [174, 175]

As NGS techniques are becoming more sensitive, more clinical research laboratories are developing targeted NGS based panel ctDNA tests [176, 177], as well as methods to measure DNA methylation and fragment length, some which incorporate other biomarkers such as proteins [178-180]. In addition, tests for circulating RNA are also being investigated [181]. These are all in the research arena at present. Overall it is likely that different test technologies will be suitable for different applications of ctDNA testing.

As ctDNA testing expands, there will be a demand for regular testing and a need for rapid turnaround that could greatly increase demand on existing services. Increased use of gene panels will also increase analytical demands on laboratories. Clear guidelines are also required on when and how clinicians should request ctDNA testing, as well as how to interpret and deliver the results.

As the evidence for ctDNA use accumulates and more targeted and histology independent therapies are approved, it is likely that it will be more widely used as a companion diagnostic for targeted therapies across multiple cancer types. Patients with cancers such as brain tumours where biopsies are particularly difficult or unfeasible may benefit the most from liquid biopsy testing for treatment selection. Uses of ctDNA as a monitoring tool for tumour response to treatment, including for the detection of relapse and the emergence of resistance mutations, show clear potential in several cancers but are not yet available outside of clinical research studies [182]. Similarly, uses of ctDNA as a prognostic

indicator to help predict the likely aggressiveness of different cancers, as well as a tool for more accurate diagnosis, show promise but are not yet available in the clinic.

### *Histology-independent products*

Tumours that occur in different parts of the body can have the same genetic characteristics, also known as histology-independent markers. Histology-independent cancer drugs work in a range of tumour types each of which harbour the genetic mutation targeted by the drug. At the end of 2019 the first two products entered NICE appraisal (entrectinib and larotrectinib) with six more anticipated for market entry in 2020 – larotrectinib received approval from NICE in April 2020 [183]. These drugs present challenges in that the incidence of actionable genome aberrations is low overall and highly variable across tumour types, necessitating the testing of large numbers of tumours with significant use of resources [164]. Further evidence gathering is needed, as such, the NIHR is funding research that will explore the level to which NICE's existing approaches for assessing clinical and economic value can be applied for these drugs [184].

### *Single cell and long read sequencing*

Long-read sequencing efforts have focused on RNA sequencing and have discovered novel fusion and splicing isoforms that are relevant to tumour progression or treatment resistance. Currently, however, the error rates for long-read technologies are too high for somatic variant detection [164].

Single-cell analysis of circulating tumour cells (CTC) enables the generation of patient-specific tumour models. Viable CTC are thought to be involved in the formation of tumour metastasis and thus may reflect a metastatic genotype. Therefore, single-cell analyses can lead to the detection of actionable aberrations that are implicated in metastatic spread, whereas tumour models derived from CTC might serve to test novel drugs or sensitivity to current standard therapies [164].

## **9.7 Conclusions**

The field of precision oncology is moving from isolated genomic analyses towards a multi-omic approach to achieve a better understanding of tumour biology and to increase treatment opportunities. Beyond single gene analyses, mutational signatures, RNA-based gene expression profiling, immunophenotyping, and tumour mutational burden determination have proven to be useful prognostic and predictive biomarkers of response to anticancer therapies, but whether they will lead to an increase in treatment opportunities is still unclear [164].

Current experience with existing tests suggests that whether this demand manifests in reality is highly dependent on the availability of testing in clinically meaningful time frames; clinician awareness of the availability of testing; guidelines for use; and access to targeted therapies. Current use of genomic analysis in cancer is restricted to identification of the small number of specific mutations that can inform the most appropriate care

pathway for patients. As more targeted therapies are developed demand for molecular testing for the specific targets will be increasing. Alternative technologies, such as immunophenotyping through immunohistochemistry, which are simple, quick, and direct could have an increased role in the diagnostic pathway.

Tumour WGS requires a significant quantity of high quality DNA, preferably from fresh tissue sample, and sample heterogeneity issues necessitate both high depth sequencing and a normal germline genome for comparison. These issues significantly complicate sampling and analysis and requires changes to current pathology sample pathways. Strategies for logistical issues around achieving reliable, high quality tumour whole genome analysis are being addressed and being introduced into clinical care pathways with the introduction of the Genomic Medicine Service and the National Genomic Test Directory in England. Assuming that sample throughput increases over time, this should in turn shift the balance of considerations for laboratories around which techniques are most cost-effective for delivering genomic analysis, most likely leading to a transition to wider use of NGS for refining diagnosis, prognosis and treatment decisions using targeted cancer genomic analysis.

## 10 The future clinical genomics landscape

This review has outlined genomics technologies that are already in use in clinical services or will be more widely implemented in the next three years. However, with the rapid pace of development not just in genomics technologies but also in their application, there are a number of developments that are likely to have an impact towards the end of the time frame of this review, most likely in the next five to ten years.

Here we briefly outline what these advances are and describe the potential impact of more extensive use of NGS, including WES and WGS, in the mid- to long-term.

### 10.1 Technological developments in sequencing technologies

#### *Whole genome and whole exome sequencing*

The cost of DNA sequencing continues to fall, albeit at a slower pace than in previous years. The result is that higher coverage sequencing is more accessible. It is likely that in clinical settings, changes to sequencing equipment will be slow. Currently the landscape of sequencing equipment is heterogeneous and some equipment no longer produced by manufacturers is still in use in the UK. Renewal or replacement of sequencing equipment can be extremely expensive, although in circumstances where NGS is required on a large number of samples, newer equipment in general offers cheaper sequencing per sample.

Newer and cheaper short read sequencers designed for lower output sequencing of a small number of samples are also becoming more prevalent; these may be useful for targeted sequencing but are unlikely to reduce the need for large sequencing-at-scale machines in centralised sequencing centres. Instead, high cost, high output machines such as the Illumina NovaSeq now provide more efficient short-read sequencing at scale, although significant initial investment is required if the sequencing is to be performed in-house.

The cost and efficiency of library and sample preparation also impacts upon the usefulness of WGS and WES in clinical situations. It is likely that the use of WES will expand to replace some CES and targeted sequencing, providing greater genome coverage as improved library preparation and capture kits for WES are becoming available.

The cost of analysis, data storage and the ability to perform clinically useful analysis on WGS data remain a key point for consideration when determining whether to use WES or WGS more extensively in clinical settings. Although pioneering sequencing projects such as the 100,000 Genomes Project utilised WGS on patient samples, much of the subsequent analysis has been restricted to protein-coding regions of the genome.

There is potential for the use of carefully developed AI in the analysis of genomic data, most immediately in the identification of variants and prediction of their effects. A number of the potential benefits from the use of AI and machine learning in genomics pertain to developing our understanding of the genome and genetic variants for health and disease, but streamlining of analysis through systems designed specifically for genetic variant detection may have a role to play in future clinical genomic analysis [\[185\]](#).

## *Long read sequencing*

As outlined in chapter 6, long read sequencing (LRS) is a relatively new approach encompassing several technologies that are currently being used as a research tool and being most notably applied to pathogen genome sequencing. In clinical genomics, long read sequencing could provide several benefits for the diagnosis and investigation of some rare diseases and cancers.

The production of longer reads better allows for improved detection of large mutations, complex structural features or variants, highly homologous regions, repetitive regions and highly polymorphic regions such as the HLA region, all of which can cause significant disease. Long reads are more easily distinguished from one another than short reads are, and can also permit allele phasing (assigning alleles to maternal or paternal chromosomes); longer reads are therefore more easily mapped to the correct part of the patient genome, and can therefore detect changes such as gene fusion events and chromosomal rearrangements, which are particularly prevalent in cancer.

Long read sequencers, such as those produced by Oxford Nanopore Technologies, allow for reads to be examined in real time and sequencing can be terminated as required, for example when enough information has been collected. This can permit faster turnaround times for individual samples than more conventional short-read sequencing approaches.

Much like short read sequencers, long read sequencing platforms are not equivalent (chapter 6.3). Alterations and improvements needed to see wider use of long-read sequencing are unique to the system in question, however, there are broader points to consider.

One of the hurdles to the use of long read sequencing in clinical genomics is the need for test validation. Long read technologies are developing and evolving quickly, which makes validation of a test difficult. Test validation can take time, and a software or hardware change or substantial alteration to analytical pipelines may require the test developed using these techniques be re-validated.

Questions remain around the accuracy of results retrieved from some long-read sequencers and actual cost of sequencing runs required to obtain reliable results. Although cost and accuracy of long read sequencing is being improved upon across the board, different platforms raise different concerns - PacBio provides high accuracy, long reads at a higher cost, whilst Oxford Nanopore offers lower accuracy, long to ultra-long reads at a lower cost. Investment in new sequencing equipment (different from more established short read technologies) and re-training – as with many new tools – is also required.

As a broad generalisation, academic and clinical researchers are, as yet, not as familiar with long-read technologies and techniques as with short-read sequencing, meaning there is a shorter and less rich history of first-hand experience to contribute to shared knowledge about their use. However, this is changing, and collaborative efforts between researchers, the private and commercial sectors are being made to improve bioinformatic pipelines, associated techniques, skills and protocols for long-read sequencing.

In the next five years or more, we are likely to see continued efficiency gains from these technologies, both in terms of accuracy and output. More substantial LRS machines, which are more suited to sequencing multiple genomes than older LRS systems are beginning to rival more expensive short-read sequencing technologies in terms of data output.

Specific technological developments include an ONT-developed automated DNA preparation system, called VolTRAX, and the recently released the MinION Mk1C, an all-in-one sequencing and analysis platform. Their smallest sequencing device, designed to be used as an attachment to a mobile phone and dubbed 'SmidgION', has been in development for some time is due to be released. ONT also provide a cloud-based analysis system called METRICHOR; whilst many other bioinformatics tools and pipelines are being developed by researchers. The MinION has been used by clinical researchers, mainly in the context of pathogen genomics, for example, to trace the spread of pathogens around a hospital [186, 187].

In the UK, collaborators at NIHR Guy's and St Thomas' Biomedical Research Centre, King's College London, the London South Genomic Laboratory Hub, and Viapath used the ONT MinION to develop a genetic test for Huntington's disease. As stated in a related press release in May 2019: 'This is the first time that Oxford Nanopore Technology has been used in an NHS laboratory accredited by the United Kingdom Accreditation Service (UKAS)' [188]. However, the test is now not in clinical use owing to rapid changes in technology since its development. This technology is also being investigated for the detection of breakpoints in mitochondrial DNA by the Oxford Medical Genetics Laboratories to provide patients with a more detailed diagnosis [189].

Developments in PacBio technologies include the research program 'SOLVE-RD', a five-year diagnostics-focused project aiming to 'solve' many rare disease cases and funded by the European Commission [190], which will use PacBio's sequencing technologies to sequence the genomes of around 500 participants by 2022 [191]. It is hoped that long read sequencing will reveal disease-causing variants that other approaches may be unable or poorly equipped to identify. It has also been used for microbial genomes, and to examine epigenomic changes, such as patterns of DNA methylation.

## 10.2 Applications of genomic technologies

There are two key themes that emerge in terms of the future clinical applications of genomic technologies:

- » More extensive use of genomic technologies in the areas already described in this report, both through the life course and in the management of cancer.
- » Use of genomic testing and analysis in healthy individuals to inform potential future care and disease prevention efforts.

## Expansion of genetic testing provision

### Pre-natal

NIPT is currently planned to be used as a second line test, however as evidence accumulates over its effectiveness and the cost of the service, it may in future be recommended for use as a first line test. This would require the capacity to expand NIPT testing to all pregnant women, compared to the current estimate of 1.9%-2.4% women in England [192] who have a 1 in 150 chance of a fetus with T21/T18/T13 and so would be eligible for NIPT as a second line test. Once NIPT is established as either a first or second-line test it may be possible that it is expanded beyond the common trisomies to test for a wider variety of genetic disorders. This service is already available for certain disorders on the NHS when clinically indicated, however as the evidence basis expands it seems likely that these tests may become suitable for use in a broader categories of patients, potentially extending to use in screening of 'healthy' fetuses [193]. Increased use of NGS methods to sequence entire genes rather than specific regions may deliver more results for fetuses with *de novo* mutations, especially in those with clinically indicated genetic abnormalities or a family history of monogenic disorders [194]. In the longer term use of WES or WGS to analyse cfDNA have also been proposed as an alternative option to testing multiple individual genes is another application that has the potential to transform NIPT, if shown to be feasible and cost effective for use in the clinic [132].

Aside from NIPT, trio exome sequencing of fetuses with structural abnormalities unexplained by standard genetic tests (of samples obtained from amniocentesis or chorionic villus sampling (CVS) is also highly likely to be implemented, with the potential for WGS in future if this is shown to improve diagnostic yield [195].

### Post-natal

WES trio analysis of sick babies in PICU/NICU is already underway via the South West Genomic Laboratory Hub in England and testing of this type is likely to expand as further evidence is gathered as to its effectiveness in aiding diagnosis in seriously ill infants and children. In future it is possible that WGS may be recommended over WES, especially as costs of sequencing and interpretation decrease. Aside from the sequencing of critically ill children, it is anticipated that in the near future, patients in England with an undiagnosed rare disease will have access to WGS through the Genomics Medicine Service [155]. In Scotland, the focus is currently on use of WES trio analysis for critically ill babies and on clinical exome sequencing for diagnosis of other cases of rare disease. However, the utility of NGS for patients with rare disease is also being evaluated, and it seems likely that the use of all these sequencing methods is going to steadily expand, with preparations already underway [86].

Applications that are further away from implementation but which should be considered as part of longer-term thinking are the potential use of genetic testing as part of the newborn blood spot programme, including the possibility of sequencing healthy newborns. Use of genetic testing could allow more disorders to be screened for than is available through the current newborn blood spot programme, as well as improving the diagnostic and prognostic utility of screening [135, 196]. Questions remain over which method of screening would be most suitable for use, ranging from targeted sequencing of specific genes from dried blood spots to some calls for WES or WGS of all newborns [39, 135, 197]. The use of genome wide sequencing in healthy newborns is currently

controversial due to the many complexities in interpreting results and managing ethical concerns and data privacy [198, 199]. However research continues to evaluate its use, and in 2019 it was suggested that genome sequencing should be offered to all children in the UK at birth through a pilot study, though no further details have been announced [198, 200, 201].

## Cancer and ctDNA

Circulating tumour DNA testing is already available on the NHS in the UK as a companion diagnostic test for *EGFR* mutation status in NSCLC to determine eligibility for tyrosine kinase inhibitor therapy. In Wales only, ctDNA testing is used to determine *KRAS* and *NRAS* mutation status and eligibility for treatment with cetuximab in advanced colorectal cancer. It is expected that further companion diagnostic tests will become available as further evidence is gathered.

Looking further ahead, the use of ctDNA for detection of minimal residual disease and monitoring is currently in the clinical research phase and further evidence is needed of clinical utility. Research studies have demonstrated that it is possible to detect residual disease after surgery or other treatment where conventional methods such as imaging are unable to do so [202]. There is also evidence that ctDNA testing can be used to detect cancer recurrence earlier than other methods such as imaging up to months ahead in breast, colorectal and lung cancers [202].

While there has been much publicity around the use of ctDNA testing as a screening or early diagnosis test, research is at a very early stage and there is currently no evidence of clinical utility and a number of trials are ongoing [203-206].

## *Clinical genome analysis of healthy individuals*

### Polygenic scores

There has been a recent surge of research activity regarding polygenic scores for risk prediction of common diseases such as cardiovascular disease, cancer, diabetes and neurological disorders such as Alzheimer's and schizophrenia. Polygenic scores are based on combining the effect sizes of many common DNA variants that are found to have an association with, sometimes extremely small, disease risk. These variant associations are typically identified from genome wide association studies, where up to millions of single nucleotide polymorphisms are identified across large cohorts of people using either microarrays or genome sequencing methods. These can then be examined for prevalence in both diseased and healthy individuals. By adding up the effect sizes of common variants that have even a tiny effect, it is possible to generate a polygenic score that could potentially be used to predict an individual's risk of developing a disease. Due to their potential ability to contribute to the prediction of common diseases at a population level, there has been considerable interest surrounding their use in clinical practice.

Potential applications of polygenic scores range from prediction of common diseases, improvement of risk prediction, refining a diagnosis for a current disease, informing therapeutic selection, to informing disease screening, and, on a personal level, informing life planning. Work is ongoing to gather the clinical evidence needed to support the use of

polygenic scores and evaluation in clinical settings is only just starting. Initiatives include the CRISP-DNA project [207] which is exploring the feasibility of assessing bowel cancer risk using genomic testing in primary care and screening, and BOADICEA which is looking at the high-risk/heritable breast cancer settings [208].

The Accelerated Detection of Disease (ADD)/healthy lives cohort plans to recruit up to 5 million healthy individuals with the intention of developing polygenic scores for 16 common diseases. It is anticipated this will also allow for refinement and further development of the scores themselves [209].

In the NHS in England there has been particular interest in the use of polygenic scores as a risk factor for coronary artery disease (CAD), particularly when used alongside other conventional risk factors in for example the existing Q-RISK<sup>©</sup> risk prediction tool. However, polygenic scores are still being developed, their potential uses explored, and although there may be some evidence that they could be useful in specific contexts, they still need to be evaluated and their clinical utility determined [210]. In addition, no polygenic score test has yet been designed for clinical use.

### Pharmacogenomics – including reactive and pre-emptive testing

It is likely that in the future more pharmacogenomic tests will be approved and recommended for routine use in the NHS. The family of cytochrome P450 (CYP) enzymes alone metabolize a large number of clinically used drugs, with genetic variation in the *CYP2C9*, *CYP2C19*, and *CYP2D6* enzymes in particular associated with variable drug reactions [211]. For example, there is early evidence that testing the genetic variants in *CYP2C9* as well as the *VKORC1* gene could be used to determine dosing of the anticoagulant warfarin, reducing the incidence of adverse drug reactions such as internal or external bleeding [212].

A study within English primary care found that three pharmacogenes – *CYP2D6*, *CYP2C19* and *SLCO1B1* – accounted for >95% of the drugs prescribed. The same study found that multiple exposure to drugs associated with pharmacogenomics is extremely common, with 60% of patients being prescribed  $\geq 2$  and 18%  $\geq 5$  of these drugs [213].

This evidence points to potential increase in testing demand, should these gene-drug pairs be implemented. Whilst testing in the shorter term is likely to be reactive it is anticipated that in the longer term there will be a move towards pre-emptive testing for a comprehensive suite of targets. This would require use of large panels for a number of different genetic variants, or potentially the use of WGS.

Pre-emptive testing of healthy people raises a number of considerations in terms of:

- » Circumstances of testing, it is currently unknown under which circumstances pre-emptive testing would best meet the needs of the patient and the health system
- » The type of health system interaction that would trigger an offer of pre-emptive testing to a healthy person
- » Comprehensiveness of testing e.g. WGS, panel or a microarray test of all known gene-drug pairs

- » The form of data to inform future care e.g. in an electronic or other health care record
- » Circumstances under which testing would be repeated e.g. as information evolves, or further gene-drug pairs are identified.

### 10.3 Wider implications for the health system

There are a number of implications for the health system that arise from the more extensive use of next generation sequencing and other genomic technologies.

#### *Infrastructure and provision of testing*

Increasing test demand will have an impact on laboratory infrastructure and provision of testing. While it is not yet clear how demand will rise in the next five to ten years, certain proposed target areas for sequencing illustrate just how high potential demand could be. The ambitions to sequence up to 500,000 whole genomes by 2023/24 via the Genomic Medicine Service, as outlined in the NHS Long Term Plan would have an impact on patients with cancer and rare disease and represent a significant increase in the number of genomes sequenced per year [8]. Looking even further ahead, for example, in 2018 in the UK there were 731,213 live births [214] – should sequencing of healthy newborns become more commonplace, even 10% uptake of testing would result in year on year demand of over 70,000 whole genomes. A survey conducted in 2013 suggested that over 70% of parents in the US would be either definitely or somewhat interested in their newborns undergoing whole genome sequencing [215].

#### Data storage and analysis

One of the greatest challenges is around use and management of the volume of genomic data produced as sequencing provision increases, considerations include:

- » **Infrastructure** – for data storage and planning for growth of storage needs
- » **Type of data stored** – raw sequence data and/or analysed data
- » **Longevity of data stored** – anticipation of future compatibility with software
- » **Length of data storage** – determine useful period of data storage. Should patient-related data be stored beyond the lifetime of that individual?
- » **Reanalysis or resequencing** – whether only some data should be stored for interrogation at a later date, or whether resequencing of stored DNA as needed is a more viable option
- » **Curation** – appropriate management of any data stored
- » **Methods of storing data** – determining needs for patient data storage, taking into account security, capacity, accessibility and more. Cloud-based storage or locally-stored records

### *Recontacting as genomic knowledge develops*

Proposed expansion of sequencing to earlier in the life-course and to healthy individuals raises considerations around recontact about genetic results for a number of reasons:

- » Consenting people who were sequenced as babies or children once they reach adulthood, to give them choices around further options for storage and/or analysis of their data
- » Informing individuals of important developments in genomic knowledge, for example when VUSs have been more accurately classified
- » Alerting those sequenced as children to any findings relevant to them as adults

Each of these situations will be affected by the type of analysis done on the initial sequence and what type of information has been stored e.g. whole genome sequence, whole exome sequence, test specific results, raw data vs. analysed data.

### *Downstream effects on clinical pathways*

As genomic information and knowledge increases, more patients will receive genetic information or a genetic diagnosis. This will have differing impacts on other clinical services depending on the reason the testing was carried out and knowledge about expected diagnostic yield of these tests. Where a rare disease is suspected and WES/WGS is being carried out to support diagnosis, it is unclear what the effects of a positive diagnosis will be on downstream clinical pathways. Consideration will be necessary as to the wider context in which testing will be delivered, and what the impact might be on clinical services that will be responsible for coordinating ongoing care.

### *Links between clinical practice and research*

One key concept that is driving forward the development of health systems is that of further integration between clinical practice and research, through:

- » More extensive patient participation in research
- » More clinician participation in research activities
- » Quicker integration of research and innovation into healthcare as a result of these activities

For genomics, the Accelerating Detection of Disease project in England is one example of a major research initiative where research and health data are to be linked from the start, to support efforts to accelerate research in approaches to early diagnosis of disease.

## 11 Conclusions

There have been significant developments in the provision of clinical genomics services in the UK in the past five years. Genomics is being embedded into mainstream clinical practice, and national efforts such as the 100,000 Genomes Project have contributed to the development of the infrastructure that will be required to deliver these services.

Our knowledge of the genetic basis of disease is increasing and new clinical interventions are being developed for a specific genetic status on an ongoing basis. This is increasing the usefulness and impact of clinical genome analysis. As testing volume increases, in particular CES, WES and WGS, laboratory capacity for such analysis is expected to increase still further in order to provide the population health benefits of these developments. In the last five years developments have included an increase in both the number and/or size of testing panels offered in cancer and rare disease, more extensive trio-based WES to support rapid diagnosis of rare disease in babies and children, and implementation of new approaches such as ctDNA testing in cancer.

Developments since 2016 show that as sequencing has become cheaper, and with higher throughput, WES and WGS are being used more extensively. This is a trend that is likely to continue in the next 5-10 years, however what is less clear is if or when a 'tipping point' will be reached whereby WES/WGS becomes the default assay. Despite aspirations for sequencing a genome for \$1000, micro-costing studies carried out in the UK show that sequencing costs were £6,841 per cancer case (tumour and germline sample – two genomes) and £7,050 per rare disease case (trio – three genomes), taking into account sample collection and preparation, staff, consumables, analysis and reporting costs [\[67\]](#).

Of particular note is the increase in scale of genomic data being produced as more WES and WGS are being carried out. This will create challenges not only in terms of data storage but also for the infrastructure and expertise needed to carry out genome analysis. Improvements in bioinformatics pipelines are likely to occur, including further automation of some analysis processes, which will contribute to data management efforts. There is still a debate to be had around whether long term storage of genomic data is a suitable model, or whether resequencing is more appropriate.

As highlighted by this review, one single technology cannot provide all the necessary clinical genome analyses, and so retaining capacity for existing techniques will be necessary. While a shift towards WES/WGS has been proposed as the likely direction of travel longer term it is clear that the situation is more nuanced in terms of the usefulness of different sequencing technologies – in particular non-sequencing methods which still constitute a significant proportion of testing carried out by clinical genetics laboratories. Therefore, consideration still needs to be given to the distribution of test volume across the different methods.

This will continue to be the case as NGS technologies and other sequencing methods, such as long read sequencing, continue to develop and increase in their ability to be clinically useful tools. The timescale for such developments is uncertain, for example while there has been much progress in the accuracy and reliability of long read sequencing technology, no application is currently validated or accredited for clinical use. It is possible that specific tests that take advantage of the strengths of long read sequencing could have clinical

utility in the next three years, for example in rare disease applications where sequencing of repetitive regions is required, a task that is more challenging for short read technologies.

Looking ahead, proposed initiatives to expand genomic testing of both ill and healthy newborns and infants, and testing of otherwise healthy people through the use of polygenic scores, present a number of data challenges for the health system. Longer term data storage raises a number of ethical issues in terms of privacy and consent, such as which stakeholders can have access to data and under what circumstances, and how patient consent given in terms of data use can reliably and reasonably be updated as knowledge develops and clinical needs change. For example, there are particular ethical considerations around testing of adult-onset conditions in children. There will be a need for public engagement efforts to help the health system navigate these issues and to build long-term public trust. Issues around recontacting individuals in light of them coming of age (if they were sequenced as children) or due to updated knowledge about their genomic variants, will also need to be considered.

Technological developments in genomics, along with our evolving knowledge of the relationship between genes and disease will continue to facilitate more accurate, sophisticated and cost-effective clinical genome analysis in the next three years. Increased use of sequencing technologies has already been harnessed by the NHS to improve healthcare in a range of disease areas, in particular in the diagnosis of rare disease and in the management of cancer. In order to support this ongoing progress, it is vital that health systems are able to rapidly exploit advances in genomic analysis and research as they emerge from both the biomedical research sector and from the collaborative healthcare research ecosystem. This will ensure that they can implement the most appropriate genomic laboratory approaches in a timely manner.

If the health system is able to harness current and future genomic technologies to their full potential, it will be able to meet expected increased testing volume and demand, through the provision of equitable services that meet unmet clinical needs and provide improved quality of care to NHS patients and their families.

## 12 Appendix

### 12.1 Appendix 1: Acknowledgements

We thank the following individuals for sharing their knowledge and insight:

- » **Dr Joo Wook Ahn** – Bioinformatics lead, East Midlands and East of England Genomic Laboratory Hub
- » **Dr David Baty** – Consultant clinical scientist and consortium lead scientist, East of Scotland Regional Genetic Service
- » **Prof Rachel Butler** – Operational Director, South West Genomic Laboratory Hub
- » **Prof Zosia Miedzybrodzka** – Professor of Medical Genetics, University of Aberdeen and Service Clinical Director of Genetics, NHS Grampian
- » **Dr Shehla Mohammed** – Consultant clinical geneticist, Guy’s and St. Thomas’ NHS Foundation Trust
- » **Sian Morgan** – Head of Laboratory, All Wales Medical Genetics Service
- » **Dominic McMullan** – Consultant clinical scientist, Birmingham Women’s and Children’s NHS Foundation Trust
- » **Dr Anca Oniscu** – Consultant Histopathologist, Clinical lead molecular pathology, Royal Infirmary of Edinburgh
- » **Alexandra Pickard** – Policy and strategy lead, Genomics Unit, NHS England
- » **Becky Treacy** – Deputy Director for GenQA (molecular genetics)

We would also like to acknowledge the support from colleagues at NHS national services Scotland, national specialist and screening directorate (NSD):

- » **Peter Croan** – Programme Associate Director, Head of Finance and Operations
- » **Louise Mathew** – Programme Support Officer
- » **Karina O’Rourke** – Programme Manager

## 12.2 Appendix 2: Glossary of key terms

<b>Allele phasing</b>	The analytical process of assigning alleles to maternal or paternal chromosomes, which can be relevant in some cases of inherited disease
<b>Amplicon</b>	A piece of DNA that is generated as a result of the amplification step that occurs during polymerase chain reaction (PCR)
<b>Aneuploidy</b>	Human diploid cells normally contain 46 chromosomes. When there is an abnormal number of chromosomes this is referred to as aneuploidy
<b>Array Comparative Genome Hybridisation (Array CGH)</b>	Compares the quantity of patient DNA bound to an array to that of a reference DNA sample, to detect genetic imbalances such as copy number variations and deletions
<b>Consanguinity</b>	Being descended from the same ancestor; two consanguineous people (usually second cousins or closer) are more likely to have similar DNA variants. Children arising from consanguineous relationships have a greater likelihood of inheriting identical recessive disease-causing variants
<b>Copy number variations (CNVs)</b>	Copy number variations are a form of structural variation in the genome caused by duplication, deletion or translocation of multiple bases, resulting in differences in the number of copies of a particular gene or segment of DNA. CNVs are sometimes referred to as copy number alterations (CNAs)
<b>Circulating tumour DNA (ctDNA)</b>	DNA that is released from cells in tumour, most commonly into the circulation, but also into other bodily fluids such as urine
<b>Fluorescent <i>in situ</i> Hybridisation (FISH)</b>	A molecular karyotyping technique that uses fluorescent probes designed to detect regions of interest in a patient's DNA, such as copy number variations
<b>Formalin fixed-paraffin embedded (FFPE)</b>	A form of preparation of biopsy samples that involves preserving the tissue in formalin before embedding it in a paraffin wax block, which facilitates slicing the sample to mount on microscope slides for further examination
<b>G-banding</b>	Chromosomes are treated with trypsin and then a Giemsa stain, resulting in characteristic banding patterns. Analysis of alterations in these banding patterns allows identification of large-scale chromosomal abnormalities during karyotyping
<b>Heterozygosity</b>	When two different alleles of a gene are present
<b>Homozygosity</b>	Where two identical alleles of a gene are present
<b>Immunohistochemistry</b>	Method for the identification of specific proteins in a tissue section through binding of complementary molecules to provide a visual signal

<b>INDEL</b>	INsertion and DEletions, the most common form of genomic variation, ranging in size from 1-1000 base pairs
<b>Karyotyping</b>	A classical low resolution cytogenetic technique using microscopy to analyse large scale changes in the number or structure of chromosomes (the karyotype), such as whole chromosomal deletions or duplications
<b>Loss of heterozygosity</b>	The deleterious effects of a mutant allele may be silenced by the presence of an existing normal allele. If this normal allele is then disabled through an additional deletion or mutation, this is referred to as loss of heterozygosity. This is of particular relevance in cancer where, for example, mutations in tumour suppressor genes may be compounded by subsequent loss of heterozygosity, giving rise to disease
<b>Metaphase</b>	A stage of cell division where chromosomes are distinct and condensed and most amenable to visual analysis. For cytogenetic analysis cells are arrested at the metaphase stage of cell division to allow chromosomes to be examined
<b>Methylation</b>	The process whereby methyl groups are added to a DNA molecule, without changing the underlying sequence. Methylation can have an impact on gene expression with implications for some forms of inherited disease and cancer
<b>Microarray</b>	A microarray is a solid platform such as a microscope slide, containing thousands of short, single stranded DNA fragments, arranged in precise microscopic locations. The DNA sequences act as probes which can hybridise with complementary DNA sequences present in patient DNA. They can be used to interrogate thousands of potential DNA sites at the same time. E.g. Array CGH, SNP arrays
<b>MLPA</b>	In PCR, probes are designed to bind to specific DNA sequences and initiate the amplification process. Multiplex Ligation dependent Probe Amplification (MLPA) makes use of split probes which can only ligate and function when a specific DNA sequence is present. The PCR product is therefore only visible when the complete target sequence is present
<b>Mosaicism</b>	A situation where cells within one individual contain different genetic complements e.g. mosaic Down's syndrome, where some cells contain the normal 46 chromosomes, and others contain 47, with the additional chromosome 21
<b>miRNA</b>	Micro RNA. Class of RNAs which include those that do not code for proteins and are very small, normally less than 30 bases in length

<b>mRNA</b>	Messenger RNA. Class of RNAs produced from protein-coding regions of DNA and can be translated to produce proteins
<b>Non-invasive prenatal testing (NIPT)</b>	A form of testing where fragments of fetal DNA are extracted from a maternal blood sample and analysed to determine if the fetus has a chromosomal disorder. NIPT is not diagnostic, follow up testing is required to confirm results
<b>Non-invasive prenatal diagnosis (NIPD)</b>	A form of testing where fragments of fetal DNA are extracted from maternal blood and genotyped to determine if the fetus has one of a small number of rare genetic disorders
<b>PCR</b>	Polymerase chain reaction. A method for amplifying the quantity of DNA, by making multiple identical copies of a DNA molecule from a small amount of template DNA, such as a sample of DNA from a patient
<b>PGD</b>	Preimplantation genetic diagnosis (PGD), a technique that allows couples with a genetic disorder to check the genetic status of embryos created through IVF, prior to implantation. One or two cells are biopsied from the embryo at blastocyst stage and the genetic material is analysed with a view to selection of embryos unaffected by the disorder for implantation
<b>PGH</b>	PGH is a form of PGD that relies upon the identification of panel of informative microsatellite markers which are associated with the disease-causing mutant allele. This set of markers is referred to as a haplotype, and the technique is therefore known as preimplantation genetic haplotyping. These markers are analysed in the cells biopsied from the embryo, and the presence or absence of the mutation can be inferred from the markers. This allows unaffected embryos to be selected for implantation
<b>Pharmacogenomics</b>	The study of the role of the genome in drug response and the application of this information in clinical practice
<b>Pyrosequencing</b>	A method of DNA sequencing by synthesis which detects the release of light-emitting pyrophosphate as each base pair is incorporated into the DNA sequence. As each nucleotide is added in a sequential fashion, the sequence of the template DNA can be determined from the sequence of signals emitted
<b>qPCR</b>	Quantitative polymerase chain reaction. The same principle of PCR but performed in a way that allows the DNA product to be quantified, in order to understand the amount of DNA originally in a sample. Often performed in real time, using fluorescently labelled molecules to allow DNA quantification

<b>RT-qPCR</b>	Reverse transcription quantitative polymerase chain reaction. Method for the amplification and quantification of specific sections of RNA. Devised of two stages – converting RNA into cDNA (reverse transcription) followed by quantification
<b>Single nucleotide polymorphism (SNP)</b>	Single nucleotide polymorphisms are the most common type of genetic variant. Each represents a difference in a single base pair in the DNA sequence. SNPs in genes have been identified as causal in disease, SNPs may also act as markers which segregate with causal genetic variants, and can therefore be used to deduce that a linked variant is present
<b>Transcription</b>	The copying of DNA into RNA (particularly mRNA), which is the first step in gene expression
<b>Translation</b>	The synthesis of protein by a cell using information coded in an mRNA template
<b>Uniparental disomy</b>	A situation where both copies of a chromosome or chromosomal region have been inherited from only one parent, as opposed to one copy from each parent
<b>Whole exome sequencing (WES)</b>	Sequencing of the protein coding genes in the genome only
<b>Whole genome sequencing (WGS)</b>	Sequencing of the entire genetic sequence of an organism, including both the protein-coding and non-protein coding regions of the genome

## 13 References

1. Davies, S. C. Chief Medical Officer annual report 2016: generation genome. Department of Health and Social Care. 2017; Available from: <https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2016-generation-genome>
2. Life sciences: sector deal. Office for Life Sciences, Department for Business, Energy and Industrial Strategy, Life Sciences Organisation. 2018; Available from: <https://www.gov.uk/government/publications/life-sciences-sector-deal>
3. NHS Scotland laboratory genetic services. NHS Scotland. 2020; Available from: <https://www.sglc.scot.nhs.uk>
4. NSS Strategy 2019-2024. NHS National Services Scotland. 2019; Available from: <https://nhsnss.org/media/3869/nhs-national-services-scotland-strategy-2019-2024.pdf>
5. The 100,000 genomes project. Genomics England. 2020; Available from: <https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/>
6. Informing the future of genomic medicine in Scotland. Scottish Science Advisory Council. 2019; Available from: <https://www.scottishscience.org.uk/sites/default/files/article-attachments/Genomics%20Full%20Report.pdf>
7. Next steps on the NHS five year forward view. NHS England. 2019; Available from: <https://www.england.nhs.uk/publication/next-steps-on-the-nhs-five-year-forward-view/>
8. NHS long term plan. National Health Service. 2019; Available from: <https://www.longtermplan.nhs.uk>
9. DNA testing on the NHS to fast track diagnosis for critically ill babies and children. NHS England. 2020; Available from: <https://www.england.nhs.uk/2020/01/dna-testing-on-the-nhs/>
10. National genomics board. UK Government. 2020; Available from: <https://www.gov.uk/government/groups/national-genomics-board>
11. Life sciences: industrial strategy. Office for Life Sciences. 2017; Available from: <https://www.gov.uk/government/publications/life-sciences-industrial-strategy>
12. Life Sciences Industrial Strategy: Who's driving the bus? House of Lords Science and Technology Committee. 2018; Available from: <https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/115/11502.htm>

13. Genomics and genome editing in the NHS. House of Commons Science and technology committee. 2018; Available from: <https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2017/genomics-genome-editing-nhs-17-19/>
14. Accelerating detection of disease. UK Research and Innovation. 2019; Available from: <https://www.ukri.org/innovation/industrial-strategy-challenge-fund/accelerating-detection-of-disease/>
15. The UK strategy for rare diseases: 2019 update to the implementation plan for England. Department of Health and Social Care. 2019; Available from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/781472/2019-update-to-the-rare-diseases-implementation-plan-for-england.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/781472/2019-update-to-the-rare-diseases-implementation-plan-for-england.pdf)
16. Health minister: NHS must lead the world in genomic healthcare. Department of Health and Social Care. 2019; Available from: <https://www.gov.uk/government/news/health-minister-nhs-must-lead-the-world-in-genomic-healthcare>
17. The Topol review: Preparing the healthcare workforce to deliver the digital future. Health Education England. 2019; Available from: <https://topol.hee.nhs.uk/the-topol-review/>
18. Advancing our health: prevention in the 2020s. Department of Health and Social Care. 2019; Available from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/819766/advancing-our-health-prevention-in-the-2020s-accessible.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819766/advancing-our-health-prevention-in-the-2020s-accessible.pdf)
19. Report of the independent review of adult screening programmes in England. NHS England. 2019; Available from: <https://www.england.nhs.uk/wp-content/uploads/2019/02/report-of-the-independent-review-of-adult-screening-programme-in-england.pdf>
20. Wilson, C. Sequencing the genome of every UK baby would be an ethical minefield. New Scientist. 2019; Available from: <https://www.newscientist.com/article/2222754-sequencing-the-genome-of-every-uk-baby-would-be-an-ethical-minefield/>
21. Life sciences industrial strategy update. Office for Life Sciences. 2020; Available from: <https://www.gov.uk/government/publications/life-sciences-industrial-strategy-update>
22. Genomics for precision medicine strategy. Welsh Government. 2017; Available from: <https://gov.wales/genomics-precision-medicine-strategy-0>
23. Prudent Healthcare. Bevan Commission. 2020; Available from: <http://www.bevancommission.org/en/prudent-healthcare>

24. A healthier Wales: long term plan for health and social care. Welsh Government. 2019; Available from: <https://gov.wales/healthier-wales-long-term-plan-health-and-social-care>
25. Pathology statement of intent. NHS Wales. 2019; Available from: <https://gov.wales/sites/default/files/inline-documents/2019-04/pathology-statement-of-intent.pdf>
26. Imaging statement of intent. NHS Wales. 2019; Available from: <https://gov.wales/sites/default/files/publications/2019-03/imaging-statement-of-intent.pdf>
27. Advanced therapies statement of intent. NHS Wales. 2019; Available from: <https://gov.wales/sites/default/files/inline-documents/2019-04/190409%20-%20VG%20-%20Advanced%20Therapies%20Statement%20of%20Intent%20-%20English.pdf>
28. Welsh government digital action plan 2017-2020. Welsh Government. 2017; Available from: <https://gov.wales/sites/default/files/publications/2019-07/digital-action-plan-2017-to-2020.pdf>
29. Gething, V. Written statement: precision medicine. Welsh Government. 2019; Available from: <https://gov.wales/written-statement-precision-medicine>
30. Healthcare science in NHS Wales. NHS Wales. 2018; Available from: <https://gov.wales/healthcare-science-nhs-wales>
31. Wales cancer network: single cancer pathway. NHS Wales. 2019; Available from: <http://www.walescanet.wales.nhs.uk/single-cancer-pathway>
32. Welsh rare diseases implementation plan. NHS Wales. 2017; Available from: [https://gov.wales/sites/default/files/publications/2019-01/welsh-rare-diseases-implementation-plan-july-2017\\_0.pdf](https://gov.wales/sites/default/files/publications/2019-01/welsh-rare-diseases-implementation-plan-july-2017_0.pdf)
33. Gething, V. Written statement: Update on the review on the future structure and governance for health informatics in Wales. Welsh Government. 2019; Available from: <https://gov.wales/written-statement-update-review-future-structure-and-governance-health-informatics-wales>
34. About Health Technology Wales. Health Technology Wales. 2020; Available from: <https://www.healthtechnology.wales/about/>
35. About MediWales - the life science network for Wales. MediWales. 2020; Available from: <https://www.mediwales.com/about/>
36. What we do: life sciences hub Wales. Life Sciences Hub Wales. 2020; Available from: <https://lshubwales.com/what-we-do>

37. Genomes Project, Consortium, Auton, A., Brooks, L. D., *et al.* A global reference for human genetic variation. *Nature*. 2015. 526(7571): pp. 68-74.
38. Claussnitzer, M., Cho, J. H., Collins, R., *et al.* A brief history of human disease genetics. *Nature*. 2020. 577(7789): pp. 179-189.
39. Fixing the present, building for the future. Newborn screening for rare conditions Genetic Alliance UK. 2019; Available from: <https://geneticalliance.org.uk/wp-content/uploads/2020/01/Newborn-Screening-Report.pdf>
40. Shekhani, R., Steinacher, L., Swen, J. J., *et al.* Evaluation of Current Regulation and Guidelines of Pharmacogenomic Drug Labels: Opportunities for Improvements. *Clinical pharmacology and therapeutics*. 2019. p. 10.1002/cpt.1720.
41. List of Cleared or Approved Companion Diagnostic Devices. US Food and Drug Administration 2019; Available from: <https://www.fda.gov/medical-devices/vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-vitro-and-imaging-tools>
42. Sanger, F., Coulson, A. R., Friedmann, T., *et al.* The nucleotide sequence of bacteriophage phiX174. *J Mol Biol*. 1978. 125(2): pp. 225-46.
43. Marco-Puche, G., Lois, S., Benitez, J., *et al.* RNA-Seq Perspectives to Improve Clinical Diagnosis. *Front Genet*. 2019. 10: p. 1152.
44. Delaney, C., Garg, S. K., Yung, R. Analysis of DNA Methylation by Pyrosequencing. *Methods Mol Biol*. 2015. 1343: pp. 249-64.
45. Constitutional cyto- and molecular genetics: Karyotyping, FISH and CGH array. Eurofins Scientific. 2016; Available from: [https://www.eurofins-biomnis.com/wp-content/uploads/2016/04/56-INTGB-Focus\\_Karyotyping\\_SNP\\_array.pdf](https://www.eurofins-biomnis.com/wp-content/uploads/2016/04/56-INTGB-Focus_Karyotyping_SNP_array.pdf)
46. Bejjani, B. A., Shaffer, L. G. Application of array-based comparative genomic hybridization to clinical diagnostics. *J Mol Diagn*. 2006. 8(5): pp. 528-33.
47. Mason, J., Griffiths, M. Detection of minimal residual disease in leukaemia by RT-PCR. *Methods Mol Biol*. 2011. 688: pp. 269-80.
48. Shen, C.-H. , 'Ed' Diagnostic Molecular Biology. Elsevier; 2018.
49. Shi, Z., Yu, H., Wu, Y., *et al.* Systematic evaluation of cancer-specific genetic risk score for 11 types of cancer in The Cancer Genome Atlas and Electronic Medical Records and Genomics cohorts. *Cancer Med*. 2019. 8(6): pp. 3196-3205.

50. Stuppia, L., Antonucci, I., Palka, G., *et al.* Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci.* 2012. 13(3): pp. 3245-76.
51. Hernandez, H. G., Tse, M. Y., Pang, S. C., *et al.* Optimizing methodologies for PCR-based DNA methylation analysis. *Biotechniques.* 2013. 55(4): pp. 181-97.
52. Olmedillas-Lopez, S., Garcia-Arranz, M., Garcia-Olmo, D. Current and Emerging Applications of Droplet Digital PCR in Oncology. *Mol Diagn Ther.* 2017. 21(5): pp. 493-510.
53. Elazezy, M., Joosse, S. A. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. *Computational and Structural Biotechnology Journal.* 2018. 16: pp. 370-378.
54. NE Thames regional genetics service: Prader-Willi syndrome (PWS) Great Ormond Street Hospital for Children NHS Foundation Trust. 2011; Available from: <http://www.labs.gosh.nhs.uk/media/384195/PWS%20v4.5.pdf>
55. Karow, J. Clinical Utility of Polygenic Risk Scores Questioned in Commentary GenomeWeb. 2019; Available from: <https://www.genomeweb.com/molecular-diagnostics/clinical-utility-polygenic-risk-scores-questioned-commentary>
56. Smith-Woolley, E., Selzam, S., Plomin, R. Polygenic Score for Educational Attainment Captures DNA Variants Shared Between Personality Traits and Educational Achievement. *Journal of Personality and Social Psychology.* 2019. 117(6): pp. 1145-1163.
57. DNA Fragmentation. New England BioLabs Inc. 2020; Available from: <https://international.neb.com/applications/ngs-sample-prep-and-target-enrichment/dna-fragmentation>
58. Terry, M. Top 10 Gene Sequencing Companies by Revenue. BioSpace. 2019; Available from: <https://www.biospace.com/article/top-10-gene-sequencing-companies-by-revenue/>
59. Lander, E. , Linton, L. , Birren, B. , *et al.* Initial sequencing and analysis of the human genome. *Nature.* 2001. (409): pp. 860–921.
60. DNA extraction and sequencing. Cambridge University Hospitals NHS Foundation Trust. 2020; Available from: <https://www.cuh.nhs.uk/addenbrookes-hospital/services/genetics-laboratories/services-and-tests/dna-extraction-and-sequencing>
61. Schedule of Accreditation - Birmingham Women's and Children's NHS Foundation Trust. UK National Accreditation Body. 2019; Available from: [https://www.ukas.com/wp-content/uploads/schedule\\_uploads/00007/8176%20Medical%20Single.pdf](https://www.ukas.com/wp-content/uploads/schedule_uploads/00007/8176%20Medical%20Single.pdf)

62. Instruments for Sanger Sequencing and Fragment Analysis by Capillary Electrophoresis. Thermo Fisher Scientific. 2020; Available from: <https://www.thermofisher.com/uk/en/home/life-science/sequencing/sanger-sequencing/sanger-sequencing-technology-accessories.html>
63. NovaSeq 6000. University of Utah - Health. 2020; Available from: <https://uofuhealth.utah.edu/huntsman/shared-resources/gba/htg/illumina/novaseq.php>
64. NovaSeq Specifications. Illumina. 2020; Available from: <https://emea.illumina.com/systems/sequencing-platforms/novaseq/specifications.html>
65. NovaSeq 6000 Sequencing. University of Minnesota Genomics Center. 2020; Available from: <http://genomics.umn.edu/nextgen-novaseq.php>
66. DNA sequencing costs: Data. NIH National Human Genome Research Institute. 2019; Available from: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
67. Schwarze, K., Buchanan, J., Fermont, J. M., *et al.* The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genet Med.* 2020. 22(1): pp. 85-94.
68. Jain, M., Koren, S., Miga, K. H., *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol.* 2018. 36(4): pp. 338-345.
69. Oxford Nanopore Technologies - Devices. Oxford Nanopore Technologies. 2020; Available from: <https://store.nanoporetech.com/uk/devices/>
70. Heger, M. PacBio Shares Early-Access Customer Experiences, New Applications for Sequel II. 2019; Available from: <https://www.genomeweb.com/sequencing/pacbio-shares-early-access-customer-experiences-new-applications-sequel-ii#.Xs0mjzpKhPZ>
71. Heger, M. PacBio Launches Higher-Throughput, Lower-Cost Single-Molecule Sequencing System. 2015; Available from: <https://www.genomeweb.com/business-news/pacbio-launches-higher-throughput-lower-cost-single-molecule-sequencing-system#.Xr253WhKhPZ>
72. Award-Winning Sequel II System Sets a New Standard for Long-Read Sequencing. Pacific Biosciences. 2019; Available from: <https://www.pacb.com/blog/award-winning-sequel-ii-system/>
73. Chan, E. K. F., Cameron, D. L., Petersen, D. C., *et al.* Optical mapping reveals a higher level of genomic architecture of chained fusions in cancer. *Genome Research.* 2018. 28(5): pp. 726-738.

74. Fillerova, R., Minarik, J., Petrackova, A., *et al.* Next-generation optical mapping reveals numerous previously unrecognizable structural variants in multiple myeloma. *Clinical Lymphoma, Myeloma and Leukemia*. 2019. 19(10): p. e64.
75. Kanagal-Shamanna, R., Montalban-Bravo, G., Sasaki, K., *et al.* High-Resolution Next-Generation Whole Genome Optical Mapping As a Novel Molecular Diagnostic Tool for Comprehensive Assessment of Structural Chromosomal Variations in Myelodysplastic Syndromes. *Blood*. 2019. 134(Supplement\_1): pp. 5438-5438.
76. Zhang, Q., Xu, X., Ding, L., *et al.* Clinical application of single-molecule optical mapping to a multigeneration FSHD1 pedigree. *Mol Genet Genomic Med*. 2019. 7(3): p. e565.
77. Loop Genomics. Loop Genomics. 2020; Available from: <https://www.loopgenomics.com/>
78. What is bioinformatics? European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI). 2020; Available from: <https://www.ebi.ac.uk/training/online/course/bioinformatics-terrified/what-bioinformatics-0>
79. NGS bioinformatics best practice guidelines. Association for Clinical Genomic Science. 2016; Available from: [https://www.acgs.uk.com/media/10790/ngs\\_bioinformatics\\_bpg\\_final\\_version\\_2016.pdf](https://www.acgs.uk.com/media/10790/ngs_bioinformatics_bpg_final_version_2016.pdf)
80. Targeted next generation sequencing analysis and interpretation best practice guidelines. Association for Clinical Genomic Science. 2015; Available from: [https://www.acgs.uk.com/media/10789/bpg\\_for\\_targeted\\_next\\_generation\\_sequencing\\_-\\_approved\\_dec\\_2015.pdf](https://www.acgs.uk.com/media/10789/bpg_for_targeted_next_generation_sequencing_-_approved_dec_2015.pdf)
81. Lightbody, G., Haberland, V., Browne, F., *et al.* Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Brief Bioinform*. 2019. 20(5): pp. 1795-1811.
82. DNA Sequencing Costs: Data. National Human Genome Research Institute. 2020; Available from: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
83. Deans, Z., Watson, C. M., Charlton, R., *et al.* Practice guidelines for Targeted Next Generation Sequencing Analysis and Interpretation. Association for Clinical Genetic Science. 2015; Available from: [https://www.acgs.uk.com/media/10789/bpg\\_for\\_targeted\\_next\\_generation\\_sequencing\\_-\\_approved\\_dec\\_2015.pdf](https://www.acgs.uk.com/media/10789/bpg_for_targeted_next_generation_sequencing_-_approved_dec_2015.pdf)
84. Clinical exome (GOSHHome). Great Ormand Street Hospital for Children NHS Foundation Trust. 2020; Available from: [http://labs.gosh.nhs.uk/laboratory-services/genetics/tests/clinical-exome-\(goshome\)](http://labs.gosh.nhs.uk/laboratory-services/genetics/tests/clinical-exome-(goshome))

85. TruSight Sequencing Panels. Illumina. 2020; Available from: <https://emea.illumina.com/products/trusight-panels.html>
86. Genomics for Rare Disease in NHS Scotland - The bridge to a Scottish Strategy for Genomics: Report 2. NHS Scotland. 2019; Available from: <https://www.sglc.scot.nhs.uk/wp-content/uploads/2020/02/Genomics-Bridge-Report-2-Dec-2019-v1.0-FINAL.pdf>
87. Sakharkar, M. K., Chow, V. T., Kanguane, P. Distributions of exons and introns in the human genome. *In Silico Biol.* 2004. 4(4): pp. 387-93.
88. Clark, M. M., Stark, Z., Farnaes, L., *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med.* 2018. 3: p. 16.
89. Kingsmore, S. F., Cakici, J. A., Clark, M. M., *et al.* A Randomized, Controlled Trial of the Analytic and Diagnostic Performance of Singleton and Trio, Rapid Genome and Exome Sequencing in Ill Infants. *Am J Hum Genet.* 2019. 105(4): pp. 719-733.
90. Cherot, E., Keren, B., Dubourg, C., *et al.* Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin Genet.* 2018. 93(3): pp. 567-576.
91. Han, J. Y., Jang, J. H., Park, J., *et al.* Targeted Next-Generation Sequencing of Korean Patients With Developmental Delay and/or Intellectual Disability. *Front Pediatr.* 2018. 6: p. 391.
92. PanelApp User Guide. Genomics England. 2019; Available from: [https://panelapp.genomicsengland.co.uk/media/files/PanelApp\\_User\\_Guide.pdf](https://panelapp.genomicsengland.co.uk/media/files/PanelApp_User_Guide.pdf)
93. Caspar, S. M., Dubacher, N., Kopps, A. M., *et al.* Clinical sequencing: From raw data to diagnosis with lifetime value. *Clin Genet.* 2018. 93(3): pp. 508-519.
94. Welcome to the Deciphering Developmental Disorders (DDD) study. Deciphering Developmental Disorders (DDD) study. 2020; Available from: <https://www.ddduk.org/>
95. Mestek-Boukhibar, L., Clement, E., Jones, W. D., *et al.* Rapid Paediatric Sequencing (RaPS): comprehensive real-life workflow for rapid diagnosis of critically ill children. *J Med Genet.* 2018. 55(11): pp. 721-728.
96. Lucassen, A., Hall, A. Consent and confidentiality in genomic medicine, 3rd edition. Joint Committee on Genomics in Medicine. 2019; Available from: <https://www.rcplondon.ac.uk/projects/outputs/consent-and-confidentiality-genomic-medicine>

97. Hoffman-Andrews, L. The known unknown: the challenges of genetic variants of uncertain significance in clinical practice. *J Law Biosci.* 2017. 4(3): pp. 648-657.
98. BRCA Exchange. 2020; Available from: <https://brcaexchange.org/>
99. CanGene CanVar. 2020; Available from: <https://www.cangene-canvaruk.org/>
100. Welcome to the Human Functional Genomics Project. HFGP. 2020; Available from: <http://www.humanfunctionalgenomics.org/site/>
101. Kalia, S. , Adelman, K., Bale, S. J., *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med.* 2017; 19(2): pp. 249-255.
102. Information for Participants - What can participants find out? Genomics England. 2020; Available from: <https://www.genomicsengland.co.uk/information-for-participants/findings/>
103. Caulfield, M., Davies, J., Dennys, M., *et al.* The 100,000 Genomes Project Protocol v4. Genomics England. 2017; Available from: [https://figshare.com/articles/GenomicEnglandProtocol\\_pdf/4530893/4](https://figshare.com/articles/GenomicEnglandProtocol_pdf/4530893/4)
104. Saelaert, M., Mertes, H., Moerenhout, T., *et al.* Ethical values supporting the disclosure of incidental and secondary findings in clinical genomic testing: a qualitative study. *BMC Medical Ethics.* 2020. 21(1): p. 9.
105. What are secondary findings from genetic testing? US National Library of Medicine - Genetics Home Reference. 2020; Available from: <https://ghr.nlm.nih.gov/primer/testing/secondaryfindings>
106. ACGS 2015-2016 Genetic Test Activity Audit. Association for Clinical Genomic Science. 2017; Available from: <https://www.acgs.uk.com/quality/>
107. Kroese, M., Deller, J., Dew, C., *et al.* Genetic test activity in England & Scotland 2016/17. NHS Digital 2018; Available from: <https://files.digital.nhs.uk/20/41CA7D/Genetic%20test%20activity%20in%20England%20and%20Scotland%20Report%20for%20%20201617.pdf>
108. Plans for introduction of Non-Invasive Pre-Natal Testing into pregnancy screening programme: FOI release. Scottish Government. 2019; Available from: <https://www.gov.scot/publications/foi-19-00366/>
109. McHugh, A. Update on NIPT implementation — microarray testing technology and twins. Public Health England 2019; Available from: <https://phescreening.blog.gov.uk/2019/06/17/update-on-nipt-implementation-microarray-testing-technology-and-twins/>

110. Joynson, C. Our concerns about non-invasive prenatal testing (NIPT) in the private healthcare sector. Nuffield Council on Bioethics. 2019; Available from: <https://www.nuffieldbioethics.org/blog/nipt-private>
111. Rowe, C. A., Wright, C. F. Expanded universal carrier screening and its implementation within a publicly funded healthcare service. *Journal of Community Genetics*. 2020. 11(1): pp. 21-38.
112. Hill, S., Viner, R. Finding the missing pieces – how genomics is helping to diagnose and treat rare diseases. NHS England 2020; Available from: <https://www.england.nhs.uk/blog/finding-the-missing-pieces-how-genomics-is-helping-to-diagnose-and-treat-rare-diseases/>
113. Preimplantation genetic diagnosis (PGD) Information for patients Leaflet no: 028, Version 5. NHS Lothian. 2020; Available from: <https://services.nhsllothian.scot/clinicalgeneticservice/Information%20Leaflets/Documents/Preimplantation%20Genetic%20Diagnosis.pdf>
114. Pre-implantation genetic diagnosis. NHS National Services Scotland 2020; Available from: <https://www.nss.nhs.scot/specialist-healthcare/specialist-services/pre-implantation-genetic-diagnosis/>
115. Preimplantation Genetic Diagnosis (PGD) NHS Lothian. 2016; Available from: [https://services.nhsllothian.scot/EdinburghAssistedConceptionProgramme/Documents/Preimplantation%20Genetic%20Diagnosis%20\(PGD\).pdf](https://services.nhsllothian.scot/EdinburghAssistedConceptionProgramme/Documents/Preimplantation%20Genetic%20Diagnosis%20(PGD).pdf)
116. Pre-implantation genetic diagnosis (PGD). Human Fertilisation and Embryology Authority 2020; Available from: <https://www.hfea.gov.uk/treatments/embryo-testing-and-treatments-for-disease/pre-implantation-genetic-diagnosis-pgd/>
117. Li, G., Niu, W., Jin, H., *et al.* Importance of embryo aneuploidy screening in preimplantation genetic diagnosis for monogenic diseases using the karyomap gene chip. *Scientific Reports*. 2018. 8(1): p. 3139.
118. Hao, Y., Chen, D., Zhang, Z., *et al.* Successful preimplantation genetic diagnosis by targeted next-generation sequencing on an ion torrent personal genome machine platform. *Oncol Lett*. 2018. 15(4): pp. 4296-4302.
119. Backenroth, D., Zahdeh, F., Kling, Y., *et al.* Haploseek: a 24-hour all-in-one method for preimplantation genetic diagnosis (PGD) of monogenic disease and aneuploidy. *Genetics in Medicine*. 2019. 21(6): pp. 1390-1399.
120. Masset, H., Zamani Esteki, M., Dimitriadou, E., *et al.* Multi-centre evaluation of a comprehensive preimplantation genetic test through haplotyping-by-sequencing. *Hum Reprod*. 2019. 34(8): pp. 1608-1619.

121. Outline business case: T13/T18 & non-invasive pre-natal testing implementation in Scotland. NHS Scotland. 2017; Available from: <https://www.gov.scot/binaries/content/documents/govscot/publications/foi-eir-release/2018/01/foi-17-02965/documents/foi-17-02965-nipt-outline-business-case-pdf/foi-17-02965-nipt-outline-business-case-pdf/govscot%3Adocument/FOI-17-02965-NIPT%20Outline%20Business%20Case.pdf>
122. NHS Fetal Anomaly Screening Programme care pathway: Down's (T21), Edwards' (T18) and Patau's (T13). Public Health England. 2020; Available from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/772823/FASP\\_Downs\\_T21\\_Edwards\\_T18\\_and\\_Pataus\\_T13\\_pathway.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/772823/FASP_Downs_T21_Edwards_T18_and_Pataus_T13_pathway.pdf)
123. Hill, M., Wright, D., Daley, R., *et al.* Evaluation of non-invasive prenatal testing (NIPT) for aneuploidy in an NHS setting: a reliable accurate prenatal non-invasive diagnosis (RAPID) protocol. *BMC Pregnancy Childbirth*. 2014. 14: p. 229.
124. Chitty, L. S., Cameron, L., Daley, R., *et al.* RAPID Non-invasive prenatal testing (NIPT) evaluation study - Executive summary Public Health England. 2015; Available from: [https://legacyscreening.phe.org.uk/policydb\\_download.php?doc=551](https://legacyscreening.phe.org.uk/policydb_download.php?doc=551)
125. Chitty, L. S., Wright, D., Hill, M., *et al.* Uptake, outcomes, and costs of implementing non-invasive prenatal testing for Down's syndrome into NHS maternity care: prospective cohort study in eight diverse maternity units. *BMJ*. 2016. 354: p. i3426.
126. Antenatal Screening Wales Annual Report 2018-2019. Public Health Wales. 2019; Available from: <http://www.antenatalscreening.wales.nhs.uk/sitesplus/documents/989/Antenatal%20Screening%20Wales%20Annual%20Report%202018-2019%20V1.0.pdf>
127. UK NSC: Notes of the meeting held on the 27 February 2019 in Holiday Inn, London. UK National Screening Committee. 2019; Available from: <https://drive.google.com/drive/folders/0B8eopFA9myQefk9EalVQLWs5RlFBSmlUdzNvaUxuS2FhdWV2VWJaSTZ2YXITdVZDYXdyNDA>
128. Skrzypek, H., Hui, L. Noninvasive prenatal testing for fetal aneuploidy and single gene disorders. *Best Pract Res Clin Obstet Gynaecol*. 2017. 42: pp. 26-38.
129. Recommendations for the use of chromosome microarray in pregnancy The Royal College of Pathologists. 2015; Available from: <https://www.rcpath.org/uploads/assets/06664c28-0f90-4230-86158c91fea14be6/Recommendations-for-the-use-of-chromosome-microarray-in-pregnancy.pdf>
130. Health Quality Ontario. Noninvasive Prenatal Testing for Trisomies 21, 18, and 13, Sex Chromosome Aneuploidies, and Microdeletions: A Health Technology Assessment. *Ont Health Technol Assess Ser*. 2019. 19(4): pp. 1-166.

131. Kostenko, E., Chantraine, F., Vandeweyer, K., *et al.* Clinical and Economic Impact of Adopting Noninvasive Prenatal Testing as a Primary Screening Method for Fetal Aneuploidies in the General Pregnancy Population. *Fetal Diagn Ther.* 2019. 45(6): pp. 413-423.
132. van der Meij, K. R. M., Sistermans, E. A., Macville, M. V. E., *et al.* TRIDENT-2: National Implementation of Genome-wide Non-invasive Prenatal Testing as a First-Tier Screening Test in the Netherlands. *Am J Hum Genet.* 2019. 105(6): pp. 1091-1101.
133. Lord, J., McMullan, D. J., Eberhardt, R. Y., *et al.* Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): a cohort study. *Lancet.* 2019. 393(10173): pp. 747-757.
134. Kodabuckus, S. S., Quinlan-Jones, E., McMullan, D. J., *et al.* Exome Sequencing for Prenatal Detection of Genetic Abnormalities in Fetal Ultrasound Anomalies: An Economic Evaluation. *Fetal Diagn Ther.* 2020. pp. 1-11.
135. van Campen, J. C., Sollars, E. S. A., Thomas, R. C., *et al.* Next Generation Sequencing in Newborn Screening in the United Kingdom National Health Service. *Int J Neonatal Screen.* 2019. 5(4): p. 40.
136. PerkinElmer SCID Test Chosen for Newborn Screening Evaluation in England. *GenomeWeb.* 2020; Available from: <https://www.genomeweb.com/business-news/perkinelmer-scid-test-chosen-newborn-screening-evaluation-england>
137. UK NSC public consultation: screening for severe combined immunodeficiency (SCID). *Public Health England.* 2017; Available from: <https://phscreening.blog.gov.uk/2017/10/13/uk-nsc-public-consultation-screening-for-severe-combined-immunodeficiency-scid/>
138. Newborn blood spot evaluation update — screening for SCID. *Public Health England.* 2019; Available from: <https://phscreening.blog.gov.uk/2019/10/29/scid-update-bloodspot-screening>
139. French, C. E., Delon, I., Dolling, H., *et al.* Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med.* 2019. 45(5): pp. 627-636.
140. Guidelines on autopsy practice: Fetal autopsy (2nd trimester fetal loss and termination of pregnancy for congenital anomaly). *Royal College of Pathologists.* 2017; Available from: <https://www.rcpath.org/uploads/assets/b20ea503-7799-433c-99160653762f896c/Fetal-autopsy-2nd-trimester-fetal-loss-and-termination-of-pregnancy-for-congenital-anomaly.pdf>
141. Transforming pathology and diagnostic practice. *Genomics England.* 2019; Available from: <https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/information-for-gmc-staff/cancer-programme/pathology-in-the-nhs/>

142. Fellmann, F., van El, C. G., Charron, P., *et al.* European recommendations integrating genetic testing into multidisciplinary management of sudden cardiac death. *Eur J Hum Genet.* 2019. 27(12): pp. 1763-1773.
143. Guidelines on autopsy practice: Sudden death with likely cardiac pathology. Royal College of Pathologists. 2015; Available from: <https://www.rcpath.org/uploads/assets/823dfcf4-8eba-40f7-81b7e174675ecdd9/Guidelines-on-autopsy-practice-Sudden-death-with-likely-cardiac-pathology.pdf>
144. Lahrouchi, N., Raju, H., Lodder, E. M., *et al.* The yield of postmortem genetic testing in sudden death cases with structural findings at autopsy. *Eur J Hum Genet.* 2020. 28(1): pp. 17-22.
145. Alfirevic, A., Pirmohamed, M. Genomics of Adverse Drug Reactions. *Trends Pharmacol Sci.* 2017. 38(1): pp. 100-109.
146. Lumacaftor-ivacaftor for treating cystic fibrosis homozygous for the f508del mutation. National Institute for Health and Care Excellence (NICE). 2016; Available from: <https://www.nice.org.uk/guidance/ta398>
147. Alirocumab for treating primary hypercholesterolaemia and mixed dyslipidaemia. National Institute for Health and Care Excellence (NICE). 2016; Available from: <https://www.nice.org.uk/guidance/ta393>
148. Meulendijks, D., Henricks, L. M., Sonke, G. S., *et al.* Clinical relevance of DPYD variants c. 1679T> G, c. 1236G> A/HapB3, and c. 1601G> A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *The Lancet Oncology.* 2015. 16(16): pp. 1639-1650.
149. Stavrika, C., Pouptsis, A., Okonta, L., *et al.* Clinical implementation of pre-treatment DPYD genotyping in capecitabine-treated metastatic breast cancer patients. *Breast cancer research and treatment.* 2019. 175(2): pp. 511-517.
150. Mehta, D., Uber, R., Ingle, T., *et al.* Study of pharmacogenomic information in FDA-approved drug labeling to facilitate application of precision medicine. *Drug Discov Today.* 2020.
151. British National Formulary (BNF). National Institute for Health and Care Excellence (NICE). 2020; Available from: <https://bnf.nice.org.uk/>
152. The NIHR BioResource on behalf of the 100, 000 Genomes Project. Whole-genome sequencing of rare disease patients in a national healthcare system. *bioRxiv.* 2019. p. 10.1101/507244.
153. Martin, A. R., Williams, E., Foulger, R. E., *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet.* 2019. 51(11): pp. 1560-1565.

154. Taylor, A., Brady, A. F., Frayling, I. M., *et al.* Consensus for genes to be included on cancer panel tests offered by UK genetics services: guidelines of the UK Cancer Genetics Group. *Journal of Medical Genetics*. 2018. 55(6): pp. 372-377.
155. 2019/2020 National Genomic Test Directory for rare and inherited diseases. . NHS England. 2019; Available from: <https://www.england.nhs.uk/publication/national-genomic-test-directories/>
156. 2019/2020 National Genomic Test Directory for Cancer. NHS England. 2019; Available from: <https://www.england.nhs.uk/publication/national-genomic-test-directories/>
157. Ravandi, F., Walter, R.B., Freeman, S.D. Evaluating measurable residual disease in acute myeloid leukemia. *Blood Advances*. 2018. 2(11): p. 11.
158. Kruse, A., Abdel-Azim, N., Kim, H. N., *et al.* Minimal Residual Disease Detection in Acute Lymphoblastic Leukemia. *Int J Mol Sci*. 2020. 21(3).
159. Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer-after chemotherapy. TA428. National Institute for Health and Care Excellence (NICE). 2017; Available from: <https://www.nice.org.uk/guidance/ta428>
160. Wan, J. C. M., Massie, C., Garcia-Corbacho, J., *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017. 17(4): pp. 223-238.
161. Diagnostics guidance DG9: EGFR-TK mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer. National Institute for Health and Care Excellence (NICE). 2013; Available from: <https://www.nice.org.uk/guidance/dg9>
162. Howells, M. Colorectal cancer – RAS Mutation Testing on ctDNA. All Wales Genetics Laboratory 2018; Available from: <http://www.wales.nhs.uk/sites3/Documents/525/MI-GEN-RASctDNA.pdf>
163. Gorgannezhad, L., Umer, M., Islam, M. N., *et al.* Circulating tumor DNA and liquid biopsy: opportunities, challenges, and recent advances in detection technologies. *Lab Chip*. 2018. 18(8): pp. 1174-1196.
164. Malone, E. R., Oliva, M., Sabatini, P. J. B., *et al.* Molecular profiling for precision cancer therapies. *Genome Med*. 2020. 12(1): p. 8.
165. Doherty, G. J., Petruzzelli, M., Beddowes, E., *et al.* Cancer Treatment in the Genomic Era. *Annu Rev Biochem*. 2019. 88: pp. 247-280.
166. Robbe, P., Popitsch, N., Knight, S. J. L., *et al.* Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. *Genet Med*. 2018. 20(10): pp. 1196-1205.

167. Berner, A. M., Morrissey, G. J., Murugaesu, N. Clinical Analysis of Whole Genome Sequencing in Cancer Patients. *Current Genetic Medicine Reports*. 2019. 7(2): pp. 136-143.
168. Jones, L., Craig, C. Tissue handling for molecular pathology. *The Bulletin of The Royal College of Pathologists*. 2017. 180: p. 3.
169. ACMG Board of Directors. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2015. 17(1): p. 68.
170. Mandelker, D., Donoghue, M., Talukdar, S., *et al*. Germline-focussed analysis of tumour-only sequencing: recommendations from the ESMO Precision Medicine Working Group. *Annals of Oncology*. 2019. 30(8): pp. 1221-1231.
171. Hironaka-Mitsuhashi, A., Sanchez Calle, A., Ochiya, T., *et al*. Towards Circulating-Tumor DNA-Based Precision Medicine. *J Clin Med*. 2019. 8(9).
172. Mari, R., Mamessier, E., Lambaudie, E., *et al*. Liquid Biopsies for Ovarian Carcinoma: How Blood Tests May Improve the Clinical Management of a Deadly Disease. *Cancers*. 2019. 11(6).
173. Mattox, A. K., Bettgowda, C., Zhou, S., *et al*. Applications of liquid biopsies for cancer. *Science Translational Medicine*. 2019. 11(507): p. eaay1984.
174. Turner, N., Bye, H., Kernaghan, S., *et al*. Abstract OT1-06-03: The plasmaMATCH trial: A multiple parallel cohort, open-label, multi-centre phase II clinical trial of ctDNA screening to direct targeted therapies in patients with advanced breast cancer (CRUK/15/010). *Cancer Research*. 2018. 78(4 Supplement): pp. OT1-06-03.
175. Rothwell, D. G., Ayub, M., Cook, N., *et al*. Utility of ctDNA to support patient selection for early phase clinical trials: the TARGET study. *Nat Med*. 2019. 25(5): pp. 738-743.
176. Provencio, M., Perez-Barrios, C., Barquin, M., *et al*. Next-generation sequencing for tumor mutation quantification using liquid biopsies. *Clin Chem Lab Med*. 2020; 58(2): pp. 306-313.
177. Zhang, H., Liu, R., Yan, C., *et al*. Advantage of Next-Generation Sequencing in Dynamic Monitoring of Circulating Tumor DNA over Droplet Digital PCR in Cetuximab Treated Colorectal Cancer Patients. *Transl Oncol*. 2019. 12(3): pp. 426-431.
178. Cohen, J. D., Li, L., Wang, Y., *et al*. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018. 359(6378): pp. 926-930.

179. Gilmore, J. Sweden's Saga Diagnostics Eyes Clinical Oncology Space With Ultrasensitive Mutation Detection Method. GenomeWeb. 2019; Available from: <https://www.genomeweb.com/liquid-biopsy/swedens-saga-diagnostics-eyes-clinical-oncology-space-ultrasensitive-mutation>
180. Dutch Startup Cyclomics Developing Liquid Biopsy Assay Using Nanopore Sequencing. GenomeWeb. 2019; Available from: <https://www.genomeweb.com/liquid-biopsy/dutch-startup-cyclomics-developing-liquid-biopsy-assay-using-nanopore-sequencing>
181. Sunakawa, Y., Usher, J. L., Jaimes, Y. S., *et al.* Clinical verification of circulating tumor RNA (ctRNA) as novel pretreatment predictor and tool for quantitative monitoring of treatment response in metastatic colorectal cancer (mCRC): A biomarker study of the DEEPER trial. *Journal of Clinical Oncology*. 2019. 37(15).
182. Cescon, D. W., Bratman, S., Chan, S. M., *et al.* Circulating tumor DNA and liquid biopsy in oncology. *Nature Cancer*. 2020. 1: pp. 276-90.
183. Larotrectinib for treating NTRK fusion-positive solid tumours. National Institute for Health and Care Excellence (NICE). 2020; Available from: <https://www.nice.org.uk/guidance/ta630>
184. Modelling approaches for site-agnostic cancer drugs to inform NICE appraisal. HTA - NIHR127852. National Institute for Health Research 2019; Available from: <https://www.journalslibrary.nihr.ac.uk/programmes/hta/NIHR127852/#/documentation>
185. Raza, S. Artificial intelligence for genomic medicine. PHG Foundation. 2020; Available from: <https://www.phgfoundation.org/report/artificial-intelligence-for-genomic-medicine>
186. Xu, Y., Lewandowski, K., Jeffery, K., *et al.* Nanopore metagenomic sequencing to investigate nosocomial transmission of human metapneumovirus from a unique genetic group among haematology patients in the United Kingdom. *J Infect*. 2020. 80(5): pp. 571-577.
187. Rhodes, J., Abdolrasouli, A., Farrer, R. A., *et al.* Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*. *Emerg Microbes Infect*. 2018. 7(1): p. 43.
188. New technology could speed up diagnosis for Huntington's disease. Guy's and St Thomas' NHS Foundation Trust. 2019; Available from: <https://www.guysandstthomas.nhs.uk/news-and-events/2019-news/may/New-technology-could-speed-up-diagnosis-for-Huntingtons-disease.aspx>
189. Wood, E. The Characterisation of Mitochondrial DNA Rearrangements using the Illumina MiSeq and Oxford Nanopore Technologies MinION. Health Education England. 2019; Available from: <https://nshcs.hee.nhs.uk/research/the-characterisation-of-mitochondrial-dna-rearrangements-using-the-illumina-miseq-and-oxford-nanopore-technologies-minion/>

190. Solve RD - Solving the Unsolved Rare Diseases. Solve RD. 2020; Available from: <http://solve-rd.eu/>
191. European Research Initiative Adopts PacBio Sequel II System to Solve Rare Diseases. 2019; Available from: [https://www.pacb.com/press\\_releases/european-research-initiative-adopts-pacbio-sequel-ii-system-to-solve-rare-diseases/](https://www.pacb.com/press_releases/european-research-initiative-adopts-pacbio-sequel-ii-system-to-solve-rare-diseases/)
192. FASP-S03a: test: screen positive rate T21/T18/T13 screening. Public Health England. 2020; Available from: <https://www.gov.uk/government/publications/fetal-anomaly-screening-programme-standards/standards-valid-from-1-april-2018#fasp-s03a-test-screen-positive-rate-t21t18t13-screening>
193. Prospective parents should be prepared for a surge in genetic data. *Nature*. 2019. 565(7741): p. 536.
194. Zhang, J., Li, J., Saucier, J. B., *et al.* Non-invasive prenatal sequencing for multiple Mendelian monogenic disorders using circulating cell-free fetal DNA. *Nat Med*. 2019. 25(3): pp. 439-447.
195. Vora, N. L., Gilmore, K., Brandt, A., *et al.* An approach to integrating exome sequencing for fetal structural anomalies into clinical practice. *Genet Med*. 2020. 22(5): pp. 954-961.
196. Newborn screening: time to expand the list? . Health Education England. 2019; Available from: <https://www.genomicseducation.hee.nhs.uk/blog/newborn-screening-time-to-expand-the-list/>
197. Johnston, J., Lantos, J. D., Goldenberg, A., *et al.* Sequencing Newborns: A Call for Nuanced Use of Genomic Technologies. *Hastings Cent Rep*. 2018. 48 Suppl 2: pp. S2-S6.
198. Genome sequencing of healthy newborns: BSGM response. British Society for Genetic Medicine. 2019; Available from: <https://www.bsgm.org.uk/news/genome-sequencing-of-healthy-newborns-bsgm-response/>
199. Bioethics briefing note: whole genome sequencing of babies. Nuffield Council on Bioethics. 2018; Available from: <https://www.nuffieldbioethics.org/wp-content/uploads/Nuffield-Council-on-Bioethics-briefing-note-whole-genome-sequencing-of-babies.pdf>
200. Ceyhan-Birsoy, O., Murry, J. B., Machini, K., *et al.* Interpretation of Genomic Sequencing Results in Healthy and Ill Newborns: Results from the BabySeq Project. *Am J Hum Genet*. 2019. 104(1): pp. 76-93.
201. Milko, L. V., Rini, C., Lewis, M. A., *et al.* Evaluating parents' decisions about next-generation sequencing for their child in the NC NEXUS (North Carolina Newborn Exome Sequencing for Universal Screening) study: a randomized controlled trial protocol. *Trials*. 2018. 19(1): p. 344.

202. Chin, R. I., Chen, K., Usmani, A., *et al.* Detection of Solid Tumor Molecular Residual Disease (MRD) Using Circulating Tumor DNA (ctDNA). *Mol Diagn Ther.* 2019. 23(3): pp. 311-331.
203. Tracking Mutations in Cell Free Tumour DNA to Predict Relapse in Early Colorectal Cancer (TRACC). US National Library of Medicine - Clinical Trials. 2020; Available from: <https://clinicaltrials.gov/ct2/show/NCT04050345>
204. Clinical trial: c-TRAK TN. The Institute of Cancer Research. 2020; Available from: <https://www.icr.ac.uk/our-research/centres-and-collaborations/centres-at-the-icr/clinical-trials-and-statistics-unit/clinical-trials/c-trak-tn>
205. GRAIL Announces Positive New Data with Multi-Cancer Early Detection Blood Test from CCGA Study GRAIL, Inc. 2019; Available from: <https://grail.com/press-releases/grail-announces-positive-new-data-with-multi-cancer-early-detection-blood-test-from-ccga-study/>
206. Han, A.P. Guardant Health Announces Plans for Prospective Colorectal Cancer Screening Study GenomeWeb. 2019; Available from: <https://www.genomeweb.com/molecular-diagnostics/guardant-health-announces-plans-prospective-colorectal-cancer-screening-study>
207. CRISP-DNA: The feasibility of assessing bowel cancer risk using genomic testing to target screening in general practice. University of Melbourne. 2020; Available from: <https://medicine.unimelb.edu.au/research-groups/general-practice-research/cancer-research-group/crisp-dna-the-feasibility-of-assessing-bowel-cancer-risk-using-genomic-testing-to-target-screening-in-general-practice>
208. The CanRisk web tool. University of Cambridge. 2020; Available from: <https://www.canrisk.org>
209. Genomics Plc analysis demonstrated potential for genomic data to identify those most at risk for 16 common diseases. Genomics Plc. 2019; Available from: <https://www.genomicsplc.com/genomics-plc-analysis-demonstrates-potential-for-genomic-data-to-identify-those-most-at-risk-for-16-common-diseases/>
210. Moorthie, S., Babb de Villiers, C., Bridgen, T., *et al.* Polygenic scores, risk and cardiovascular disease. PHG Foundation. 2019; Available from: <https://www.phgfoundation.org/report/polygenic-scores-cardiovascular>
211. Tornio, A., Backman, J.T. Cytochrome P450 in Pharmacogenetics: An Update. *Adv Pharmacol.* 2018. 83: pp. 3-32.
212. Pirmohamed, M., Burnside, G., Eriksson, N., *et al.* A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med.* 2013. 369(24): pp. 2294-303.

213. Kimpton, J. E., Carey, I. M., Threapleton, C. J. D., *et al.* Longitudinal exposure of English primary care patients to pharmacogenomic drugs: An analysis to inform design of pre-emptive pharmacogenomic testing. *Br J Clin Pharmacol.* 2019. 85(12): pp. 2734-2746.
214. Vital statistics in the UK: births, deaths and marriages. Office for National Statistics. 2019; Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/vitalstatisticspopulationandhealthreferencetables>
215. Goldenberg, A. J., Dodson, D. S., Davis, M. M., *et al.* Parents' interest in whole-genome sequencing of newborns. *Genet Med.* 2014. 16(1): pp. 78-84.

