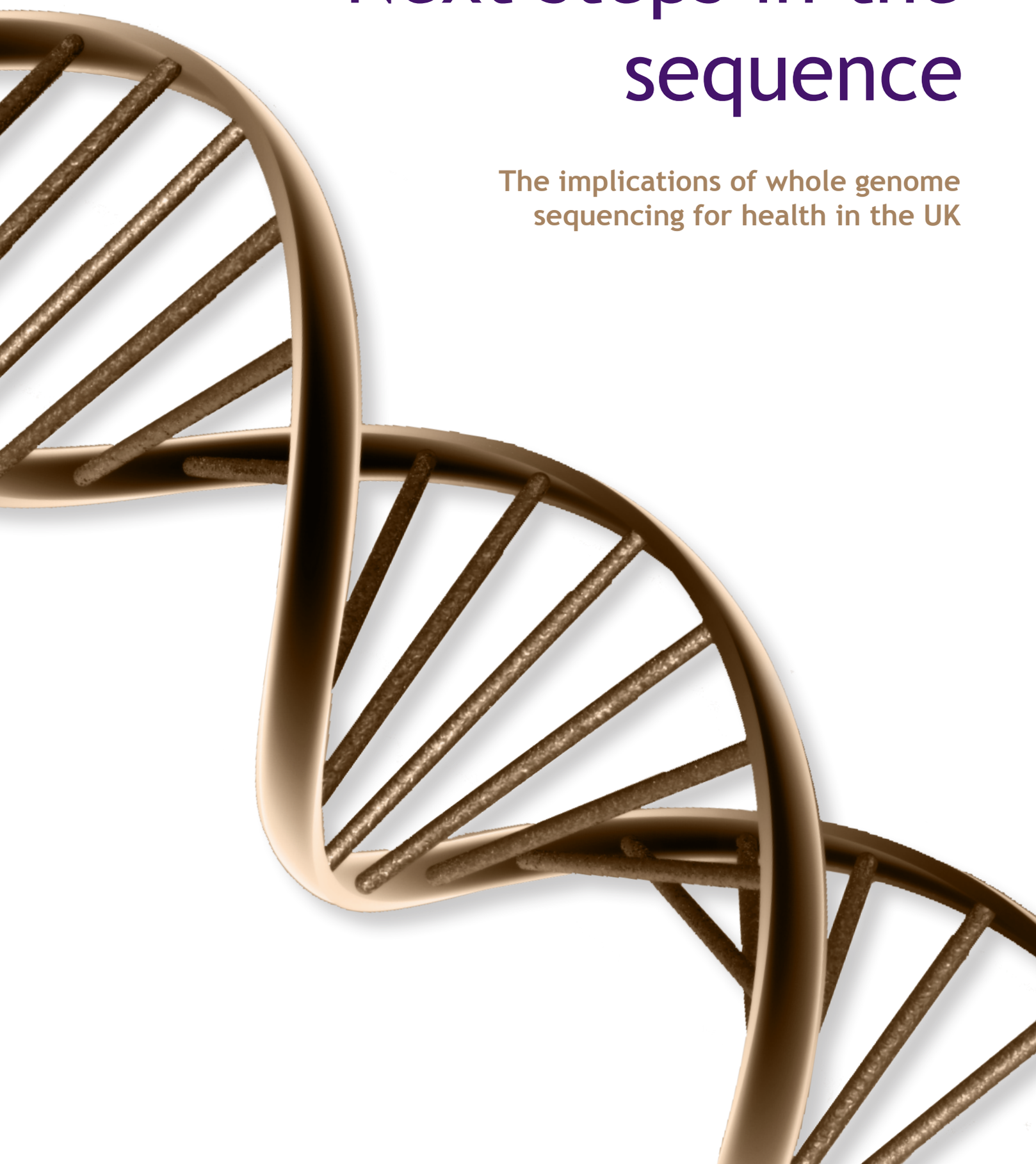
phg foundation
making science work for health

Next steps in the sequence

The implications of whole genome sequencing for health in the UK



PHG Foundation project team and authors

Dr Caroline Wright*	*Programme Associate*
Dr Hilary Burton	*Director*
Ms Alison Hall	*Project Manager (Law)*
Dr Sowmiya Moorthie	*Project Coordinator*
Dr Anna Pokorska-Bocci‡	*Project Coordinator*
Dr Gurdeep Sagoo	*Epidemiologist*
Dr Simon Sanderson	*Associate*
Dr Rosalind Skinner	*Senior Fellow*

‡ *Corresponding author:* anna.pokorska-bocci@phgfoundation.org

* *Lead author*

Disclaimer

This report is accurate as of September 2011, but readers should be aware that genome sequencing technology and applications will continue to develop rapidly from this point.

This report is available from

www.phgfoundation.org

Published by PHG Foundation
2 Worts Causeway
Cambridge
CB1 8RN
UK
Tel: +44 (0)1223 740200
Fax: +44 (0)1223 740892

October 2011

ISBN 978-1-907198-08-3

Cover photo: http://www.sxc.hu/photo/914335

Acknowledgements

Dr Tom Dent contributed the section on risk prediction in the *Other Applications* chapter. Dr Richard Fordham contributed to the *Health Economics* chapter. Professor Anneke Lucassen, Dr Kathleen Liddell and Professor Jenny Hewison contributed to the *Ethical, Legal and Social Implications* chapter. Dr Ron Zimmern contributed to the *Policy Analysis* and *Recommendations* chapters.

The PHG Foundation is grateful for the expert guidance and advice provided by the external Steering Group, in particular to Dr Jo Whittaker, Mr Chris Mattocks and Dr Paul Flicek for their extensive comments on the manuscript. The project team would also like to thank Dr Sarah Wordsworth for her initial advice on the health economics content, and all the participants of the two invited expert workshops (see *Appendices*), whose valuable contributions helped shape the recommendations within this report.

Funding

This project was funded by the PHG Foundation, with additional financial support from Illumina. The second expert workshop for the project was supported by the University of Cambridge Centre for Science and Policy (CSaP) and the Wellcome Trust. The contributions of these supporters are gratefully acknowledged.

Contents

Executive summary

Context and process

The rapid development of new high-throughput and massively parallel DNA sequencing technologies has substantially reduced both the cost and the time required to sequence an entire human genome. It may soon be easier and cheaper to sequence an entire genome than to sequence a single gene or genotype a series of known mutations. Whole genome sequencing, together with our evolving knowledge of genes and disease, is likely to change the current practice of medicine and public health by facilitating more accurate, sophisticated and cost-effective genetic testing.

The development of genomic medicine was considered in a House of Lords Report in 2009, which identified the need for a strategic vision for implementation of genomic technologies within the National Health Service (NHS) to maximise both their health and cost benefits. The PHG Foundation – an organisation dedicated to the evaluation of genome-based technologies for the benefit of health and healthcare – formulated a response to that report, and has subsequently undertaken a year-long project to evaluate the implications of whole genome sequencing for health services and develop recommendations that will facilitate its adoption into the NHS to improve patient care. During the preparation of this report, the authors were guided by the following key questions:

a) What is the role of new DNA sequencing technologies in medicine and population health, in the short to medium term, and how will they alter or augment routine clinical practice?
b) What new ethical, legal, social and economic issues are raised by whole genome sequencing?
c) What are the implications of whole genome sequencing for diagnostic services within the UK and health services more generally?
d) What operational barriers exist to adoption within health services and how should they be tackled?

In this report, we have brought together a diverse range of relevant information and stakeholder opinions to produce a comprehensive overview of the clinical impact of sequencing whole human genomes in the short to medium term. Although the scope of this project is large, we believe that it is crucial to consider the numerous different aspects together in order to make informed recommendations about the possible implementation of such a wide-ranging and potentially disruptive technology. To the best of our knowledge, no other group has produced such a comprehensive overview of the *status quo* and subsequently developed realistic future scenarios for medical human genome sequencing.

The first ten chapters of this report present background research undertaken throughout 2010-11 by the PHG Foundation and supervised by an expert steering group. The chapters cover a broad range of topics relevant to the implementation of whole genome sequencing:

1) Project methodology and terms of reference
2) Introduction to the human genome
3) DNA sequencing technologies
4) Informatics and analysis
5) Applications of whole genome sequencing in inherited and heritable diseases
6) Applications of whole genome sequencing in cancer
7) Other health applications (specifically risk prediction, pharmacogenetics and tissue typing)
8) Ethical, legal and social issues
9) Health economics
10) Laboratory service delivery models

The final two chapters present an analysis of discussions at two invited stakeholder workshops, held in March and June 2011, in which we developed scenarios and recommendations for implementation of whole genome sequencing within the UK.

Summary of key findings

Whole genome sequencing (WGS) has novel implications beyond traditional genetic testing because of the sheer volume and complexity of data generated. Each individual genome contains around 3-4 million variants that are not present in the reference human genome sequence. Only a small number of these variants will have important clinical consequences, and although some may also have implications for family relatedness, the vast majority will be of uncertain or no clinical or personal significance. In addition, cancer genomes have numerous additional somatic alterations superimposed on the patient's inherited genome. These acquired changes constitute the 'genetic fingerprint' of each cancer type in an individual patient, but again, only a small number will be clinically significant. Understanding the health impact of individual genomic variants presents a considerable challenge for analysis, interpretation and management of the data.

Although most WGS to date has been carried out in a research setting (or through consumer genomics companies), here we focus specifically on service delivery within the NHS for the purpose of improving patient diagnosis, clinical outcomes or disease prevention. In this scenario, it is useful to make a distinction between an *assay* (sequencing an individual's genome) and a *test* (analysing any portion of that genome sequence for a specific purpose). The former contains no medical information *per se,* but is essentially a translation of the genetic code from DNA into a string of data stored on a computer; next generation sequencing (NGS) technologies have radically reduced both the cost and time required to accomplish this process. In contrast, the latter is an interpretive step which requires purposeful analysis of the genome sequence, directed by clinical symptoms/phenotypes or disease risk and informed by prior knowledge of the likely pathogenicity of genomic variants. Although the cost and time to *sequence* a human genome has decreased dramatically, the opposite has occurred for *analysis* and interpretation, which is now the major bottleneck for wider application of WGS. Substantial uncertainty still exists over the clinical or phenotypic impact of the majority of genomic variants, making interpretation of test results extremely difficult against a background of constantly evolving scientific understanding. However, interpretation can be substantially simplified by using a directed analysis approach to target just relevant variants, which will require both new bioinformatics tools and a robust evidence base.

Within an individual (or familial) diagnostic setting, the analysis of WGS data will necessarily be undertaken in a directed manner in order to maximise sensitivity (*i.e.* diagnostic yield) and avoid overwhelming clinicians and patients with irrelevant or uninterpretable information. In practice, this involves confining the analysis to answering a specific clinical question, using a process of targeting, prioritising and selecting variants based on an understanding of pathogenicity and frequency. Where a specific diagnosis is suspected, only known disease-associated variants in specific genes with appropriate levels of evidence need to be selected for interpretation; here, the analysis is essentially a closed-ended test and can be evaluated in the same way as any other diagnostic test. For example, for many genetically heterogeneous inherited diseases, where a single mutation in any one of a number of genes may cause the disorder, multigene testing using next generation sequencing technology is already more cost effective than standard techniques and has a higher diagnostic yield. Where the molecular diagnosis is unknown, although genome-wide investigation will be required, only novel or rare variants need to be selected for interpretation; here, the analysis could be thought of as an open-ended test, which has implications for evaluation due to the absence of a single gold standard diagnostic test for comparison. While it is likely that the former will increasingly by used by numerous mainstream medical specialties, we anticipate that the latter will remain within the realm of clinical genetics.

Within oncology, WGS offers a unique opportunity to better understand the somatic mutations responsible for cancer and to improve its diagnosis, management and prognosis, especially in common, solid tumours (such as breast, lung and colorectal cancer) which to date have been largely intractable. The main benefits offered by next generation sequencing technologies in cancer management are: detailed molecular (rather than morphological) diagnosis and classification of cancer types in individual patients; individualisation of cancer prognosis,

treatment, and disease monitoring to an extent not previously possible; and streamlining of clinical and diagnostic pathways. Again, a targeted analysis strategy will be needed to quickly identify clinically important variants in an individual's cancer genome. Once a comprehensive catalogue of cancer-causing mutations has been compiled, and a robust evidence base of the clinical validity and utility of somatic genomic testing developed, this application could offer substantial health benefits for the majority of the population at some point in their lives.

In either situation, by using a clinically targeted analytical strategy, the majority of results will be pertinent to the disease, and incidental findings that are not relevant to the clinical question will be minimised. Although an individual's genome may contain other variants of medical significance, only those that are co-incident with the potentially pertinent results will be unavoidably uncovered. There may be good arguments for directing genomic analysis independent of phenotype, at specific variants associated with known, highly predictive variants with treatable or preventable outcomes, including carrier status for recessive disease. These extended analyses could have an impact on both individual and population health. However, deliberately and pre-emptively analysing the genome for different conditions constitutes *opportunistic screening*, and the results should not be considered as incidental findings of clinically directed analyses. Such health analyses could be undertaken on a population basis so as to maximise specificity (*i.e.* minimise false positives) and therefore could only be considered for a small, agreed list of variants for which the evidence of clinical utility is uncontroversial.

We do not believe that the NHS has an explicit duty of care to screen an individual's genome or offer tests beyond evidence-based analyses relating to the specific clinical or population health purpose(s) consented to by the patient. Moreover, it is advisable to minimise incidental findings where possible; health care professionals should not have an obligation to feedback findings that do not relate to the clinical question, except in cases where they are unavoidably discovered and have high predictive value. It follows that the NHS does not have an obligation to provide patients with their raw genome sequence data for further analysis outside of the NHS. We make no judgement here about whether the individual should be able to purchase and analyse their genome sequence independently; however, if this course of action is pursued, the NHS should provide follow-up advice and care only when additional findings are considered to be of significant clinical relevance in that individual.

The clinically-directed genomic analysis approach outlined here presents a pragmatic and realistic method for integrating whole genome data into medical practice. Nonetheless, the application of WGS has major implications for both human and IT resources within the NHS. It is dependent on robust, extensive and transparent databases of population genome variation and genotype-phenotype correlations, in addition to standardised, purpose-built algorithms for selecting variants of likely relevance to the clinical question. It is not dependent on storing genome sequences for individual medical use or future genome screening, about which there is still extensive debate, but on storing aggregated anonymised data to aid and improve interpretation of novel variants. Crucially, health care professionals, patients and the public must be clear about the purpose and limitations of whole genome analysis services offered by the NHS.

Recommendations for the use of WGS within the NHS:

(1) NHS use of genome sequencing

We recommend that next-generation sequencing technology should be implemented within the UK NHS in the short to medium term for applications where it offers clear clinical or cost benefits over existing tests – specifically, for the diagnosis of diseases with a strong heritable component and the management of cancer.

(2) Clinically targeted analysis

The analytical approach for clinical interrogation of genome-wide sequence data should be clinically directed, such that only variants of relevance to the specific condition are analysed and shared with patients. At this time, we do not recommend interrogating genomic data more extensively for preventive purposes in the absence of a clinical indication.

(3) Biomedical informatics

The NHS should urgently seek to develop clinical bioinformatics expertise and infrastructure, to ensure the availability of sufficient technical support to allow clinical interrogation of genomic sequence data. This may best be achieved through the establishment of a National Biomedical Informatics Institute, in addition to employing bioinformaticians embedded in local clinical services.

(4) Developing the evidence base

An evidence base is needed to allow clinical interpretation of genome-wide sequence data, with standardised databases of normal and pathogenic genomic variation at its core, and linked analytical tools to facilitate clinical use. Construction and maintenance of this evidence base should be an urgent and on-going task of a National Biomedical Informatics Institute.

(5) Policy development

Policy research is needed to define the evolving relationship between the health service and patient, their respective rights and responsibilities in the context of genome-wide analysis, and to develop professional guidance for clinical use of WGS.

(6) Competences and best practice guidelines

Competences and best practice guidelines should be developed for health care professionals to facilitate the responsible and equitable translation of WGS into the NHS.

(7) Service provision

A modular approach to service provision should be taken, with a small number of sequencing laboratories (or providers) acting as regional hubs to provide national coverage to the NHS, in addition to maintaining local expertise for interpretation.

(8) Health economics

Outcomes evidence and health economic modelling of the impact of genomic analyses within the NHS is urgently needed to identify costs and savings both for diagnostic applications and in order to identify where it can be used to stratify interventions.

(9) Commissioning

Rational, clear and transparent commissioning pathways need to be developed and agreed between all relevant stakeholders to enable NGS technologies to be accessed and delivered effectively and equitably.

(10) Genomic screening

Policy research is needed to consider under what circumstances wider screening of the genome might be offered, how relevant health policy should be developed and what issues might arise.

Glossary of acronyms

ACC	Association of Clinical Cytogenetics
ACCE	Analytical validity, Clinical validity, Clinical utility, Ethical legal & social issues
ADR	Adverse Drug Reaction
arrayCGH	Array Comparative Genomic Hybridisation
CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
cDNA	Complementary DNA
cffDNA	cell-free fetal DNA
CHD	Coronary Heart Disease
CMA	Cost-Minimisation Analysis
CMGS	Clinical Molecular Genetics Society
CML	Chronic Myeloid Leukaemia
CNV	Copy Number Variation
CPA	Clinical Pathology Accreditation
CVS	Chorionic Villus Sampling
CUA	Cost-Utility Analysis
CUP	Carcinoma of Unknown Primary
DCE	Discrete Choice Experiments
dNTP	DeoxyriboNucleotide Triphosphates
ddNTP	DiDeoxyriboNucleotide Triphosphates
DME	Drug Metabolising Enzymes
DTC	Direct-To-Consumer
EBI	European Bioinformatics Institute
FISH	Fluorescence *In Situ* Hybridisation
GenCAG	Genetic Commissioning Advisory Group
GINA	Genetic Information and Nondiscrimination Act
GWAS	Genome-Wide Association Study
HGC	Human Genetics Commission
HGMD	Human Gene Mutation Database
HGP	Human Genome Project
HGSG	Human Genomics Strategy Group
HLA	Human Leucocyte Antigen
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
ICC	Inherited Cardiovascular Condition
ICGC	International Cancer Genome Consortium

IF	Incidental Finding
IP	Intellectual Property
IVDD	*In Vitro* Diagnostics Directive
MCDA	Multi-Criteria Decision Analysis
MHRA	Medicines and Healthcare products Regulatory Agency
MLPA	Multiplex Ligation-dependant Probe Amplification
MRC	Medical Research Council
NGS	Next Generation Sequencing
NGRL	National Genetics Reference Laboratory
NHGRI	National Human Genome Research Institute
NICE	National Institute for Clinical Excellence
NIPD	Non-Invasive Prenatal Diagnosis
NSC	National Screening Committee
OECD	Organisation for Economic Co-operation and Development
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase Chain Reaction
PGD	Preimplantation Genetic Diagnosis
PSA	Probabilistic Sensitivity Analysis
QA	Quality Assurance
QALY	Quality-Adjusted Life Year
RCPath	Royal College of Pathologists
RCT	Randomised Control Trial
RGS	Regional Genetic Services
SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedure
TS	Tumour Suppressor
UKGTN	UK Genetic Testing Network
WGS	Whole Genome Sequencing
WTP	Willingness-To-Pay

Glossary of genetic terms

Allele	Variant form of the same gene
Base-pair	Pair of complementary nucleotides
CNV	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.
Diploid	Chromosome complement of normal body cells, which have two copies of each chromosome (one inherited from the mother and one from the father). The size of the diploid human genome is around 6 billion bases arranged on 23 pairs of chromosomes.
DNA methylation	Chemical modification of the DNA molecule which does not change the genetic sequence but results in heritable gene silencing
Epigenetic	Heritable influence on gene activity that does not involve changes to the DNA sequence itself
Exome	All the exons in the genome (around 1-2% of the human genome)
Exon	The region of the gene that codes for protein
Genome-wide association study	A study in which specific genetic markers (usually SNPs) across the entire genome of multiple people are genotyped in **(GWAS)** order to find genetic variations associated with disease. By comparing DNA samples from a group of patients who share a particular disease with those who do not, genome-wide association studies aim to pinpoint the genetic differences that correlate with and may therefore play a causative role in that disease
Genotype	Specific genetic constitution of an individual
Germ-line	The sex cells (sperm and egg), which transmit genetic information from one generation to the next
Haploid	Chromosome complement of sex cells (*i.e.* sperm or egg cells), which have only one copy of each chromosome, and therefore have half the number of chromosomes found in other cells of the body. The size of the haploid human genome is around 3 billion bases arranged on 23 chromosomes.
Haplotype	A specific set of alleles that are physically located close to each other on the same chromosome and are commonly inherited together
Heterozygous	Two different alleles, one on each of a pair of chromosomes, at a particular position in the genome of an individual
Homozygous	Two identical alleles, one on each of a pair of chromosomes, at a particular position in the genome of an individual
Locus	The location of a gene or a marker on a chromosome
Mutation	Relatively rare change in the DNA sequence from the normal sequence

Nucleotide	Molecular unit from which DNA is made, consisting of a nucleobase: adenine (A), guanine (G), cytosine (C), or thymine (T), a sugar molecule and one to three phosphate group
Oncogene	A gene that has a potential to cause cancer
PCR	Polymerase chain reaction; molecular biology technique in which a fragment of DNA with a specific sequence is copied or amplified exponentially
Phenotype	The observable traits of an organism
Point mutation	A DNA sequence variation that involves substitution of a single base A, C, G or T (also see SNP)
Polymorphism	Common variation in a region of DNA sequence
SNP	Single nucleotide polymorphism; common variation in a single base at a particular position in the genome
Tumour suppressor	A gene that protects the cell from cancer

Introduction

1.1 Motivation for the project

The development of new high-throughput and massively parallel DNA sequencing technologies has radically reduced both the cost and the time required to sequence an entire human genome. The Human Genome Project took around 13 years to sequence the first human genome and cost several billion dollars; today, the same process can be completed within weeks for a few thousand dollars using next generation sequencing technologies (NGS). Just a few years from now, whole genome sequencing (WGS) will be possible in a matter of days, for less than the current cost of sequencing a small portion of the genome. In addition, using the same technologies, gene expression profiling and epigenetic analyses are becoming simpler and cheaper.

The development of next generation sequencing technologies has led to an enormous increase in the volume of sequence data generated. WGS is likely to have a major impact on the practice of medicine and population health through the ability to facilitate a more accurate, flexible and cost-effective means of genetic testing. It may soon be cheaper to sequence an entire genome than sequence a single gene, or genotype a series of known mutations. WGS has several important clinical applications in the short to medium term: improved diagnosis and management of diseases with a strong heritable component, and personalised diagnosis and stratified treatment of cancer through tumour profiling. In the longer term, it may have many other applications including tissue matching, risk prediction and pharmacogenetics. However, although there is an opportunity to substantially improve health through the application of WGS, there are numerous evaluation steps and ethical issues that must be considered before it can be responsibly and effectively translated into the NHS.

The impact that new genomic technologies could have on healthcare was recognised in the House of Lords Genomic Medicine Report in 2009, along with the need to develop a strategic vision for implementation within the NHS. The potential benefits are many, requiring a focused approach on clinically actionable and appropriate health applications of NGS to maximise both the health and cost benefits of this technology. The PHG Foundation – an organisation dedicated to the application and evaluation of genome-based technologies for the benefit of health – therefore undertook to write a Report analysing the implications of NGS and WGS for medicine and population health, to be presented to the UK Human Genomics Strategy Group (HGSG).

1.2 Project scope and aims

The aim, scope and objectives of the project were set out as follows:

Aim

To evaluate the implications of next generation sequencing technologies and whole genome sequencing for health services and develop recommendations that will facilitate their adoption into the NHS to improve patient care.

Scope

1. To review existing and emerging 2nd and 3rd generation sequencing technologies, including different methods of targeted and bioinformatics analysis, and their clinical applications to date

2. To analyse issues relating to the development, maintenance and validation of appropriate informatics platforms for storage, access, annotation and clinical interpretation of genomic data

3. To evaluate the utility and impact of NGS technologies for the prediction, diagnosis and management of diseases with a strong heritable component, and the implications for pathology services, clinical genetics and other medical specialties

4. To evaluate the utility and impact of NGS technologies for the diagnosis, monitoring and treatment of cancer, including the potential for personalised and stratified medicine, and the implications for pathology services, oncology and other medical specialties

5. To outline service models and workflow pathways describing how NGS technologies might be used in clinical practice, and their economic implications within the UK NHS

6. To explore the economic, ethical, legal and social implications of implementing widespread WGS for specific medical applications

7. To consider the possible impact of private companies offering WGS and analysis services to healthcare providers and on a direct-to-consumer basis

Although the scope of this project is broad, we believe that it is necessary to consider all these factors together in order to make informed recommendations about the possible implementation of such a wide-ranging and potentially disruptive technology. Despite the rapid scientific progress, to the best of our knowledge, no other group has brought all these areas together to produce an overview of the current *status quo* and develop realistic future scenarios for medical genome sequencing.

1.3 Method of operation

The work was directed by a small steering group (Appendix 1), and undertaken by PHG Foundation (with external advisors as appropriate). Stakeholder workshops of invited experts were held in March and June 2011, to inform the policy analysis and recommendations in this report. Workshop participants are listed in Appendix 2 and 3. The work was carried out in 2010-11.

1.4 The report

The report is set out in twelve chapters; Chapters 2-10 present background research across a broad range of topics including the technology, its applications in inherited and heritable diseases and cancer, ethical, legal and social issues, health economics and laboratory service provision. The final two chapters present an analysis of discussion at two invited stakeholder workshops, in which scenarios and recommendations for implementation were developed.

2 The human genome

2.1 Introduction to the human genome

The human genome is stored in 23 chromosome pairs with the entire diploid genome consisting of around six billion base pairs of DNA. It was initially sequenced as part of the Human Genome Project, and since then technological advances have made it possible to determine the genome sequences of a large number of individuals, to date a number of human genomes have been reported[1-9] along with those of cancer cells[10;11]. Technical progress and the availability of reference sequences have made genome assembly and interpretation much faster. This Chapter aims to give a broad overview of how a genome is pieced together, techniques for studying genes/genomes and information that is obtained from sequencing and how this can be interpreted in terms of health and disease.

2.2 Constructing a genome

2.2.1 The Human Genome Project

The Human Genome Project (HGP) was an international effort started in 1990 to sequence all three billion base pairs of the haploid human genome. The culmination of this work was the publication of a draft sequence in 2001[12]; this was compiled using DNA from a number of anonymous individuals of mainly European ancestry. A parallel effort by Celera Genomics was also published at the same time[13]. A high-quality reference sequence became available in 2004, and work is on-going by the Genome Reference Consortium to improve the quality and coverage of this reference sequence. Updated assemblies of the reference sequence are regularly released; the current reference sequence (GRCh37) covers much of the euchromatic genome (lightly packed gene dense region of DNA) and is >99.99% accurate; missing regions include centromeric and telomeric regions as well as several loci that contain members of multigene families. The presence of a high degree of repetitive sequence makes these regions difficult to sequence, especially with current methods of whole genome assembly.

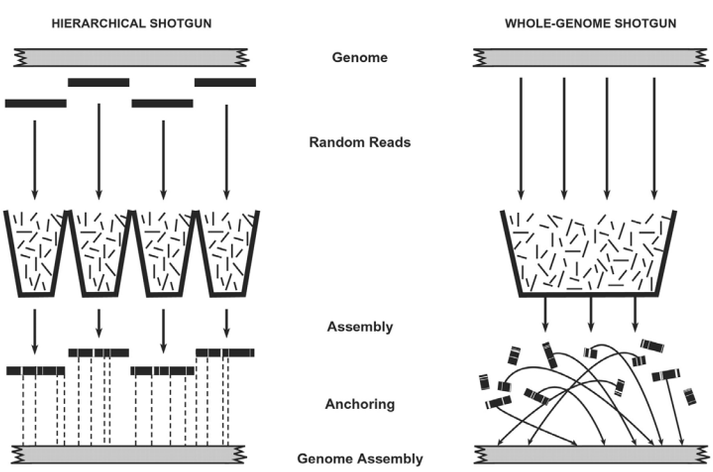
2.2.2 Hierarchical versus shotgun assembly

The human genome is too large to be sequenced or analysed intact, so the approach taken is to sequence small pieces which are brought together to produce a full genome. Whole genome sequencing employs two main strategies, either hierarchical shotgun or simply shotgun sequencing. The former is the top down approach taken by the HGP and involves breaking the genome into large overlapping pieces (150Mb); knowledge of markers is then used to order these fragments and map them to specific chromosomes. Subsequently, each of these larger fragments are randomly broken into pieces small enough for sequencing and reads aligned with identical sequences overlapping to produce a contiguous sequence (known as a contig). These smaller contigs are lined up to produce larger and larger contigs as more and more of the larger pieces are sequenced. Information gained from the initial mapping process is then used to produce an ordered finished genome (Figure 2.1).

A different approach, known simply as shotgun sequencing, was taken by the private sequencing company Celera. This involves randomly breaking the whole genome into a huge number of small pieces, sequencing these pieces separately, and using computing techniques to put the entire sequence together by analysing overlaps between the fragments (Figure 2.1). The advantage of this technique is that it is much quicker and cheaper than the hierarchical approach, due to the fact that the initial mapping process is by-passed. However, it is more prone to errors, due to the large number of random pieces that have to be assembled, difficulties in ordering repeat sequences, and gaps where no fragment has been found.

Although the hierarchical shotgun approach was crucial for *de novo* assembly of the human genome, the availability of a reference genome, combined with powerful computational methods and short-read (next generation) sequencers, has made the shotgun approach the preferred method currently employed in genome resequencing.

Figure 2.1: Hierarchical versus shotgun sequencing (from Waterston *et al*.[14])



2.2.3 Genome content

Following sequencing of the human genome, analysis was carried out to identify functionally important sequences. Approximately 5% of the genome is made up of these sequences which include protein-coding genes, RNA genes and regulatory sequences. Although the total number of genes is unknown, it is thought that there are approximately 21,000 protein-coding genes in the human genome[15] (~1% of genome).The majority of the genome is made up of repetitive DNA sequences whose function is currently unknown, including repeat elements present in heterochromatin (tightly packed DNA).

2.3 Variation in the human genome

Many different types of DNA variation can occur between human genomes which together with environmental factors account for differences between individuals. These variations can range from single nucleotide changes to gain or loss of whole chromosomes, and may be inherited or occur spontaneously (*de novo*). Many DNA variants arise as a result of failure to repair damage or correct replication errors, and as a result of recombination events.

2.3.1 Single base changes

The most common type of variation present in the human genome is the single nucleotide polymorphism (SNP) – also known as a point mutation or single base substitution – where a single nucleotide in the genome differs between individuals (or paired chromosomes), such that there are (at least) two alleles. SNPs occur every 1000bp on average throughout the genome. SNPs that occur in more than 1% of the population are classified as common variants; these are often located in non-coding or regulatory sequences and tend to have little or no phenotypic effects. Numerous common SNPs from many individuals of different ethnic backgrounds were typed by the International HapMap Project, and form a valuable database of common variation and a resource for identifying suitable SNPs for genome-wide association studies. SNPs that occur in <1% of the population are classified as rare variants or mutations, and may have a profound phenotypic effect.

Single bases may also be added (insertions) or removed (deletions), which can have a substantial effect in coding regions where they may result in a 'frameshift' in the downstream genetic code.

2.3.2 Multiple base changes

Genomic variation can also be caused by multiple base changes, the most common of which is in the form of INDELs - INsertions and DELetions (which may be co-localised) that range in size from 1-1000 base pairs. Larger insertions or deletions are referred to as copy number variants (CNVs), and include both common and rare variants, though CNVs greater than 5Mb are rare. Translocation, inversion and duplication events also contribute towards CNVs, which can result in large structural changes affecting many genes that may be visible under microscope.

2.3.3 Somatic variation

The different types of variation described above can be present in either germline or somatic cells. Variation in germline cells is either inherited or occurs *de novo* during meiosis or just after fertilisation, and is present in all (or most) of the individual's cells. Variation that occurs in somatic cells is termed acquired and can arise randomly or as a result of external, environmental factors. Extensive somatic mutations are a hallmark of cancer, and tumour genomes frequently contain a multitude of novel changes including small mutations, large rearrangements and polyploidy (*i.e.* presence of multiple copies of the same chromosome). Mutations can also occur in a single somatic cell of a developing fetus and be passed onto subsequent daughter cells, thereby leading to some cells with a mutation and others without. This phenomenon is known as mosaicism. Aside from cancer, somatic mutations have been also been implicated in neurodegenerative, autoimmune and cardiac disease, as well as in aging[16].

2.3.4 Epigenetic variation

In addition to variation that occurs at the DNA sequence level, epigenetic variation also contributes to heritable differences in gene expression. Epigenetic modifications such as DNA methylation, histone modifications and non-coding RNA influence gene expression and function without altering the DNA sequence itself. Epigenetic modifications are tissue specific, can change over time or in response to environmental stimuli, and may have an important impact on disease[17]. For example, an epigenetic mutation that switches off an otherwise functional tumour suppressor gene could result in the same strong predisposition to cancer as a genetic mutation in the same gene.

2.4 Analysis of genes and genomes

Traditionally, genetic analysis has been split into two separate but related disciplines: molecular genetics (targeted testing of specific small variants) and cytogenetics (whole genome analysis to detect large structural variation). Examples of the former include determining the allele of a single nucleotide in a specific gene, and sequencing part of a specific region to identify any variants present; examples of the latter include karyotyping (*i.e.* visually characterising the chromosomes) and using DNA microarrays to look for regions with a large copy number change. Whole genome sequencing bridges the gap between the two approaches, by providing genome-wide molecular detail revealing both small and large variation.

Most current genome sequencing efforts are resequencing projects, *i.e.* determination of a new sequence relative to the reference sequence. For large or novel regions, resequencing uses the same DNA sequencing chemistry but employs the reference genome to enable fast and accurate assembly of the sequenced fragments. However, in many cases, establishing the identity of specific bases is more important than determining the sequence of multiple consecutive bases. Analysis to determine which variants an individual possesses is known as genotyping and this can be achieved through a number of non-sequencing based methods. Sequencing is the most comprehensive method of genotyping, as it allows determination of the exact sequence of bases, whereas other techniques rely on inferring particular genotypes based on the results obtained from the assay.

2.4.1 Genotyping known variants

Characterising the allele at a particular position in the genome is known as genotyping. Numerous techniques for genotyping known variants are used both diagnostically and for research purposes, and range from those that scan entire genomes to identify regions that can be investigated further, to those that analyse specific sequence changes. Analysis of known sequence changes can be carried out using PCR, multiplex ligation-dependent probe amplification (MLPA), fluorescence *in situ* hybridisation (FISH), DNA microarrays and mass spectrometry among other techniques.

These methods have a number of advantages over sequencing such as cost, specificity, simplicity and their ability to more accurately classify certain variants (*e.g.* repeat sequences). However, each assay has limitations in the size and type of sequence that can be analysed; furthermore, these are indirect methods of genotyping, as they involve inferring a particular genotype based on the assay results. For example, with SNP microarrays, sample DNA will bind to complementary probe sequences on the array surface, and the absence or presence of hybridised DNA allows the presence or absence of particular sequences to be inferred.

2.4.2 Imputation

Genotype information can also be gathered through analysis of a limited number of markers. This is the basis of haplotype marking, where analysis of a few 'tag' or 'marker' SNPs can provide most of the information on the pattern of genetic variation in a chromosomal region. A haplotype is a set of closely linked alleles (*e.g.* SNPs) in a chromosomal region that tend to be inherited together; therefore, information on a few SNPs in this region can provide information on the likely presence of other alleles. This method has been developed further through computational methods allowing *in silico* genotyping, where computational analysis is used to impute missing genotypes[18]. This involves genotyping a set of genetic markers and comparing it to a reference panel of haplotypes that includes information on a larger number of markers such as the HapMap Consortium Database. Missing genotypes can be filled in by identifying matching reference haplotypes. Currently genotype imputation is mainly used to study SNPs and the reference panel used is HapMap data. It is hoped that it will soon become possible to study other types of genetic variants and that many more non-European reference panels will become available.

Genotype imputation techniques can also make whole genome sequencing more cost-effective by reducing the depth of sequencing needed. Using new technologies, target bases are resequenced up to 40 times in order to accurately call bases. The approach taken by the 1000 Genomes Project (Box 2.1) is to resequence each target base 2-4 times and use imputation techniques to combine information across individuals who share particular haplotypes to call polymorphisms. However this approach is only applicable when a large number of individuals are being sequenced, as the

chances of sampling all alleles are greater in such cases. Therefore its use is currently restricted to a research setting and is unlikely ever to be used diagnostically.

2.4.3 Discovery of new genotypes

Genotyping techniques described above tend to analyse sequences that have already been classified and are not always suitable for identification of novel variants, such as those that are rare, specific to certain populations or families ('private' mutations) or come from understudied populations. Instead, mutation scanning is used to look for unknown mutations in specific loci, *e.g.* sequencing the *BRCA* genes to identify rare private mutations in a particular family.

Technological advances in the field of sequencing have facilitated the discovery of novel variants in both research and diagnostic setting. Targeted resequencing has been used to discover novel variants related to conditions such as familial breast cancer[19] and mental retardation[20] among others. However, the discovery of new genotypes and their relation to specific phenotypes is complicated by the enormous amount of variation present in human genomes, and it is extremely difficult to predict the effect of previously unknown variants.

Box 2.1: 1000 Genomes Project

The 1000 Genomes Project is an international collaborative project involving the Wellcome Trust Sanger Institute – WTSI (UK), the Beijing Genomics Institute – BGI (China) and the National Human Genome Research Institute – NHGRI (US). The goal of the project is to create a public reference database of human genetic variation that can be used by researchers. Samples from 2500 anonymous individuals from different populations worldwide will be used to generate this data through a combination of methods: low-coverage whole genome sequencing (4x), array-based genotyping and targeted sequencing of coding regions. The primary goals of the project are threefold: to discover single nucleotide variants at frequencies of 1% or higher in diverse populations; to uncover variants down to frequencies of 0.1–0.5% in functional gene regions; and to reveal structural variants, such as copy number variants, insertions and deletions. The results of a pilot project comparing different strategies for sequencing have already been published[21] and the sequencing of over 1000 genomes was completed in May 2011. This publicly available resource can be used by researchers to both impute genotypes and to identify variants in the regions that they suspect of being associated with disease. By identifying and cataloguing most common genetic variants in the populations studied, the data generated from this project will serve as an invaluable reference for clinical interpretation of genomic variation.

2.4.4 Types of findings from genetic analyses

Regardless of the type of genetic analysis undertaken, some of the results may not relate to the purpose of the analysis. These so-called 'incidental findings' are endemic to all human-subject research and clinical practice, but can be minimised by carefully tailoring the test to answer only the clinical or research question posed. Within genetics, these findings fall into three broad categories:

- Variants with unknown or no clinical significance
- Health-related variants, which may have implications for the individual and/or their family but vary substantially in their certainty, predictive value and the extent to which they are clinically actionable
- Findings that have a personal or legal significance such as ancestry, misattributed parentage or consanguinity

The incidence of such findings will vary depending on the breadth of analysis and its nature. Targeted analysis of specific disease-associated genes is likely to minimise such findings, whereas genome-wide analysis is more likely to maximise them due to the broad analysis that is carried out. Analysis of germline DNA is more likely to identify variants that may have a personal, familial or legal significance, whereas this is less likely to be the case when analysing somatic variation.

2.5 Genes and disease

Like many environmental factors, genetic variation is an important determinant of health, and knowledge of how genetic factors contribute to disease is crucial for understanding disease risk and aetiology. The genomes of apparently healthy people as well as those with a genetic disease will contain around 3-4 million sequence variants, which can be broadly classified into three groups:

- Those that have no effect (*i.e.* neutral variation)
- Those that have an effect on the normal phenotype (*e.g.* height, eye colour)
- Pathogenic variants or mutations which either cause or predispose to disease

Determining if a variant is related to disease can be an arduous process, especially when the effect of the variant is subtle. Single gene disorders are usually associated with rare, highly penetrant genetic mutations which affect the function of a gene. Mutations may cause disease by disrupting the coding region of a gene resulting in the production of a malfunctional or truncated protein or no protein at all. Deleting or duplicating gene(s) may also have a marked effect, particularly for dosage-sensitive genes where inappropriate levels and/or forms of protein are produced. The severity of the phenotype relates to both the penetrance and the expressivity of the mutation, which may in turn be influenced by other modifier genes. Thus, even conditions which appear to be monogenic may in fact be influenced by multiple genes.

In addition to rare variants, common variants found in all human populations may contribute to susceptibility to complex polygenic diseases. These variants are often in non-coding regions of the genome and each variant influencing a complex disease has a small (additive or multiplicative) effect on the disease phenotype which is also heavily dependent on the environment. Genetic variation can also influence individuals' response to drugs, either by affecting their rate of absorption or metabolism or by affecting the response of a drug target. One of the aims of personalised medicine is to tailor drugs and their dosage to an individual's genotype.

Ultra high-throughput DNA sequencing strategies are already being used to study human genomes. The potential usefulness of such data for clinical practice is of great interest, and early examples of how whole-genome sequencing (WGS) can be used for diagnosis and choice of therapy have already been presented[22-24]. However, along with the utility of these technologies in aiding specific molecular diagnoses, consideration must also be given to the incidental findings that can arise as a result of DNA analysis. The applications of WGS to diseases with a strong heritable component and cancer are discussed in more detail in later chapters.

3 DNA sequencing technologies

3.1 Introduction to DNA sequencing

Sequencing has been possible since the 1970s when the Sanger method was developed and it has undergone various modifications and developments in the intervening years to facilitate automation and increase throughput. Despite these modifications, the technique is still too laborious and expensive for routinely sequencing whole genomes. More recently, a number of sequencing technologies have been developed (or are in development) that are radically reducing both the cost and time required for sequencing[25-28]. These post-Sanger technologies can be collectively described as next generation sequencing (NGS) technologies and have been developed with the primary aim of allowing whole genomes to be sequenced at a much faster pace. However, this is not their sole function and they can be employed for a number of other purposes such as the analysis of specific genes and other genetic elements including RNA.

Although many NGS platforms rely on the same underlying sequencing chemistry, they have different performance characteristics and costs and require different levels of upstream and downstream processing. It is not yet clear which of the various platforms will be most appropriate for medical diagnostic use, or which will offer the cheapest or most accurate method for whole genome sequencing.

3.2 Sequencing generations

3.2.1 Overview

Over the last three decades, enormous advances have been made in the technologies available for sequencing DNA, allowing increasing automation and higher throughput (see Figure 3.1).

The terminology surrounding newer sequencing technologies can be diverse and confusing so for the purposes of this report we have divided DNA sequencing technologies into three generations[29] (see Figure 3.2): first generation, or Sanger sequencing (from the 1970s); second or 'next generation' massively parallel sequencing (current); and third or 'next-next generation' (future). Despite these somewhat artificial divisions between generations, the underlying principle of massive parallelisation of sequencing reactions allowing enormously higher throughput is the key unifying feature of all post-Sanger DNA sequencing technologies, and the chemistry of many of the technologies works on very similar principles. The main difference between second and third generation technologies is the elimination of the need to amplify molecules prior to detection.

Figure 3.1: Evolution of sequencing technologies (adapted from[30]).

Over this time, the terminology used to describe a sequenced base has changed from a *band* on a gel, to a *peak* on a capillary electrophoresis machine, to intensity on a next generation platform; a fragment of sequenced DNA is known as a *read*.

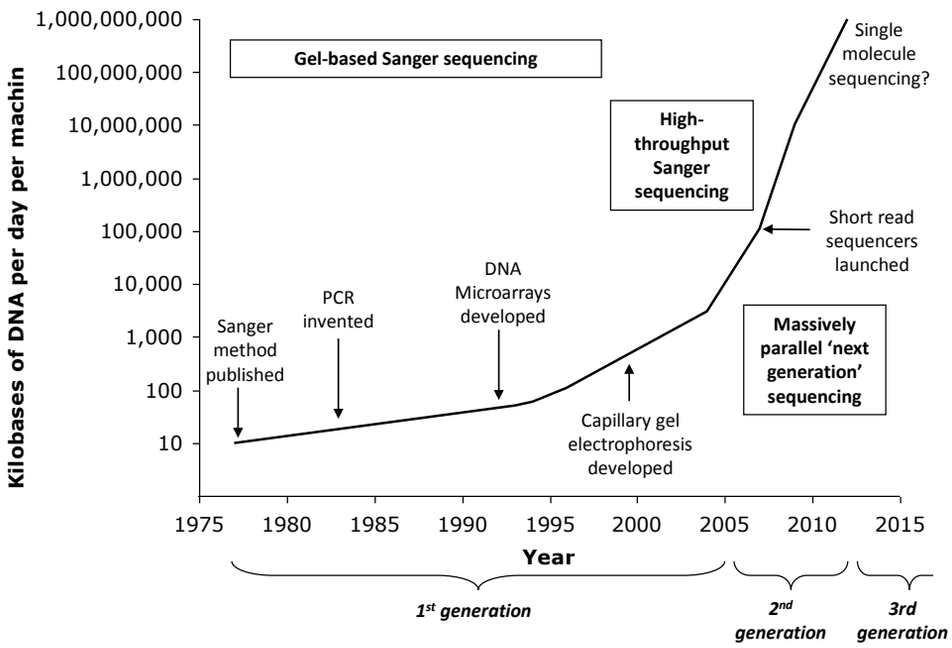
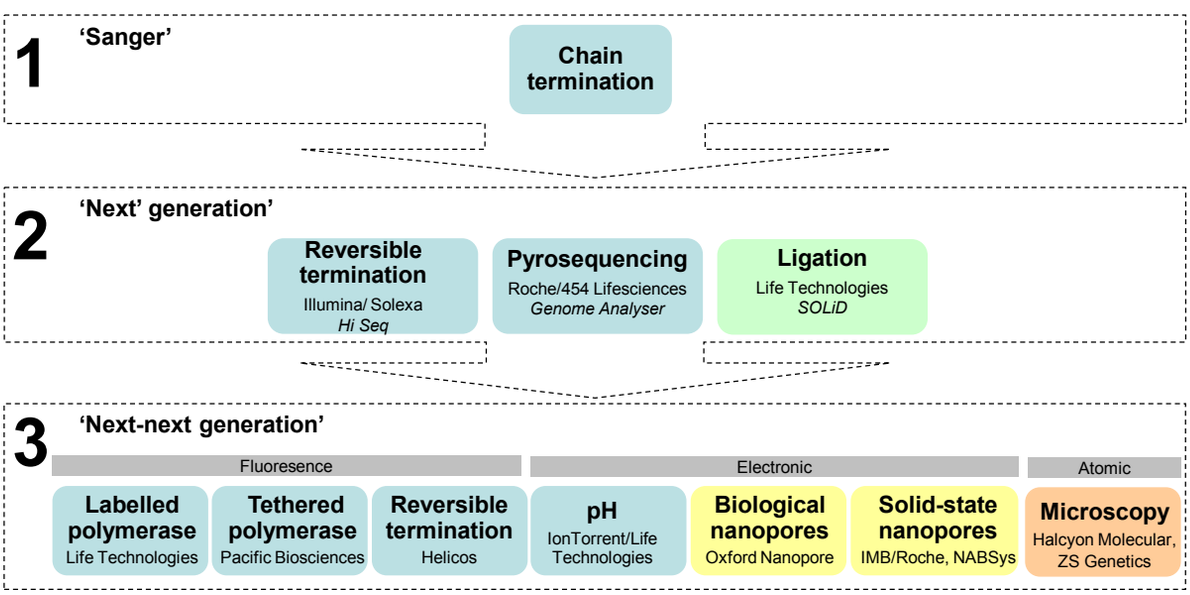


Figure 3.2: Summary of first, second and third DNA sequencing generations

Sequencing chemistries and some of the leading commercial developers are indicated



Key: blue = sequencing by synthesis; green = sequencing by ligation; yellow = nanopore seqencing; orange = sequencing by electron microscope

3.2.2 First generation (Sanger) sequencing

Chain termination is the method invented by Fred Sanger in the 1970s for sequencing DNA by synthesis using DNA polymerase[31], and is commonly referred to as Sanger sequencing[1].

The initial step involves annealing of a primer (a fragment of DNA that acts as a starting point for synthesis) to a single stranded template DNA molecule and the initiation of DNA synthesis by the enzyme DNA polymerase. The reaction mix also contains all four nucleotides (deoxynucleotide triphosphates, dNTPs, corresponding to the bases A, C, G and T) and a lower concentration of four chemically modified terminator nucleotides (dideoxynucleotide triphosphates, ddNTPs) with a fluorescent label attached. As strand elongation proceeds, DNA polymerase incorporates either a complementary dNTP or occasionally a dye labelled ddNTP into the growing strand. Incorporation of a ddNTP terminates synthesis of the chain at that position. Because there are many millions of DNA molecules, and incorporation of a terminating base is rare, this results in the production of multiple different lengths of DNA ending with different fluorescently labelled bases. These fragments can easily be separated by size using electrophoresis, and detected sequentially to allow the sequence of the template molecule to be determined.

Since its inception the process of Sanger sequencing has undergone a number of modifications and improvements to allow automation and increase throughput. Current methods involve the use of four different base-specific fluorescent dyes rather than radioactive labelled nucleotides for chain termination[32] and capillary electrophoresis rather than gel electrophoresis for separation of fragments[33].

Sequencing reactions now typically take place in a 96-well plate, with each well containing a single amplified template DNA molecule. This technique is still widely employed in both research and diagnostic settings and is considered the gold standard in DNA sequencing today. Moreover, the same 'sequencing-by-synthesis' methodology forms the core of numerous second and third generation platforms.

3.2.3 Second generation sequencing

NGS platforms have allowed massive parallelisation of sequencing reactions, due to their ability to simultaneously sequence a large number of DNA molecules. Unlike Sanger sequencing, which takes place in the liquid phase and each sequencing reaction represents a single predefined target, the DNA molecules in second generation platforms are bound to a solid surface upon which they are amplified and sequenced *in situ*. This has allowed many millions of target molecules to be sequenced in parallel, resulting in substantial reductions in cost. However, the length of DNA that can be sequenced in an uninterrupted read on these platforms is substantially lower than using traditional Sanger sequencing.

Current second generation sequencing platforms use clonal amplification of template DNA to generate clusters of identical DNA at high densities on a solid surface, followed by sequencing through stepwise incorporation of either fluorescently labelled nucleotides bases or short pieces of DNA (oligonucleotides). Since around 2005, there are three main massively parallel next generation sequencing platforms which have been commercialised, based on three different sequencing chemistries (see Table 3.1). Their specifications are given in Table 3.2.

Reversible termination

This method is closely based on the Sanger sequencing-by-synthesis method, but uses special fluorescently labelled terminator nucleotides in which the chemical modification can be removed, rendering the chain termination process reversible[2]. It has been commercialised by *Solexa/Illumina* through the *Genome Analyser* and *HiSeq* systems.

Template DNA molecules are immobilised onto a glass surface at high density, on which both amplification and sequencing take place. The tethered fragments are subjected to clonal amplification using surface PCR to create dense clusters of identical DNA templates across the

1 See www.phgfoundation.org/tutorials/dna/5.html for a tutorial on Sanger sequencing.

plate. The sequencing reaction then begins with the addition of a universal primer, DNA polymerase and four reversible nucleotide terminators labelled with different coloured dyes. Incorporation of a complementary nucleotide into the first position results in termination of polymerisation. At this point, unincorporated nucleotides are washed off and the first base on the template strand is identified by colour imaging. The dye and the terminating group are then cleaved chemically, and the process is repeated, allowing further extension of the DNA fragment. Repetition of this cycle allows identification of specific bases along a template DNA strand as they are incorporated, which can be built into a sequence read.

Pyrosequencing

Pyrosequencing is based on a sequencing-by-synthesis technique, but rather than measuring the growing DNA strand, it relies on detecting a by-product of the synthesis reaction to monitor the incorporation of labelled bases[34]. It has been commercialised by *Roche Diagnostics/454 Life Sciences* through the *Genome Sequencer FLX* and *FLX Junior Systems*.

Template DNA molecules are hybridised to special beads upon which both amplification and sequencing takes place; the beads are used in excess to ensure that just a single template molecule binds to one bead. The tethered fragments are subjected to clonal amplification by thermal cycling which takes place in an oil/water emulsion (known as emulsion PCR), to create millions of identical DNA templates on each bead. The beads are then loaded onto a specially fabricated 'PicoTitre Plate', such that each of the millions of tiny wells contains just a single bead.

The sequencing reaction then begins with the addition of a universal primer, DNA polymerase and one of the four nucleotides. If this nucleotide is incorporated by DNA polymerase into the growing DNA strand, it releases an inorganic phosphate ion, which in turn initialises an enzyme cascade reaction (with ATP sulfurylase and luciferase) that results in the release of a flash of light. The emitted light is detected by a camera and is proportional to the number of identical nucleotides added. Excess nucleotide is then washed off and another corresponding to a different base added, and the process is repeated. The addition of nucleotides individually and sequentially allows the correlation of a light signal with a particular nucleotide. The whole plate is imaged following the addition of each nucleotide, thereby allowing the location and intensity of light generated to be recorded. The data are then analysed to build up a sequence read.

Sequencing by ligation

Unlike the previous techniques, this method does not involve DNA synthesis, but instead uses ligation of fluorescently labelled hybridisation probes to deduce the sequence of a template DNA strand two bases at a time[35]. It has been commercialised by *Life Technologies/Applied Biosystems* through the *SOLiD* (***S**equencing by **O**ligonucleotide **Li**gation and **D**etection*) system[2].

Template DNA molecules are hybridised to beads upon which both amplification and sequencing take place. The tethered fragments are subjected to clonal amplification using emulsion PCR to create millions of identical DNA templates on each bead, and the beads are then immobilised at high density on a glass slide.

The first step in the sequencing process is the addition of a universal primer, followed by fluorescently labelled oligonucleotides which act as sequence probes. Each of the probes comprises eight bases, of which the first two bases are used to probe the template sequence, whilst the following six are degenerate (*i.e.* able to pair with any nucleotide on the template strand). After the complementary probe hybridises to the template DNA, it is chemically linked by the enzyme DNA ligase, excess primers washed off and the fluorescence signal recorded. Then, the three final degenerate bases, along with the fluorescent dye, are removed from the bound probes and the process repeated multiple times, after which the newly synthesised strand is denatured and removed from the template. This sequencing round itself is then repeated, but this time the universal primer used is one base shorter, so the sequencing reaction is offset by one base. This 'primer reset' process is repeated for five rounds of sequencing, each time reducing the size of the

[2] An open platform based on sequencing-by-ligation has also been developed by *George Church/Dover Systems*, known as the *Polonator*[35], but is not described further here.

primer by one base, allowing all the bases on the template strand to be interrogated twice.

Counter-intuitively, although there are 16 possible probes defined by combinations of the first two bases, only four coloured dyes are used; thus each colour represents one of four possible two-base permutations for each of the four bases in the first position. For each colour, there is only one representation of each of the four bases at each of the two positions; therefore, if the first base is known, the second can be inferred from the colour. Since the first base in the sequence belongs to the universal primer added initially, the rest of the sequence can be inferred from the raw colour data by applying logical colour space rules.

Table 3.1: Summary of NGS platforms

Platform	**Illumina/Solexa** *GA, HiSeq*	**Roche/454** *GS FLX*	**Life Technologies/ Applied Biosystems** *SOLiD*
Loading	Adaptors on template DNA bind high density primers across surface of slide	Adaptors on template DNA bind primers on beads, one molecule per bead	Adaptors on template DNA bind primers on beads, one molecule per bead
Clonal Amplification	Surface PCR used to generate clusters by bridge amplification	Emulsion PCR used to create clusters on beads	Emulsion PCR used to create clusters on beads
Parallelisation	Random array on flow cell	Beads loaded onto high density plate	Beads immobilised on high density glass slide
Sequencing enzyme	DNA polymerase	DNA polymerase	DNA ligase
Generation of complementary strands **Detection** **Re-initiation**	4 labelled terminator nucleotides added ↓ 4 colours detected ↓ Terminator and dye removed ↺	One of 4 labelled nucleotide added, *n* incorporated, and phosphate released ↓ Flash detected proportional to *n* ↺	16 labelled 8-base oligonucleotides added and hybridised ↓ 4 colours detected ↓ Last 3 bases and label removed ↺

Table 3.2: Summary and specifications of currently available NGS platforms (data from company websites, August 2011)

Platform	Read lengths	Run time	Output (per run)
Illumina *HiSeq1000*	1×35 bp 2×50 bp 2×100 bp	~1.5 days ~4.5 days ~8.5 days	47-52 Gb 135-150 Gb 270-300 Gb
Illumina *HiSeq2000*	1×36 bp 2×50 bp 2×100 bp	~1.5 days ~4.5 days ~8.5 days	95-105 Gb 270-300 Gb 540-600 Gb
Illumina *Genome Analyzer IIx*	1×35 bp 2×50 bp 2×75 bp 2 x 100bp 2x 150bp	~2 days ~5 days ~7 days ~9.5 days ~14 days	10-12 Gb 25-30 Gb 37.5-45 Gb 54-60 85-95
Roche *GS FLX Titanium XL+*	Up to 1000bp	23 hours	700 Mb
Roche *GS FLX Titanium XLR70*	Up to 600 bp	10 hours	450 Mb
SOLiD *5500 System*	Mate-paired: 2 x 60 bp Paired-end: 75 bp x 35 bp Fragment: 75 bp	2 -7 days depending on sample	7-9 Gb/day
SOLiD *5500xl System*			10-15 Gb/day

3.2.4 Third generation sequencing

A large number of companies are involved in developing much faster and higher throughput third or next-next generation DNA sequencing systems. Most of these are focused on sequencing single molecules of DNA, often in real-time; many are based on the principle of sequencing-by-synthesis, though there are also several novel methodologies, such as monitoring the passage of DNA through nanopores. The advantage of single molecule sequencing is that there is no library preparation or amplification steps, therefore biases and errors introduced at this point can be avoided. In addition, improvements are expected to be made in the length of DNA molecule that can be sequenced, resulting in higher throughput and parallelisation, as well as the speed of sequencing by using real-time detection. Moreover, some of the technologies can also simultaneously identify DNA modifications such as methylation.

The third generation sequencing platforms can be divided into categories based on the method they use to detect the DNA sequence:

Fluorescence

Most of the third generation sequencing platforms under development using fluorescence detection are based on the standard sequencing-by-synthesis method. The first single molecule sequencing platform to market was the *Heliscope* from *Helicos Bioscience*[36], launched in 2009, which is based on a similar methodology to that described for the *Solexa/Illumina* second generation platform. The main differences are that the template DNA is not amplified prior to sequencing and the dye labelled nucleotides are added individually and sequentially, as the presence of a dye on the nucleotide itself allows for the addition of only one nucleotide at a time onto the growing strand by the DNA polymerase[37]. Following washing of excess nucleotides and polymerase, the slide

is imaged to identify where a base was incorporated. The dye is then cleaved and the process repeated for each nucleotide, with each cycle extending the DNA strand. The data are then analysed to build up a sequence read.

In contrast, the *SMRT*™ platform developed by *Pacific Biosciences* monitors the addition of four different dye-labelled nucleotides in specialised wells containing an immobilised DNA polymerase[38]. The template DNA is added and forms a complex with the polymerase, and the particular nucleotide incorporated into the growing strand is monitored by laser excitation and detection in the well. The difference between this method and others that use fluorophores is that the dye is attached to the phosphate of the nucleotide rather than the base itself. Thus it is cleaved and released as a natural part of DNA synthesis, resulting in release of the dye without interruption to the sequencing process. The sequencing is therefore both single molecule and real-time.

The single molecule sequencing method developed by *Life Technologies* uses a modified DNA polymerase with a quantum dot (Qdot®) attached[39]. The quantum dot is a tiny nanocrystal that absorbs photons of light, then re-emits photons at a different wavelength, producing a very strong localised fluorescence signal (around 100-fold greater than standard dyes) which enables real-time single molecule detection. Like the second generation platforms, the template DNA is immobilised to the surface of a glass slide, and sequencing is initiated by the addition of DNA polymerase and four differently labelled nucleotides. As bases are incorporated by the modified polymerase, the dyes are energised by photons transferred from the laser-activated quantum dot before being cleaved off, generating a characteristic flash of coloured fluorescence light. The flashes of dye colour and quantum dot signal are detected by a laser for each individual DNA strand and used to determine the sequence. Both the polymerase and newly synthesised DNA strand can be then removed, allowing the immobilised template DNA to be sequenced repeatedly, thereby increasing the number of reads with minimal sample preparation.

There are also numerous other smaller companies developing third generation DNA sequencing platforms based on fluorescence detection, such as *GnuBio,* which is developing a microfluidics device that uses microdroplets as miniature reaction vessels, thus vastly reducing the cost of the reagents.

Electronic

Various third generation DNA sequencing platforms are being developed that monitor changes in electrical current, offering label-free sequencing. The system being developed by *Ion Torrent/ Life Technologies* is again based on the sequencing-by-synthesis method and uses proprietary semi-conductor technology to monitor the synthesis reaction[40]. Unlike other platforms, sequencing is based on monitoring the release of hydrogen ions (H^+), which are another by-product of DNA synthesis. DNA templates are held in specialised wells which are designed as ion sensors, and nucleotides are added sequentially to each well. If a particular nucleotide is incorporated into a growing strand by DNA polymerase the result will be a release of H^+ into solution and a concomitant change in acidity (pH). The change in pH is detected as a voltage shift by sensors and can be related to the number of molecules of a particular base incorporated. Currently this system does not detect single molecules and amplification is required prior to sequencing, but the synthesis reaction is detected in real-time. This technology has been used to sequence bacterial genomes and produce a draft human genome sequence[41].

However, the majority of platforms under development for single molecule DNA sequencing using electronic detection are not based on the sequencing-by-synthesis method, but on an entirely new method using either biological or solid state nanopores. These technologies monitor changes in electrical current following the passage of DNA strands or individual bases through a nanopore, which is a small hole in an electrically insulating membrane that can be used as a single molecule detector. The nanopores themselves can either be simply small holes in an inorganic membrane (solid-state nanopores), such as silicon nitride[42] or graphene[43], or specific channels made from modified natural proteins[44] embedded in a lipid bilayer or synthetic membrane. Nanopore sequencing technologies are based on one of two approaches – either strand sequencing as the DNA strand itself passes through the nanopore, or base sequencing as DNA bases are cut off from the end of the strand then individually and sequentially fed into the nanopore.

A voltage is placed across the membrane to drive the translocation of negatively charged DNA molecules across the pore; the nanopore is the only point at which current can flow across the membrane. As DNA bases pass through the pore, each base blocks the current by a different amount, so changes in electrical current can be monitored and related back to strand composition.

Numerous companies are currently developing nanopore-based DNA sequencing platforms, including *Oxford Nanopore Technologies*, *NABSys*, *base4innovation*, and *IBM/Roche*. Key challenges for this promising technology include fine control of the passage of DNA through the nanopore, accurate reading of DNA bases during this dynamic translocation, and parallelisation[45;46]. Perhaps the most advanced to date is the platform under development by Oxford Nanopore Technologies, which uses a scalable electronics platform to stream and analyse sequence data in real time from nanopores set in array chips. Alternative sequencing chemistries are in development: 'strand' sequencing (where a strand of DNA is passed intact through a nanopore) and 'exonuclease sequencing' (where individual bases are cleaved from a strand). In each case, individual DNA bases are identified by characteristic disruptions in an electronic current as they pass through the nanopore[47;48]. Furthermore, base modifications such as methylation also produce a characteristic current change, so both the sequence and the modification can be read directly and simultaneously[49].

Another novel technique for DNA sequencing is based on detection of the bases at atomic resolution, using transmission electron microscope to visualise strands of DNA directly[50]. The latter method is being developed and commercialised by several companies, including *Halcyon Molecular* and *ZSGenetics*, using heavy metal atoms to label the different bases in the sequence to make them visible by electron microscopy.

3.2.5 Bench-top sequencers

Technological innovations in sequencing have been aimed at increasing throughput leading to the development of the platforms described in this chapter. However, it has been recognised that such platforms currently have greater utility in large sequencing centres rather than smaller clinical laboratories, where throughput required may be lower and large systems unaffordable. This has led many companies to develop smaller sequencing platforms (based on the same chemistry) which are more affordable by smaller laboratories and have a lower throughput (see Table 3.3). Currently, these platforms are not suitable for whole genome or exome sequencing, however, they do have a number of applications such as small genome sequencing (*e.g.* microbial), sequencing of PCR products, SNP discovery and analysis of structural variation. It is unclear at the moment if and when "bench-top" sequencers with the ability to carry out whole genome sequencing will become available, as these may influence laboratory structuring. In the near term it may be that individual laboratories will use bench-top sequencers for specific applications and larger regional hubs provide the whole genome sequencing capabilities (see Chapter 10 for more on laboratory service organisation).

Table 3.3: Summary and specifications of currently available bench top sequencers (data from company websites, August 2011)

Platform	Read lengths	Run time	Output (per run)
Mi-Seq (Illumina)	1×35 bp 2×100 bp 2×150 bp	4 hours 19 hours 27 hours	120 Mb 680 Mb 1 Gb
GS Junior (Roche)	400 bp	10 hours	35 Mb
PGM (Ion Torrent)	100 bp	1-2 Hours	1 Gb

3.3 Sequencing pipeline

The process of taking an *in vitro* sample through to a clinically useful genetic result involves a number of steps both upstream and downstream of the actual sequencing itself, summarised in Figure 3.3. Upstream processes include isolation of DNA and preparation of samples that will subsequently be loaded onto the sequencing instrument, and may include targeting specific genes as appropriate (see Section 3.4); downstream processes mainly relate to the steps involved in converting raw data into clinically relevant information (see Chapter 4 on Bioinformatics). The requirements of both the upstream and downstream process in terms of time, cost and expertise varies substantially between sequencing generations and platforms.

Although the exact details of the upstream stages of sample preparation vary with different sequencing platforms and applications, they involve broadly similar steps[51]. Initially, genomic DNA must be extracted and isolated from the biological sample (*e.g.* blood, saliva, tissue, *etc.*). For most medically relevant sequencing reactions, the starting material is double-stranded DNA in the form of genomic DNA, which must be broken into smaller fragments prior to sequencing. (Note that this step is not required for DNA from PCR reactions or short RNA molecules.) There are two principle methods of DNA fragmentation: mechanical force (including nebulisation and ultrasound) or enzyme digestion. Manufacturers of second generation sequencing technologies may recommend different methods of fragmentation, as different fragmentation methods influence the fragment size distribution; this in turn has consequences for the amount of starting material required.

Following sample preparation, a library is prepared from the fragmented/digested genomic DNA or PCR products. A DNA library is simply a collection of DNA fragments that is representative of the entire DNA sample to be sequenced. Previously, library preparation was carried out using biological vectors, such as cloning into bacterial chromosomes (BAC) or plasmids, but current methods for use with NGS platforms are usually carried out *in vitro* and involve a number of steps.

First, because the fragmentation process can produce DNA molecules with terminal ends that may either be damaged or incompatible with downstream processes, the ends of the DNA strands must be repaired either by filling in or removing protruding ends (blunt ending). This is followed by linking short, synthesised DNA molecules (adaptors) to the ends of the genomic DNA fragments by ligation. These have the dual purpose of acting as primers to initiate subsequent reactions, and tethers to a solid surface to which the DNA template fragment(s) will be subsequently attached in all the current NGS platforms. Finally, size selection is carried out to enrich for the correct template DNA, followed by quality control and amplification steps to ensure that correct amounts of the right template DNA are obtained.

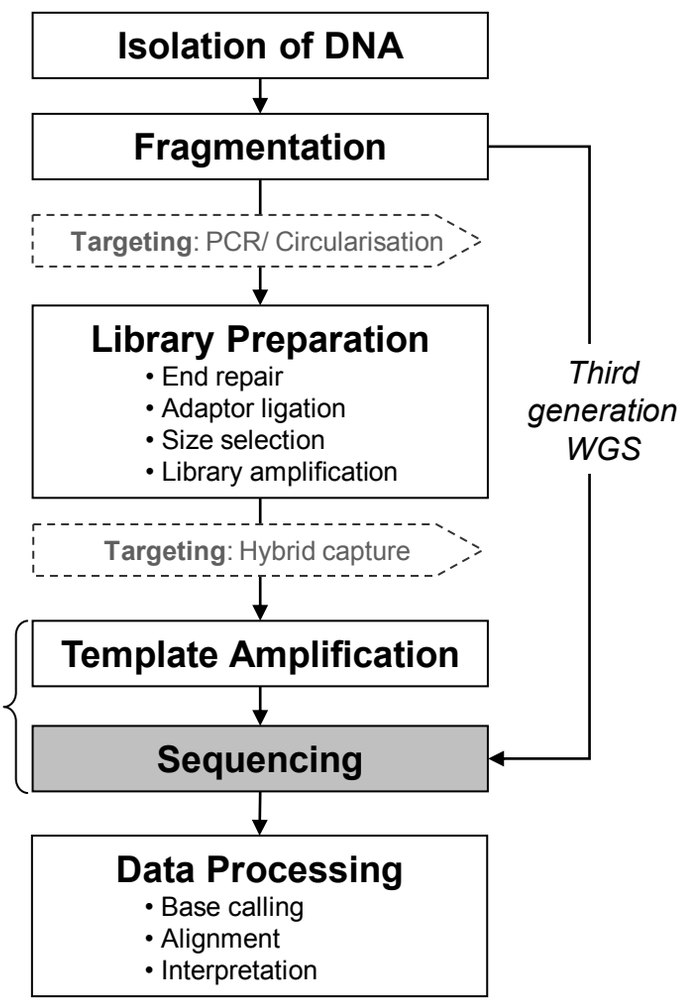
Prior to sequencing using either Sanger or NGS platforms, the template DNA must be amplified in order to produce a large number of identical DNA template fragments to ensure a high signal to noise ratio. Library amplification can either be *in vivo* in the case of bacterially cloned DNA fragments, or *in vitro* as with NGS technologies. Second generation platforms employ a method known as clonal amplification *in situ* in order to increase the number of copies of template DNA.

Following amplification, the sample is sequenced to determine the order of bases on each template fragment, using one of a number of different sequencing chemistries (see Section 3.1.3). Although sequencing a single read to determine the order of the bases is usually only performed in one direction, it is also possible to sequence both ends of a fragment of DNA by modifying the library preparation protocol slightly, thus artificially increasing read length. This method is known as ‘paired-end’ sequencing (or ‘mate-pair’ sequencing for large fragments), and is very useful for alignment and detection of structural variants[26;52]. Analysis of the genome coordinates of the two ends of paired-end reads along with knowledge of the length of DNA between the ends in the sequenced molecule allows a wide range of structural variants to be identified, including deletions, insertions, translocation, and tandem or inverted duplications.

3.4 Targeting methods

Although second generation sequencing machines have enhanced throughput, sequencing whole genomes still requires considerable effort and involves high reagent costs. Moreover, often only a specific subset of genes or portion of the genome is required, particularly within a clinical context. As a consequence, current studies usually employ new sequencing technologies for the analysis of specific regions of interest as opposed to whole genomes. This ranges from the analysis of gene families or regions that are associated with a specific disease or pharmacogenetic effects, to the analysis of all coding regions in the genome (the 'exome'). However, this strategy requires the parallel development of methods that are capable of targeting and enriching for large numbers of specific gene-encoding regions which can subsequently be sequenced. Although this approach has substantial clinical (and also currently financial) advantages, it is a major bottleneck in the sequencing pipeline (see Figure 3.3).

Figure 3.3: Current sequencing pipeline
If required, targeting can be carried out either before or after library preparation depending on the method being used



There are a number of approaches to targeting, including several alternative chemical methods of enriching specific genes prior to sequencing[53;54], as well as computational methods following whole genome sequencing (see Chapter 4 on Bioinformatics). The chemical enrichment methods can be divided into three broad groups: PCR-based, circularisation methods and hybrid capture. Each of these methods is suitable for a different target size, type and sample number. In addition, each has its advantages and disadvantages in relation to performance, ease of use and costs, hence different approaches may be undertaken for different projects or applications[55].

Ideally any method of targeting should allow enrichment of multiple different loci, independent of their size, sequence composition or spatial distribution, and should be amenable to automation so that it can match the sequencing capacity of NGS machines. However, this is currently not the case; different targeting approaches have their own biases (Table 3.4), which relate both to the types of sequences that they are able to capture as well as their ease of use. Because of the repetitive nature of the genome, none of the methods is able to uniformly and efficiently capture all the target regions, and additionally may capture off-target regions. This can lead to errors in the analysis due to non-uniform coverage of targeted regions leading to variable read depth, and off-target capture resulting in mapping errors (see Chapter 4).

Further improvements to these approaches and their automation are still under development and no standardised approach to sequence capture is available yet.

3.4.1 PCR

Polymerase chain reaction (PCR)[3] is a molecular technique developed in the 1980's for exponential amplification of specific DNA sequences, using targeted primers, DNA polymerase and repeated thermocycling to alternately denature and extend DNA[56]. With this approach, DNA regions of interest are amplified following shearing of genomic DNA, and the resulting products are then subjected to library construction.

The most common method of targeting is traditional PCR, which can be used to generate DNA oligonucleotides that are well suited for Sanger sequencing. Although PCR can also be used to generate samples for use with NGS platforms, this would require a large number of reactions to make optimum use of their high-throughput. This has been partially overcome by the development of a number of techniques to multiplex PCR[57;58], including using primers immobilised on beads and high-throughput megaplex PCR using immobilised primers tagged with a universal sequence, thus allowing the products to be subsequently amplified using a single universal primer[59]. *RainDance Technologies* have reported the simultaneous amplification of up to 4000 short DNA sequences in separate microdroplets, each supporting an independent PCR reaction[60].

Table 3.4: Advantages and disadvantages of different chemical targeting approaches[55]

Method	Advantages	Disadvantages
PCR	High sensitivity, specificity, reproducibility and uniformity.	High cost, low throughput, and cannot be used for large regions or a very large number of genes.
Circularisation	Low cost (if many samples), easy to use, high sensitivity and specificity.	Uniformity and sensitivity depends on design of probes. Cannot be used for a very large number of genes.
Hybrid capture	Medium cost, easy to use, high sensitivity and specificity. Can target large sections of DNA and large numbers of genes.	Uniformity and sensitivity depends on design of probes. Array design may be rather inflexible.

3.4.2 Circularisation methods

Circularisation methods are based on targeting DNA using a combination of DNA hybridisation and PCR amplification, and are suitable for targeting small to medium sized regions of interest. A number of slightly varying methods of producing circularised probes have been described[61-66], however, they are all based on similar principles and involve the use of probes to produce circular libraries by hybridising to template genomic DNA.

[3] See www.phgfoundation.org/tutorials/dna/4.html for a tutorial on PCR.

The probe comprises a linker sequence containing universal primer sequences flanked by sequences that are complementary to a specific target sequence on the genomic DNA. When added to sheared genomic DNA, the complementary sequences bind either end of the target sequence forming a circle, with a gap in the middle. The gap is filled by the enzymes DNA polymerase and DNA ligase, resulting in the formation of either a circular probe hybridised to the region of interest, or circular genomic DNA fragments. Treatment of these samples with exonucleases that specifically only digest linear DNA allows for enrichment of circularised DNA containing the target region(s) of interest, which can subsequently be amplified to generate target material for sequencing.

3.4.3 Hybrid capture

The principle of this method of targeting is the same as DNA microarray technology, where specific oligonucleotide probes are used to pull-down sequences of interest in the template genomic DNA. Unwanted DNA is then washed off, and captured material subsequently eluted for sequencing. Numerous custom hybridisation platforms specifically for use prior to next generation sequencing have been developed by microarray companies (such as Agilent, Roche Nimblegen and Illumina), using both conventional arrays for 'on-array' capture and beads for 'in-solution' capture[67-69]. These allow multiple large sections of DNA or even the entire exome to be targeted prior to sequencing[70;71].

3.5 Performance metrics

The performance metrics of current sequencing platforms can be delineated by a number of parameters such as read length, accuracy, run time, genome coverage and cost (see Tables 3.2 and 3.3). However, these parameters are changing rapidly as technical improvements are made, such as increasing automation, modifying the chemistry and improving the detection equipment. Moreover, some of the major companies have launched multiple platforms aimed at different markets, including high and low throughput applications.

A number of factors influence the utility of next generation sequencing technologies for particular applications. These include:

Analytical accuracy, systematic errors and quality of base calls

In addition to amplification errors (which should largely be eliminated by single molecule sequencing), all sequencing methodologies suffer from both random and systematic errors. The raw accuracy of the sequencing process and quality of base calling are critically important factors when applied to medical diagnostics. As a result, a quality score is assigned to each base, which is a representation of the probability that the base is called correctly. Errors can include overcalls (insertions) and undercalls (deletions) as well as miscalls (incorrect base assigned)[72].

Different sequencing technologies are prone to different systematic errors, which influence their utility for different applications; for example, accurately sequencing homopolymeric regions can be difficult using pyrosequencing due to intermediate fluorescence signal intensities resulting from the incorporation of n identical nucleotides. Sanger sequencing has a low (but non-zero) error rate for single calls (around 10^{-4}-10^{-5}, one error per 1,000-10,000 bases), but the accuracy for detecting heterozygous variants is much more difficult to assess. When it comes to detecting low level variants, for example mosaic or somatic mutations, the limit of detection in terms of minor allele representation is only around 20%. Still it remains the gold standard diagnostic test for sequencing.

Current NGS platforms have a somewhat higher raw error rate (around 10^{-2}-10^{-3}, one error per 100-1000 bases, depending on read length), which improves substantially with increasing depth of coverage and the assembly of a consensus sequence. In contrast to Sanger sequencing detection of low level variants is basically limited by the raw error rate and would be substantially below 0.1%.

Read depth, genome coverage and uniformity

The read depth or depth of coverage refers simply to the number of times a base is sequenced in a single run of the machine. The required read depth varies depending upon the specific application and level of certainty required for the result. However, coverage of the genome is non-uniform, due to factors such as repetitive elements, non-uniform targeting and variable GC content that affect both amplification and sequencing efficiency[73], therefore a high average read depth is required to ensure that every base has been covered to sufficient depth for the required application (*e.g.* to detect heterozygocity).

The consensus accuracy of base calling is a combination of the raw accuracy and the depth of coverage, so much lower error rates can be achieved using NGS technologies by increasing read depth. However, for diagnostic purposes, simply reporting the average read depth may be insufficient, as critical (potentially heterozygous) loci may be significantly underrepresented; thus, the actual read depth at each base is required. For resequencing, the assembly is guided by a reference sequence and requires much lower coverage (8–12x) than assembling genomes *de novo* (25–70x)[74].

Read length

Read length (number of bases per read) is an important factor in certain applications, such as sequencing through repetitive regions, identifying genomic rearrangements and getting short range haplotype information. In addition, it makes alignment to the reference sequence a substantially easier task by reducing the number of sequence matches throughout the genome. Sanger sequencing still offers the longest reads at around 800 bases, with the current NGS platforms offering between 35-400 bases per read[25;54], but this is rapidly improving. It is anticipated that the third generation platforms will have substantially longer read lengths; *Oxford Nanopore Technologies*, *Pacific Biosciences* and *Life Technologies* claim that their platforms will have very long read lengths (>1kb), however, the existing *Helicos* single molecule sequencer has very short reads of 30-35 bases.

Throughput, capacity, run time and multiplexing

Factors such as the number of bases of DNA that can be sequenced per run, the number of different samples that can be sequenced simultaneously, and the length of the run itself all have a major impact on the suitability of a particular platform for different laboratories. These factors vary substantially between machines and applications, ranging from 1-14 days per run with a throughput of 1-25 Gb per day, depending upon read length and sequencing protocol (single fragment, paired-end or mate-pair).

Sample multiplexing is also crucial for cost-effective use of NGS, as although NGS has substantially reduced the per-base cost of sequencing, this is only if the capacity of the instrument is effectively used. Often individual experiments do not require the full capacity of a machine; consequently, methods for analysing multiple samples in a single run are important.

Although DNA sequencing machines are unable to differentiate between matching target DNA isolated from different individuals, there are a number of methods that allow multiple samples to be sequenced simultaneously[54]. For targeted sequencing, it is possible to mix multiple different tests so that results from each specific test relate to only one individual patient. In addition, many sequencing platforms allow physical separation between samples, by having multiple separate channels. Finally, DNA ‘barcode tags’ (multiplex identifiers, MIDs) are under development, which are added to the ends of DNA fragments during library preparation, and provide a unique DNA signature to mark and track individual patients[75;76].

Reagent and instrument cost

Reagent cost for sequencing has plummeted over the last decade, from a cost of around $500/Mb for Sanger sequencing reagents, to less than $0.50/Mb for reagents on the newest NGS platforms[77]. However, the sequencing machines themselves are often fairly expensive (ranging from US$0.2-1 million each), and in addition, upstream and downstream costs of sample preparation and data analysis can be sizeable.

4 Informatics

4.1 Informatics pipeline

The enormous volume of data generated by massively parallel sequencing means that the bulk of the workload is shifted downstream from the laboratory towards data analysis processes. Analysis of WGS data requires substantial computational power, purpose-built informatics tools and accurate databases of genomic variation. This creates enormous challenges for using NGS technologies within the NHS, particularly for generating, standardising, analysing, validating and interpreting genomic information.

The sequencing reaction is only the beginning of the process of converting a sample of DNA into meaningful genetic information. The next step of data collection and analysis involves extensive use of various computational methods for converting raw data into sequence information, and the application of bioinformatics techniques for the interpretation of that sequence. Although this basic pipeline has always been an important part of analysing genetic test results, the enormous amount of data generated by massively parallel NGS technologies has shifted the workload substantially away from upstream sample preparation to downstream analysis processes.

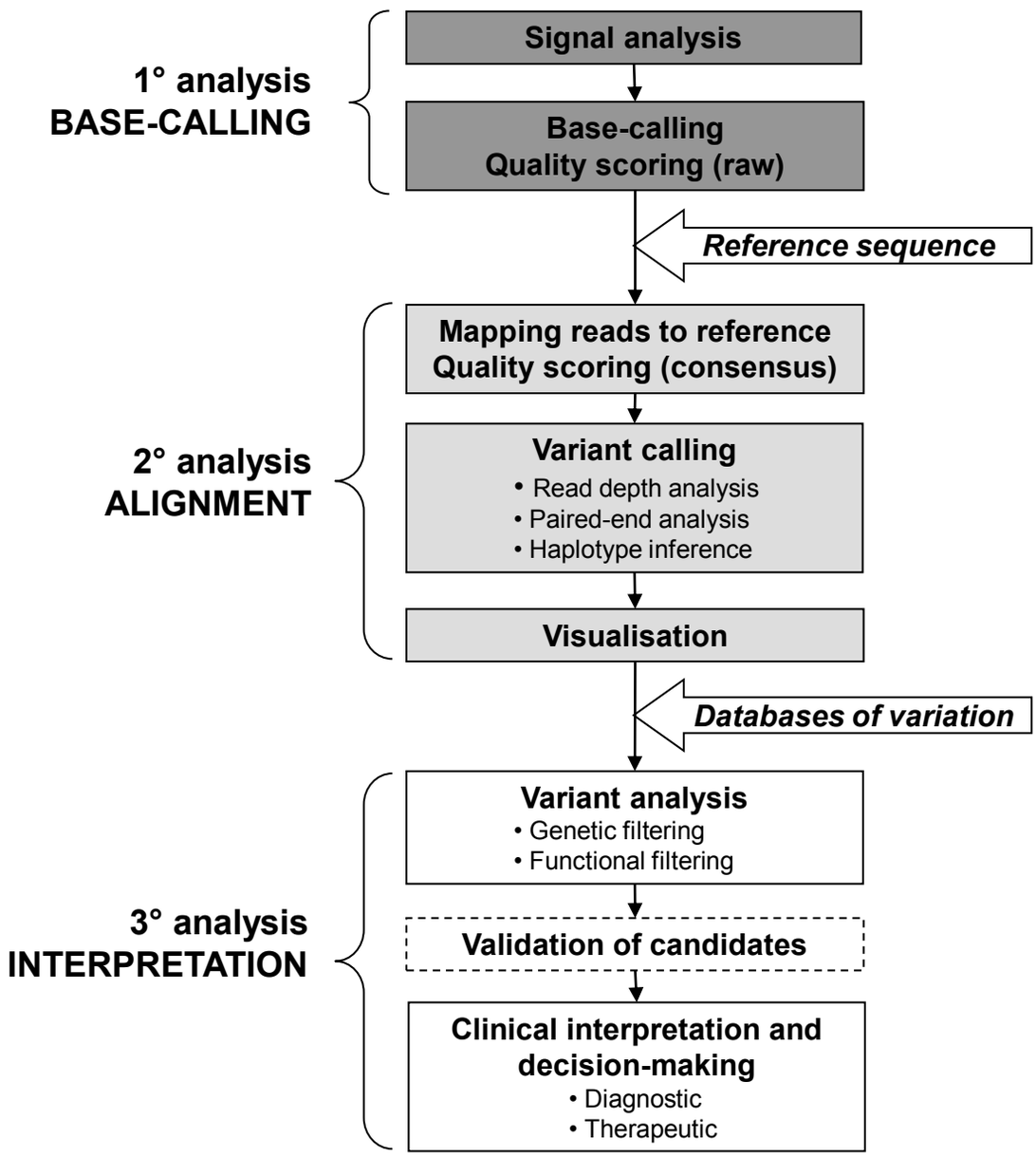
The informatics pipeline for genome resequencing using NGS technology can broadly be divided into three analytical steps (see Figure 4.1):

- **Primary analysis: base calling** – converting light signal intensities into a sequence of nucleotides (generally performed by software on the sequencing machine itself)
- **Secondary analysis: alignment and variant calling** – mapping DNA reads to an annotated reference sequence and determining the extent of variation from the reference (for which numerous algorithms and software packages have been developed)
- **Tertiary analysis: interpretation** – analysing variants to assess their origin, uniqueness and likely functional impact (for which numerous databases, algorithms and software packages can be used, though more need to be developed and integrated)

Currently tools and techniques are being developed in all three categories, and the first two steps are rapidly becoming increasingly automated and reliable. However, providing a clinically valid and meaningful interpretation is likely to remain a major challenge for the foreseeable future. Although the steps involved in the interpretation and application of results to a specific clinical setting varies substantially based on the purpose of testing and the precise clinical question, particularly with regards to germline versus somatic sequencing, the basic analysis pipeline and many of the tools involved are common.

Similarly, while some of the details may change with the introduction of third generation technologies, such as those that involve electronic rather than optical detection, the data analysis challenge posed by massively parallel sequencing technologies will remain essentially the same.

Figure 4.1: Outline of informatics pipeline for processing and analysing data from massively parallel sequencing platforms



4.1.1 Base calling

Most NGS platforms produce raw data in the form of images or signals which are associated with interrogation of individual nucleotides. Primary analysis involves the conversion of these images (which may be light intensities or colour calls) into a sequence 'read' for each DNA fragment. This process is carried out on the sequencing platform itself using computational software developed by instrument manufacturers, and is largely a solved problem for existing NGS technologies. Image files need not be directly accessed or removed from the sequencing machine; the output of this process is a substantially smaller text file for further analysis, containing base sequences and quality scores assigned to each base. The standard quality score is based on the logarithmic *phred* quality score[4], which is a measure of the probability that the base has been called correctly[78;79], which depends upon factors such as signal intensity, cross-talk between reactions, overloading

[4] Q10 = 1 in 10 error rate, Q20 = 1 in 100 error rate, Q30 = 1 in 1000 error rate, *etc*.

(*i.e.* more than one cluster/bead per signal), imperfect chemistry, signal decay and background noise. Reads that do not reach a certain quality threshold will not pass internal quality control filters and will not be listed in the text file. Given that the raw accuracy of each individual sequencing reaction using massively parallel NGS platforms is substantially less accurate than in standard Sanger sequencing, providing a quality score associated with each base to assess the likelihood that the call is correct is important in later analysis steps for both alignment and variant calling.

4.1.2 Alignment

Once sequence reads have been generated, the next essential step is to assemble them into a genome sequence. In the case of resequencing, this process involves aligning the reads against a relevant reference sequence and mapping them to the correct location in the genome. Alignment is made challenging by true differences between the reference and sample sequences, the highly repetitive nature of substantial portions of the genome, and (in the case of NGS platforms) short read lengths which mean that individual reads may map to multiple locations. The reference sequence itself does not represent any single individual and is still incomplete due to highly repetitive regions that are resistant to sequencing. Moreover, it is regularly updated based on new higher resolution data, which can cause specific regions to map to different locations or disappear entirely between versions.

Because it is not always possible to unambiguously align a read to a unique position in the reference genome, a mapping quality score may be used to provide a measure of the likelihood that the read is mapped correctly[80]. This score is based on the number of mismatches between the reference and the sample in a defined genomic window, using a gapped alignment (*i.e.* allowing for insertions and deletions) if required. The mapping reliability can be substantially improved by using paired-end (or mate-pair) reads, which involve sequencing both ends of a single DNA molecule and mapping each end unambiguously to a particular location in the genome. Because the ends are physically paired, the alignment of one end can be cross-referenced against the other, improving the reliability of the alignment and providing information about structural rearrangements. In practice, regions with multiple ambiguously placed reads can be excluded, leaving around 85% of the reference genome assembly being 'accessible' for further analysis[21].

Since every base is independently sequenced multiple times with massively parallel sequencing technologies, a consensus sequence is produced by aligning multiple overlapping reads, and heterozygous loci are distinguished by having a ~1:1 ratio of two alternative bases at the same position. The accuracy of this consensus sequence can be assessed with a calibrated quality score for each base, which depends largely upon the number of reads (*i.e.* depth of coverage) mapped to each position.

There are essentially four reasons for differences between the reference and sample genome sequence: inaccuracies in the reference, incorrect base calls in the sample sequence, incorrect alignment between the two, or true genetic variation in the sample. Every human genome is estimated to differ from the reference sequence in around 3-4 million sites[21;81], thus variant calling is a major challenge that forms part of the mapping process. Single base changes and structural variants must be accurately inferred from the alignment, whilst minimising the number of false-positive variants due to error. Variant calling is achieved through a combination of read depth analysis and paired-end analysis, to assess the reliability of each base call and the presence of insertions and deletions.

Determining which variants are derived from the same physical chromosome (haplotype inference) can be difficult from short reads, and must be assembled from clustered local alignments of individual reads with shared alleles[82], or imputed from reference haplotypes[18]. Using paired-end reads for mapping is particularly important for identifying structural variation, and is critical for cancer genome sequencing due to the presence of extensive large structural rearrangements relative to the matched germline genome, including both intra- and inter-chromosomal rearrangements[83].

In the last few years, numerous programmes have been developed specifically for genome assembly, alignment and variant calling based on DNA sequence reads from NGS platforms (see Table 4.1)[84;85]. These include software developed for use with a particular sequencing platform, open access software with a variety of functionalities and platform compatibilities, and proprietary software designed for specific purposes such as diagnostics (see Table 4.2). A major issue for standard alignment programmes is the interpretation of small insertions and deletions (INDELs), which has been partly addressed by the development of new programmes for this purpose[86]. However, current technology does not y*et allow* for confident analysis of INDELs of several hundred base pairs, especially those due to repeat sequences. Various dedicated software packages have also been developed specifically for cancer genome assembly and variant calling, which take into account factors such as genetic heterogeneity in the sample[87].

In the final stage of the alignment phase, sequence data are annotated with biological information and presented visually through a graphical interface or genome browser. This allows direct interrogation of the location of variants throughout the genome and comparison against a fully annotated reference. The ability to provide accurate and comprehensive genome annotation is critical for interpretation, and typically involves a mixture of automatic annotation by computational prediction and accurate manual annotation (curation) by applying human expertise and experimental data. The process of adding biological information to sequence data consists of two main steps[88]:

- **Structural annotation**: identification of functional elements in the genome, *e.g.* gene location, structure, coding regions and regulatory motifs
- **Functional annotation**: attaching biological information to these elements, *e.g.* protein structure, function, interactions and expression

An enormous amount of information can be included in a fully annotated sequence, including comparisons with other species[89], known genetic and epigenetic variants, regulatory features, expression data, and links to protein databases. However, annotation is currently both incomplete and imperfect in the human genome. Projects such as the **Enc**yclopedia **O**f **D**NA **E**lements (ENCODE)[90] and related Encyclopedia of genes and gene variants (GENCODE)[91;92] aim to identify all functional elements in the human genome sequence and will ultimately be used to annotate all evidence-based gene features in the entire human genome at a high accuracy.

Several mainstream genome browsers already exist, which are used extensively for research purposes and have largely overlapping functionality: Ensembl (developed and maintained by the European Bioinformatics Institute and the Wellcome Trust Sanger Institute, available at http://www.ensembl.org)[93] and the UCSC Genome Browser (developed and maintained by the Genome Bioinformatics Group at the University of California Santa Cruz, available at http://genome.ucsc.edu)[94]. These browsers provide the most recently assembled build of the human genome from the Genome Reference Consortium, with information on gene location, intron/exon boundaries, data on RNA expression, common SNPs and CNVs, mutations, and alignments with other species. They are extensively hyperlinked to external databases, are actively curated and regularly updated, and essentially act as portals for accessing and exploring annotated reference genomes and databases.

Various proprietary genome browsers have also been developed by numerous companies that enable a sample genome to be viewed, annotated and compared directly against the reference genome; some browsers accompany particular sequencing platforms, some proprietary browsers have been developed for specific markets (such as the medical diagnostics industry), and some are freely available and can be adapted by the user to fit various purposes.

Table 4.1: Selected examples of open access bioinformatics software for alignment, viewing and interpretation of NGS data

Software	Function
MAQ, Bowtie, SSAHA2, GATK, BWA, SOAP2, MOSAIK, SAMtools	Analysis, short read alignment and SNP calling
BreakDancer, Pindel, Dindel, PEMer, VariationHunter,	Structural variant (INDELs, CNVs) calling
SNVmix, VarScan, SomaticSniper, TIGRA	Cancer-specific genome assembly and variant calling
SAMtools, EagleView, SAMtools, MaqView, Tablet, MapView, IGV	Alignment viewers
Pairoscope	Visualisation of paired-end data
BLAST, BLAT, phyloP, PHAST	Analysis of evolutionary conservation
SIFT, PolyPhen-2, SNPs3D, PMUT, TopoSNP, PANTHER, Align GVGD, MAPP, PhDSNP, nsSNPA, Parepro, Mutation Taster	Prediction of the effect of amino-acid substitution

Table 4.2: Selected examples of companies developing dedicated bioinformatics packages for alignment, browsing and clinical interpretation of NGS data

Company	Software	Comment
Cartagenia	BENCHlab, BENCHclinic	Analysis tools
CLC Bio	CLC Genomics Workbench	Various tools and integrated software packages
GATC Biotech	DNASTAR	Service provider
GenoLogics	Geneus	Laboratory Information Management Systems
Illumina	CASAVA, ELAND	Platform specific
Life Technologies/ Applied Biosystems	BioScope	Platform specific
Real Time Genomics	RTG mapx, RTG cgmap	Various tools and integrated software packages
Roche/454	Newbler, GS Reference Mapper	Platform specific
SoftGenetics	Mutation Surveyor, NextGENe	Specific analysis tools

4.1.3 Interpretation

A substantial gulf currently exists between the relative ease of generating a fully assembled genome sequence, with confidence values assigned to variant calls, and the immense difficulty of interpreting the data in the context of an individual. The challenge includes algorithm development and curation, population and management of effective databases to allow evidence-based interpretation of genomic variants.

Clinical interpretation of whole or targeted genome resequencing data depends heavily upon the context and application in a particular setting, and it is usually the case that the analyses performed on the data are designed to allow the researcher or clinician to develop and test a hypothesis that can be explored using empirical methods or to help to reach a clinical diagnosis. There are currently few standardised methods for data analysis, and although there are shared approaches, neither a single unified reference dataset nor a comprehensive integrated analysis platform exist, and different groups have tended to develop their own customised pipeline. A common genomic analysis for clinical purposes is to attempt to identify likely pathogenic mutations that account for a specific phenotype. The first step in such an analysis is to exclude known (or suspected) non-pathogenic variation, by applying a number of different filters[95]:

- **'Genetic' filter** – exclusion of common, non-pathogenic or irrelevant variants, and inclusion of unique, pathogenic and candidate variants, by comparison with databases of genomic variation (see next section), or the individual's germline genome in the case of somatic sequencing; additional analysis of family members (to determine inheritance) or unrelated individuals with the same phenotype may also be required
- **'Functional' filter** – exclusion of variants that are expected to have no functional effect by analysis of the genomic location of the variant, evolutionary conservation and predicted effect on protein structure, function or interactions

Once the majority of variants have been excluded, a small number of potentially interesting variants remain that may or may not have been linked to a disease phenotype previously. (For the purposes of diagnostic testing currently, these variants are typically validated using traditional genetic testing methods, such as Sanger sequencing.)

Even after filtering, there may be too many variants of unknown significance to deal with, so computational prioritisation of variants is likely to be important[96]. Loss-of-function variants are likely candidates for being causally involved in disease, and in theory can act by disrupting any essential genetic element including non-coding regulatory motifs[97]. In practice, partly because so little is currently understood about the rest of the genome, coding variants that introduce changes in their corresponding proteins are usually identified as being the most likely candidates to have functional consequences.

There are numerous analyses that can be performed to filter and validate candidate variants and predict the likely contribution to the phenotype, including:

- Evaluation of the **evolutionary conservation** at that site[98], assuming that variants in a highly conserved region are more likely to be pathogenic
- Prediction of the effect of **splice site disruptions**, including insertions, deletions, and frame shift mutations, that might cause the protein chain to be substantially truncated or even entirely eliminated
- Prediction of **haploinsufficiency** status of genes[99]
- Investigation of the **expression** of the RNA or protein in the relevant tissue, using databases such as RefExA, Expression Profiler and the Gene Expression Omnibus
- Assessment of the role of the protein in relevant biochemical **networks and pathways**, using databases such as the NCBI Gene Expression Omnibus, IntAct or KEGG
- Prediction of the effect of **amino acid substitutions** caused by non-synonymous changes on protein stability, structure and function based on physical and comparative methods[100-102] such as Grantham scores, assuming that major changes to the physicochemical properties of the amino acid itself are likely to be more detrimental than conservative changes (depending upon their location in the 3D structure of the protein)

Various sequence analysis platforms have been developed that integrate and automate many of these processes, including those for use in medical diagnostics. However, biological support for disease causality is the most convincing evidence, such correlating inheritance with phenotype,

or a direct functional assay if available. A final clinical interpretation must integrate biological knowledge with relevant phenotypic and clinical information to assess the relevance of the candidate variant(s) to decisions regarding appropriate interventions. This might include the heritability, likely penetrance and expressivity of the variant, as well as implications for therapeutic options and treatment regimes.

Outside of specialist clinical genetics services, an appropriate decision support framework supported by a robust evidence base may be required to allow physicians to access genomic information at an appropriate level. For most purposes, the information required by (and potentially provided to) the clinician will be guided by the purpose of testing, and the majority of genomic information generated by WGS will be irrelevant to any given clinical question. Therefore, as an alternative to using targeting prior to sequencing (see Chapter 3), comprehensive masking and computational targeting could be used to limit the analysis to regions of the genome known to be relevant to the specific phenotype of interest. However, this approach has the disadvantage that it may prevent novel causal mutations from being identified in non-candidate genes, and would preclude wider application of personal genomic information.

4.2 Databases

Information about genes is organised and made accessible on the internet in many different ways. Resources range from annotated registries of all publicly available DNA sequences (such as the European Nucleotide Archive, GenBank and the DNA Data Bank of Japan), to searchable gene information portals such as Ensembl, GeneCards, BioGPS and WikiGenes. Numerous online databases of human genomic variation also exist, which are absolutely central to the task of filtering and interpreting whole genome sequence data[103]. Existing databases can be broadly divided into the following categories, containing data relating to:

- **Databases of genomic variation**, *e.g.*
  - 1000Genomes
  - HapMap
  - dbSNP (database of Single Nucleotide Polymorphisms)
  - dbVar/DGVa (peer databases of large scale genomic variants)
  - DGV (Database of Genomic Variants)
- **Databases containing potentially identifiable human subject data,** *e.g.*
  - bGaP (Database of Genotypes and Phenotypes)
  - European Genome-Phenome Archive
- **Databases containing variant-disease associations across the genome**, *e.g.*
  - OMIM (Online Mendelian Inheritance in Man) and the more clinically focused OMIM-Morbid
  - HGMD (Human Gene Mutation Database)
  - DMuDB (Diagnostic Mutation Database)
  - CDC HuGENavigator
  - NHGRI catalogue of genome-wide association studies
  - DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources)
  - ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations)
  - HGVbaseG2P (Human Genome Variation Genotype-to-Phenotype database)
- **Locus-specific databases (LSDB)**, see www.hgvs.org/dblist/glsdb.html, www.gen2phen.org/data/lsdbs or www.hgmd.cf.ac.uk/docs/oth_mut.html for a list, *e.g.*
  - Cystic Fibrosis Mutation Database
  - TP53 database
  - LDLR Familial Hypercholesterolemia database
  - IMGT/HLA database
- **Disease specific databases**, *e.g.*
  - AlzGene (Alzheimer's disease)
  - PDGene (Parkinson's disease)
  - T1DBase (type I diabetes)

- **Databases of somatic cancer genome variation**, *e.g.*
  - COSMIC (Catalogue Of Somatic Mutations In Cancer)
  - ICGC (International Cancer Consortium)
  - TCGA (The Cancer Genome Atlas)
  - HGMD (Human Gene Mutation Database)
- **Databases of pharmacogenetic associations**, *e.g.*
  - PharmKB (Pharmacogenomics Knowledge Base)
- **Databases of existing clinical genetic tests**, *e.g.*
  - EuroGenTest
  - GTR (NIH Genetic Test Registry)
  - UKGTN (Genetic Testing Network) gene dossiers
  - Orphanet

Over the next few years, as the majority of human variation across different ethnic groups is mapped and catalogued, these databases will become increasingly powerful tools for interpreting individual genomic data and improving genetic diagnoses. However, the existence of so many databases of both common and rare genomic variation, with varying levels of associated phenotypic information, is in itself a barrier to clinical translation. Ultimately, if NGS technologies are to achieve their full potential, there will need to be a unified catalogue of all known human variation and its functional or phenotypic consequences.

Access to the numerous private LSDBs that currently exist in isolation within different laboratories will be needed, through projects such as Gen2Phen, Mutadatabase and the Leiden Open Variation Database (LOVD), so that the results can be integrated into publicly accessible genome-wide databases. Similarly, databases containing genetic variation specific to a particular ethnic group or geographic location will need to be incorporated so that appropriate reference genomes can be used for alignment and interpretation purposes. Although some steps have been taken towards integration and unification of existing databases, there is still much work to be done as their scope, format and content varies enormously, and no standard method of collating, referencing or presenting the data has yet been adopted.

4.3 Interoperability and decision support

The informatics and database resources described above are constantly developing and evolving, and while they may be sufficient for research purposes, the application of whole genome sequencing to routine diagnostics use will require a stable ‘clinical grade’ analysis pipeline to ensure reliable performance and accuracy. Standard operating procedures and algorithms must be agreed for processing genomic information; quality filters should be developed to enable high-confidence variant detection; and appropriate informatics tools for interpreting and accessing genomic data will have to be built, validated, standardised and maintained as well as integrated with laboratory information management systems (LIMS) specific to massively parallel sequencing platforms.

One of the major hurdles that must be overcome is ensuring interoperability between platforms, databases and analysis tools. This requires computer-interpretable nomenclature, ontologies and syntax to be agreed and universally adopted for both genomic and phenotypic data. Organisations such as the Human Genome Organisation nomenclature committee[104], the Human Genome Variation Society, the US National Center for Biomedical Ontology, the Genome Ontology consortium[105] and the Human Phenotype Ontology project have already had a substantial impact amongst the research community on standardising the representation of gene and gene products across databases. In addition, initiatives such as the Human Variome Project and the EU-FP7 Gen2Phen project are working on models and standards in data description, storage and integration for biomedical databases, although such attempts at standardisation are still lacking in cancer genomics.

Meanwhile, the Unified Medical Language System, Systematised Nomenclature of Medicine-Clinical Terms (SNOMED-CT), Logical Observation Identifiers Names and Codes (LOINC) and Health Level 7 initiatives have been integral in the development of a common language for electronic health records to allow the appropriate retention, integration, processing and exchange of unambiguous medical data. Ultimately, in order to use genomic information routinely in a clinical setting, these initiatives (and many more) will need to be harmonised into a unified, computer-processable representation of patient data. This will clearly be a considerable task.

It seems likely that the level of informatics support needed in a clinical setting will depend heavily on both the specialty and training of the healthcare professional accessing information derived from these technologies. Although members of the specialist clinical genetics service may wish to access genomic data directly through a genome browser, specialists in other areas of clinical medicine and general practitioners are likely to want a more problem-oriented means of accessing limited genomic information of direct relevance to the clinical question. Therefore, downstream of the sequencing and informatics pipelines, a decision support system will be needed to help clinicians integrate genomic information into the patient pathway and guide preventive and therapeutic options, both for diagnosis and personalised/stratified treatments. Most clinical decision support systems consist of three parts: a dynamic knowledge base, an inference engine based on an agreed set of rules, and an appropriate mechanism for communication with the healthcare professional (or patient)[106]. Standardised representation of genomic and non-genomic patient data is essential to ensure reliable computer-based interpretation and processing[107], and robust epidemiological data and statistical methods are required to ensure evidence-based analysis.

Ultimately, the value of any clinical decision support system is dependent on the robustness of the knowledge base, which must be regularly updated and maintained. Genetic variants are likely to fall into three categories: those with a clear clinical interpretation (mostly relating to monogenic disorders), those associated with disease but with unknown clinical significance, and those with no known association with disease. There will be regular movement of variants between these categories as new discoveries are made and genotype-phenotype associations catalogued. However, currently no standardised process or system for assigning and annotating this categorisation exists, and frequently a lack of data upon which to make an evidence-based assessment of the clinical validity and utility of any individual test or analysis.

4.4 Informatics provision in the NHS

The National Genetics Reference Laboratory (NGRL) Manchester currently develops and maintains informatics resources that NHS diagnostic genetic services require in order to operate and develop effectively. However, the majority of bioinformatics expertise and storage capacity for biomedical data in the UK currently lies outside of the NHS, in research centres such as the Wellcome Trust Sanger Institute and European Bioinformatics Institute (EBI), and the four Medical Research Council (MRC) high-throughput sequencing hubs. These institutes have access to high performance computing clusters, high-speed networks and dedicated data storage facilities with enormous capacity. Although it is unlikely that hospitals would aim to match these, nonetheless if NGS technologies are to be translated into clinical use within the NHS, the provision of informatics and IT within the NHS will need a major overhaul in order to provide appropriate support for NGS platforms and whole genome sequencing.

The UK House of Lords Science and Technology Committee Report on Genomic Medicine (2009) recommended that a National Institute of Biomedical Informatics be set up to fulfil this requirement, which is now under consideration by the Human Genomics Strategy Group Bioinformatics Expert group.

One of the distinguishing features of massively parallel sequencing platforms relative to other medical or genetic tests is the sheer volume of data that is produced, requiring substantial processing power, storage capacity and network speeds[108]. Although a single run on an NGS platform can create many terabytes (1TB = 10^{12} bytes) of digital data, there is an outstanding question about how much of this genomic data is required to be kept for clinical purposes, given that there are a number of different options with variable data storage implications. The volume of data decreases substantially upon moving from raw image files, to processed intensity traces for each read, to a consensus sequence, to the set of variants for a given individual. Assuming that neither raw image data nor intermediate pipeline processing data will be stored for clinical use once a consensus sequence has been determined, it is worth considering two storage options more explicitly: storing an individual's entire genome sequence (~3 billion bases), versus storing a 'minimal genome' limited to only differences from the reference sequence (3-4 million bases for a germline genome).

In either case, the results output file would need to contain the following information as a minimum at every position: chromosome, location (reference coordinate), reference genotype, sample genotype (including heterozygotes), and a measure of confidence in the variant calls. Ignoring the effects of data compression algorithms, a full genome sequence including this information could be stored in full as a single text file of ~3GB (which would easily fit on a DVD-ROM), whilst a 'minimal' genome might require only ~30MB[5]. Thus storage of a whole assembled genome is possible and the 'minimal genome' is easy.

The size of the stored data file also has major implications for data transfer. Sequence data must be transferred off the sequencing machine, and may also need to be moved to several other locations for further analysis and safe archiving. It is likely that capacity for transferring terabytes of data will be required, which is substantially beyond the ability of the existing NHS network; use of removable hard disks is currently a common solution to this issue, but it does not facilitate controlled data sharing between numerous individuals who need varying levels of access to some or all of the data for different purposes. The use of cloud computing over the internet to distribute and process data 'virtually' across multiple large hardware clusters has been suggested as a potential solution to many of these issues[109]. However network bandwidth could still pose a barrier, and the importance of keeping highly sensitive, identifiable medical data secure and confidential means that 'virtual' storage of data and use of shared resources distributed across multiple locations (and jurisdictions) may be impossible.

A number of initiatives are already underway to try to both centralise and streamline the management and processing of patient data between multiple providers of health and social care. All those who process patient identifiable data within the NHS are required to conform to the Information Governance Toolkit[110] and as part of the accreditation process led by the Care Quality Commission, are required to comply with relevant data handling guidance[111]. The Secondary Uses Service was established in 2007 following recommendations from various policy groups[112;113] to provide a single, authoritative and comprehensive source of data which can be used for non-clinical applications including the systematic use of patient data in research, clinical audit and health surveillance and monitoring. All secondary care providers are required to complete and return Commissioning Data Sets that contain details of the care provided. Increasingly these data sets are populated from other services provided by the National Programme for IT including Chose and Book and the Electronic Prescribing Systems. Access to these returns is determined on the basis of a role-based access approach to ensure that data security is proportionate, and some users are limited to pseudonymised data.

The NHS Connecting for Health programme[110] managed the National Programme for IT, which aimed to facilitate the sharing of individual electronic records between providers of primary and secondary care to improve patient care across geographical and institutional boundaries. Initial ambitious plans to create summary and detailed care records to be held on a centralised spine foundered following criticism from professional bodies that patients were not given enough power to veto the creation of an electronic record[114]. Ultimately these criticisms were addressed by ensuring that patients gave an opt-in consent to the creation of the summary care records, and that detailed care records were held locally rather than on a national database.

[5] See http://www.genetic-future.com/2008/06/how-much-data-is-human-genome-it.html.

As a result of severe financial constraints, there has also been a delay in the implementation of the programme across the UK and empirical evidence of the anticipated benefits of these programmes has been sparse[115;116]. It also seems likely that the accumulation and sharing of electronic patient records in the UK will be affected by plans to reform the NHS (including the reforms envisaged as part of the Information Revolution and Quality, Innovation, Productivity and Prevention initiatives)[117]. These include more proposals for patient held records and for online access by patients to relevant medical records. The detail of how these proposals might apply to predictive genetic information, or to multiple family members who share genetic risk to a known monogenic disease, have yet to be addressed and it is clear that much more research is needed on the potential impact of WGS data on electronic patient systems and records.

Finally, provision of bioinformaticians within the NHS is currently very limited, and regardless of the laboratory service model (see Chapter 10) and support systems, a substantial number of informaticians will be needed. Personnel requirements will include process-level informaticians responsible for data quality control, base-calling and alignment, molecular bioinformaticians responsible for variant calling and annotation, and clinical informaticians responsible for interpretation of genomic information in the context of an individual patient. This latter group will need to integrate epidemiological and statistical expertise with medical and genomics knowledge. This need was highlighted in the House of Lords Science and Technology Report on Genomic Medicine (2009), and is being considered further by the Human Genomics Strategy Group, and addressed by the National Healthcare Science School of Genetics in developing the Modernising Scientific Careers training programme for clinical scientists in the NHS.

5 Applications in inherited and heritable diseases

5.1 Introduction to inherited and heritable diseases

5.1.1 Pathology

Diseases with a strong heritable component[118] are often known as Mendelian disorders as their inheritance pattern is fairly predictable. In many cases, a mutation in just a single gene causes the disease, which can be inherited in either a dominant, recessive, or sex-linked manner. Several thousand monogenic disorders are known, and although their inheritance is predictable, the resulting phenotype may be hard to predict due to genetic and environmental factors. Moreover, while some inherited diseases have well defined invariable phenotypes, others may vary enormously in their onset, severity, symptoms and prognosis due to underlying genetic heterogeneity and modifiers.

Increasingly, so-called single gene subsets of complex disorders are being discovered, in which a strongly heritable factor is responsible for an apparently multifactorial disease phenotype in a small number of individuals. In particular, numerous inherited forms of cancer are now known, many of which are incompletely penetrant such that the probability that a mutation carrier will develop cancer may be much less than 100%.

In addition to these inherited diseases, many 'genetic' disorders are caused by *de novo* mutations and are often described instead as heritable. Recent studies have indicated a relatively high per-generation mutation rate of ~50 new mutations per new generation[23;119], which may occur either in the maternal or paternal germ cells or in somatic cells during early embryonic development. Because *de novo* pathogenic mutations would manifest as sporadic (rather than familial) disease, and are therefore not amenable to standard methods of familial linkage analysis, most known *de novo* (spontaneous) pathogenic changes are associated with severe phenotypes and involve relatively gross chromosomal changes that can be detected reasonably easily.

Presence of an abnormal number of chromosomes (aneuploidy) is the most obvious example – such as Down's syndrome, which is caused by having three copies of chromosome 21 – though most aneuploidies result in spontaneous abortion of the fetus. Many cases of childhood learning disability, congenital malformation and dysmorphology are caused by much smaller *de novo* chromosomal rearrangements (most commonly deletions, see Table 5.1). The occurrence of heritable diseases caused by *de novo* pathogenic changes generally increases with parental age.

Inherited and heritable diseases usually involve mutations in the coding regions of the genome, which have a direct deleterious effect on the resulting protein. Pathogenic variants may affect a single base or multiple bases, and may involve substitution, insertion, deletion or translocation of bases (see Table 5.1). Of those known, by far the most common pathogenic mutation that is compatible with life is a single base substitution that results in either an amino acid alteration in the resultant protein (missense) or premature truncation of the protein (nonsense).

Table 5.1: Relative frequency of different types of mutations underlying disease phenotypes (data from the Human Gene Mutation Database, September 2011).

Mutation type	Number	% of Total
Micro lesions:		
Missense/nonsense substitutions	63,313	55.9
Splicing substitutions	10,653	9.4
Regulatory substitutions	2,049	1.8
Small deletions	17,807	15.7
Small insertions	7,346	6.5
Small INDELS*	1,671	1.5
Gross lesions:		
Repeat variations	353	0.3
Gross insertions/duplications	1,583	1.4
Complex rearrangements	1,089	1.0
Gross deletions	7,383	6.5
Total (in 3,888 genes)	**113,247**	100

* Co-localised insertions and deletions

5.1.2 Epidemiology

Mendelian disorders where the gene(s) is known are catalogued in the Online Mendelian Inheritance in Man database (OMIM) which contains information on the relationship between phenotype and genotype of over 13,000 genes (see Table 5.2)[120]. Although individual Mendelian disorders are uncommon, with more than 4,000 known disorders, they collectively account for significant amounts of infant mortality and morbidity[121]. One in 17 people in England (approx. 3 million people) are estimated to have a rare disease, as well as a further 30 million Europeans and 25 million North Americans[122]. The vast majority of these diseases are thought to be single gene disorders, although the underlying genes are often unknown. Many of these disorders show vast genetic heterogeneity, with multiple mutations within a single gene or mutations across hundreds of individual genes already identified. In many cases, they may be commonly described as single gene subsets of complex diseases, and as such relate to any organ in the body and may be

encountered in every medical specialty. As whole genome approaches becomes more widespread, the proportion of common diseases explained by numerous rare, highly penetrant mutations is likely to increase, making a genetic diagnosis crucial to understanding aetiology and managing the disease appropriately. Unless these disorders can be distinguished phenotypically, there is no *a priori* way to assess which individual genes should be targeted for testing.

Table 5.2: Number of entries in OMIM (totalling 20,813 in September 2011)

	Autosomal	X-linked	Y-linked	Mitochondrial
Gene with known sequence	12,916	635	48	35
Gene with known sequence and phenotype	197	7	0	2
Phenotype description with molecular basis unknown	2,975	254	4	28
Mendelian phenotype or locus with molecular basis unknown	1,633	133	5	0
Other phenotypes with suspected Mendelian basis	1,810	129	2	0

Below we briefly explore several examples of such inherited disorders, which are intended to be illustrative rather than exhaustive.

Example 1: Inherited cardiovascular conditions

Inherited cardiovascular conditions (ICCs) are a group of more than 50 disorders that include primary electrical and structural heart diseases, and vascular diseases. An understanding of their epidemiology is hampered by the fact that many of the individual conditions are rare, they have a high degree of genetic heterogeneity, incomplete penetrance and can sometimes be difficult to accurately diagnose as they can be hard to distinguish from common multifactorial conditions with similar symptoms. There is no national database to identify clearly the mortality and morbidity due to ICCs, and sudden death is often the first clinical presentation. Thus the true prevalence is difficult to determine, but they are collectively thought to affect around 1 in 240 individuals in the UK[123]. More than 1,000 mutations in over 30 genes are known, with more than 1,900 genetic tests for ICCs carried out in the UK per annum[123]. The majority of ICCs demonstrate a pattern of Mendelian autosomal dominant inheritance although severe disease has also been seen in the less common autosomal recessive and X-linked subtypes. With variable penetrance and expression, family members carrying the same genetic mutation may experience varying degrees of disease severity.

Example 2: Monogenic eye disorders

The monogenic eye disorders are a clinically and genetically heterogeneous group of over 50 conditions in which specific gene defects, with variable penetrance, lead to the abnormal structure or function of the eye. The majority of these conditions demonstrate a Mendelian autosomal dominant, recessive or X-linked pattern of inheritance. It has been estimated that up to one third of children diagnosed in the UK with sight loss (including partial sight and blindness) and 10% of adults diagnosed with sight loss are due to an inherited disorder[124]. National active surveillance for severe visual impairment/blindness in the UK has revealed a cumulative incidence by 16 years of age of 5.9 per 10,000[125]. Of these, one third were inherited. Retinitis pigmentosa is the most common individual hereditary retinal degeneration, caused by numerous mutations in over 40 genes, and has a prevalence of 1 in 3,500 to 1 in 4,000 individuals[124]. Assuming 10% of the adult population diagnosed with sight loss (approximately 2 million individuals) are due to inherited conditions, this represents an estimated prevalence of 1 in 309 individuals in the UK.

Example 3: Inherited cancers

Although the majority of cancers are sporadic, caused by somatic mutations (see Chapter 6), some 5% to 10% may be inherited as mutations within genes that markedly increase susceptibility to one or more types of inherited cancer. Approximately 5% of breast cancer cases have inherited genetic variants in the *BRCA1* and *BRCA2* genes (tumour suppressor genes)[126]. Inherited variants in *BRCA1* and *BRCA2* also increase risk of ovarian cancer and are implicated in prostate, bowel and pancreatic cancers. Familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC also known as Lynch Syndrome) are two further examples of familial cancer syndromes. FAP is caused by mutations within the *APC* gene with penetrance close to 100% and accounts for less than 1% of all colorectal cancers. HNPCC is more common, accounting for around 3%, and is caused by mutations in a gene from the family of mismatch repair family with around 90% of familial cases being caused by mutations in the *MLH1* and *MLH2* genes[118]. The inherited components of several common cancers are listed in Table 5.3. Some are associated with very rare familial cancer syndromes such as TP53 and Li Fraumeni syndrome, a disorder of a wide spectrum of cancers occurring in children and young adults including breast and adrenocortical cancers[118]. It is estimated that around 1 in 300 people have a monogenic inherited cancer predisposition in the UK (J Burn, personal communication).

Table 5.3: Common cancers with an inherited component[127]

Gene(s)	Principal tumours	Mutation prevalence	Penetrance (by age 70 years)
BRCA1	Breast, ovarian, bowel, prostate	1 in 850 *(1 in 100 Ashkenazi Jews)*	65%-85% (breast cancer) 39%-45% (ovarian cancer)
BRCA2	Breast (including male), ovarian, prostate, pancreatic	1 in 500 *(1 in 100 Ashkenazi Jews)*	45%-80% (breast cancer - female) 5% (breast cancer – male) 7.5% (prostate cancer)
APC	Familial adenomatous polyposis (multiple bowel polyps with malignant potential)	1 in 8,000[128]	~100%
Mismatch repair family including *MLH1*, *MSH2*, *MSH6*, and *PMS1*	Hereditary non-polyposis colorectal cancer (mostly bowel but also endometrial, ovarian, and gastric).	Between 1 in 200 and 1 in 3,000	Up to 80% (bowel cancer for males although lower for females) 40%-60% (endometrial cancer)

Example 4: Learning disorders

Learning disability (LD) or mental retardation (MR) is a common condition affecting 1% to 3% of individuals worldwide[129] with genetic factors estimated to be the main cause in around half of all patients with severe LD and around 15% of patients with mild LD[130], though the true proportion may be substantially higher[20]. In the UK, the number of individuals with severe LD has been estimated at 210,000 with a further 1.2 million estimated to have mild to moderate LD[131]. Assuming 50% of cases with severe LD and 15% of cases with mild to moderate LD are due to inherited genetic factors suggests at least 285,000 individuals in the UK with an inherited form of LD (1 in 217). In many cases, LD is caused by *de novo* structural variants.

Example 5: Deafness

Deafness is a condition where an individual's ability to detect sound is completely or partially impaired. Severe or profound deafness occurs in about 1 in 1,000 children with a further 2/3 in 1,000 affected by moderate or progressive deafness[118]. More than half of these are caused by

inherited genetic mutations favouring an autosomal recessive pattern of inheritance although autosomal dominant, X-linked, and mitochondrial inheritance is also seen[118;132]. Deafness has been associated with over 100 genes[133] including mutations in the *CX26* and *CX30* genes at the *DFNB* locus on chromosome 13[118] and many others[134].

5.1.3 Genetic testing

There are various different reasons for an individual or family to be referred for (germline) genetic testing, primarily due either to symptoms or family history of a particular disease. Testing may be diagnostic, predictive, presymptomatic or to affirm carrier status of recessive diseases; it may be offered just to an individual, close family members or cascaded out to relatives. There are various different times and reasons when genetic testing might be indicated, which can be grouped as follows:

- **Preconception**, *e.g.* a couple with a family history of an inherited condition, wanting to know if they are carriers and their reproductive options
- **Preimplantation**, *e.g.* a couple who know they are at risk of passing on an inherited condition, and wish to undergo preimplantation genetic diagnosis (PGD) to prevent the birth of an affected child[135]
- **Prenatal**, *e.g.* a pregnant couple who know they are at risk of carrying a child with a particular identifiable genetic condition, and opt to test the fetus for this condition
- **Neonatal**, *e.g.* a newborn child with a congenital abnormality, if the parents want information about the condition and recurrence risks
- **Childhood**, *e.g.* a child with learning difficulties that may be due to a genetic condition, referred for investigation and diagnosis
- **Adulthood**, *e.g.* an individual with a strong family history of cancer, wanting to know if they are at increased risk and what options they have

Diagnosis in clinical genetics allows rational management and surveillance of individuals and families. In some cases, genetic testing can result in a substantial decrease in mortality and morbidity through directed interventions, *e.g.* colonoscopy and removal of polyps to prevent hereditary nonpolyposis colorectal cancer (HNPCC). However, even where no treatment is available, achieving a molecular diagnosis through genetic testing may be crucial for reproductive choice, avoiding a further diagnostic odyssey, or accessing relevant services.

There are broadly two types of genetic tests currently used in clinical diagnostics[136]: targeted tests for analysing specific genetic variants (molecular genetics), and whole genome tests for analysing chromosomal rearrangements and copy number changes (cytogenetics). The former includes genotyping to detect the presence of particular mutations in a specific location (typically by PCR), sequencing to determine the sequence of an entire gene or region, or targeted dosage analysis to ascertain if insertions or deletions have occurred in a specific region, such as a triplet repeat expansion. The latter includes analysis of G-banded chromosomes (karyotyping), fluorescent in situ hybridisation (FISH) and array comparative genomic hybridisation (array CGH) to detect and localise different sizes of chromosomal rearrangements and copy number changes across the genome.

Traditionally separate, these two disciplines are being brought closer together by new technologies that are common to both disciplines, such as DNA microarrays and real-time PCR. The convergence is reflected in the imminent merger of the two professional societies in the UK, the Association of Clinical Cytogeneticists (ACC) and the Clinical Molecular Genetics Society (CMGS).

5.1.4 NHS genetics services

NHS genetics services include a range of clinical, laboratory and screening services which are delivered by a network of 23 integrated Regional Genetics Services (RGS) across the UK (see Chapter 10) and in a number of specialist centres around the country. They provide an effective, coordinated service to patients and families with inherited diseases. The delivery of clinical genetics services revolves primarily around the provision of genetics clinics for affected families, the principal activities of which include:

- Careful enquiry to obtain detailed relevant family information and a pedigree
- Clinical assessment and investigation of affected individuals and other family members as appropriate to confirm the diagnosis
- Pedigree analysis and genetic risk assessment for patients and family members
- Genetic counselling, which is the process of communicating clinical and genetic risk information to individuals and to family members who may be at risk, to allow them to make informed choices and decisions
- Discussion of the options that are available to help individuals and their families to enable them to choose a course of action appropriate for them. These options may include pre-implantation or prenatal diagnosis or predictive genetic tests

Subsequent medical management of affected individuals is usually provided by other clinical services, though clinical geneticists may help coordinate care in multisystem genetic disorders.

Unlike other medical specialties, the genetics service revolves around managing families rather than individual patients. Patients and families may be referred to the genetics service from a variety of sources, but mainly by other specialists (such as paediatrics) or from primary care. Though consultant-led, these services are delivered by a multi-disciplinary team which includes consultant clinical geneticists, genetic counsellors and specialist nurses.

In addition to routine genetics advisory clinics, most centres are also involved in specialist clinics. These may be in particular aspects of genetics such as dysmorphology or cancer genetics (reflecting an area of specialist expertise and interest of the clinical geneticist) or may be joint clinics in association with other medical specialties such as neurology, paediatrics or ophthalmology. Consultant clinical geneticists and genetic counsellors are the cornerstone for training and education in genetics in both under- and post-graduate medical education, as well as in relation to other healthcare professions.

The clinical NHS genetics service is supported by a network of NHS molecular genetic and cytogenetic laboratories. In addition, there are usually close links with the laboratories providing Down's syndrome screening, newborn screening laboratories and with specialist paediatric biochemistry laboratories involved in the diagnosis and management of children with inherited metabolic disorders. In most RGSs the molecular and cytogenetic laboratories are co-located with the clinical genetics service, either in the same building or on the same campus. Molecular genetics laboratories play an important role in providing DNA-based diagnostic testing for patients with inherited disorders, pre-symptomatic testing for specific genetic disorders of adult life, and pre-natal diagnosis of inherited disorders for families at risk that require this service. Increasingly molecular genetic tests are being used for patients' management in other clinical specialties, *e.g.* clotting factors in haematology, mutation testing in diabetic patients with MODY (Maturity Onset Diabetes of the Young) or patients with specific forms of cancer as an aid to target treatments.

Cytogenetic laboratories investigate suspected cases of chromosomal imbalance in a variety of different situations. These include the investigation of children with congenital disorders or learning disability, prenatal samples from pregnancies at increased risk of foetal aneuploidy and increasingly in the diagnosis and management of a variety of cancers. Though conventional karyotyping is still the standard test in most centres, as previously discussed, traditional karyotyping is increasingly being combined with or replaced by DNA-based approaches such as FISH, PCR and array CGH. Several regional centres have already merged their cytogenetics and

molecular diagnostics laboratories, and more will no doubt follow.

Because of the implications of the information generated by DNA tests, there is considerable emphasis on ensuring that DNA tests are only carried out when appropriate. DNA testing should only be undertaken when it will assist in clinical management. The UKGTN has identified five situations where genetic test results may inform management: to confirm a diagnosis; to change a treatment plan; to advise on prognosis and management; to allow pre-symptomatic testing; and to provide an accurate assessment of genetic risk. Clinical geneticists and their laboratory colleagues are heavily involved in advising colleagues in other specialities on the appropriateness of ordering a DNA test. Additionally, although diagnostic genetic testing is organised and conducted within the clinical genetics services, it may also have major implications for other medical specialties in terms of management of disease. For example, inherited breast cancer caused by *BRCA1* or *2* mutations has a different histopathology from other breast cancers, which substantially alters the management from a standard lumpectomy to a bilateral mastectomy.

5.2 Diagnostic applications of NGS and WGS

5.2.1 Clinical diagnosis

To date, much of the expertise of the clinical geneticist has been to achieve a diagnosis through recognition of specific phenotypes coupled with targeted molecular testing. However, the diagnostic yield from phenotypic diagnosis has now largely plateaued, and high resolution genome-wide approaches are needed to uncover novel causal variants that might explain rare phenotypes. In order to capitalise on the high-throughput advantages of massively parallel sequencing relative to traditional Sanger sequencing, there are currently two major diagnostic applications of NGS technologies: multi-gene diagnostic panels, for large or multiple gene sequencing in genetically heterogeneous conditions; and achieving a molecular diagnosis where the condition had previously only been characterised clinically.

The former application – using NGS for targeted multi-gene diagnostic panels – is likely to become standard practice, as it essentially treats NGS as a cost-effective replacement technology in specific cases where defects in any one of multiple different genes could be responsible for the phenotype. Examples include inherited cardiac conditions[123], X-linked mental retardation[137], retinitis pigmentosa[138] and non-syndromic hearing loss[139], all of which are highly heterogeneous conditions associated with hundreds of different loci. In such cases, serially testing each candidate gene using Sanger sequencing is not feasible and would be expensive, so a substantial portion of patients currently remain undiagnosed.

A similar application involves sequencing large genes, where multiple genetic variants within the gene may be responsible for the phenotype. One of the clearest early examples of this in the context of diseases with a strong heritable component is the use of NGS for targeted sequencing of *BRCA1* and *2* tumour suppressor genes, where the high cost and limited throughput of diagnostic sequencing restricts testing to individuals identified as being at very high risk of breast cancer[140].

The latter application – using NGS for gene discovery and molecular diagnosis – has become a recognised method since late 2009, since when numerous studies have been published using exome or whole genome sequencing to diagnose the cause of rare inherited diseases[141]. This is particularly relevant for very rare or seemingly sporadic disorders, and conditions that directly impair fertility, which are intractable to classical methods such as of positional cloning and linkage analysis due to the absence of large families[142]. Although WGS is able to overcome this obstacle by testing unrelated individuals with the same clinical phenotype, identification of the causal variants is much more challenging as the *a priori* likelihood that any given variant is causal is much lower than using linkage analysis to focus on a small candidate region.

Even targeting just the exome still involves sequencing ~30Mb of DNA and produces ~20,000 variants relative to the reference sequence[143]. Nonetheless, the method has been successfully applied to autosomal dominant[144], autosomal recessive[22] and X-linked disorders[145], caused by both inherited or *de novo*[20] mutations, by studying families and/or unrelated individuals with the same condition versus controls. Moreover, because so much genetic information is available about every individual, it may sometimes be possible to determine if complications are part of the clinical phenotype or independent, which is otherwise extremely difficult in rare diseases[146].

Although details vary between studies, a common filtering method has developed for identifying the disease causing genes from the plethora of individual genetic variants, based on comparison against databases of genomic variation, and functional analyses of candidate causal variants (see Chapter 4)[95]. Causal mutations may be found in known disease-associated genes or novel genes based on the predicted functional effects of variants, but the process is complicated by two main areas of uncertainty[95]: technical, due to inhomogeneous coverage of the target region, incorrect mapping of reads, low depth of coverage and relatively high false-positive rates in functional regions due to sequencing errors[97]; and genetic/biological, due to locus heterogeneity (*i.e.* the presence of multiple distinct disease genes or loci for one clinical phenotype) and the challenge of proving causality of a novel mutation.

Although this application falls between the boundary of research and clinical practice, it is likely to be used increasingly for both diagnosis and to inform clinical management of patients. This has been demonstrated clinically in a child with intractable inflammatory bowel disease, who after receiving many failed interventions was successfully treated following diagnosis of an X-linked inherited immunodeficiency through whole exome sequencing[147]. Whilst currently the analysis is very complex and individual whole exome or genome sequencing is not cost-effective in a clinical setting, as the informatics pipeline becomes standardised and the costs continue to plummet it is likely that NGS technologies will increasingly be used to achieve a molecular diagnosis in individual patients across all areas of medicine.

5.2.2 Antenatal testing

Antenatal diagnostic testing currently involves invasive removal of a sample of fetal cells directly from the uterus for genetic analysis, using either chorionic villus sampling (CVS) between 11 and 14 weeks gestation or amniocentesis after 15 weeks, which carries a risk of miscarriage of around 1%[148]. A method of non-invasive prenatal diagnosis would therefore be desirable, and most recent work in this area has focused on fragments of cell-free fetal DNA (cffDNA) circulating in maternal plasma during pregnancy[149]. CffDNA originates from placental cells, which expel their DNA into the maternal circulation during normal cell death, breaking up the chromosomes into short fragments. Although the exact amount of cffDNA varies between individuals, it has been detected as early as four weeks gestation[150], is rapidly cleared from the maternal circulation such that it is undetectable two hours after delivery[151], and comprises only around 10% of the cell-free DNA in maternal plasma (the rest being maternal)[152;153].

There are several clinical applications where early non-invasive diagnosis using cffDNA testing could replace or reduce the need for invasive testing[154], including determination of fetal sex in pregnancies at risk of an X-linked disorder, and diagnosis of specific single gene disorders. Although the former application has been successfully achieved through targeted PCR amplification of Y chromosome DNA from male fetuses[155], the latter is substantially less well developed and more technically challenging if the mother also carries the mutation of interest[156].

Recent work indicates that using NGS to sequence all cell-free DNA in maternal plasma (both maternal and fetal) could be used to provide a prenatal diagnosis of almost any inherited genetic condition[153]. Although fetal DNA is in the minority, a genome-wide genetic map of the fetus can be constructed in a stepwise fashion by comparison with parental genotypes. Paternally inherited alleles can be detected directly where they differ from maternal variants, whilst maternally inherited alleles must be deduced as a series of inheritance blocks, based on a slight imbalance in the relative dosage of maternal haplotypes in the plasma caused by the presence of fetal DNA. This genetic map can then be used to determine the mutational status of the fetus at a particular disease locus.

To date, only a single proof-of-principle study has been published using this technique to diagnose the status of a fetus at risk of compound heterozygosity for β-thalassaemia[153], and further large studies are needed before this application could be offered widely. However, this technique still requires extensive further research and development, and there is no realistic prospect of it being introduced as a routine test throughout the NHS in the next few years.

5.2.3 Ongoing NGS projects

There are an increasing number of projects in the UK using NGS to improve the understanding of diseases with a strong heritable component. At this point, most projects are planning to use targeted exome resequencing rather than whole genome sequencing for two main reasons: first, targeted sequencing is currently significantly cheaper, and second, most known disease-causing variants are located in the coding region of the genome[157]. There are at least four major sequencing projects currently in the UK that aim to uncover rare genetic variants causing disease:

- **UK10K**

  Funded by the Wellcome Trust to study rare genetic variants and their role in health. By performing exome sequencing on 6,000 people with extreme health problems, and comparing the results against 4,000 individuals with well documented phenotypic characteristics, including the TwinsUK and Avon Longitudinal Study of Parents and Children (ALSPAC) cohorts, the project aims to uncover novel rare genetic variants responsible for disease.

- **Deciphering Developmental Disorders (DDD)**

  Co-funded by the Wellcome Trust and the Department of Health to identify the genetic causes of developmental disorders. A collaboration between the 23 NHS Regional Genetic Services and the Sanger Institute, the project aims to study 12,000 children affected by developmental disorders and their parents, using microarray technology and genome-wide sequencing to identify mutations that might explain the disorder and provide a diagnosis. Pertinent results will be fed-back to clinical teams for validation and made publicly available online through the existing DECIPHER database[158].

- **Cardiovascular Biomedical Research Unit (BRU) at Royal Brompton Hospital**

  Funded by the National Institute for Health Research to investigate the links between specific genetic variants and forms of cardiomyopathy. The project will perform exome sequencing on 10,000 patients to establish new genetic factors causing heart disease in individual patients, with a view to developing personalised treatments for this group of cardiac conditions.

- **East Anglian Sequencing and Informatics Hub (EASIH)**

  Funded by the Medical Research Council to apply NGS to routine medical diagnostic uses. Located at Addenbrooke's hospital in Cambridge, this project will apply X chromosome exome sequencing to investigate genetic causes of X-linked mental retardation and develop a diagnostic service for these patients.

There are too many sequencing projects underway globally to attempt to make a comprehensive list. Nonetheless, two examples of US-based projects are perhaps particularly noteworthy:

- **The Exome Project**

  Jointly funded by National Heart, Lung, and Blood Institute and National Human Genome Research Institute to develop cost-effective, high-throughput and integrated exome sequencing pipeline in well-phenotyped populations, to improve the diagnosis, management and treatment of heart, lung and blood disorders.

- **ClinSeq**

  Funded by the National Human Genome Research Institute to pilot large-scale medical sequencing in a clinical research setting, with the aim of developing the technological and procedural infrastructure to facilitate this type of research and demonstrating that it is feasible to sequence and interpret large amounts of genomic sequence data and return individual results to subjects.

5.3 Potential screening applications of WGS

According to the UK National Screening Committee (NSC) website, "*screening is a process of identifying apparently healthy people who may be at increased risk of a disease or condition. They can then be offered information, further tests and appropriate treatment to reduce their risk and/or any complications arising from the disease or condition.*" There are a series of programme appraisal criteria relating to the condition, the test, the treatment and the screening programme itself that have to be met with robust evidence before screening for a condition should be initiated (see www.screening.nhs.uk/criteria).

Genetic screening programmes involve a systematic offer of a test or inquiry to a defined population to identify individuals at increased risk for a specified disorder or at risk for passing the disorder to offspring, so that the individual(s) can be informed of their risk and the options for diagnosis. There are various stages during an individual's reproductive history when genetic screening could be offered to lessen the impact of diseases with a strong heritable component: preconception, prenatal and neonatal. Each presents a potential opportunity for the application of a WGS approach if the technology offers benefits or improvements over existing molecular (DNA-based) or biochemical (protein-based) tests.

5.3.1 Preconception carrier screening

Preconception genetic testing identifies carriers of genetic mutations responsible for a range of genetic conditions, knowledge of which can inform – and may qualitatively affect – the reproductive choices of individuals tested or those of their close biological relatives[159]. It has the advantage of identifying at risk individuals at a point when they have the widest range of personal and reproductive choices, including choosing not to have biological children, undergoing preimplantation genetic diagnosis, having prenatal testing (with the option to terminate), or accepting the chance of having an affected child.

Preconception carrier screening could include offering preconception carrier testing either to a high risk group where there is a high incidence of a particular disease, or to the general population for a range of conditions. Although offering genetic screening as early as possible in an individual's potential reproductive life allows for a greater range of options, it may be harder to deliver preconception screening systematically at a population level as individuals may not yet have engaged with reproductive health services. There is no consensus on when would be best time to offer carrier screening.

At a population level there are two main factors that influence the chance of a fetus inheriting a recessive condition: consanguinity and ethnicity. A consanguineous couple is usually defined as being related as second cousins or closer[160]. Consanguinity is encountered all over the world, and recent estimates indicate that some 10.4% of the world population are either married to a close biological relative or are the progeny of a consanguineous union[161]. Whilst it may have some social and economic advantages[162], the risk of inheriting recessive conditions from a shared common ancestor is substantially increased through cousin marriage. Although consanguineous marriage may also have an effect on reproductive behaviour and fertility, recent research suggests that there is an excess of stillbirth, neonatal, and infant deaths at first cousin level compared to progeny of non-consanguineous couples of 1.5%, 1.1%, and 1.1%, respectively[163]. Autosomal recessive disorders are present at higher frequencies in the offspring of consanguineous marriages, with the rarer the disorder, the greater the relative influence of consanguinity on its expression[163].

Some conditions are more common among specific ethnic groups, or occur more often in a particular geographic location due to founder effects. For example, cystic fibrosis is more common among people from Western Europe, the Mediterranean region and the Middle East; sickle cell disease is more common among people of African origin; thalassaemia is more common among people of Mediterranean origin, the Middle East and South Asian origin; and Tay Sachs disease is more common among people of Ashkenazi Jewish origin.

Although there are very few examples of preconception genetic screening programmes, those that do exist are generally targeted at ethnic groups with a high incidence of a particular condition. A number of Middle Eastern and Mediterranean countries have put in place thalassaemia preconception carrier screening programmes[164], and data from Iran (where preconception carrier screening is part of a mandatory premarital blood test) suggests a 70% reduction in the thalassaemia birth rate is achievable[165]. Tay Sachs carrier screening programmes also exist in a number of countries with a large Ashkenazi Jewish population[166]. In the UK, carrier screening for Tay Sachs was approved in 1999 by the NSC and funded as an NHS service, but the coverage is currently rather low[167] and preconception carrier testing is principally provided by the voluntary sector.

There are numerous private tests available for carrier screening, either for specific diseases or a range of conditions, usually based on testing for a panel of relatively common known mutations. For example, various private organisations (such as Jewish Care in the UK and the international Dor Yeshorim) offer an Ashkenazi Jewish panel, which tests for carrier status of numerous disorders with an increased incidence amongst that population (including Bloom syndrome, cystic fibrosis and Gaucher disease, as well as Tay Sachs disease). Perhaps the most comprehensive example of preconception carrier screening currently available is offered by the American company Counsyl, which tests individuals or couples for over 100 autosomal recessive and X-linked Mendelian diseases and calculates a post-test probability of fetus inheriting each condition[168]. Counsyl estimate that although around 35% of individuals will be carriers of at least one disease on their panel, the frequency of carrier couples (in whom there would be a risk of conceiving an affected fetus) is only 0.6-0.8%[168].

An even more ‘universal’ carrier screening test is being developed by the National Center for Genome Resources in the US, in collaboration with the Beyond Batten Disease Foundation, which uses targeted enrichment of over 400 genes followed by deep sequencing to identify causal mutations for 448 autosomal recessive and X-linked disorders[169;170].

Although full genome sequencing may offer the only realistic option for a truly universal carrier test, filtering out unwanted, uninterpretable and irrelevant variants would present a major challenge. Moreover, because many inherited diseases are so individually rare, sequencing errors may be more frequent than the pathogenic variants themselves. Nonetheless, NGS technologies offer an unprecedented opportunity for comprehensive carrier screening.

5.3.2 Antenatal screening

There are already several well established antenatal genetic screening programmes in the UK, including the sickle cell and thalassaemia screening programme and the fetal anomaly programme. The latter is primarily aimed at screening for Down’s syndrome (trisomy 21), though also detects other clinically significant cases of fetal aneuploidy including Patau’s (trisomy 13) and Edward’s (trisomy 18) syndromes. The current screening protocol comprises several stages, involving blood tests for numerous maternal protein markers and a nuchal translucency ultrasound scan of the fetus[171]. If invasive diagnostic testing is indicated, fetal genetic material is assayed directly for the presence of an extra chromosome using either real-time PCR or karyotyping. Current uptake for screening in the UK is around 70%, and for diagnosis is around 80%. In 2007/8, 1,843 new cases of Down’s syndrome were diagnosed in England and Wales, 60% prenatally of which 93% were subsequently terminated[172]. Although the performance of the screening test has been improved substantially by earlier testing and better methods for combining the results, the false-positive rate is still relatively high[173] and it is likely that over 200 healthy fetuses are lost per annum due to miscarriage following invasive testing offered as a result of screening.

The existence of circulating cffDNA offers an alternative method of screening for Down's syndrome, which may be more accurate than the existing process. Several different methods have been proposed to allow quantification of cffDNA and determine if there has been a net increase (or decrease) of fetal DNA originating from a particular chromosome. Perhaps the most promising method is based on using NGS technology to sequence all cell-free DNA in maternal plasma[174;175]. By counting the number of reads that map to the chromosome of interest relative to a reference chromosome, a ratio can be calculated, which should be ~1:1 (for most chromosomes) in a normal fetus, and ~2:3 in an aneuploid fetus. However, because fetal DNA is only a small fraction of the total cell-free DNA circulating in the maternal plasma, a large number of reads is needed to confidently detect a small overall change in chromosome dosage, which makes the use of massively parallel sequencing technologies necessary. The same methodology could be used for simultaneous detection of fetal sex, any chromosomal aneuploidy and potentially small chromosomal abnormalities that result in a change in dosage.

Early indications are that this technique is very accurate[176], however further large-scale studies will be needed to determine its technical accuracy, clinical performance, feasibility, reliability and cost-effectiveness. Initially it is likely that this method will be used to augment or complement the current screening process, and may be offered only to individuals deemed to be at high risk. However, in the longer term, it is possible that it will eventually be proven to be accurate enough to be used diagnostically, thus replacing both the existing screening method and the need for invasive testing. This would involve a major change from a complex multi-step, probabilistic screening method where individuals must decide whether to undergo invasive testing following screening, to a single, potentially diagnostic test where individuals may be presented with a decision to terminate.

Additionally, because a considerable number of Down's syndrome fetuses could be diagnosed early in pregnancy that would otherwise have miscarried spontaneously prior to (invasive) diagnosis, the technique could increase both the number of women facing a decision regarding termination and the total number of terminations for fetal aneuploidy. Nonetheless, a non-invasive method for Down's syndrome testing is likely to be welcomed by both physicians and patients alike.

5.3.3 Neonatal screening

The NHS newborn bloodspot screening programme identifies babies who may have rare but serious conditions for which early treatment can improve their health and prevent severe disability or even death. The screening test uses filter paper cards to collect blood samples from a heel prick and transport them to laboratories, to determine if the level of various specific protein-based biomarkers are outside of their normal range. The UK NSC recommends that all babies in the UK are offered screening for phenylketonuria, congenital hypothyroidism, sickle cell disorders, cystic fibrosis and medium-chain acyl-CoA dehydrogenase deficiency. Other countries screen for a wider range of diseases – such as the United States which has as mandatory core panel of 29 conditions plus an additional 25 conditions that could be identified in the course of screening[177] – and consideration is now being given to somewhat expanding the set of inherited metabolic diseases included in the UK newborn screening programme[178].

Though controversial, many people believe that once DNA sequencing technology is sufficiently robust and affordable, all babies will have their genomes sequenced at birth replacing both newborn bloodspot screening and additional genetic tests required later in life. Disagreements over the value of genomic profiling of newborns have centred on the pace of technological development, delivery of clinical value, major ethical, legal and social implications such as safeguarding the autonomy of the future adult, and issues around data storage and access.

The concept was thoroughly explored by the Human Genetics Commission in their 2005 Report *Profiling the Newborn*[179], which concluded that genetic profiling could not be applied as an NHS screening programme in the near future (although the topic should be kept under review). At this time, we do not consider WGS of newborns to be a high clinical priority, and do not anticipate it replacing newborn screening within the UK within the next ten years.

5.3.4 Adult population screening

There are currently no established screening programmes for diseases with a strong heritable component in adults within the UK. However, some conditions may be sufficiently common, and have adult-onset preventable or treatable outcomes, to justify population screening, *e.g.* hereditary nonpolyposis colorectal cancer (HNPCC), hereditary haemachromatosis, adult polycystic kidney disease (APKD) or familial hypercholesterolaemia (FH). This could substantially reduce unnecessary morbidity and mortality; for example, although cascade testing is available for FH, the majority of individuals still remain undiagnosed in the UK. If WGS were to be used for testing, a panel of validated variants associated with the most common, serious and treatable Mendelian diseases could potentially be offered to the adult population.

5.4 Specific operational issues

Major barriers still exist to the implementation of WGS within clinical genetics practice including workflow issues, uneven or insufficient coverage of the genome, development of robust analysis pipelines, high capital costs and the need for training. Even when these issues are solved, WGS will not replace all genetic testing within clinical genetics practice and there will be a need to maintain standard targeted tests for some conditions such as Huntington's disease and other repeat expansion disorders. Nonetheless, it is likely that NGS technologies will be applied in clinical practice as soon as it becomes the most efficient and cost-effective way to analyse the genes of interest in individual patients.

It is probable that targeted *assays*, such as multigene panels and exome sequencing, are just a transition phase and whole genome sequencing will become the method of choice for genome-wide analysis once the price of sequencing drops below a certain threshold. However, targeted tests or *analyses* that answer specific clinical questions will likely still be preferred by the clinical community, as this approach fits with current clinical practice. An automated filtering system will therefore need to be developed that offers packages of tests based on clinical phenotypes, symptoms or specific age ranges, *e.g.* analysis of the 'dysmorphome' would include all variants that might be causally related to dysmorphic features observed in a child.

These clinical 'windows' into the genome would restrict analysis to relevant variants (based on contemporary scientific knowledge) and would largely solve issues around incidental findings and medical negligence. Such suites of analyses would need to able to evolve rapidly as knowledge grows, and will only be possible through the development of a reliable, curated database of variants and provision of sufficient bioinformatics expertise within the NHS.

Ultimately, the clinical geneticist may become more of a molecular pathologist, determining the molecular aetiology of a wide variety of diseases with a strong heritable component, and providing advice to other specialists and multidisciplinary teams. However, it is still unclear exactly how the clinical service will evolve, or how and by whom the laboratory and bioinformatics services will be organised, provided and accredited.

6 Applications in cancer

6.1 Cancer epidemiology

Around 200 different cancer types have been identified, each with its own unique epidemiology and characteristics. Cancers are leading causes of death worldwide, accounting for one in eight deaths (approximately 7.4 million deaths in 2004)[180]; the World Health Organisation estimates that this number will rise to over 12 million deaths by 2030[181].

In the UK, the average annual incidence of new cases of cancer was 298,000 (149,500 males and 148,500 females) for the years 2005-2007[182]. The majority (85%) of these cancers are solid epithelial tumours (for example, carcinomas of the lung, breast, and prostate) as opposed to haematological or connective tissue cancers. This is important because these common solid epithelial cancers are likely to be the main focus for NGS research and clinical applications, as haematological cancers are less genetically complex and are more amenable to alternative genetic diagnostics (see later).

The three most commonly diagnosed cancers in men are prostate, lung, and colorectal cancer, accounting for over half of all diagnoses. For women, the three most commonly diagnosed cancers are breast, lung, and colorectal, again accounting for more than half of all diagnoses in women. Improvements in cancer diagnosis and therapy over the last two decades have led to increased survival rates. At any one time, there are approximately two million people with cancer in the UK[183]. On average, there were around 154,500 UK deaths per year from cancer with 80,437 deaths in males and 74,056 in females during 2005-2007[182]. Lung cancer is the commonest cause of cancer death in both men and women. In men, death from lung cancer is followed by death from prostate and colorectal cancers. For females, breast and colorectal cancer are the second and third most common causes of cancer death.

6.2 Cancer as a disease of the genome

All cancers arise from the accumulation of pathological alterations to the genomes of somatic cells, resulting in disruption of normal cellular networks and control pathways. The chances of accumulating such mutations rise with increasing age; most cancers occur in people aged over 50. Most (around 90%) of these mutations are not present in the germline (or constitutional) DNA and are not passed on to the next generation[30]. Cancer is believed to develop when cellular DNA is altered or damaged in specific ways, allowing such cells to escape normal regulatory control mechanisms, and to proliferate, invade, and spread to distant sites of the body in an uncontrolled manner[184;185]. The acquisition of specific somatic mutations dictates how each cancer occurs and behaves.

Early evidence supporting the central role of the genome in the pathogenesis of cancer was based on the observation of extensive chromosomal aberrations in dividing cancer cells. This led to the proposal that cancers are abnormal clones of cells characterised by defects of hereditary material. This hypothesis was confirmed after the discovery of DNA and its identification as the agent of inheritance, and the fact that mutation-causing factors (mutagens such as irradiation and chemicals) could also cause cancer. Finally, the introduction of human cancer cell genomic DNA into specific mouse cell lines, was shown to convert phenotypically normal NIH3T3 cells into cancer cells. This led to the discovery of the first human cancer-causing gene, *HRAS*, in 1982[30]. Such cancer-causing genes are called oncogenes and many others have subsequently been discovered. Similarly, work on the very rare retinoblastoma (a malignancy affecting the retina of the eye) and other hereditary cancers led to the discovery of cancer-preventing genes, which are called tumour suppressor (TS) genes.

6.2.1 Cancer-associated genes

Cancer-associated genomic alterations tend to affect genes that control cell cycling or death (apoptosis) or protect the integrity of the cell's genome. To date, over 350 human cancer-associated genes have been identified, although the final number is likely to be much larger[30]. Most cancers are multifactorial diseases that are caused by the combined effects of many genes interacting with lifestyle and environmental factors. Such cancers occur haphazardly and are often called 'sporadic' for this reason[186]. However, it is worth noting that some of the genes identified in hereditary cancers are also important in the corresponding sporadic cancers (for example, the *APC* and *TP53* genes)[187]. This chapter focuses on genetic analysis of the somatic cells of sporadic cancers, whereas cancers with an inherited predisposition are dealt with in Chapter 5.

There are two main groups of cancer-associated genes, oncogenes and TS genes, although such a classification is likely to be an over-simplification. Approximately 90% of the known cancer-associated genes are dominantly-acting oncogenes[30]; mutation in only one copy of the gene leads to gene activation and is sufficient to transform the cell's normal phenotype to a malignant phenotype. The rest are recessively-acting TS genes; both copies of the gene must be inactivated to cause loss of function (one of which may be inherited). This is known as Knudson's two-hit hypothesis, as two successive mutations ('hits') are required to convert a normal cell to a malignant one[187].

Cancer-causing genes (oncogenes)

The normal function of these genes is to control cell proliferation, cycling and apoptosis. Activated oncogenes stimulate abnormal cell division and survival, leading to the development of a malignant cellular phenotype. Five broad classes of oncogenes can be distinguished (see Box 6.1). Oncogene activation occurs by a number of mechanisms, including:

- **Amplification**
  Many cancer cells contain multiple copies of structurally normal oncogenes, resulting in greatly increased gene expression. Examples include the *ERBB2* gene in breast cancer and *NMYC* in late-stage neuroblastomas.

- **Point mutations**
  Specific point mutations in the *RAS* family of genes are found in a number of human cancers including lung, breast, colon and bladder. These genes are responsible for cell signalling and mutations lead to an excessive cellular response from receptor signals.

- **Structural genomic alterations creating novel fusion genes**
  Cancer cells typically have very abnormal karyotypes, reflecting a general genomic instability, which is a hallmark of cancer. These rearrangements can include insertions, deletions, inversions and translocations. The most well-known is the Philadelphia translocation in chronic myeloid leukaemia (CML), which creates a novel fusion (or chimeric) gene (*BCR-ABL*) which does not respond to normal cellular controls.

- **Transposition away from normal regulatory controls**
  In Burkitt's lymphoma, a common feature is activation of the *MYC* oncogene. A series of characteristic translocations move the oncogene to a new location, thus releasing it from its normal upstream controls and leading to abnormally high gene expression in susceptible cells (in this case B-cell immunocytes).

This huge variety in cancer-causing genomic alterations means that any testing method has to be able to detect the full repertoire; this is one reason why NGS is so attractive because it is capable of detecting the full range of genome alterations in somatic cells.

Box 6.1 Functional types of oncogene products

1. **Growth factors**
   These stimulate abnormal cell growth when activated *e.g.* platelet derived growth factor (PGDF)
2. **Growth factor receptors**
   These generally make cells abnormally sensitive to growth factors *e.g.* ERBB2
3. **Signal transducers**
   These relay messages between growth factor receptors and the cell nucleus *e.g.* the *RAS* gene family
4. **Transcription factors**
   These control gene expression and protein production *e.g.* MYC
5. **Programmed cell death regulators**
   These prevent normal cell cycling when activated *e.g.* TP53

- **Cancer-preventing genes (tumour suppressing genes)**

  TS genes are responsible for restraining abnormal cell growth and division, causing cellular apoptosis and protecting the integrity of the cell's genome. Loss of function mutations in these genes contribute to the development of cancer by allowing uncontrolled cell growth. Examples include the *RB1* gene, *TP53* gene and the *MLH1* gene[187]. TS genes can also be silenced by epigenetic changes, such as methylation, which alter gene chromatin structure and gene expression. Important examples include the *CDKN2A*, *RB1* and *MLH1* genes. Genome-wide hypermethylation is seen in many cancers and it should be noted that standard DNA sequencing techniques for mutation screening do not detect such epigenetic changes. There are specialized methods for detecting such changes, although some third-generation technologies may be able to detect DNA methylation directly[188].

- **Other important genes in cancer**

  DNA repair and mismatch repair genes help to maintain the stability of the genome and prevent errors during replication or caused by external factors, such as irradiation, ultraviolet light or chemical mutagens. Loss of function impairs cells' ability to repair mutations and thus accelerates the progression to malignancy. Mismatch repair gene mutations are a common feature in hereditary non-polyposis colorectal cancer (HNPCC) and in about 13% of sporadic colorectal, gastric and endometrial cancers[187]. There is also a group of genes that are responsible for maintaining chromosomal stability. Loss of function in these genes is thought to cause the chromosomal instability commonly seen in cancer cells, which often have highly rearranged genomes.

6.2.2 Clonal evolution

One of the strongest pieces of evidence for the genetic basis of cancer at the cellular level is that nearly all cancers are clonal; that is, they arise from a single common ancestor and thus share a common genotype. From an evolutionary perspective, cancers can be viewed as large (epi)-genetically heterogeneous populations of cells (clones). Changes that are beneficial to malignant clones are usually harmful to the host, ultimately causing death of both the host and the neoplasm. This microevolution occurs in several stages, with each successive mutation leading to a survival advantage for the cell's descendants. Clonal evolution normally selects for survival and proliferation and might lead to invasion, metastasis and treatment resistance[189].

Molecular analyses of leukaemias and lymphomas usually demonstrate that all of the cancer cells contain the same rearrangement of immunoglobulin or T-cell receptor genes, indicating descent from a common ancestor. Similarly, certain cancer types show consistent genomic rearrangements at the chromosome level, such as the 'Philadelphia' translocation between chromosomes 9 and 22, which is found in over 90% of patients with chronic myeloid leukaemia (CML). The situation in solid tumours is more complex; whilst cancer remains clonal, a number of clones can be found within the same neoplasm and within an individual person[189]. Because cancer is the inevitable end-result of clonal evolution rather than the result of a specific disease process, cancers of a given type do not all have the same mutations in a standard set of genes, although there may be certain commonalities[187]. However, the requirements of genomic instability, clonal survival and proliferation impose a degree of regularity on the type (and possibly sequence) of necessary genomic alterations.

Two recent studies have suggested that cancer genomes evolve as the disease progresses. Sequencing of two primary breast cancers and their metastases have shown novel mutations in metastases or enrichment of previously low frequency mutations. These may provide new targets for the development of therapeutics that could be effective in cancer relapses[190].

6.2.3 The cancer cell genome

A wide range of cytogenetic, genetic and epigenetic alterations are found in cancer genomes, all of which are acquired over the cancer patient's lifetime. The cancer genome comprises a copy of the patient's constitutional (germline) DNA upon which a unique set of acquired somatic genomic alterations have been superimposed. These changes constitute the 'genetic fingerprint' of each cancer type in individual patients. They include point mutations, copy number variations, insertions, deletions, and rearrangements. Previously, each of these genomic alterations would require its own method for detection but one of the principal advantages of NGS is that it can detect the full range of alterations in a single assay with high sensitivity and increasing specificity.

Certain genes are frequently mutated in a diverse range of cancer types (such as *KRAS* and *TP53*) whereas others are restricted to only one cancer type (such as *RB1*). In some cancer types, abnormalities in several cancer-associated genes are common, whereas in others (such as gastric cancer) relatively few mutations in known cancer genes are found. Genetic heterogeneity has important practical implications. For example, the lack of recurrent mutations across cancers of a given type may hamper the development of new therapeutic agents targeted at specific cancer cell processes. However, heterogeneity allows the development of personalised biomarkers that can be used to individualise prediction of a patient's prognosis and likely response to treatment which may ultimately improve clinical outcomes (see below).

- **Cancer-associated mutations: drivers and passengers**

  Some somatic mutations will be cancer-causing and are known as 'driver' mutations whilst others are not (so-called 'passenger' mutations). Driver mutations confer growth or survival advantage to the cell and have been positively selected for during clonal evolution. Conversely, passenger mutations are neutral somatic mutations that have been acquired from errors in cell division, DNA replication and repair. They have not been selected for, and are simply carried along in subsequent clonal expansions. Some of these mutations are acquired by the cell when it is phenotypically normal whilst others are acquired after there is evidence of neoplastic change. Although passenger mutations may not influence tumour growth, they may help to provide insights into the nature of mutagenic exposures and for the quantification of cancer burden[191;192].

  One important subclass of a driver is a mutation which confers resistance to chemotherapeutic agents. Treatment-resistance mutations are typically found in recurrences of cancers which responded to initial treatment but became resistant later on. Some resistance mutations may initially be classified as passengers in certain early clonal populations because they do not contribute directly to survival or proliferation. However, when the patient is subjected to chemotherapy, clones without the resistance mutation die, whilst those with the mutation survive. The passenger becomes a driver and the resistant clone preferentially expands, leading to later recurrence. The implication is that early intervention may be more likely to prevent the emergence of treatment-resistance mutations, although some cancer treatments may be directly mutagenic[30].

  One of the central challenges of cancer genome analysis is distinguishing driver from passenger mutations, a problem that will be exacerbated by NGS methods and particularly whole-cancer genome sequencing, as most of the cancer genomes already sequenced have not been complete. For example, the full genome sequence of a lung cancer cell line yielded nearly 23,000 point mutations, of which 134 were in protein-coding regions[11]; deciding which are drivers and which are passengers is an immense bioinformatics and interpretational challenge, especially as the number of drivers is likely to be very small (see below).

- **How many mutations are needed to cause cancer?**

  It is generally accepted that the development of cancer is driven by only few, but significant, genetic changes[189]. It is not known how many driver mutations are necessary and sufficient for the development of cancer, although estimates range from 3-12 for different cancer types[193]. There is evidence suggesting that the common adult epithelial cancers (such as breast, colorectal and prostate) require more mutational events than childhood cancers and blood malignancies[193]. These estimates are supported by cell biology experiments, which require changes in the functions of five to seven genes to transform normal human somatic cells into cancer cells[30].

- **Does the rate of mutation increase in cancer clones?**

  The existence of mutation-increasing mutator phenotypes has been found in some colorectal and endometrial cancers, as a result of defective DNA repair caused by mutations or epigenetic changes in the MLH1 and MSH2 genes[194]. This genetic instability may allow a sudden widening of diverse cell populations upon which clonal selection can act and thus increase fitness. However, other studies suggest that normal mutation rates are sufficient to account for the development of many cancers[195;196]. The acquisition of somatic mutations over time is not linear and sudden increases in mutation rates can occur (crisis episodes). These episodes may be caused by the sudden attrition of telomeres which normally protect the ends of chromosomes and require substantial genomic rearrangement for the cell to be able to survive[197]. Somatic mutation rates are not well understood but it is likely that they are characteristic of each structural class of mutations and that they differ within cell types. Better understanding of the mutation rate might provide a biomarker that could be used to direct patient surveillance, assess prognosis, and to measure the effect of therapeutic interventions[189].

6.3 Cancer genome profiling using NGS

The increasing recognition that cancer is fundamentally a disease of the genome has led to the wider use of genome-based technologies in diagnosis, management and prognosis. The application of microarrays and capillary-based Sanger sequencing, has led to a better understanding of genomic alterations in cancer, but NGS methods have the potential to further our understanding of cancer genomes. However, cancer samples and genomes have a number of specific problems that are distinct from other tissue samples and from the genomic analysis of constitutional (germline) DNA[198].

6.3.1 Sample-related issues

Cancer samples differ considerably in both quantity and quality from tissue samples typically used for analysis of germline DNA. 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. DNA obtained from cancer samples may also be of a poorer quality, either because histopathology samples are fixed in formalin and paraffin-embedded, or because of biological factors, such as the extensive tissue necrosis found in many cancers. The purity of cancer samples is further compromised by admixture between malignant and non-malignant cells (such as normal tissue, fibroblasts, and infiltrating lymphocytes). Some of these problems may be ameliorated by micro-dissection and laser capture, but they often yield tiny amounts of DNA. Nevertheless, the many-fold coverage made possible by NGS methods (especially targeted sequencing using sample enrichment) means that high-quality data can be obtained from such samples, although this adds to cost and analytical complexity. Another advantage is that a wide range of body sample types (including blood, urine and faeces) can also be investigated using NGS.

6.3.2 Genetic heterogeneity and clonal evolution

Cancers are often highly heterogeneous. This heterogeneity exists between cancer types and between cancers of the same type in different individuals, which may contain different clones and different genomes. A single biopsy sample in a patient may not be representative of the tumour as a whole and the tumour will continue to change after the biopsy is taken[189]. Analytical models must therefore be able to take account of different types of heterogeneity: cancer versus normal heterogeneity and within-cancer heterogeneity. Again, the many-fold coverage made possible by NGS allows even rare clones to be detected. For the foreseeable future, this means that a matched DNA sample from the same individual is an essential and integral part of cancer genome profiling, increasing the complexity, cost and ethical implications of the analytical process. Cancer genomes are not static and evolve; this may lead to the need for later resequencing when tumours recur or metastasise, especially if these occurrences are caused by new mutations that could be used to guide clinical management and assessment of prognosis.

6.3.3 Bioinformatics

As noted above, NGS methods can identify the full repertoire of cancer-associated genome alterations as well as the identification of clonal and subclonal diversity within cancer-cell populations. This may be important in identifying clones with treatment-resistant mutations (for example). However, this tremendous capability creates an immense analytical and interpretational challenge because of the very large number of somatic mutations NGS can identify. Separating causal alterations from non-causal alterations in an unstable and evolving genome is problematic. The main bioinformatics challenges for cancer genome profiling and NGS include[198-200]:

- The need to simultaneously analyse cancer and patient-matched normal genomes, compared to reference sequences. The choice of assembly and alignment method depends on the sequencing platform, data quality and computational resources.
- The ability to analyse the very different and highly rearranged genomes of cancer cells. The uniqueness of every cancer genome and the difficulty of correctly assigning sequences from homologous regions relative to constitutional reference DNA mean that *de novo* assembly of cancer genomes may be the most powerful approach, although it is computationally extremely intensive and laborious.

- The ability to account for unknown levels of normal tissue admixture and within-tumour variability. The variable purity and heterogenous karyotypes of cancer samples makes somatic mutation calling much more difficult than germline mutation calling.
- The ability to distinguish driver from passenger mutations. NGS brings a totally new dimension to this problem. Part of the challenge is that passenger mutations far outnumber drivers (as noted above). Large numbers of samples will need to be analysed to distinguish between infrequently mutated cancer genes and genes with clusters of passenger mutations.

6.3.4 Experimental approaches

The application of NGS to cancer genome profiling has allowed cancer genomics to move from focused approaches – such as single-gene sequencing or array-based analyses – to comprehensive genome-wide analysis. As described in Chapter 3, NGS methods vary by the type of input material, the extent of the genome that is analysed and the type of genomic alteration studied. Three main types of NGS are used for cancer genome profiling: whole genome sequencing (WGS), targeted genome sequencing and sequencing of transcribed RNA (RNA-Seq). Their main features are summarised in Table 6.1.

Table 6.1: NGS and cancer genome profiling

	Whole genome sequencing (WGS)	**Targeted sequencing**	**Transcriptome sequencing (RNA-Seq)**
Scope	Whole genome	Regions of strong interest (*e.g.* exons, non-coding RNAs)	Transcribed regions
Analyte	DNA	Enriched DNA or cDNA	cDNA
Typical applications	Detection of all genomic alterations (somatic mutations, insertions/ deletions, copy number variants, genomic rearrangements, non-human DNA)	Detection of all genomic alterations in sequenced regions	Detection of somatic mutations, fusion genes, alternative splicing, and abnormal allele expression at the transcript level
Main advantages	Complete identification of all somatic cancer genomic alterations Most powerful method for detecting rearrangements	Much higher coverage of targeted regions Lower cost than WGS Very useful in cancer samples of mixed purity and quality	Digital gene expression profiling Detection of novel transcripts with low levels of expression Detection of non-mutational events Not limited to known genes
Main disadvantages	Cost and time Bioinformatics complexity because of large number of detected alterations Only a few complete cancer genomes have been sequenced to date	More input material required compared to WGS Does not provide complete picture of all genomic arrangements Enrichment method determines range of detectable alterations (*e.g.* PCR cannot detect novel structural variants)	More input material required Does not provide complete picture of all genomic arrangements Sensitivity is limited by expression level of altered gene(s) Finding appropriate matched normal samples for cancer cells

WGS has the potential to identify all genomic alterations in a cancer sample using a single experimental approach. However, although it is the most comprehensive approach, it is also the most complex and technically challenging, especially in the identification of putative causal mutations.

Potential advantages of WGS in cancer are that it can identify chromosomal rearrangements in complex karyotypes with base-level resolution and genomic alterations undetectable using previous methods (such as PCR). These include somatic mutations in non-coding regions and rearrangements of repetitive elements[199]. As with all NGS methods, its other principle advantages over earlier technologies are digital measurement and the ability to oversample the genome, providing highly accurate sequence information which was previously impossible. Only a limited number of complete cancer genomes have been sequenced to date, including acute myeloid leukaemia, lobular and basal-like breast cancer, small cell lung cancer, malignant melanoma, and glioblastoma (amongst others)[200]. This is likely to change as a result of formation of the International Cancer Genome Consortium (ICGC), which was established in 2008 to coordinate international efforts to sequence cancer genomes (see Box 6.2). Based on the Human Genome Project's collaborative model, it is hoped that this project will be able to accelerate cancer genome sequencing and spread the costs.

The current cost and complexity of WGS means that targeted sequencing approaches are popular, by focusing on regions of particular interest. These include strategies for enriching the input to sequencing, by targeting specific regions using PCR or hybridisation (see Chapter 3). This means that areas of interest can be analysed with much higher coverage, which can overcome the problems of sample quality and purity often found in cancer samples. However, targeted sequencing requires more input material than WGS and is, by definition, limited to detecting alterations only in the targeted regions, which may themselves be imperfectly defined. Although PCR-based methods are well understood, they are unable to detect novel structural arrangements in cancer genomes because of the unknown surrounding sequence[199]. Increasingly PCR is therefore being replaced by hybridisation-based approaches, which can detect the full range of genomic alterations in targeted regions. However, coverage of targeted regions can be variable and there are also specific analytical issues that may cause interpretational difficulties[198-200]. Also, given the decreasing costs of WGS, it is likely that targeted sequencing will only be a transitional phase until WGS can be fully implemented. In the future, it is likely that only the bioinformatics analysis will be targeted, as it will be much quicker and cheaper to characterise the whole genome in one assay.

Box 6.2: International cancer genome sequencing and application projects

The International Cancer Genome Consortium (ICGC)

The primary goal of the ICGC is to generate comprehensive catalogues of genomic alterations in human cancers. It aims to sequence 500 cancer genomes from each of 50 cancer types. The estimated cost of the project is around US$1 billion. It comprises two older projects, the UK-based Cancer Genome Project (CGP) and the US-based The Cancer Genome Atlas (TGCA). The CGP is working primarily on breast cancer, whilst TGCA has recently completed a three-year, three-cancer (glioblastoma, lung, and ovarian) pilot. Eleven countries have now joined, and work is proceeding on around 20 different cancer genomes.

The Genomics of Drug Sensitivity in Cancer Project

This is part of a five-year collaboration between the CGP and the Center for Molecular Therapeutics, Massachusetts. The aim of this project is to discover new cancer biomarkers that define subsets of drug-sensitive patients. As part of this collaboration, a range of anti-cancer therapeutics will be screened against a large number of genetically characterized human cancer cell lines and drug sensitivity will be correlated with extensive genetic data. More than 1000 human cancer cell lines will be screened with a wide range of anti-cancer treatments.

Catalogue Of Somatic Mutations In Cancer (COSMIC)

COSMIC has been established by the Sanger Institute (Cambridge, UK) to store and display information about somatic mutations in cancer. Information about cancer genes and mutations is extracted from published scientific literature and stored in a publicly available online database.

Another evolving technology is transcriptome sequencing (RNA-Seq) which is a powerful tool for examining gene expression and can detect somatic mutations, gene fusions, and other non-mutational events. Its most notable success has been the discovery of the FOXL2 gene in ovarian granulosa cell tumours[199;201]. RNA-Seq is able to analyse gene expression with a much greater sensitivity than microarray analysis and is not limited to known genes. However, finding normal control cells can be a challenge, as they are unlikely to express exactly the same genes as a cancer cell[198]. RNA-Seq may soon be able to compete with microarray based gene expression profiling in terms of the cost and efficiency of analysis. One advantage is that PCR-based sequence amplification is not required so it will be possible to sequence single molecules and provide the most accurate quantification of transcripts. Expression profiling is likely to remain as an important tool in cancer management (especially in assessing prognosis), as it provides different information about cancer behaviour than the sequence data.

6.3.5 NGS versus microarrays

Microarrays are generally used for the simultaneous analysis of very large numbers of genes (which can be genome-wide) and their expression. They have been widely used in cancer genomics research and more recently, in clinical cancer management. In addition, they are highly customisable, have mature bioinformatics pipelines, and are inexpensive. However, they do have some important limitations[188]:

- Microarray design requires *a priori* knowledge of the genome or genomic features. This is an important limitation if knowledge is incomplete, incorrect, or out of date. This is especially problematic if the genomic landscape is unique or bizarre (as in many cancer genomes).
- Cross-hybridisation between repetitive sequences limits analysis to non-repetitive genomic regions and can limit analysis of related genes, alternatively spliced transcripts, allelic gene variants and SNPs.
- The greater noise introduced by analogue measurement (relative signal intensity) limits the dynamic detection range, meaning that low abundance sequence detection and quantitative resolution is challenging. This may be an issue in cancer, where low-level expression of certain genes might be important.
- Micrograms of DNA are required and PCR-based amplification can introduce bias. Although cancer resection samples may be large, they may be of poor quality. Diagnostic biopsy samples can be very small and of very variable quality, and are often admixed with non-cancer cells.

NGS offers potential solutions to some of these problems:

- Although helpful, genome annotation is not required; it has been possible to assemble cancer genomes *de novo*, but this is a laborious and costly process.
- Direct digital sequence analysis removes potential bias introduced by cross-hybridisation of user-defined sequences. NGS also offers single nucleotide resolution. As read lengths increase, the ability to examine repetitive genome sequences will increase and coverage requirements will be reduced.
- Quantification of signal from sequence-based approaches is digital because it is based on single nucleotide sequencing resolution (for detecting genetic variants) or on counting the numbers of sequence tags (for analysing gene transcripts and gene expression), rather than relative measures. It thus gives unlimited and fully quantitative dynamic signal range.
- Only nanograms of input material are required, eliminating the reliance on PCR amplification (with its inherent biases and higher error rates); cancer samples may only contain very small amounts of input material. Also, even small cancer samples contain heterogeneous clonal cell populations, which is problematic for first-generation sequencing technologies, where error rates are high. Because NGS does not require PCR amplification and because NGS-based analysis is digital, it provides much more accurate analysis of small, heterogeneous cancer samples than first-generation sequencing technologies.

However, until the practical, bioinformatics and cost challenges of providing NGS-derived cancer genome data within clinically relevant timeframes have been resolved, it is likely that microarrays and NGS may be used together, at least in the shorter term[188]. For human cancers that are homogeneous and have common, recurrent genomic alterations (such as many haematological malignancies) the use of NGS might be considered 'overkill', because other cytogenetic technologies (either array-based, cytogenetics, or FISH) can provide the required data more cost-effectively. For more complex cancers, with much more genomic variation and very few recurrent mutations (such as most solid epithelial cancers), NGS is likely to be more useful.

For example, it may be possible for targeted, cancer type-specific genome profiling microarrays to be developed from WGS of large numbers of cancer genomes. These arrays could be used standalone or for initial 'screening' of cancer samples to identify those that contain more frequent recurrent mutations and those that do not; such samples might benefit from deeper sequencing using NGS. Roche already produces 'Exome tiling arrays' that can be used to capture and release all annotated exons in the human genome, which can enrich protein-encoding regions before NGS[188]. Although these issues may be negated by the rate of technological progress and the rapidly decreasing costs of NGS, it is vital that the added value of NGS over other types of genomic analysis and conventional diagnostics are thoroughly evaluated before widespread clinical implementation.

6.4 NGS and cancer management

There is an unprecedented potential for somatic genetics and NGS to revolutionise the classification, diagnosis, and management of human cancers in individual patients[201]. There are a number of examples where classification and treatment protocols are already defined by the presence of abnormal genes (see Box 6.3). Most success has been achieved in haematological malignancies, which typically have highly recurrent and cancer-specific somatic mutations. Less success has been achieved in solid epithelial tumours, although there are notable exceptions (for example, in breast cancer). The availability of new technologies, such as NGS, is likely to increase the range of possible applications in many different cancer types.

Box 6.3: Examples of current uses of genetics in clinical practice

Diagnosis and classification
Acute myeloid leukaemia is now defined by mutations in genes such as *FTL3* and *KIT* and other cytogenetic abnormalities. These subgroups influence both prognosis and treatment.

Drug treatment
Around 15% of breast cancers over-express the *HER2* gene. The risk of disease recurrence in women with HER2-positive tumours, and treated with Herceptin, is markedly reduced. Similarly, the use of tyrosine kinase inhibitors (such as Imatinib) in patients with certain haematological malignancies and some lung cancers are guided by the results of gene expression profiling.

Assessing response to treatment
A number of patient-specific immunoglobulin and T-cell receptor gene rearrangements can be used during the treatment of acute lymphoblastic leukaemia to guide the choice and intensity of subsequent treatment.

Detection of recurrence
Levels of the *BCR-ABL1* fusion gene are monitored in patients with CML to predict disease relapse and to enable earlier intervention.

Cancer management typically comprises a number of inter-linked processes:

- Diagnosis, classification, and grading of cancer type

- Measurement of the extent and severity of disease (often called 'staging') and assessment of prognosis
- Predicting and monitoring response to the chosen therapies
- Post-treatment surveillance to detect cancer recurrence

Cancer management is complex, requiring careful coordination between many different clinical and laboratory specialties (see below). NGS may have three main, interrelated benefits for cancer management:

- Molecular diagnosis and classification of cancer types in individual patients, based on comprehensive and complete molecular profiling of their cancer's genome (so-called molecular fingerprinting)
- Individualisation of cancer management and prognosis: current management and prognostic assessments are based on probabilistic measures derived from large samples. A major goal of cancer management is to obtain more accurate measures of individual prognosis and likely treatment response than currently available
- Streamlined diagnostic and management pathways: faster, less invasive but more clinically useful diagnostics, less invasive treatment monitoring, personalised detection of recurrence

6.4.1 Cancer diagnosis and classification: digital molecular diagnostics

Accurate diagnosis of cancer type is essential for the accurate estimation of prognosis and the choice and intensity of treatments known to be effective. Histo-pathological analysis of morphological characteristics has been the mainstay of diagnostic and prognostic assessment of cancer samples, especially of solid epithelial cancers, revealing valuable information about tumour differentiation, aggressiveness, and recurrence risk. More recently, this has been supplemented by advances in immuno-histochemistry (such as detection of oestrogen receptor status and HER2 immunoreactivity in breast cancer). However, although morphology and individual biomarkers can provide important information, they are limited in their ability to predict individual clinical outcomes, even in histologically similar tumours. They can only identify a small fraction of the genomic changes that occur in a typical cancer[201]. The highly individual genomic nature of cancer, and our ability to measure it, means that cancer genome profiling is beginning to revolutionise cancer diagnostics.

6.4.2 Molecular Subtyping

The variability of clinical outcomes in patients with morphologically similar disease may be explained by underlying differences in cancer genomes. As we have already noted, each cancer is unique and even cancers of the same type often show extensive genomic heterogeneity. Gene expression studies have led to the identification of molecular subtypes in a number of cancers; for example, breast cancer and diffuse B-cell lymphoma. Breast cancer has been classified into five molecular subtypes that are strongly associated with clinical outcome[190]. It has also been possible to allocate 11 of 17 'special' breast cancer subtypes (which together account for around 25% of all clinical cases) to just one molecular subtype. In most common solid epithelial tumours (such as lung cancer) although molecular subtypes have been identified, there is no apparent prognostic significance. Such studies have been limited by sample size and the extent of genomic heterogeneity[201].

It is possible that NGS of many cancer genomes will provide a much wider range of relevant cancer-associated genomic alterations for classifying subtypes and prognostic signatures, especially as less *a priori* genomic knowledge is required than other approaches (such as microarrays)[188;198;199].

6.4.3 Cancer of unknown primary

Surprisingly, carcinoma of unknown primary (CUP) has been the most commercially developed profiling application after breast cancer. CUP is one of the commonest causes of death in Western countries, and clinical outcomes are very poor, often because diagnosis occurs after metastasis, probably because such cancers may be unusually aggressive. Lack of identification of the likely primary site means that type-specific treatment cannot be given. It is possible that molecular profiling may provide one way of improving the diagnosis and ultimately, outcomes. It may also streamline the diagnostic process and require less intensive, painful and costly investigations using conventional diagnostics[190]. Current CUP assays use microarrays based on genes identified from primary tumours of known origin. In this situation, NGS and especially WGS, may prove to be more informative and powerful as it will identify the full range of genomic alterations present in such cases, rather than relying on a limited range of known genomic changes[201].

6.4.4 Individualised prognosis and treatment

Clinically useful cancer genome profiling tools could be an important step towards individualising patient care and estimating prognosis. Prognosis is currently estimated using indices that combine clinical and pathological characteristics. However it is possible that genome profiling tools could herald a shift from morphological to molecular diagnosis and prognostication. This would also entail streamlining of clinical, diagnostic and treatment pathways by using a single, highly informative genetic test.

6.4.5 Individualised prognosis

A number of commercially available diagnostic tools are already available for breast cancer, including MammaPrint® (a microarray-based, 70-gene test) and OncotypeDx® (a 21-gene PCR-based test), amongst others. Interestingly, there is little overlap between the genes identified as prognostic signatures, an issue that is unlikely to be problematic if WGS were to be used instead. MammaPrint® is being trialed as a replacement prognostic tool instead of using imaging, histopathology, and staging in women with node-negative early breast cancer (the MINDACT trial); a similar trial is underway for OncotypeDx® (the TAILORx trial)[201]. The aim is to identify women at high risk of relapse and who may benefit from adjuvant treatment. However, there has been less success in generating similar tools for other solid epithelial tumours, but this may change with the advent of NGS, which can provide access to the complete mutational landscape of tumours, and international consortia working on sequencing large numbers of cancer genomes. All such tools will need extensive evaluation in large, prospective randomised trials before being introduced into routine clinical practice, to establish whether they provide greater clinical utility than conventional indices.

6.4.6 Individualised treatment

It is anticipated that molecular profiling will not only be useful for individualised prognosis but also for selecting individualised, optimal treatments from the array of available interventions. Such information could be used to guide the mode, intensity, duration of treatment and the detection of tumour recurrence. Given that many cancer treatments have unpleasant and debilitating side effects, and can be very expensive, it is important that patients with a good prognosis are not over treated; conversely, it is also critical that patients with poorer prognoses are not undertreated. Thus, individualised treatment requires individualised diagnostics. We have already given some examples where somatic genetic information is being used to guide chemotherapy decisions (such as HER2, ABL, KIT, KRAS, and EGFR)[202]. It is likely that the increased power of NGS will identify novel cancer-associated genes that may be suitable therapeutic targets. For example, NGS was used to find the IDH1 gene in a number of tumours (colorectal, glioblastoma and acute myeloid leukaemia)[203;204]. Mutation of this gene leads to accumulation of cancer-promoting metabolites in cells and at least one pharmaceutical company is aiming to discover a drug that will stop the process. It is also possible that as NGS is applied to more cancer types and genomes, new low-frequency cancer-associated genes will represent important therapeutic targets.

Another area where NGS may have an important role is in identifying cancer treatment resistance mutations. These are typically found in recurrences of cancers that initially responded to treatment but are now resistant. Treatment resistant clones have been discovered in lung cancer, CML and colorectal cancer. There is evidence that some of these pre-date initiation of treatment, perhaps existing as passenger mutations in cancer sub clones until the selective environment is changed by chemotherapy. If these mutations could be discovered at an early stage, it may be possible to adapt treatment regimens to compensate, by using different multi-drug regimens (for example). NGS could have an important role because it can map all cancer genome changes and so detect resistant mutations earlier[30;189].

6.4.7 Individualised cancer biomarkers for monitoring treatment and recurrence

To ensure optimal treatment, clinicians require accurate and sensitive methods for quantifying post-treatment residual disease burden and for detecting early recurrence in individual patients. This information can be used to guide both the mode and intensity of treatment and long-term clinical management. Highly sensitive assays for monitoring minimal residual disease have become standard practice in many haematological malignancies, allowing personalised treatment decisions. This has been possible because of the highly recurrent somatic genomic rearrangements found in such cancers (such as the BCR-ABL gene and T-cell receptor genes). At diagnosis, PCR-based assays for specific, personalised genomic alterations are developed from the leukaemic cells which are then used to identify residual circulating cells carrying the targeted alterations. These are then used to monitor treatment outcome and risks of recurrence[204].

Such approaches have yet not been implemented in solid epithelial tumours because they generally do not have the highly recurrent genomic rearrangements found in haematological cancers. However, the genomes of most tumours do carry unique rearrangements which could be used as personalised biomarkers. A recent genome-wide analysis of 24 breast cancers showed that each one contained at least one genomic rearrangement that could be detected using NGS[205]. Many types of solid epithelial tumours release cell-free DNA fragments into plasma, which could be detected by a method with sufficient sensitivity. The use of NGS has allowed the detection of genome-wide rearrangements down to base-pair resolution and given that it requires only small amounts of input material, it is possible that these rearrangements could be detected in peripheral blood samples. McBride and colleagues have already reported such an approach in two patients with breast cancer and one with osteosarcoma[191] whilst Leary and colleagues used a different method in patients with four patients with colorectal and two breast cancer samples[192].

Such techniques have an array of applications with immediate clinical utility; for example, they could be used to measure the amount of circulating tumour DNA after surgery, radiotherapy or chemotherapy, whilst short term monitoring could be used to guide treatment intensity. This could be especially important when toxic, experimental or expensive treatments are being undertaken; analyses could also be undertaken in real time. The earlier detection of recurrence afforded by personalised biomarkers would enable more effective pre-emptive action. The methods could be extended to assess adequacy of surgical resection and the analysis of regional lymph nodes, which could be very useful in tumour staging. These methods also have important benefits over conventional diagnostics:

1. They are minimally invasive
2. They are highly sensitive; for example, radiological imaging is limited by the size of detectable lesions (approximately 0.5 cm)
3. They are highly specific; for example, serum biomarkers, such as PSA, are only available for a limited range of cancers and are often non-specific
4. Conventional diagnostics are analogue, not digital, measurements

These techniques also have a number of benefits over point mutation detection in their ease of implementation, accuracy, sensitivity and specificity. Although initial rearrangement screens need to be undertaken for each patient, once detected, subsequent analyses could be performed quickly (within a day) and in most molecular pathology laboratories.

These particular applications of NGS are likely to be ones that will be translated to clinical practice at a greater speed than some of the others. For example, the Wellcome Sanger Trust has recently obtained funding from the Health Innovation Challenge Fund to develop this application in patients with breast and colorectal cancer[6]. Finally, as noted earlier, it is likely that it will be possible to extract DNA from a range of body tissues and fluids (such as urine and faeces), providing an additional source of data for monitoring disease burden. Variations between cancers will give different sensitivities.

6.4.8 Screening

Detecting cancer at a pre-symptomatic stage is one way in which clinical outcomes could be improved. A number of cancer screening programmes are in use, but there are still a number of common epithelial tumours which cannot be screened for. It is possible that the approaches described above could eventually be extended to screen for cancer in peripheral blood samples, urine or faeces (for example, for detecting early bladder and colorectal cancers). Although the method described above can only be used in cancer patients with a known genomic rearrangement, if recurrent mutations or stable chromosomal rearrangements were discovered in specific cancer types, similar methods could be used for screening. Of course, the major limitation with this approach is that such recurrent mutations are apparently not common in solid epithelial tumours. More widespread use of NGS in sequencing large numbers of genomes (as envisioned by the ICGC) may change this. The discovery of the IDH1 gene, using NGS, in a number of human cancers may provide one example of the type of recurrent mutation which could be a target for screening applications. However, we do not envisage NGS having an impact on cancer screening in the short term.

6.5 NGS and NHS cancer services

NGS applications in cancer management must be thoroughly evaluated before they can be adopted by the NHS. Evaluation will comprise ensuring that the correlation between biomarkers and cancer behaviour are adequately determined, as well as conducting high-quality randomised trials and comprehensive assessments of the health economic implications. If NGS is subsequently shown to be an effective and cost-effective technology in cancer diagnostics and management, only then can it be integrated into clinical pathways. This would have important implications for the provision and organisation of both clinical and laboratory services.

6.5.1 Cancer services, NHS organisation and policy

The NHS Cancer Programme was established as a result of the previous government's strategies to improve the prevention and management of cancer. Three bodies currently support this programme: the National Cancer Action Team (which oversees overall policy implementation), NHS Improvement (which oversees implementation of National Service Frameworks), and the National Cancer Intelligence Network (which aims to identify relevant research to improve cancer outcomes). At a regional level, service commissioners and providers are supported by Cancer Networks. Networks are organised on a regional basis (populations of 1-2 million) and they bring together local providers, service commissioners, voluntary groups and local authorities. Often, there is an Experimental Cancer Medicine Centre based within these networks. These 19 centres, funded in partnership with Cancer Research UK (CRUK), are working to introduce new cancer therapies (for example, CRUK's Stratified Medicines Programme – see Box 6.4).

Over the last decade, cancer has been the subject of extensive strategic reviews, most recently with the publication of the Coalition Government's "Improving Outcomes: A Strategy for Cancer"[206] in January 2011. This new strategy will lead to further changes, as will the extensive reforms in NHS organisation and funding that the new government is also planning. Although these changes are beyond the scope of this document, the introduction of new technologies, such as NGS, are

[6] www.wellcome.ac.uk/Funding/Technology-transfer/Funded-projects/Health-Innovation-Challenge-Fund/index.htm

further drivers for change, and will most likely result in radical changes to clinical pathways, service provision and organisation.

Box 6.4: Cancer Research UK's Stratified Medicines Programme

The primary aim of this programme is to establish a national service that makes standardised, high-quality, and cost-effective cancer somatic genetic testing available for people with cancer. This should enable better targeting of both currently available and new treatments to individual cancer patients. It will also establish a national database of somatic cancer genetic information that is linked to individual patients, cancer treatments, and clinical outcomes. This will facilitate the design of new and more effective ways of dealing with cancer. In partnership with the government, Astra Zeneca, and Pfizer, CRUK have designed a two-phase programme to implement these aims. In Phase One of the programme, running from 2011-2013, CRUK will develop a model of standardised testing that can be adopted by the wider NHS during Phase Two. Up to 9,000 people, with six cancer types (breast, bowel, lung, ovarian, prostate and melanoma), will be recruited from Experimental Cancer Medicine Centres. Cancer samples from these patients will be tested in three experimental technology hubs. It is anticipated that this programme will provide a platform for piloting and evaluating the adoption of NGS technologies in cancer diagnostics by the NHS.

6.5.2 Clinical services

The core concept of cancer management is multidisciplinary care, which comprises cooperation primarily between patients, clinicians, and laboratory scientists. There are many clinical and nursing specialties involved in cancer care, including surgeons, palliative care physicians, radiologists, pathologists, haematologists, medical oncologists, specialist nurses and other healthcare professionals (such as genetic counsellors). Clinical oncologists primarily manage solid epithelial cancers whereas haematologists typically manage lymphomas, leukaemias and myelomas. There are also separate paediatric subspecialties. Many cancer clinicians are becoming increasingly specialised and are usually based in regional cancer centres. The adoption of NGS will increase the complexity and specialisation of cancer management, and these issues must be considered when considering any future implementation. Similarly, clinical guidelines, pathways and service frameworks will need to be appropriately adapted, requiring consequential changes in service provision, funding, and staff training.

6.5.3 Laboratory services

The morphological assessment of cancer samples using histopathology is central to cancer management. However, the widespread adoption of molecular diagnostics would have major service implications. This will depend primarily on the extent to which molecular diagnostics can replace conventional cancer diagnostics. Although it is too early to predict the relative contributions of morphological and molecular diagnostics in cancer management, their respective roles will need to be carefully evaluated to ensure an optimal balance and the most cost-effective use of limited resources. This is very important as nearly 300,000 new cancers are diagnosed every year in the UK. Most of these are solid epithelial cancers, and are the cancers in which NGS technologies are likely to have the most clinical utility. The future provision of molecular diagnostics in the UK has been addressed in a recent report by the Royal College of Pathologists[207]. They found that provision of molecular diagnostics (outside of the context of disease susceptibility testing) is currently patchy and varies both between and within NHS regions.

Decisions about what tests are provided, and who performs them, are often made locally and are influenced by history, infrastructure constraints, availability of funding (which may come from the NHS, research bodies or private companies) and specialty interests. Similarly, cancer diagnostics are undertaken by different personnel in different localities. For example, cancer cytogenetics may be undertaken by cytogeneticists, haematologists or histopathologists. The private sector is becoming increasingly prominent in molecular diagnostics, including both 'embedded' privatised NHS laboratories as well standalone commercial laboratories. The report made a number of recommendations, which are highly relevant to the adoption of NGS in the NHS and to cancer in particular (Box 6.4).

Box 6.5: Key recommendations from the RCPath regarding the future of molecular diagnostics in the NHS

- Molecular tests should be approved by a national body
- The training of appropriate skilled scientists, especially for large-scale molecular diagnostics of cancers), was challenging and needed to be addressed
- Molecular testing should only be undertaken in regional centres or larger laboratories
- Funding for molecular tests should be an explicit part of clinical budgets
- Greater local integration of clinical and pathology services was required to coordinate tests, share expertise and access to equipment
- Outsourcing of NHS pathology should be considered, especially as molecular diagnostic was a prime candidate for this.

This report is one of a series of policy initiatives regarding the provision of molecular diagnostics and these are covered in detail in Chapter 10.

7 Other potential applications

7.1 Risk prediction in complex diseases

7.1.1 Introduction

One important potential application of whole genome sequencing (WGS) is in improving the prediction of the risk of individuals developing chronic complex disease. This section describes how such risk prediction can be performed, how WGS may contribute and what challenges remain in using genetic information for risk prediction.

7.1.2 Risk prediction in medicine

Nearly all the important threats to population health, at least in industrialised countries, have complex multifactorial aetiologies. The epidemiological knowledge that underpins understanding of the causes of these diseases also permits the estimation of individuals' risks of developing them, often using complex models which incorporate and weigh the relevant risk factors.

Medical risk prediction models apply a mathematical algorithm to a combination of risk factors to estimate the probability of an individual developing a particular health outcome in a specified period. Many already exist within medicine – in some cases, several being available for the same condition – and more will doubtless be developed in the next few years as more genetic and novel biomolecular risk factors are discovered. Risk prediction models are already well-developed for coronary heart disease, stroke, type 2 diabetes and breast, colorectal and prostate cancers. The scoring systems use data such as personal and family history, lifestyle, physical examination findings, the results of psychometric testing and molecular and genetic biomarkers. They can be applied in population or opportunistic screening to predict future disease before its onset, in the early diagnosis of disease before the development of symptoms, or in gauging prognosis.

Scoring systems are likely to grow in importance. More candidate risk factors are being identified every year and more interventions are available to reduce risk, via both primary and secondary prevention. There is increasing pressure, partly fuelled by recent advances in human genetics, for a shift in both medicine and public health from detection and cure to prediction and prevention of disease. Furthermore, there is rising societal and professional interest in personalised medicine, the tailoring of care to the specific characteristics of individuals; this approach is as relevant to prevention as it is to treatment, and will increase interest in risk prediction models. The identification of more risk factors, and perhaps new statistical techniques, will mean that the models themselves will become more complex.

7.1.3 Application of WGS to risk prediction

Given the complex and incompletely understood causation of chronic disease, the associated risk prediction models sometimes rely on many biomarkers. Hitherto, these have nearly all been environmental, because of the lack of knowledge of the exact genetic contribution to aetiology. However, using genetic biomarkers to estimate risk is in some ways more straightforward than using non-genetic ones, because the former can be measured almost without error and do not vary in an individual over time. Also, they need only be ascertained once, and this can be early in life, whereas other risk factors may not manifest until later in life. This allows lifestyle and other interventions to begin earlier, potentially increasing their effects.

Moreover, the application of genome-wide profiling or WGS means that multiple genetic markers can be measured simultaneously for many different diseases in the same assay, which is not true for other biomarkers. To date, risk prediction for common complex diseases has focused on combining risks from common variants[208], identified through genome-wide association studies using genotyping arrays, although in principle it could also combine rare variants identified through WGS.

There is little doubt that it is technically possible to use genetic risk factors to predict the risk of disease. Such an approach is already offered by many direct-to-consumer services, which use a genome-wide approach to estimate an individual's risk of multiple complex diseases by combining the population absolute risk (*i.e.* disease incidence) with the individual's relative genetic risk[209]. However, each common susceptibility variant confers only a small risk of disease, and it is currently far from clear whether the resulting risk prediction is clinically valid for individuals, or in the population of interest, and whether it is useful for targeting interventions. The predictive power of genomic risk models, measured by the calibration of the model and their ability to discriminate between affected and unaffected individuals, varies substantially between diseases due to variations in heritability, allele frequencies, population disease prevalence and odds ratios. To the best of our knowledge, to date, no genomic risk prediction model based solely on common variants is used in routine clinical practice, due to insufficient evidence of clinical validity or utility for individual risk prediction.

7.1.4 Case study: the role of genes in coronary heart disease

Coronary heart disease (CHD) is a major public health burden, and the commonest cause of death in the UK. Risk prediction models exist for CHD, most notably Framingham and QRISK, and are used to identify individuals above a particular risk threshold to whom cholesterol lowering drugs are prescribed to lower their risk.

Heritable effects on CHD operate in two ways. Some conventional risk factors are in part genetically determined traits; for example, both plasma cholesterol concentration and hypertension are heritable[210;211]. However these effects are not fully explanatory. Family history is an important independent risk factor for CHD; in the Framingham studies, an important series of cohort studies in CHD epidemiology, a family history of premature atherosclerotic vascular disease increased CHD risk, even after adjustment for conventional risk factors, by a coefficient of 2.0 in men and 1.7 in women[212]. So, it is likely that some of the unexplained risk is attributable to genes operating other than via the effects of conventional (non-genetic) risk factors.

Recent years have seen great advances in understanding the genetic basis of less common cardiac disorders. Causative mutations have been found in about two-thirds of cases of hypertrophic cardiomyopathy, nearly as high a proportion of cases of dilated cardiomyopathy and most cases of familial cardiac arrhythmias (see Chapter 5 on inherited disease). However, CHD is proving a much less tractable problem[213]. It is a multifactorial disease, not attributable to any single genetic or environmental cause. Risk alleles are incompletely penetrant and do not co-segregate with the disease phenotype.

As for many complex diseases, existing research into the genetic basis of CHD falls into two categories: early studies of candidate genes, and more recent hypothesis-free genome-wide association studies investigating many variants across the genome. The overwhelming majority of this research has failed to identify convincing associations between genetic factors and CHD.

However, as studies become more numerous and larger, statistically significant positive findings are emerging.

Could genomic risk prediction work in CHD? Single nucleotide polymorphisms (SNPs) on chromosome 9p21.3 have been consistently associated with CHD, but individually their effect on the estimation of risk of CHD is minimal: despite a hazard ratio of 1.60 for a single SNP, the discrimination of a traditional risk prediction model power is not significantly improved by its inclusion[214]. A proof-of-principle multigenic study of CHD performed better, based on ten genes with odds ratios between 1.13 and 1.42 and allele frequencies between 11% and 80%[215]. Compared with those carrying no risk alleles (only 0.8% of the population), those with all ten had an odds ratio for CHD of 8.25. Compared with the mean risk of the population, those with six or more alleles had a significantly higher risk of CHD; they constituted only 4.9% of the population but had a combined odds ratio of 3.73.

7.1.5 Screening programmes stratified by risk

Another potential application of genomic risk prediction using WGS is as a preliminary to screening. If individuals' risk of developing specific cancers could be estimated with precision, they could then be grouped or stratified according to risk, and screening targeted at those above a risk threshold. Stratification holds the prospect of achieving high rates of diagnosis and effective early treatment, while sparing those at lower risk, likely to remain disease-free, from the harms and inconvenience of screening. Because screening is targeted at those at highest risk, overall costs of screening are likely to be reduced.

Crude risk stratification is already used in population mammographic screening for breast cancer, using age as a risk factor. Because breast cancer becomes more common with increasing age, this works fairly well – but how would things be if we used a more sophisticated means of gauging a woman's risk, taking into account her genetic profile and environmental and lifestyle risk factors? Significant progress has been made in understanding the contribution of genes to breast cancer, and genome-wide association studies have so far reported a number of common susceptibility loci, each of which confers a small increase in risk. Knowledge of a woman's genotype could be used to adjust the baseline risk of cancer implied by her age, and thus improve the targeting of screening and other risk reduction measures[216]. This approach is currently under evaluation in the Collaborative Oncology Gene-Environment Study (COGS), whose goal is to identify individuals at increased risk of breast, ovary and prostate cancer in the population by quantifying the risks associated with genetic and environmental/lifestyle factors in the largest dataset of its kind ever generated. The COGS project also aims to model the effects of targeting screening using genetic variants in this way, and to characterise the ethical, social and organisational challenges that might result.

7.1.6 Issues to be resolved

There are some important outstanding issues which may limit the value of risk prediction based on WGS. Risk is determined in a multifactorial way for most chronic diseases, so the apparent association between SNPs and risk will therefore vary between studies, depending on the environmental risk factors present and their interaction with genetic risk. Therefore, we cannot be sure of the extent to which such estimates are applicable outside the population in which the information was originally ascertained.

The importance of environmental risk factors has another implication. The most accurate risk prediction models are likely to include both genetic and environmental risk factors, yet there are considerable difficulties in bringing the two together in one model.

Firstly, the odds ratios derived from most genome-wide studies are unadjusted for environmental risk factors. They will need to be adjusted before they could be added to existing risk models, to avoid violation of the assumption of independence of effects which underlies the models. Since in some cases the alleles are known to be associated with phenotypic risk factors, and often have modest odds ratios to begin with, their apparent effect in some studies may be entirely the result of residual confounding.

Secondly, to some extent, genes affect risk by modifying the effect of behaviour and environment on phenotype: for example, cholesterol concentration is influenced by genes and diet. The research goal is to create a model in which genotypic and phenotypic information is merged to improve CHD risk estimation, but substantial research would be needed to secure enough data on how the risk associated with each genotype is influenced by environment and behaviour. Yet without such research, the modelling of risk from genotypic and phenotypic information cannot be merged. If the genes' effects are dependent on carrier-specific environment and behaviour, the problem becomes even less tractable, since risk estimates from genes will be less generalisable over time and between populations.

Thirdly, it is increasingly clear that genes interact with each other, as well as with the environment; how this process occurs is less clear. However, most chronic diseases such as coronary heart disease and cancer arise as a result of a long sequence of steps, so genes probably have overlapping effects, by which one modifies the effect of others, or other forms of gene-gene interaction occur. This increases the complexity of the problem considerably. The way the risks are treated arithmetically is central to the outputs of the model, yet gene-gene interaction is a poorly understood area of genomics, there are many genes involved, more will be identified in future and the way they interact may well be conditioned by the environment.

Finally, most genetic research is conducted on people of predominantly European ancestry. It is not clear the extent to which information gathered in this way will be transferable to those of other ethnicities. Another potential confounder is change over time in disease risk and incidence. For example, the last seventy years have witnessed a substantial rise, then fall, in the incidence of coronary heart disease in industrialised countries. The reasons for these changes are not fully understood, but attempts to model and predict risk need to take account of complex relationships between genes, lifestyle and factors such as socio-economic status.

7.2 Pharmacogenetics

7.2.1 Basic principles

Genetic variation can influence individuals' response to drugs, either by affecting their rate of absorption or metabolism or by affecting the response of a drug target (see Box 7.1). Advances in genetic diagnostic and profiling technologies allow the development of stratified and personalised medicine where an individual genetic profile drives the therapeutic choice and disease management. This in turn requires the development and application of tailor-made drugs. Traditionally, drugs have been tested on large, essentially randomly selected populations and average responses are reported and used for clinical application. In pharmacogenomics, genomic information is used to analyse individual responses to drugs, and hence chose the drug or its dosage to suit the individual. For example, a particular (positive or adverse) response to a drug in an individual patient that is associated with specific gene variants allows the therapy to be tailored to that patient by adjusting the drug dosage or choosing a drug with fewer side-effects.

The increasing knowledge of genetic variation and therefore the possible prediction of protein structure of drug targets also allow a targeted approach to drug discovery. The implementation of next generation sequencing technologies will undoubtedly considerably accelerate the development of this type of approach. Historically, pharmacogenetics involved investigating genetic and biochemical variation in drug response in a small number of genes and proteins. However, current technologies allow assessment of entire pathways relevant to diseases or drug responses through a genomic approach. As a result, disease genetics and pharmacogenetics has undergone a shift in focus from Mendelian disorders to more complex examples of genetic causation [217].

Two terms are commonly used to describe the interaction between an individual's genetic composition and therapeutic drugs: pharmacogenetics and pharmacogenomics. Although the two terms are used interchangeably they differ in the initial scientific approach to the problem. Pharmacogenetics looks at an unexpected drug response and investigates its genetic cause; pharmacogenomics looks at genetic differences in a population to explain observed responses to a drug or susceptibility to diseases.

Box 7.1 Pharmacokinetics versus pharmacodynamics

There are two main aspects of a drug action in an individual's body:

1. **Pharmacokinetics**

   The study of what the body does to the drug, which affects the amount of drug needed to exert its action in the body

   Pharmacokinetics includes four processes (abbreviated as ADME):

   - **A**bsorption refers to how a drug enters the bloodstream
   - **D**istribution shows the dispersion of the drug throughout the body and the amount of the drug that reaches its target organ or cells
   - **M**etabolism is the breaking down of the parent drug into daughter metabolites
   - **E**xcretion describes the elimination of the drug and its metabolites from the body

2. **Pharmacodynamics**

   The study of what the drug does to the body, which affects the response of the target cells and organs to the drug

   Pharmacodynamics describes the molecular action of the drug on its target. The target can be either on the cell surface (*e.g.* a receptor), an ion channel or an intracellular protein (*e.g.* an enzyme). It studies the relationship between drug concentration and effect (*e.g.* drug-receptor interactions).

7.2.2 Genotype, phenotype and adverse drug reactions

There are already many known examples of candidate genes where variation is relevant to variable drug response. Two categories of genes can be defined: those which encode drug-metabolising enzymes (DMEs) and drug transporters, which influence the pharmacokinetic properties of the drug; the second group of genes include drug targets and elements of associated pathways, and is involved in the pharmacodynamics aspect.

Drug-metabolising enzymes and drug transporters

The majority of drug metabolism occurs in the liver and that is where most drug-metabolising enzymes are located. Many DMEs are also active in other tissues and are part of the blood-brain barrier. DMEs have large effects on the degree or rate at which a drug is converted to its metabolites, and hence the levels of active drug present in the body over time. Understanding the relationship between a patient's genetic makeup of DMEs and the efficacy or toxicity of a drug offers the potential for optimising therapy. Early discovery of genes encoding DMEs was driven by analysis of response to drugs such as debrisoquine which is associated with a frequent poor metaboliser phenotype[217]. In most people this drug is quickly broken down to metabolites and eliminated through urine. The metabolism of this drug is driven by a subfamily of cytochrome P450 (*CYP2D6*). People who carry polymorphisms that reduce *CYP2P6* activity are poor metabolisers of debrisoquine and suffer serious side effects in response to the drug.

Drug transport between or out of body compartments is another process which can affect drug response. Variation in genes encoding specialised drug transporters can account for variability in drug response. A well-studied example is the *ABCB1* drug transporter at the blood-brain barrier which is responsible for the transport of the anti-epileptic drug phenytoin[217].

Drug targets

The second category of genes where variation can influence drug response comprises genes encoding specific drug targets, proteins modulated by the drug and more broadly proteins of associated pathways. Examples of such variation include polymorphisms in the beta-2-adrenergic receptor (*ADRB2*) that have been associated with response to beta-agonists in the treatment of asthma[218] and beta-blockers in the management of congestive heart failure[219].

Drug safety

Serious side effects from medicines – adverse drug reactions (ADRs) – are relatively common and have a major impact on public health. According to the Medicines and Healthcare products Regulatory Agency (MHRA) and a study of hospital admissions in UK, around 6.5% of hospital admissions in England were for drug-related reactions and ADRs accounted for 2% of fatalities among those patients[220]. Variation in genes encoding enzymes involved in metabolic pathways, drug receptors and drug transporters have been associated with individual responses affecting the efficacy and toxicity of drugs[221]. Decreasing the burden of adverse drug reactions is one of the big challenges of pharmacogenomics research. This problem is particularly severe with anticancer drugs which mostly present a narrow therapeutic index; small changes in dosage can cause toxicity, and non-response to treatments can lead to high level of mortality.

7.2.3 Examples of clinical pharmacogenetic tests

Warfarin

One of the best described examples of clinically used pharmacogenetic data is warfarin, the most widely prescribed anticoagulant. It is a potent and effective tool in the prevention of thrombosis but its therapeutic window is very limited. Dosage has to be titrated carefully to prevent serious side-effects including haemorrhage or undesired coagulation. Important progress in the pharmacogenetics of warfarin has been made in the last seven years, since the target for warfarin-based coagulants has been identified to be a subunit of the vitamin K epoxide reductase. In 2009, a genome-wide association study identified a number of SNPs closely associated with the activity of the reductase[222]. This has led to the FDA providing genotype-specific ranges of warfarin doses in 2010 and suggesting that genotypes be taken in account prior to prescribing the drug[223]. There is strong evidence of clinical validity for testing *CYP2C9* and *VKORC* to determine the appropriate warfarin dosage, and preliminary data on clinical utility suggests that genotyping warfarin patients resulted in a ~30% reduction in hospitalisation[224].

Flucloxacillin

Flucloxacillin, an antibiotic widely used in Europe to treat staphylococcal infections, has been associated with an unusual form of hepatitis with an estimated incidence of approximately 8.5 cases per 100,000 patients[225]. Results from a genome-wide association study reported in 2009[226] showed a strong association between the presence of a SNP in the major histocompatibility complex closely linked with *HLA-B*5701* and induced hepatic injury. Development of a diagnostic test could lead to easy identification of the population group who might develop liver toxicity as an adverse event.

Abacavir: first example of prospective clinical trial

The genetic association between *HLA-B*5701* and an adverse drug effect has been already shown for abacavir, a nucleoside analogue inhibitor of HIV reverse transcriptase used in multidrug regimens to treat HIV-1 infection[227]. The reported findings resulted in modification of the abacavir label to include an FDA recommendation that patients undergo genotyping for *HLA-B*5701* before starting therapy. Further investigation of abacavir hypersensitivity recently concluded in a prospective clinical trial which demonstrated the utility of pharmacogenetic screening before administering this drug[228].

Cancer management

Cancer pharmacogenetics has two components: drug responses driven by variations in the tumour genome and those associated with the germline genome. There are already many examples of known association between tumour genetic traits and response to cancer therapy. *HER2* overexpression in patients with breast cancer is known to be directly correlated with the response to treatment with trastuzumab (Herceptin)[229]. Patients with non-small-cell lung cancer who have activating mutation in the gene encoding the epidermal growth factor receptor (EGFR) respond much better to the EGFR antagonist gefitinib. A recently identified mutation in the BRAF kinase has been associated with better survival rates upon treatment with a specific inhibitor of this kinase in patients carrying the mutation.

An example of a clinically relevant variation in a patient's germline genome affecting cancer therapy is the altered structure of the promoter region of the gene encoding an UDP-glucuronosyltransferase which results in lower enzyme activity and therefore decreased rate of metabolism and so higher toxic effects of irinotecan, a chemotherapeutic agent mainly used in colon cancer[223].

7.2.4 Barriers to progress

Technological advances in the last ten years, and particularly the development of DNA sequencing methodologies, has allowed immense progress in characterising human genetic variation and has had a huge impact on pharmacogenetic studies. However, substantial challenges still remain in discovering which of the known or novel variants are associated with differences in individual response to drugs, and to apply this knowledge to clinical practice in order to allow prescription of the safest and most effective drug for each individual patient.

Studies of pharmacogenetic effects of genetic variation require collecting appropriate sample sets. This can prove to be difficult, especially for rare phenotypes or variants, which is often the case with adverse drug effects. Although post-marketing surveillance should be a good source of patients with adverse effects, current systems rely on voluntary submission of adverse event reports and therefore do not provide sufficient data[230].

Once data from a cohort of patients has been gathered, variants have to be genotyped. Until recently, pharmacogenetic studies relied on a so-called 'candidate gene' approach. Those studies usually involved only a few genes, and few variants of these genes, selected for analysis based on existing knowledge of the drug target or pathway of drug metabolism. This approach is successful when correct candidate genes are known and chosen but fails when there is insufficient knowledge of drug targets or metabolic pathways.

Recent progress in understanding genetic variation and progress in multiplexed genotyping technologies brought to existence an alternative approach: the genome-wide association studies (GWAS). In this approach no knowledge of candidate genes is required and an unbiased search of genetic polymorphism associated with the phenotype of interest is conducted across the entire genome.

Genome-wide association studies have allowed rapid progress in pharmacogenetic studies but do not solve all the challenges. Associations uncovered by GWAS require replication which in turn involves identifying an appropriate sample set. The difficulty of finding a sufficient number of cases of adverse drug reaction to conduct replication studies presents a challenge for regulators. To date, the FDA includes pharmacogenetic information even in cases when the association between a genetic variant and a drug response has not been retested. This warns physicians of a possible risk but puts a burden of responsibility upon them to conduct genotyping investigation prior to prescription of the treatment. However, most guidelines require demonstration of clinical utility in order to justify genotyping procedures. In the absence of this relevant evidence, the pharmacogenetic tests will not be reimbursed.

Interpretation of pharmacogenetic data represents another important challenge and there is a clear need to develop simple clinical algorithms to aid clinicians to use this data to its full potential. There are an ever increasing number of modified drug labels which include pharmacogenetic information, but guidelines that link results of a pharmacogenetic test to specific dose recommendations are still very sparse[231].The education of healthcare providers and the encouragement to revisit suboptimal drug regimes in disease management also remains a fundamental step for pharmacogenetics progress.

There are many other dimensions to the progress and utilisation of pharmacogenetics, not least the availability of drug alternatives. Development of new medicines in response to increasing knowledge of variation in drug targets and metabolic pathways is an essential element of personalised treatments. Progress in stratified and personalised medicine also requires rapid development of diagnostic tests. The many and complex issues surrounding the development and implementation of companion diagnostics and stratified medicines are beyond the scope of this report.

7.3 Tissue typing and transplantation

Tissue typing is a procedure whereby the tissues of a prospective donor and recipient are tested for compatibility prior to transplantation. Achieving a good match between donor and recipient is important for solid organ, stem cell and bone marrow transplants, to prevent rejection of the donated tissue and improve overall transplant survival. An embryo can also be tissue typed (so called preimplantation tissue typing, or PTT) to ensure that the embryo implanted can act as a cord-blood stem cell donor for a sick sibling[232]. Creation of such ‘saviour siblings’ can be done in conjunction with preimplantation genetic diagnosis (PGD) where the embryo is at risk of an inherited disease (see Chapter 5).

Historically tissue-typing was performed using low resolution serological methods of antibody detection. However, newer high resolution DNA-based methods of tissue typing increase the success rate of donor transplantation, by allowing more precise matching between donors and patients[233], and are now almost universally used. The aim of either method is to determine the type of antibodies present on the surface of immune cells, which are encoded by the human leukocyte antigen (HLA) system, the major histocompatibility complex (MHC) in humans. The HLA complex locus is a large (>4Mb) region located on chromosome 6 containing over 220 genes with thousands of alleles, and is the most polymorphic region in the human genome. HLA antigens are divided into class I (A, B and C), which present peptides from inside the cell, and class II (DR, DQ and DP), which present antigens from outside of the cell to T-lymphocytes. Only a subset of these HLA genes are important for tissue typing; minimum or low resolution matching is based on just a few of these, and may be used as an initial screen for matching donors on a database, whilst high resolution matching prior to transplantation involves assaying more loci.

HLA typing is undertaken by clinical scientists in histocompatibility and immunogenetics, and currently uses either PCR-based or sequence-based methods[234]. PCR reactions can use either panels of sequence-specific primers which amplify particular HLA alleles, or locus-specific oligomers that act as hybridisation probes. Alternatively, standard capillary Sanger sequencing technology can be used to sequence the HLA region, and is the accepted reference method for discovering new alleles. However, because of the enormous amount of variation in the region, basic Sanger sequencing cannot resolve haplotypes from heterozygous samples if multiple alleles are sequenced together[235]. Therefore, unambiguous HLA sequencing currently involves haplotype separation prior to sequencing, which is slow, cumbersome and expensive.

This intrinsic phase problem can be overcome by sequencing using NGS platforms, which allow unambiguous haplotype assembly[235-237]. The average HLA exon encoding a peptide binding site is approximately 270 base pairs, which is slightly shorter than the individual amplicons generated by PCR-based targeting methods (see Chapter 3). Therefore, each exon can be sequenced completely by sequencing both strands with sufficient overlap between reads that specific HLA alleles can be unambiguously assigned. Moreover, using sequencing platforms with long read lengths (such Roche/454) means that an individual read might encompass an entire exon. The large number

of independent sequence reads also means that rare variants present in individual samples can be detected. The highly polymorphic and repetitive structure of the HLA region makes correct alignment of the individual reads difficult, so custom HLA-matching bioinformatics software (such as Assign from Conexio Genomics) has been developed. Comparison against reference databases (such as the IMGT/HLA database[238]) allows improved sequence alignment by filtering out related sequence reads from pseudogenes and other unwanted HLA genes that were co-amplified with along with the target sequence.

Various projects are underway to develop this technique, including at the Eastern Sequence and Informatics Hub (EASIH) in Cambridge, UK. The large size of the HLA region makes targeted sequencing using NGS platforms cost-effective relative to Sanger sequencing, but at this stage it is unclear whether the use of NGS platforms for HLA typing will become robust enough for routine practice. Ultimately, if the price is cheap enough to perform high-throughput HLA sequencing using NGS technologies and sample multiplexing, it could become routine to store sequence-based HLA information on donor databases to allow immediate high-resolution matching. Assessing the clinical significance of mismatches at different HLA alleles will be increasingly important if higher resolution HLA typing becomes more commonplace. NGS can also be used to identify and then test additional histocompatibility loci beyond the MHC, to provide a genome-wide view of histocompatibility.

7.4 Microbial sequencing

An important application of NGS technologies is in the field of infectious diseases. These technologies facilitate pathogen genome sequencing in a matter of days and are likely to have a big impact on healthcare. NGS has already facilitated sequencing of microbial genomes in order to track the evolution and spread of Methicillin-resistant *Staphylococcus aureus* (*MRSA*)[239], analyse the genome sequence of *E.coli* following an outbreak in Europe[240] and tuberculosis in Canada[241]. The ability to rapidly sequence genomes and share data through the internet can allow much faster responses to disease outbreaks and could become the gold standard for infectious disease epidemiology. It is possible that routine sequencing of microbial genomes could be incorporated into clinical practice in the near future.

A detailed discussion of the impact of NGS technologies for the analysis of non-human genomes is beyond the scope of this report.

8 Ethical, legal and social implications

8.1 Introduction to overarching ethical issues

Novel genomic technologies raise a number of important ethical, legal and social challenges. The detailed and potentially powerful information generated by these technologies creates opportunities for greater scientific understanding and ultimately improved health, but this complexity raises challenges for those charged with ensuring that technologies are used in a responsible and ethical way. The nature of the data generated by whole genome sequencing (WGS) means that conventional methods of minimising harms may be inadequate to protect subjects, whether patients or research participants. The volume of data produced has necessitated new methods to organise and interpret it with the addition of a new professional layer of bioinformatics between the 'test' and the patient in order to characterise findings as pathological or not[87;198]. Finally, the possibility that WGS can provide a wide range of information that is predictive of current health status or future risk to the patient, future offspring or relatives of the subject is unprecedented in its scale[153]. This leads to the most difficult problems of all: the near impossibility of informed consent across this vast range of data, and defining the extent of responsibility of health professionals to continue feeding back relevant findings from stored data as new clinical knowledge emerges and the patient's situation changes.

In this chapter we consider a range of ethical, legal and social issues that may arise as WGS is progressively introduced into clinical practice. We explore:

- To what extent these are substantively new issues
- What issues arise from the scale of WGS data and its interpretation
- Whether changes in existing clinical guidance and practice will be required
- Whether there will be a need for further regulatory responses
- What the societal effects might be of wider availability of WGS

In order to address these questions, we start by identifying a set of general issues around the processing of data, including data storage and access, safeguarding privacy, and wider aspects arising from data misuse or loss (such as stigmatisation or discrimination). A second section has a philosophical starting point, examining the basis for individual autonomy and consent - and the extent to which existing practices might change in response to WGS. A third section examines the scope and duties of professional practice, and reflects on the implications of additional or incidental findings in both a research and clinical setting. The penultimate sections of this chapter are rather more technical in nature, addressing the regulation and governance of NGS technologies and WGS approaches as well as the intellectual property issues that are likely to arise, before turning to questions concerning the public perception of these technologies.

8.2 Dealing with data

8.2.1 Data storage and processing

Data processing issues arise from the characteristics of the data, the infrastructure for holding and managing data and the processes involved. A recurring theme that arises when genetic or genomic data are discussed is the extent to which genomic data should be treated as exceptional and accorded special protection. Most commentators acknowledge that genomic data may have a number of characteristics that might merit special protection. These include its predictive capability, its potential impact on the family of the individual undergoing testing, and the physical nature of genetic material itself (given that genetic material is usually unique, immutable and durable throughout an individual's lifetime), sometimes within a context of historical and social misuse[242].

Identifiability of genomic data/uniqueness

Assuming that raw image data and consensus data is not stored, typically the analysis of an entire human genome generates data about 3 billion bases or 6 billion in the diploid genome (if information about the whole genome is retained). Identical genomes are unlikely to arise even if a 'minimal' genome is retained[243] (*i.e.* those 3-4 million bases from the germline genome which differ from the reference sequence, comprising around 30Mb of information) (see Chapter 4). Indeed single base changes in 30-80 locations in the genome are sufficient to identify a single individual from a population of 10 billion[244;245]. *Thus the human genome is effectively a unique identifier.*

Relevant legislation

Within some jurisdictions this combination of factors has been the justification for legislation and regulation that creates exceptional protection against misuse. Examples include the Genetic Information Nondiscrimination Act (2008) (GINA) in the US and its supporting regulations[246;247]. In other jurisdictions, genomic data has been treated as a subset of all those types of data (including data on infectious disease or mental health) which could or should be regarded as particularly sensitive and thus deserving of special protection, which may be preferable on the grounds of consistency.

Use of electronic health records

Over the last decade, increasing use has been made of electronic health records in clinical care as well as in research. Technologies such as NGS seem likely to drive a more systematic use of electronic records, if only because of the data processing demands of these technologies. In practice, it is not clear whether the potential of these technologies may fully be realised within the confines of existing electronic patient records systems[248] given their existing patchy provision and difficulties with interoperability. As systems are developed, attention must be paid to confidentiality, privacy and data security; it will be necessary to consider whether special protections should be placed on the storage of WGS data[242]. As data recording becomes more systematised within electronic medical records, questions may arise about the responsibilities of health professionals to access it at relevant times in order to optimise its use for the health of the patient and family. Such times might be when the patient reaches adulthood, or a significant age (for example, when breast screening might be started), when the patient may be planning to start a family, or when new scientific knowledge arises that might be relevant to a current diagnosis or future preventive action. Although there is some experience of these issues with existing stored genetic data (for example mulitiplexed microarray data) the scale and complexity involved in WGS is many times greater.

Patient centred access

One solution might be to foster a more central role for the individual patient or research participant across the spectrum of data processing from data generation through to interpretation of results, consistent with the emerging doctrine around patient choice within the UK[7]. Proposals range from patient held results discs through to increased use of data masking or controlled access measures to allow patients to dictate data disclosure protocols[242]. These strategies have been used by a number of organisations including the Genetic Alliance[248;249] and similar initiatives are under development in clinical settings in the UK[8].

8.2.2 Access, privacy and confidentiality

Data protection legislation

The conventional regulatory approach to worries about misuse of personal data[9] (*i.e.* personal identifiable data) has been to place limits upon the processing of data that is personally identifiable but to allow unfettered access to data that is anonymous[250]. This approach, enshrined in the UK Data Protection Act, is problematic when applied to the individual human genome, because of the unique nature of the data related to an individual (as described above). WGS data related to a patient or research participant can be identified in the following ways:

- A de-identified dataset may be compared with a reference genome or matched dataset that contains personal identifiers
- Individuals and families with rare genomic disorders may be readily identified through cross-referencing between databases with different levels of access
- Genomic data generated from WGS may be linked with non-genetic data which allow inferences to be made as to the source of the data
- Genomic data may be used to infer a profile of the sample donor[243]

7 See for example the plethora of consultation papers published as part of the 'Liberating the NHS' initiative.

8 www.patientsknowbest.com

9 *i.e.* The Data Protection Act (1998) Section 1 defines 'personal data' as data 'which relate to a living individual who can be identified- (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual'.

In the research context, existing methods for de-identification such as limited release of results, statistical degradation and sequestering identifiers through key-coding have limited applicability. Other strategies such as pooling samples have also been called into question as a means of masking identity[243].

There is additional legal complexity due to a lack of consensus about the proper interpretation of 'personal data' under the Data Protection Act. The Court of Appeal Case of *Durant*[251] found that that personal data are '*biographical in a significant sense*' and have '*the putative data subject as* [their] *focus*' which provides a precedent for a narrow definition of what constitutes 'personal data'. As Curren *et al*.[252] point out, this is a decision which is at variance with other sources of legal authority[253] and may also be scientifically flawed, (since analysis of those parts of the genome that are relatively conserved such as the Y chromosome can yield meaningful information for family members despite not being 'biographical').

Thus the regulatory and professional responsibilities that flow from the fact that WGS data may be viewed as personal sensitive data under the Data Protection Act, and the extent to which legislative change may be required are as yet unclear.

Minimising the risks of unanticipated data disclosure

To date, a range of possible approaches have been used to minimise the risks associated with data processing (including unauthorised data disclosure). The most common approach is to restrict the pool of legitimate data users to those who have professional obligations to keep personal data private and confidential (by providing that data is released to bona fide researchers who are contractually required to keep participant data secure, or by using a committee of experts to vet possible users in the form of a data access committee)[243]. Other strategies include a combined approach which allows role based access to certain categories of data. Some commentators have argued that ensuring confidentiality of genomic information is impossible, and that another legitimate approach is to assume that all data will be disclosed. This is the approach used by the Personal Genome Project where groups of highly qualified participants have made very fully informed deliberate choices to disclose all their results on a public database[254;255].

Data sharing between jurisdictions

Increasingly genomic technologies and the need for validation involves multiple centres across the world suggesting that the most effective regulatory responses should be enforceable at an international level: indeed a number of international data sharing agreements are widely used for research purposes[256].

Data sharing data with other family members

Recent applications using NGS technologies often utilise samples from other family members either as a control or to inform interpretation of findings in the proband[20]. This comparison may be particularly sensitive when used in the context of reproductive choice (such as where fetal DNA from maternal plasma during pregnancy is compared with maternally and paternally derived genotypes, for example, as a means of inferring heterozygosity and homozygosity at a particular site)[153]. This comparative process could identify future susceptibility to ill health or other personally sensitive information (including incest or misattributed parentage). In many cases it is unclear whether or what information of this type should be shared and the risks associated with this need to be made transparent as part of the consent process.

Distinguishing between research and clinical care

Within different jurisdictions, there are varying approaches to protecting data obtained in the context of research and clinical care. In part this stems from the differing duties which researchers and clinicians owe to their participants or patients. Also relevant is the way in which data might be passed onto third parties such as insurers, employers or state authorities including the police and social services. In the UK, the Data Protection Act (1998) provides that where data processing is necessary for medical purposes, undertaken by a health professional or someone owing an equivalent duty of confidentiality, then the general requirements for specific consent that would normally apply to the processing of health data do not apply (Schedule 3, Paragraph 8). This legal

formulation may create problems for non-clinical medical researchers who engage in secondary research (especially for those that use clinical data for population research without the express consent of the participant).

In the US, legislation distinguishes between data processed for research and data used for clinical care. Both are protected by the Common Rule on Protection of Human Subjects and the Privacy Rule under the Health Information Portability and Accountability Act[243].

Data sharing and bioinformaticians

There is also a need for clarity about the status of bioinformaticians who process raw (identifiable) data from WGS for the purpose of research or clinical care, particularly as in the future it is possible that these roles may be centralised within health services or even contracted out beyond the clinical or research team. Relevant issues include the way in which such systems are developed, funded, managed and accredited, as well as issues about the accountability and transparency of those who do this work (and how this will relate to existing mechanisms for governance of medical practice and medical negligence).

8.2.3 Genetic discrimination

The very characteristics of genetic material that make it so clinically useful as a proxy for future ill health (its predictiveness and comparative immutability) are associated with reports of misuse by governments. There is concern that if genetic and genomic knowledge are more widely used at population level, it is more likely to be used in ways that are potentially stigmatising[257]. As a result, some jurisdictions such as Germany[10] have introduced legislation that limits the use of genetic/genomic information by third parties in an attempt to prevent the repetition of past abuses[11;258]. These provide a useful context for discussion about how policy might develop in the UK and within Europe more generally.

The use of genetic/genomic information by insurers

In the US, in addition to the state laws in 34 states that prohibit employment discrimination on the basis of genetic information, the federal Genetic Information Non-Discrimination Act (GINA) applies to private employers with 15 or more employees. Brought into force on 21 November 2009, it prohibits a health insurer from requiring an individual (or family member) to undergo a genetic test[12], or to request, require or purchase genetic information as a condition of enrolment, or to increase premiums as a result of genetic information in the context of group health insurance. The definitions used in this Act are widely drawn so that protection is maximised: thus 'family member' includes all dependents (including adopted children) from those who are related to a first to the fourth degree. The regulations acknowledge that this degree of relatedness is broader than the definition that is conventionally used in medical practice. Insurers (and employers) are free to use and disclose genetic information about an existing symptomatic disease which has a genetic component that could reasonably be detected 'by a healthcare professional with appropriate training and expertise in the field of medicine involved'.

Tests 'that determine the presence of markers associated with ancestry' are specifically included (although information about race and ethnicity that are not derived from a genetic test are not)[13]. Thus WGS data would seem to fall under the scope of GINA (unless WGS is being done within a research cohort with a known phenotype already manifesting a genetic disease[14;259]).

---

10 German Human Genetic Examination Act (2008)

11 See for example the UK Human Genetics Commission that opined a principle of non-discrimination namely that 'no person shall be unfairly discriminated against on the basis of his or her genetic characteristics'. HGC (2002) Inside Information: Balancing interests in the use of personal genetic data (p.44).

12 'Genetic test' means 'an analysis of human DNA, RNA chromosomes, proteins or metabolites, if the analysis detects genotypes, mutations or chromosomal changes' […but excluding] an analysis of proteins or metabolites that is directly related to a manifested disease, disorder or pathological condition'.

13 GINA Section 1635.3(f)

14 GINA Section 1635(g)

Within the UK, there is currently a moratorium against disclosing predictive genetic test results to insurers (subject to securing appropriate regulatory approval and to certain financial limits) which has recently been reviewed and is due to expire in 2017[260]. However insurers within the UK are free to use clinical or family history information. Currently the moratorium is expressed to cover predictive genetic tests in terms of cytogenetic and molecular tests. It is unclear whether this existing definition would also cover the results of WGS conferring predictive genetic information

Use of genetic/genomic information by employers

In the US, similar restrictions apply to employers under GINA. The protection offered by the Act has three strands: there is a general prohibition on obtaining genetic information, on using it for potentially discriminatory purposes and a requirement to keep that information confidential. There is provision for employers to keep genetic information within a safe harbour provided that the request for disclosure is worded appropriately and the respondent is aware that there is no obligation to disclose a wide range of genetic information (from prenatal test results to ancestry or paternity test results)[247;261]. GINA specifically provides that genetic information relating to employees, like other medical information, must be kept in separate personnel files after the Act came into force on 21 November 2009, and that genetic information should only be disclosed under limited circumstances[261].

In the UK, the Information Commissioner's Office has published guidance for employers setting out their rights to request or utilise genetic test information from job applicants and existing employees. Generally, employers are restricted from requiring pre-employment genetic tests without justification. It seems likely that this restriction would also apply to WGS data[262].

Anti-genetic discrimination legislation within the UK

The UK Human Genetics Commission recently explored the issue of genetic discrimination, and canvassed stakeholders in an effort to establish whether genetic discrimination was occurring, in order to provide an evidence base for a proposed change in legislation (*i.e.* to include provisions against genetic discrimination in the Equality Act 2009). They found little evidence of discrimination occurring[15] and the Equality Act does not explicitly address anti-genetic discrimination, although arguably existing legislation does allow a claim to be brought for discrimination on grounds of genetic pre-disposition[263]. This may need to be re-assessed if genetic information is used more routinely and evidence of discriminatory practice emerges.

8.3 Autonomy and informed consent

Respect for individual autonomy is widely regarded as the basis for securing an informed consent to treatment and research, but given the wide-ranging nature of WGS data, and the constantly developing state of scientific understanding of the genome, it is unclear whether the standard models of informed consent are fit for purpose in this context. Two issues have emerged as being particularly pressing: first is the paradox that individuals cannot be asked to consent to the discovery of risks the importance of which is impossible to assess. Thus from a legal perspective, there is no 'meeting of minds' and contractually the contract between researcher and participant might be void. It is also unclear whether informed consent is sufficient to deal with the feedback of incidental findings which are not pertinent[16] to the initial research or clinical question but that may have either clinical or personal significance.

In the UK, as in other jurisdictions worldwide, the trend has been to document ever more precisely, the likely risks and benefits of treatment or research in response to relevant case law[264;265] and legislation. In a research setting, participants are typically explicitly informed about various indirect harms that might arise (such as the transfer of data or samples to jurisdictions

15 Genetic Discrimination Working Group report publication pending.

16 The Oxford Dictionary defines pertinent as 'relevant or applicable to a particular matter'. In the context of genetic testing, this could include both anticipated findings relating to a clinical question, as well as unanticipated findings such as a previously unknown genetic association. The term is not used consistently in the literature.

that have a less secure data processing policy) and increasingly the consent process is used as a way of tailoring future research or care by allowing the respondent to make a choice about how their samples will be used or to determine the extent of continuing contact with researchers or third party researchers[266]. A variety of strategies have emerged for seeking consent, and as yet a consensus approach has not emerged.

Individual versus familial consent

One challenge is to clarify the obligations owed to family members. In the research setting, these potential obligations can be categorised into four different types:

- An obligation to include genetic relatives in process of obtaining consent for WGS
- Data release (to data processors including researchers, directly or via databases)
- Data analysis
- Feedback of findings (particularly those that are incidental to the clinical question)

The prevailing view is that an autonomy based view of consent (as outlined above) is both necessary and sufficient for WGS used in a research setting. Some commentators have urged that participants should consult with close (first degree) genetic relatives prior to sample donation. Although it may be desirable to seek consent from family members if the data generated by WGS is to be published, especially in parallel with other information that may make family members more identifiable (such as a family history), this raises profound ethical and professional concerns.

Whilst there are already well-established precedents within clinical genetics for communicating clinically relevant diagnostic and predictive information to genetic relatives, the same is not true in genomics research where findings are generally not fed back. In particular, there may be questions about the justifiability of researchers approaching the family members of participants direct, and the nature of the relationship that they have with those individuals. It seems likely that in the initial stages of implementation of the technologies, the majority of the information that is obtained will not be sufficiently robust (or have sufficient predictive value) to justify feeding it back to relatives[267], but this may change over time as more is understood about susceptibility to disease.

Those not competent to give consent – minors

Within clinical genetics, prevailing practice within the UK is to oppose genetic testing for adult onset diseases during childhood on the basis that the decision to test (and to forgo 'a right not to know') should be made by the child once they can make decisions for themselves even if under-age[17] [268]. This principle may be difficult to apply if NGS technologies are used more widely for a variety of uses.

To date, the cases involving WGS and children fall broadly into two types. In the first category, the rationale for using WGS has been therapeutic: to reach a diagnosis or improve treatment (sometimes in the face of intractable, and sometimes life threatening medical problems)[147;269]. In such cases, the parents of the children concerned have been approached to provide consent. In order to understand the pathological significance of the findings, this may involve comparing sequences from parents and children, especially if the parents are consanguineous[270]. Here, the use of WGS is justified on the basis that it is the approach of choice (and represents the only or best method of obtaining results that are clinically useful). A parental consent is usually sought, supported by an assent from the child (if they are old enough). In cohort studies there is also an emerging practice of seeking a re-consent from the child when they reach 16 years old. There may still be concerns about an increased likelihood of incidental findings, and if WGS data is stored and therefore accessible in the future, that this might pre-empt the child's ability to take these decisions on reaching adulthood.

[17] Children are judged sufficiently competent to decide for themselves if they are 'Gillick competent'.

The concerns might be greater in cases where there is no direct therapeutic goal. For example, in a research setting where samples have been taken from family groups, older adolescents were approached for assent after extensive counselling, and other measures were taken to reduce the potential risks arising from research (*i.e.* public accessibility to whole genome sequences was limited and potentially stigmatising data were blanked out)[271]. In a clinical setting, it seems less likely that WGS will be done on children without a therapeutic goal in mind, although this could arise within a public health screening setting.

Right not to know?

As well as the rights of children to have their future autonomy protected, others have described rights to informational privacy, and in particular, the right 'not to know' the results of a genetic test or risk status[272]. In practice it is not straightforward to operationalise this, as any enquiry as to what an individual might wish to do, might itself imply that an individual is at risk. WGS has the capacity to generate different types of information about risk, such as the distinction between pertinent and incidental findings (although there is not always a clear distinction between the two). These distinctions seem likely to evolve quickly over time, and the weight given to the right to remain in ignorance of risk needs to be carefully thought out in emerging policy.

The development of novel methodological strategies

Several research studies have described the increased predictive power of sequencing multiple family members (particularly a quartet of both parents and siblings)[23]. Comparing the genomes from different family members allows for more effective identification of recombination sites and sequencing errors, and aids in identifying candidate genes. However, the development of novel research methods involving the simultaneous sequencing of family members, particularly in healthy individuals, seems likely to make an individualistic autonomy driven consent model more and more difficult to arbitrate and administer. Conflicts may arise between the interests of family members: in clinical genetics they are currently resolved by adopting a variety of approaches to information disclosure. One commonly used strategy is to apply a familial approach to the ownership of genetic information[273].

In a minority of cases, a formal assessment of the seriousness and proximity of likely risks to health may be involved. It may be difficult for these assessments to be made, if there is little knowledge about the severity, penetrance or timing that unusual results may confer. This suggests that the only manageable or feasible option may be a move towards a simpler model of broad consent which can more effectively accommodate changing knowledge about risks and benefits of disclosure, analysis and feedback, and/or one that is based more comprehensively on the family unit[267].

Examples of existing practice such as open consent and the primacy of veracity

In what was more than a symbolic gesture, James Watson, the first person to have his whole genome sequenced, was given his entire genome on a hard disc and was offered genetic counselling to assist in interpretation. He chose to redact (not have analysed or disclosed to him) information about apolipoprotein E which has links with Alzheimer's disease, but the rest of his genome was made available on a publically accessible database[9] and it is worth noting that other commentators have been able to infer his Alzheimer's disease status even though this data was redacted.

Other projects have developed more strident open-consent approaches, such as the Personal Genomes Project which requires full disclosure and provides no guarantees of anonymity, privacy or confidentiality. Here the principle of veracity or truthfulness is offered as the prevailing ethical basis for obtaining a valid consent; other requirements within this research project are that participants have the equivalent of a master's degree in genetics as one of the entry criteria[254]. This approach is also justified on the basis that true genetic privacy is an impossible goal. But as McGuire argues, the precedent set for full disclosure of raw data, and a very high threshold for understanding the risks and benefits of research, are not applicable once the technology is rolled out more widely[267] and provides a very defensive basis for wider implementation (since in practice, the pool of eligible participants would be so small).

Relevance of ethical concepts such as genetic solidarity, altruism and justice

Moving beyond autonomy, a number of other ethical concepts are relevant to NGS technologies. In their report 'Inside Information' the Human Genetics Commission described a set of overarching principles which they argued should be taken into account. These included the principle of genetic solidarity and altruism[18], respect for persons[19], privacy, consent, confidentiality and non-discrimination[274].

8.4 Scope of duties of the health professional

8.4.1 Intended or pertinent findings

The scope of the research question in WGS

Whole genome sequencing offers a different paradigm from conventional genetic research or clinical genetic practice. Unlike genetic tests that target a particular known mutation, the scope of WGS analysis may be left deliberatively broad, the clinical or research question may be framed very widely, and findings may be generated that illuminate functional pathways or particular causes of disease. All these may be considered pertinent to the clinical enquiry. Although this open-ended approach is already common to technologies such as array CGH or even karyotyping[275] the difficulties raised by 'open' tests are likely to be much greater and more prevalent with WGS.

This approach to testing can be distinguished from situations in which the scope of the clinical enquiry is guided by a particular set of clinical circumstances. Here, WGS may be used on either a cohort basis[158] or on an individual patient basis[147] to systematically evaluate a range of possible pathological mutations[138]. It can also be used as a diagnostic aid in patients who are suspected as having a genetic condition[270] or whose symptoms arise in an atypical way[276] or in rare monogenic disorders[141]. WGS would not seem to raise distinctively novel issues where it offers an affordable 'replacement' technology for testing known mutations, or for clarifying genetic causation to an existing phenotype.

Distinguishing between causative alterations and non-pathological changes

A major challenge will be to distinguish between those changes that are pathological and aetiologically relevant to phenotype and those that are not. In particular, the context for mutations may influence their pathology, such that mutations and polymorphisms which are normally regarded as benign might have a major effect on susceptibility. Rather than categorising some mutations as pathological and others that are not, a spectrum of pathogenicity is likely to exist which will be strongly effected by interactions within and between genes and other risk factors such as those arising from the environment Even in a study of limited scope, Daiger *et al*. estimated that each individual would be heterozygous for a potential disease-causing variant every 200kb, (excluding insertions or deletions, copy number variants and repeats), amounting to several hundred to thousands of disease conferring mutations in each individual[138]. These will be difficult to evaluate clinically (given that the extent to which a mutation might be disease conferring will depend upon the background context). This challenge may be exacerbated in certain applications such as cancer genome sequencing where the cancer genome is unstable and evolving[198] (and may need to incorporate information about differential rates of repair or transcription rates to make sense of observed changes)[10]. In addition to difficulties assessing the impact of novel variants of unknown clinical significance, it is unclear how to evaluate the impact of known pathological variants in healthy individuals due to the influence of other (genetic and non-genetic) factors.

---

18 "We all share in the same basic human genome, although there are individual variations which distinguish us from other people. Most of our genetic characteristics will be present in others. This sharing of our genetic constitution not only gives rise to opportunities to help others but also highlights out common interest in the fruits of medically-based genetic research." Human Genetics Commission (2002) Inside Information, paragraph 2.11.

19 "Respect for persons affirms the equal value, dignity and moral rights of each individual. Each individual is entitled to lead a life in which genetic characteristics will not be the basis of unjust discrimination or unfair or inhuman treatment". Human Genetics Commission (2002) Inside Information, paragraph 2.20.

Professional obligations for disclosure

Experience with other emerging technologies suggests a lack of a consistent approach to uncertainty, and in particular, variations in the extent to which healthcare professionals see themselves as being under an obligation to communicate information they hold, regardless of its likely clinical utility. Arguably the politicisation of a choice agenda within the proposed NHS reforms could promote a culture of disclosure without sufficient evidence as to the possible harms that could arise. Worries about exposure to litigation (should potential abnormalities be detected but not communicated to parents) and the precedent of wrongful birth cases may also be relevant since where negligence has been proved, successful wrongful birth claims have been brought against laboratories and GPs[277;278].

Reference genomes and sample repositories

In the translational phases of NGS technologies there will be increasing reliance on repositories of samples and data of a number of different types. These may comprise a reference set of perhaps 10,000 individuals who illustrate 'knock outs' for the many human genes that are not necessary for survival[279]. In cancer genomics, these biobanks may take the form of linked genomic and transcriptomic information about individual cancers together with details of therapeutic responses[24]. More work is needed to determine the legal and ethical basis for these repositories. Recently published guidelines from the OECD on human biobanks and genetic research databases include explicit recommendations on consent, such as offering a spectrum of choices for consent, operating different strategies for engaging with the wider family or community, or offering a range of options for feeding back results[280].

Technical shortcomings associated with the technology

Since the technologies typically rely upon the analysis of short fragments, some types of genomic variation are less likely to be accurately accessed (such as structural variations, and repetitive sequences)[7] and therefore some conditions will be more difficult to detect using this technology than others. In order for these technologies to enter clinical practice, there is a need to assess the key determinants of sensitivity (proportion of mutations detected: absence of false negative results) and specificity (proportion of non-pathological variants reported: presence of false-positive results) against an existing diagnostic gold standard[140] so that these technologies can reliably be used to guide the choice of interventions and treatment[24].

8.4.2 Incidental findings

How relevant is the concept of incidental findings to WGS?

Recent technological advances have highlighted the importance of 'incidental findings' *i.e.* findings which are outside the scope of the research or clinical enquiry whether framed in a narrow or broad fashion. In the context of WGS, these span multiple clinical, genetic and social dimensions – from racial ancestry, misattributed parentage, consanguinity, disease susceptibility (a spectrum of highly predictive to minimal increased risk) through to reproductive risks. The concept has neither been well defined, or used systematically or consistently in the literature[281]. In the context of research, an incidental finding (IF) has been defined as 'a finding concerning an individual research participant that has potential health or reproductive importance and is discovered in the course of conducting research but is beyond the aims of the study'[282]. In clinical care, the IF is one which is outside the scope of the immediate clinical enquiry.

One central question is the extent to which whole genome sequencing may change existing paradigms of testing and care. Like other pioneering technological advances that have gone before, such as array CGH and single nucleotide polymorphism genotyping chips, there are broadly two alternative paradigms for use[275]. In the first scenario, targeted use of the technologies (which are applied to a cache or panel of known areas of interest) are used to identify genomic variations across a wider range of sites than would have previously been feasible or practical. If WGS is used in this way, the numbers of findings that are likely to be outside the scope of the initial clinical enquiry may be modest, although some unavoidable incidental findings are likely to arise (such as when variations disclosed through the investigation of a target area inevitably point to associations

with other diseases). Even where the clinical question is targeted, the scope of the clinical enquiry may be quite broad if technologies are used for diagnostic purposes in patients with developmental delay, behavioural abnormalities or birth defects. The possibility of co-incidentally identifying clinically actionable findings should be reflected in pre-test and post-test counselling regimes. For example, in a study of these patients using array CGH, around 0.6% (29/4805) were found to have genomic imbalances of less than 5Mb involving cancer predisposition genes which would have been undetectable using existing methods[283].
In the second scenario, the scope of enquiry is not framed in terms of existing clinical knowledge (such as in research use), and there is likely to be more scope for incidental findings to arise.

The need for a systematic approach

Future policy might be informed by the experience of dealing with incidental findings within the research setting. Whilst this topic is attracting considerable academic interest, there is a lack of consensus as to what policy direction should prevail. Some work has been done using incidental findings in imaging as a precedent for handling genomic information. In a recent review of incidental findings in "healthy" volunteers during imaging for research, Booth *et al.*[284] note prevailing inconsistent and ambiguous legal and ethical guidance, (less complete in the UK than elsewhere), and an urgent need for improved standards. A number of approaches have emerged in the context of genetics and genomics[266]: in the first, only potentially life-saving results are fed back to individuals. Justifications for limiting the disclosable results include that more extensive disclosure would blur the boundary between clinical care and research (thus promoting the therapeutic misconception – namely the belief of a research participant that the primary aim of the researcher is therapeutic rather than research). It might also impose operational constraints as a result of a lack of feasibility or expertise. This approach, exemplified by large-scale biobank research, emphasises the importance of collective responsibility and solidarity[285]. However, it may be difficult to evaluate what constitutes a potentially life-saving intervention, given that diagnosis and treatment are improving all the time and professionals may have contradictory views. A contrasting approach sanctions wider, albeit limited, disclosure of individual genomic results on the basis that this fosters beneficence, autonomy and a sense of reciprocity. Instead, it is argued, the focus should be on selecting the results to be fed back and identifying the procedures to be used.

Other groups have adopted a phased approach to incidental findings, whereby findings are categorised according to their likely age of onset, the treatability of condition and its potential impact for other family members (see Table and Boxes below). Often there is a discretionary element, which allows the research participant some discretion as to the results that are fed back. A range of approaches have been adopted within individual research projects, but more work needs to be done to identify whether a unified approach is needed or whether a flexible strategy works better.

Box 8.1: Case study

In a case study describing the exome sequencing of a child with persistent inflammatory bowel disease refractory to medical and surgical treatment, Mayer *et al.* allocated possible incidental findings into three broad categories that were framed more narrowly on the basis of the potential impact for the child[147;269]:

1. ALWAYS RETURNED: Information that will have a direct impact on the clinical care of a child in childhood
2. ELECTIVE RETURN*: Information about actionable adult onset disease (for the child in question) which may have significant implications for the parent's own current health maintenance
3. ELECTIVE RETURN: Information about non-treatable adult onset disease for the child in question

*It is unclear how generalisable this approach might be to informing risks for other genetic relatives such as siblings.

Table 8.1: Dealing with incidental findings

Wolf et al.[282] *provide a framework for dealing with IF's in research which straddles the consent process, the future analysis of archived data, and involvement of expert consultants both for verification of results and providing follow up. This framework identifies three categories of IF which are determined on the basis of their net benefit.*

Category	Relevant IF	Recommended Action
Strong net benefit	• Information revealing a condition likely to be life-threatening • Information revealing a condition likely to be grave that can be avoided or ameliorated • Genetic information revealing significant risk of a condition likely to be life-threatening • Genetic information that can be used to avoid or ameliorate a condition likely to be grave • Genetic information that can be used in reproductive decision-making (1) to avoid significant risk for offspring of a condition likely to be life-threatening or grave or (2) to ameliorate a condition likely to be life-threatening or grave	Disclose to research participant as an incidental finding unless s/he elected not to know
Possible net benefit	• Information revealing a nonfatal condition that is likely to be grave or serious but that cannot be avoided or ameliorated where a research participant is likely to deem that information important • Genetic information revealing significant risk of a condition likely to be grave or serious, when that risk cannot be modified but a research participant is likely to deem that information important • Genetic information is likely to be deemed important by a research participant and can be used in reproductive decision-making: (1) to avoid significant risk for offspring of a condition likely to be serious or (2) to ameliorate a condition likely to be serious	May disclose to research participant as an incidental finding unless s/he elected not to know
Unlikely net benefit	• Information revealing a condition that is not likely to be of serious health or reproductive importance • Information whose likely health or reproductive importance cannot be ascertained	Do not disclose to research participant as an incidental finding

The principles of scientific validity, clinical validity and utility and personal utility are increasingly being recognised as a means of identifying those results for which an ethical duty to feedback may pertain[286]. However, it may be impossible to objectively assess these criteria for many genomic variants. The need to validate results obtained within a research setting to clinically actionable standards will be impractical, at least for a transitional phase, since WGS is likely to throw up vast numbers of variants of unknown significance for which no mechanism (or capacity) exists for validation within the NHS. Moreover, it may be difficult to determine the personal utility inherent in individual decision making without compromising a right not to know. Nevertheless, at

international level, there now exists an ethical, albeit minority, view that there is a duty to return individual genetic research results subject to the existence of proof of validity, significance and benefit[286]. This approach is inconsistent with professional guidance in some countries including the US which stipulate that only valid, actionable, clinically significant results should be fed back to research participants[287]. The implications of incidental findings in clinical rather than the research settings are only beginning to be explored in the literature, and there seems to be a paucity of relevant guidance which extends beyond general professional guidance to work within established expertise and refer to appropriate experts where necessary.

Box 8.2: Alternative feedback framework

Bredenoord *et al.*[266] have proposed an alternative framework in which the default package is that life-saving data and data of immediate clinical utility are disclosed (from which research participants may opt-out)[288]. Three additional packages are identified that allow research participants to opt in to disclosure of incidental findings conferring potential or moderate clinical utility, data of reproductive significance and data of personal or recreational significance. This model allows greater discretion for research participant to have knowledge of results that do not have clinical utility.

Berg *et al.*[289] explore a possible framework for categorising WGS findings according to a combination of their clinical utility and validity (Table 8.2). In clinical practice this framework could be used for those findings that remain once possible diagnostic variants have been extracted. In a public health setting, where asymptomatic individuals are being assessed, this framework would apply to all variants.

It will be necessary to develop relevant patient pathways for referral: this is likely to have significant implications for funding and commissioning. Existing mechanisms for validating genetic tests through a structured gene dossier process, where the elements of utility, significance and benefit are clearly identified for each test, are likely to be too unwieldy and may be unworkable. Undoubtedly there will be a need to construct and maintain robust bioinformatic support systems as well as clarifying areas of clinical responsibility[106].

Misattributed parentage

As well as unintended findings related to health, incidental findings might reveal previously unanticipated misattributed parentage. This problem is not unique to the WGS approach (although it may arise more frequently than with existing approaches). One of the most challenging IF's is the finding of incest[153], consanguinity or misattributed paternity in the context of reproductive choice[290] and the possibility of this arising should be covered in general terms before testing. Clinicians should also be aware that where there is significant absence of heterozygosity (suggesting a degree of consanguinity), and if the mother is a minor, other social care issues might require further investigation.

The scope of obligations to identify and disclose harmful but preventable conditions within publicly funded health services

Another possibility is that mutations related to other harmful and preventable conditions ought to be routinely sought when WGS is undertaken for clinical reasons. In other clinical contexts this would be considered to be opportunistic screening and the debate is yet to be undertaken about how decisions of this nature will be taken.

Table 8.2: Deploying WGS in clinical practice and public health – a proposed system for classifying genes and variants (adapted from Berg *et al.*[289])

	Criteria:	Clinical utility	Clinical validity			Unknown clinical implications
Genes	Bins:	Medically actionable incidental information *e.g.* BRAC1/2 NF1	High risk incidental information *e.g.* Huntington disease	Medium risk incidental information *e.g.* APOE or carrier status for Mendelian disorders	Low risk incidental information *e.g.* Common SNPs	All other loci
	Estimated number of loci	10s	10s	1000s	10s (eventually 100-1000)	~20, 000
Alleles that would be reportable or not in a clinical context						
Variants	Known deleterious	YES	YES/NO	YES/NO	YES/NO	N/A
	Presumed deleterious	YES	N/A	YES/NO	YES/NO	NO
	Unknown significance	NO	N/A	NO	NO	NO
	Presumed benign	NO	N/A	NO	NO	NO
	Known benign	NO	NO	NO	NO	NO

8.5 Additional implications for the NHS

Dealing with incidental findings that are generated outside the NHS (such as those generated from research settings)

There are concerns that the development of NGS technologies could have a significant effect on the NHS in a number of ways: research participants who are identified as having known pathological mutations outside the scope of the research question could seek counselling and treatment (examples include conditions such as a familial susceptibility to breast cancer or another inherited cancer syndrome). These 'at-risk' individuals might already be aware of their increased susceptibility to disease on the basis of their family history. These cases are likely to be supported by a clear body of evidence as to clinical validity and utility.

However where variants are of unknown significance for an individual, the responsibility of the NHS to investigate may be less clear, and the requirement to do no harm may prevail (for example, in not subjecting an individual to years of invasive screening where a presumed cancer susceptibility variant is actually harmless). It is becoming increasingly clear that the 'normal' genome may include many 'pathological' mutations including ones that were previously thought to be incompatible with life, which may be an inherent property of the genomes themselves, or perhaps a product of genomic variation database curation and collection.

Quality assurance (QA) and clinical standards

Existing NHS laboratories might also be involved in re-testing of samples to validate results obtained in a research setting to clinical standards, which might impact upon the capacity of laboratories to deal with their existing workloads. In this developmental phase of the technology, approaches are not standardised and technical and interpretative errors are bound to occur. Given the less stringent controls on research practices, issues such as sample mix-ups, contamination and false reporting are likely, and any findings to be used in a clinical setting would therefore need to be confirmed in a diagnostic laboratory before relying upon the results to make clinical decisions. The basis for good professional practices are emerging: these include some consensus on the extent to which analysis should be repeated to provide sufficient certainty (for example as to read depth) and where clinical tests exist, the need to validate research findings to clinical specifications and standard operating procedures (SOPs)[140]. However if tests are in development in a research setting, then there may not be accredited NHS laboratories which can immediately adopt this role.

Professional education and the development of disease specific professional guidance

As more is learnt about the pathology of disease, there will be an increasing need for healthcare professionals at all tiers of the health service to have some degree of knowledge about genomic susceptibility and/or develop protocols for on-going referral. It seems likely that there will be increased reliance upon disease specific professional protocols to standardise clinical approaches that can be applied across health services (such as that formulated for retinitis pigmentosa[138]).

Ensuring equity of access

Access to NGS technologies in the initial phases will most likely be through recruitment to research projects rather than as part of clinical care. Experience has shown that integration into routine clinical practice in a particular geographic area frequently follows these 'centres of research excellence' and may lead to inequity between such regions and others that have not benefitted from research experience and enthusiasm[123].

Providing on-going care for individuals who have accessed tests on a direct-to-consumer basis

Although whole genome sequencing is available on a direct-to-consumer basis, the high cost of sequencing is likely to be a deterrent to the majority. Over time however, the reduction in sequencing costs may make it cost-effective for DTC companies to use NGS technologies instead of SNP testing. Despite the small numbers involved, direct-to-consumer genetic testing may be another driver for mainstreaming genomic knowledge. Individuals who have made use of privately available NGS will seek help from the NHS in interpreting results or further investigating possible current of future health problems raised by test results. Emerging research suggests that between 25% and 78% of participants might seek further guidance on their results from their physicians[291;292]. Providing this support could divert health service resources (given that DTC or privately funded tests are more likely to have been adopted by wealthier citizens).

8.6 Regulation of diagnostics

The regulation of whole genome sequencing approaches has many facets including laws and policies that govern the collection of data and samples[293] (such as the OECD Guidelines on Human Biobanks and Genetic Research databases), the CE marking of the devices used to deliver the test, and the accreditation of the clinical test itself. Some regulation operates at European level, other regulation at national level.

8.6.1 Quality assurance, equipment validation and CE-marking

As with other technologies used in clinical settings, there is a need to develop robust systems of quality assurance for NGS devices and processes. The proliferation of NGS systems (each with limitations of read length and depth, throughput or error rate) indicates a number of possible approaches depending upon the systems and their likely applications[294].

8.6.2 In Vitro Diagnostic Medical Devices Directive

European legislative framework

The uptake of these new technologies will also be influenced by the prevailing regulatory framework governing the development and marketing of novel diagnostic medical devices. Within member states within Europe, this process is governed by the *In Vitro Diagnostic Directive 1998*[295]. The purpose of this legislation is to provide the conditions that are 'necessary and sufficient to ensure, under the best safety conditions, free movement of the [applicable] *in vitro* diagnostic medical devices'.

Technical performance accreditation or CE marking

The In Vitro Diagnostic Directive sets out the basis for obtaining a CE mark, which is a measure of performance assessment. Most diagnostic genetic tests are classified as being 'low risk' under the Directive, which means that manufacturers are not required to solicit a third party assessment but can self-certify.

The scope of the Directive

In simple terms, the scope of the directive applies to products which are used to examine specimens from the human body for the purpose of obtaining information to be used for certain purposes[20]. The Directive sets out the essential requirements that must be satisfied, such as CE marking for higher risk devices, a risk categorisation scheme (in Annex II), a mechanism for conformity assessment (Annex III) and a national system of delegation of authority to notified bodies, to monitor compliance with the Directive. The MHRA is currently the statutory body with this responsibility for monitoring compliance in the UK.

Whole genome sequencing approaches fall within the scope of the IVDD Directive in that they comprise analysis of human specimens, which concern 'a physiological or pathological state'. However, these technologies allow previously unrecognised mutations from cellular and tumour DNA to be identified and as such, challenge the paradigm of the IVDD which is predicated upon detecting the presence or absence of a pre-determined genetic analyte[296]. To the extent that WGS approaches are used to identify known mutations, clearly they can be used for simultaneous testing for multiple applications straddling different categories of risk (but are not unlike other technologies such as array CGH in this respect). Increasing use of multiple simultaneous tests have caused some to question whether there are specific parts of the Directive that should be amended, such as the 'essential requirements', the procedure for 'conformity assessment', or whether more radical reform might be needed. Any review might also allow for a more explicit focus upon the competence of genetic test providers (health institutions, enterprises and personnel) both to deliver the tests and interpret their results to patients and their families[296].

8.6.3 Laboratory Accreditation

OECD guidelines

One of the most influential sources of guidance has been that from OECD, whose guidance has underpinned subsequent European Directives[297]. The influential nature of this guidance stems from its global reach, and the fact it can facilitate harmonisation of best practice guidance. Relevant publications include OECD Recommendations on Quality Assurance in Molecular Genetic Testing (2007)[298].

---

20 '*in vitro* diagnostic medical device' means any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, equipment, or system, whether used alone or in combination, intended by the manufacturer to be used *in vitro* for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information:
- concerning a physiological or pathological state, or
- to determine the safety and compatibility with potential recipients, or
- to monitor therapeutic measures

Laboratory accreditation (ISO 15189)

Until recently in the UK, laboratory accreditation has been organised through a voluntary network, Clinical Pathology Accreditation (CPA), which developed its own standards. However, the enactment of ISO 15189 under the aegis of European legislation called into question the extent to which the CPA accreditation should be rationalised within the demands of a European context. As a result of this debate, CPA was subsumed within the UK Accreditation Service (UKAS) which now has responsibility for delivering accreditation (initial inspection, followed by annual surveillance visits and re-accreditation every four years). A similar system exists in the US with the College of American Pathologists (CAP) accreditation.

8.6.4 Genetic test accreditation and approval

The accreditation of clinical genetic tests has evolved somewhat piecemeal over the last decade. In part this is because the system for validation is not consistent across jurisdictions. Within Europe for example, member countries have taken different approaches to implementing the IVDD Directive, such that some countries rely heavily upon the exemptions provided for the development of 'home-brew' tests. In practice, this has meant that many laboratories in the UK have independently developed tests for the same condition, which are utilised for healthcare within their immediate geographical area. Since interpretation of the relevant regulatory requirements allows member states some discretion, much of the existing process for genetic test approval has relied upon developing consensus opinion and of best practice guidelines for genetic testing.

EuroGentest validation group

A voluntary process for the validation (novel test) and verification (where an existing performance standard is available) phases of the ACCE framework[21] has been described by Mattocks *et al.* on behalf of the EuroGentest Validation Group[299]. This paper usefully summarises how the nature of the test (*i.e.* whether it is quantitative or qualitative) and the quality of the result[22] interact to give a range of possible dimensions. It remains to be seen whether NGS technologies represent an increase in complexity or whether these technologies introduce novel dimensions (in which case new validation processes might be required).

UK Genetic Testing Network

The UKGTN has led the way in pioneering systems for approving and recommending tests for NHS service provision, (of which most are single gene disorders). Evidence is provided as to each of the key components of the ACCE framework in the form of gene dossiers (namely analytical validation, clinical validation, clinical utility and any particular ethical, legal and social implications arising from the test). Endorsement by the UKGTN currently provides the mechanism for those commissioning genetic tests to fund tests. Thought needs to be given as to how these current mechanisms might be relevant for WGS approaches where multiple genes are analysed in parallel. Issues of capacity and competence also need to be addressed.

Companion diagnostics/pharmacogenetic tests

Other regulatory strategies might be to restrict the use of certain types of pharmaceutical product to those who have previously had a diagnostic test which predicts the likely safety and efficacy of the drug. Processes for developing these companion diagnostic tests are being identified, and these processes could be useful in the context of WGS.

Laboratory regulation and outsourcing across borders

In a global market place, it seems likely that NGS technologies will increasingly be outsourced across geographical boundaries so that users and patients in the UK will access WGS services from another country. Certainly health service users and research participants need to be made aware that their data and samples might be transferred overseas and their privacy might be compromised as a result. It is also important that laboratories conform to universal robust and enforceable standards.

---

21 **A**nalytical validity, **C**linical validity, **C**linical utility and **E**thical legal and social issues

22 For example, whether the result is categorical (*i.e.* placed in one of a number of defined categories)

Regulating an 'interpretation' service

In the future, data collection, analysis, bioinformatics, interpretation and clinical feedback could be provided by different organisations (with separate legal status). This has implications for how these constituent parts of the service are integrated and regulated.

8.7 Intellectual property

8.7.1 Gene patents

Patents describe the protection of intellectual property in an invention for a limited period which provides a means of recouping the time and money invested in its development. Gene patents[23] can be divided into a number of different types; product claims to isolated DNA or RNA sequences; product claims for diagnostic test kits; process claims for diagnostic testing methods and product claims to the devices used for testing (such as gene chips and microarrays)[300]. In practice, many gene patents exert multiple claims in a number of these forms. Although patents are available in many different countries, they vary in their scope and the protection they confer, resulting in a lack of harmonisation[301]. Additional uncertainty is caused by on-going high profile litigation (particularly in the US)[259;302] between patent holders, and patient and professional organisations, concerning the patentability of genes (as products) and genetic tests (as processes)[303;304]. There are also concerns that rigorous patent enforcement might have a stultifying effect on the uptake and implementation of these technologies[305].

The impact of patents on the implementation of NGS technologies is currently uncertain. Current estimates suggest that scientists have applied for patents for around 20% of the human genome. There are two questions which arise in relation to gene patents. The first, (addressed in this chapter) is the extent to which the users of NGS technologies might be liable to existing gene patent holders. A second question is the extent to which developers and manufacturers of NGS technologies can protect their inventions through patenting. This chapter only addresses the first of these questions. A potential problem is that historically patent holders have claimed rights in respect of isolated DNA sequences (including genes and mutation sites), and cDNA sequences as well as methods of associating a gene/mutation with a phenotype to diagnose an illness or disorder (so called association patent claims).

There is a real possibility that NGS technologies may infringe valid association patent claims if NGS results are used to diagnose the illnesses and disorders described in the association patents. The situation is less clear with relation to valid patents on isolated genes/mutations. Within Europe, a recent case (*Monsanto Technology LLC v Cefetra BV,* C-428/08) has suggested that DNA sequence claims are implicitly, if not explicitly use-limited[306]. This case concerned plant breeding and the implications for diagnostic testing are as yet unclear.

To date in the UK, NGS technologies have been developed in the context of research (for which a defence to infringement of a valid patent applies), so the willingness for existing patent holders to grant licenses to users of these technologies within the NHS is as yet untested. Since WGS approaches sequence the whole genome, a theoretical liability might arise in respect of every mutation site for which a patent is held. This could result in a number of liabilities arising for a single mutation site, and also for liabilities accumulating across the genome. These multiple rights (or so-called patent thickets) would be administratively burdensome and if strictly enforced, the cumulative cost of multiple licences could render WGS approaches uneconomical. Indeed, in a recent dissenting judgement in a decision of the Federal Circuit which largely reversed the District Court's decision in *Association of Molecular Pathology v. USPTO* (the *Myriad* gene patent litigation), Judge Bryson described patent thickets as "a significant obstacle to the next generation of innovation in genetic medicine – multiplex tests and whole-genome sequencing" noting that the costs of determining the scope of such claims could be prohibitive[302]. There are also doubts as to the effectiveness of strategies designed to streamline the logistical process of acquiring multiple patents, such as patent pools or royalty-collection clearinghouses[305].

[23] *I.e.* any patent which claims DNA or RNA sequences as *products* or concerns the *processes* to make or identify DNA or RNA

Current recommendations for reform include limiting the patentable subject matter and providing that licenses are both non-exclusive and easily obtainable[307;308]. Other commentators support an extension of the exemptions to patent infringement, including a broader research exemption and an exemption for activities related to genetic testing used for patient care purposes. However these reforms do not address the uncertainties which are at the heart of current litigation, such as the failure to distinguish between different types of claim (for example, between natural (non-isolated) or isolated DNA sequences) and to acknowledge that the value derived from DNA is a product of both its chemical structure and the information it confers. In Europe existing patent protection seems to bite at the point that research has clinical application[300] although within the US the evidence is somewhat more ambivalent[309]. Even so, some commentators have suggested that were reforms to be enacted, they seem unlikely to mitigate the 'near perfect storm that is developing at the confluence of clinical practice and patent law'[305]. Nor do they address how the legacy of DNA patents will apply when novel functions or purposes are revealed about existing DNA sequences.

8.7.2 Data release and publication strategies

Concerns about patent thickets prompted the Bermuda agreement and the formation of the SNP Consortium to ensure that genomic information remained in the public domain by systematically and defensively filing provisional patent applications (to establish priority and dissuade competitors) and then convert or abandon the patents and share the data on publically accessible databases[301]. Although of practical importance, this strategy does not seem to address the difficulties outlined above.

Where NGS technologies are used to identify biomarkers for a personalised approach to cancer therapy, the IP implications do not seem to have been addressed in the literature. Clearly the methods used to develop these markers (such as the PARE method described by Leary) are patentable, but at least in the initial phases, patient numbers are likely to be small and initially confined to research[192] and therefore might be less commercially significant.

8.8 Implications for the public

An important measure of the uptake of any new technology is the extent and speed with which it is translated and accepted by relevant stakeholders (including healthcare providers, funders, patients and the public). The introduction of novel technologies can be disruptive and negatively impact upon the perceived trustworthiness of healthcare providers and services. Issues of public perception are therefore profoundly important.

8.8.1 Genetic eceptionalism and genetic determinism

Some categories of genetic information have particular significance. For individuals diagnosed with a single gene disorder, for example, the predictive nature of genetic information confers evidence of an individual's past, present and future health. It may also have additional implications for relatives, future reproductive choices and for future generations[274]. Thus rational beliefs about some types of genetic information may motivate or stall individual behaviour.

However, for some, genetic information may be imbued with a more disparate 'exaggerated symbolic significance' as a 'blueprint' or 'essence' of life itself, which may be viewed as a justification for all genetic samples or data being treated as being unique or special. This erroneous belief, known as genetic exceptionalism, is sometimes used to vindicate special protection or other exceptional measures by policy makers, regulators and other stakeholders[310]. The term 'exceptionalist' was coined following the discovery of the human immunodeficiency virus (HIV), with the aim of capturing both the uniqueness of the disease and the dual belief of policy makers that testing was both essential for prevention and treatment and that individuals might reject testing if absolute confidentiality could not be guaranteed[311;312]. Others have argued that the lessons that need to be learned from exceptionalist stances to HIV and genetic information suggest a need for a paradigm shift (particularly in public health) in the way that all sensitive health information is protected[313;314].

Another commonly held set of misconceptions concern genetic determinism. This term is used in various ways to characterise the causal relationship between genetics and behaviour. One version is that genetic factors alone determine health or behaviour rather than influence it. Historically, there have been attempts to justify state sponsored eugenic policies on the basis that criminality or 'dissolute' behaviour is determined primarily by genetic factors. Another is that individual behaviour, beliefs and desires are fixed by genetic factors, and in the context of direct-to-consumer testing for example, that learning about personal genomics could adversely affect health by promoting a fatalistic attitude. These misconceptions are important, as they provide the context for policy development in this area.

8.8.2 Sources of evidence

Evidence on public perceptions of NGS and WGS technologies is sparse and dispersed through a number of different disciplines: these include literatures on the implementation of novel technologies particularly where clinical findings have generated uncertainty (such as evidence from imaging and ultrasound), and on unclassified variants and uninformative results. Evidence from the psychological literature suggests that individual responses to genetic risk communication and risk stratification are not straightforward. Historically, the presumption has been that increased information is of benefit (the deficit model) and there has been a failure to recognise the possible potential for generating harm[315].

8.8.3 Perceived utility of individual testing

The context for communicating genetic risk information has an impact on how a genetic test (including offer, uptake and results) is perceived. One tangible measure of public perception is the uptake of genetic testing or screening. Where the test has clear clinical utility (because it can be used for more effective disease management, through more effective drug treatment or other intervention) then rates of uptake may be high. Where there is less evidence of clinical utility, uptake is likely to be lower. Perceived risks may even be more important than actual risks. Where treatments or interventions are minimal, such as for the severe incurable genetic disease, Huntington's disease, the utility associated with testing may be perceived in more individualistic and personal terms, such as promoting autonomy interests.

In the context of predictive genetic testing, another tangible measure of utility might be the extent to which the results of testing may prompt changes in behaviour or lifestyle. In contrast to the fears about genetic exceptionalism, recent reviews have suggested, from the rather scanty evidence available, that feedback of predictive genetic risk information may not motivate behaviour change[316] nor encourage fatalistic beliefs[317].

As previously noted, genomic information may have significance for genetic relatives of the person being testing. A genetic test result in one family member may allow testing to be cascaded out to others at-risk. It is not clear what the threshold for disclosure might be.

8.8.4 Reproductive decision making

Another dimension for using genetic information is in the context of reproductive decision making. Many studies have examined the changing role of prenatal diagnosis, and the extent of participant engagement and disclosure in this application constitutes a useful paradigm for how technological advances have been implemented into clinical practice. For example, when first introduced, prenatal ultrasound screening was targeted at safe fetal growth rather than diagnosing structural abnormalities in the unborn child. As the scope of routine ultrasound screening increased to include multiple objectives (obstetric risk reduction, fetal diagnosis and screening for chromosomal aberrations), markers of uncertain clinical significance and markers known to have high false-positive rates were routinely reported. In other cases, soft markers that had been customarily reported to parents were subsequently shown not to be associated with abnormality[318]. As a result some commentators argue that this placed heavy burdens upon parents such that 'medical information about the unborn child, considered as value neutral within the biomedical paradigm, is thus transformed into a profound and private moral dilemma'[319].

Thus experience suggests that the problems may be two-fold: technologies may be introduced prematurely, before evidence of associations between observed findings and diseases have been robustly established, and that, in the reproductive arena at least, there is a growing tendency to place the burden of responsibility on the parent or patient[318]. Evidence also suggests that for some parents, being the recipient of adverse findings may have enduring psychological impact (whether provided during the first trimester of pregnancy[320] or once a child is born).

Even if the information that is communicated is unlikely to have any immediate implications for the child's health (such as being confirmed as a carrier of sickle cell or cystic fibrosis following newborn screening), anxiety and distress is exacerbated by poor communication practices and a failure of screening participants to understand the possibility of an untoward result before screening commences[321].

Scaling up this process to simultaneously screen for hundreds of severe recessive diseases of childhood has profound social, legal and societal implications, particularly as early pilots have suggested that about 10% of existing disease mutation annotations are incorrect, and that, on average, individuals in the general population carry three recessive childhood disease mutations (range 0-7)[170]. Past experience therefore suggests a need for a cautious, mindful approach and a need for robust evidence of clinical utility. In the context of scaling up screening technologies, research will also be needed to explore how participants interpret these risks (within a context of 'normal' variation) as well as alternative strategies for participant education and engagement.

8.8.5 Social context

The social context for testing being offered can have profound implications for patient or participant understanding. It can also influence the extent to which test results are welcomed or distrusted, particularly if there are implications for other family members. Less advantaged or vulnerable groups having low health literacy (including the elderly, disabled or socially disadvantaged) may be less able to access these technologies, through a combination of poverty, illiteracy and ignorance.

Religious or cultural factors may also influence an individual's world view: in the context of reproductive choice, a woman may regard the health of her baby as being 'God's will' and the subsequent birth of a disabled baby as a test of faith and as such, something even to be welcomed[322]. Religious and cultural beliefs may also dictate the options that are available to an individual following receipt of a test result. For example, if the identification of a person as a carrier of a genetic disease is regarded as a source of stigma or dishonour then some types of population based testing (such as carrier testing) may be culturally unacceptable, particularly if cultural norms also favour familial relationships which are likely to increase the incidence of recessive diseases (such as first cousin marriage).

Faith views on termination of pregnancy may also be important. For example, a number of Islamic states have ruled that termination of pregnancy for a fetus with a serious disorder may be permissible (within 120 days of conception if certain conditions are met, whilst the legal framework within the UK generally allows for termination up to 24 weeks gestation)[322]. However, there is a need to recognise diversity within different faith groups and avoid stereotypical views based upon people's ethnicity or religion[323].

8.8.6 Strategies for ameliorating harms arising from testing

If simultaneous testing for multiple conditions is implemented, it may prove impractical for those delivering tests to provide participants with exhaustive information about each condition. In a research context, various strategies have been adopted to attempt to simplify processes of risk communication whilst ensuring informed choice. Hewison *et al*. explored attitudes to prenatal testing and termination of pregnancy for fetal abnormality in order to ascertain whether conditions could be clustered and the consent process simplified. Although severity of the condition was identified as an important dimension, considerable variability in individual opinion was noted suggesting that it might not be a straightforward process to apply this strategy in practice[324].

8.8.7 Medicalisation of the genome

In other contexts, commentators have worried that access to genomic information of uncertain predictive value could lead to a process of 'medicalisation' whereby 'healthy' individuals become anxious about their future health. If technologies are utilised within a research setting, we have already seen that there may be concerns that participants believe that the primary purpose of the intervention is care rather than research (therapeutic misconception). In the context of reproductive medicine, there are also related concerns that if these technologies are used for prenatal genetic diagnosis or prenatal testing, that they may cause commodification of the embryo or fetus. In neonatal screening, the American President's Council on Bioethics noted an inexorable trend towards broadening screening (in response to the emergence of new technologies and to pressure from parent organisations and consumer groups), to the extent that this knowledge is increasingly described as a parental 'right to know'. This includes broad support for a parental right to have all available genetic information about their child[325] and evidence that a substantial minority would be interested in testing their youngest child for severe untreatable late onset conditions[326].

Evidence from other settings suggests broad support for NGS technologies and WGS approaches. For example, comparisons of hypothetical and actual participation rates in biobanks suggest that factual willingness is greater than hypothetical rates (12/22 studies) suggesting that factors like altruism, trust and a sense of duty are important[327]. Arguments about genetic exceptionalism in policy making may also be important in this context[254;328].

8.8.8 Personal consumer genomics

Previous sections have highlighted the increasing role of consumer genomics. Undoubtedly there are examples of bad practice: the claims made by some consumer genetics providers to potential consumers, highlight some of the risks involved. However policy makers are divided about what constitutes an appropriate response, and some have questioned the need for special privacy or discrimination rules for genetics given that bathroom scales and the blood pressure cuff may be far more predictive[329].

Emerging evidence suggests that the perception of genetic risk information may be different according to whether there is mediation via professionals (lab technicians, researchers or clinicians) or on a direct-to-consumer basis. Where testing is provided using standard genetic counselling regimes, limited evidence suggests that contrary to the fears of clinical professionals and regulators, that the majority are underwhelmed by genomic risk information: Heshka *et al*. found no differences in risk perception between carriers and non-carriers after 12 months (although some differences in screening or prophylactic intervention uptake were recorded)[330]. Where genetic risk information is accessed on a direct-to-consumer basis, the provision of genetic risk information has little influence on behaviour including psychological health, diet or exercise behaviour, or uptake of genomic profiling[292] although Bloss *et al*. found that residual anxieties may remain about some aspects of testing including breaches of privacy[331]. In this study, 44% of those enrolled, failed to complete the study: of those that did, 10% spoke to a genetic counsellor provided by the company delivering genetic testing and 26% reported sharing their results with a physician, suggesting that the knock-on effects for the NHS could be substantial if direct-to-consumer testing is adopted more widely.

More work is needed as to how the uptake and possible harms associated with direct-to-consumer genetic testing might be mediated or affected by social networking. In a review of attitudes of social networkers to direct-to-consumer provision, social networkers expressed high rates of interest. McGuire *et al*. found that 6% of those surveyed had already accessed personal genome testing services. Of the 64% surveyed who would consider using personal genome testing, 74% would use it to gain knowledge about disease in their family (including their children in 54% of cases) and 78% would seek help from their physician in interpreting results[291]. The insights offered by the Genomes Unzipped website are an excellent source of critical analysis and personal insight into personal genome testing[332].

8.8.9 Managing expectations

One of the most problematic aspects of this technology is how it might influence public expectations concerning the type of society that is possible or probable in the future. This is likely to include identifying the range of diseases for which treatment (and in the context of reproductive choice, termination of pregnancy) might be offered. This requires more than an objective assessment of risks and benefits: the public perceptions of these technologies as an agent for better targeting medical care, accessed only by the richest in society or even as a means of imposing social control (through increased knowledge of behavioural genetics) could dictate its uptake and use. The extent to which public perception may impact on the development of these technologies urgently requires research, as there is currently a lack of empirical data on issues such as incidental findings.

There are opportunity costs associated with these technologies, and their systematic uptake might be at the expense of other interventions. Thus comparative risks and benefits of alternative interventions should be explored[333]. This suggests that as professional practice is developed, a wide range of stakeholders needs to be engaged in discussions about what the limits of transparency might or should be.

8.9 Key issues raised by WGS

Whole genome sequencing technologies challenge existing paradigms of care in a number of ways, and in so doing, raise multiple ethical, legal and social issues. Firstly, the capacity for data generation from these technologies is unprecedented: this highlights practical difficulties inherent in accessing, storing, processing and interpreting the data.

As well as these logistical challenges, a set of issues arise in relation to the rights and responsibilities flowing from that data. These include factors such as whether unexpected or incidental findings should be communicated to individuals being tested, their genetic relatives, or even insurers, employers or researchers who may have commercial interests in the data. Where disagreements arise, legal claims about property interests over samples may prove to be important over the next few years, in trying to disentangle competing entitlements. Since multiple agents may be involved in the process of extracting clinically useful information from raw data, there are also questions about how the responsibilities of each professional group should interconnect, and which group should ultimately be liable to the patient or participant, particularly if things go wrong.

The ethical, legal and social implications of these technologies depend in part on the applications for which they might be used. If WGS is used as a public health screening tool for the purpose of reproductive choice, then the consequences will be very different from targeted clinical use within a closely regulated medical genetics setting. Issues of global ethics also arise, when considering how best to translate these technologies equitably to populations on a worldwide basis.

9 Health economics

9.1 Health economic evaluation

The continued growth in healthcare spending in the UK seen over the last few decades is unsustainable over the long term. This has been recognised in the 'Nicholson challenge' of preparing for harder times with the *"need to release unprecedented levels of efficiency savings between 2011 and 2014 – between £15 billion and £20 billion across the [National Health] service over the three years"*[334]. Rapid advances in both genomic knowledge and genomic technologies coupled with a difficult financial climate is leading to increased demands on the National Health Service (NHS) within already limited resources in which to meet existing demands. Clearly the two are incompatible without new ways of working, possibly with substitution between these new and traditional technologies in order to realise potential savings within the NHS.

Health economics is a discipline that attempts to tackle the problem of scarcity of resources in the healthcare setting and uses economic evaluation as a method of informing the decisions of the healthcare system on which healthcare intervention to fund from the limited resources that are available. Drummond *et al.*[335] have defined economic evaluation as "*the comparative analysis of alternative courses of action in terms of both their costs and consequences*" and at their most basic level is the need to identify, measure, value, and then compare these costs and benefits. The three types of full economic evaluations are cost-benefit analysis (CBA), cost-effectiveness analysis (CEA), and cost-utility analysis (CUA). Another commonly used partial economic evaluation is cost-minimisation analysis (CMA).

- CBA: both costs and benefits (health and non-health) are measured in monetary units and may present net monetary gains (losses) in the form of a cost-benefit ratio or a monetary value representing the net benefit (loss) of one programme over another.
- CEA: interventions with a common outcome, often in natural units (such as cases diagnosed or life years saved), are compared to determine which intervention maximises the outcome for the same input to produce a cost per outcome unit.

- CUA: measures outcomes of alternatives interventions in terms of a more generic utility measure such as quality-adjusted life-years (QALYs) which incorporates views and perceptions of value from the end-user and health state utilities combined with length of life to create a single metric presented as the cost per QALY. A limitation of using CUA is that it relies on health related quality of life surveys to capture a utility gain/loss and these may not be accurate or sensitive enough for all anticipated benefits/disbenefits.
- CMA: this method can be used when two or more evaluated alternatives produce outcomes (health effects) that can be argued to be sufficiently similar or equivalent, the choice between alternatives then comes down to costs with the least costly chosen[335].

With a lack of Randomised Control Trials (RCT) performed to date evaluating both costs and benefits derived from WGS, as well as the uncertainty surrounding these measures, there is a clear need to incorporate all available and relevant evidence into an economic analysis. This can be done by building decision analysis models that attempt to provide a systematic approach to decision making under uncertainty[336]. Economic decision models use mathematical relationships to define a set of possible consequences for a pair of alternative options under evaluation. Probabilities express the likelihood that each possible consequence occurs along with its cost and outcome measure, with each alternative option under evaluation producing an expected cost and an expected outcome which are weighted by these probabilities[337]. Probabilistic sensitivity analysis (PSA) can then be used to address the contribution of parameter uncertainty on decision uncertainty and provide estimates of expected costs and outcomes[338].

Key strengths of economic decision modelling that lend themselves to evaluating genetic technologies are that they allow for variability and uncertainty associated with all decisions and allow models to adopt a sufficiently long time horizon so that all key costs and benefits can be included and factored into decisions[337]. Because many of the psychosocial benefits/dis-benefits are highly personal in nature along with highly personalised treatment in the case of inherited cancer patients, economic evaluations may need to develop a more individualistic approach similar to that of N-of-1 clinical trials where a patient also acts as their own control.

9.2 Limitations of current economic evaluations

The methods described above can be applied to the economic evaluation of NGS. However, methodological challenges arise with each as a result of factors such as the multiple uses of NGS technology, its dynamic nature and the uncertainty over resulting health outcomes. Drummond *et al*.[339] have argued that although the general methods for economic evaluation are well developed, several methodological challenges arise when diagnostic devices (such as next generation sequencing for genetic testing for example) are evaluated, which we list below.

9.2.1 Evaluation of complex diagnostic technologies with multiplex use

Diagnostic devices often have multiple applications where the overall value of a device could be assessed as an average of its multiple uses (as NGS/WGS is able to test for many diseases or many genes for a single disease at once). This challenge may be overcome in the same manner as drug evaluation in that the value of NGS/WGS can be assessed individually for each application. However, by doing this, some of the benefits of this technology as a 'one-stop shop' may be lost. However, the attribution of cost is also difficult on a single application basis with many of the cost-savings achievable via the use of a 'one-stop' application.

9.2.2 Diagnostic odyssey

Using NGS technology for a genetic diagnosis has the potential to reduce the diagnostic "odyssey" - the protracted search for a diagnosis to a health problem – faced by many patients and their families. Again economic modelling has a potential role to play in helping to decide at what stage these quests become cost-ineffective. Earlier diagnosis and treatment can lead to cost savings through improved outcomes as well as savings made through the reduced need for additional

diagnostic tests and their associated costs. These savings may be difficult to quantify and evaluate and may require a probabilistic modelling approach.

9.2.3 Lack of evidence

The development of drugs necessitates the need for RCTs in order to determine efficacy and effectiveness and these studies provide a platform for conducting economic evaluations. Diagnostic devices are more difficult to evaluate in an RCT due to the intermediate nature of a diagnosis that then requires a treatment or course of action and also the long-term impacts of a genetic diagnosis (such as reproductive decisions and the wider implications of a genetic diagnosis that fall on family members as well as the individual). In many cases, the value of the information added by WGS to a treatment pathway still needs careful elicitation.

9.2.4 Wider economic implications/training needs

Due to the expertise required in the application of NGS there is a need for training of both laboratory scientists and healthcare professionals involved in the clinical management of the conditions tested for. This expertise is needed for improved cost-effectiveness (altering both costs and outcomes) and can be considered as a wider economic implication of NGS technology that needs to be accounted for (and factored into the evaluation). The state of readiness of lab scientists and healthcare professionals would therefore seem to be a significant precondition and hence a significant cost issue for an economic evaluation to take into consideration when conducting a one-off evaluation or a systematic review. A whole system evaluation of NGS would need to assume a common state of readiness as well as the use of similar equipment and other laboratory procedures for cost calculations, which is currently not the case across the NHS.

9.2.5 Evaluating a dynamic technology

As NGS is developing at a fast-pace this causes three further problems for economic evaluation. First, new technologies have an associated “learning curve” whereby the efficacy improves over time. Second, laboratory protocols and the knowledge created from the use of the technology as well as the creation of networks of users and evidence databases means that the technology is constantly evolving. There are also new advances in knowledge capital being made that have an inherent and tangible value which needs to be taken into consideration in any economic assessment, especially if this knowledge is not publically shared and must be paid for through normal market-based exchange. This means that there is unlikely to be a period in the next few years, as the technology continues to evolve, where the technology is stable enough to allow meaningful comparisons over a given time frame (*i.e.* we are trying to evaluate a ‘moving target’). Third, because of the fast moving nature of both the technology itself and the knowledge arising from use of the technology, many of the potential benefits and costs are not well defined yet and as such there is a lack of formal economic evaluations.

WGS holds the potential to hedge these benefits through data storage, however the costs fall into the present whilst the benefits often, no matter how significant, may not emerge for many years. As prices continue to fall, as can be seen from Figures 9.1 and 9.2, this further complicates not only economic evaluation but also procurement of such technology in the NHS. It might however make NGS/WGS better potential value for money as both more unknown benefits are discovered and tests become cheaper to carry out (without the need to re-test) in the long-term. The dramatic fall in DNA sequencing costs measured both in terms of Mb sequenced and whole genome sequencing costs from January 2008 is the result of using ‘second generation’ sequencing platforms highlighting the rapid technological progress.

Figure 9.1 Cost (in US$) per Mb of DNA sequence as determined by the National Human Genome Research Institute.

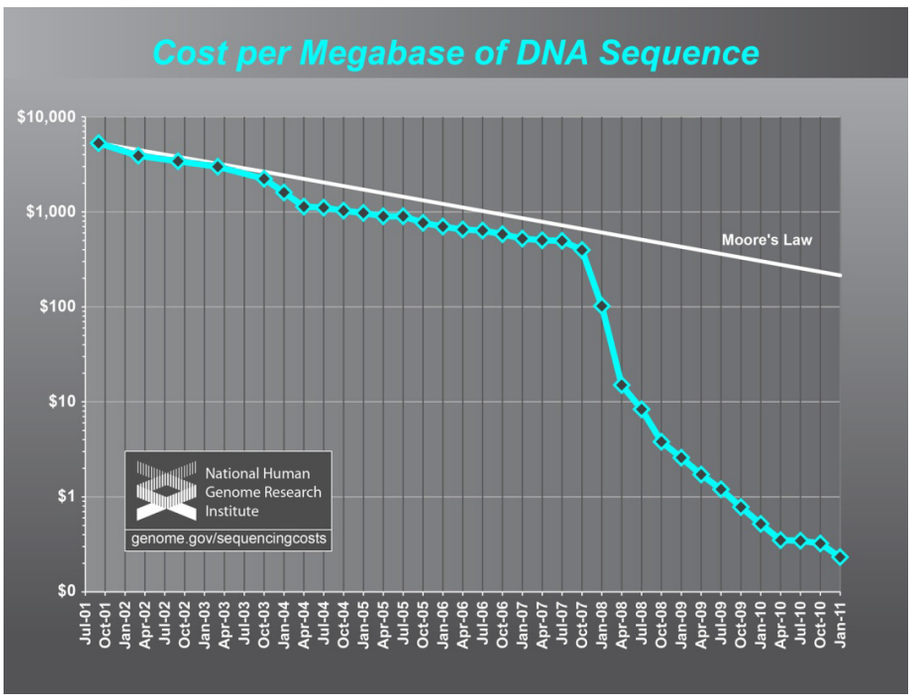
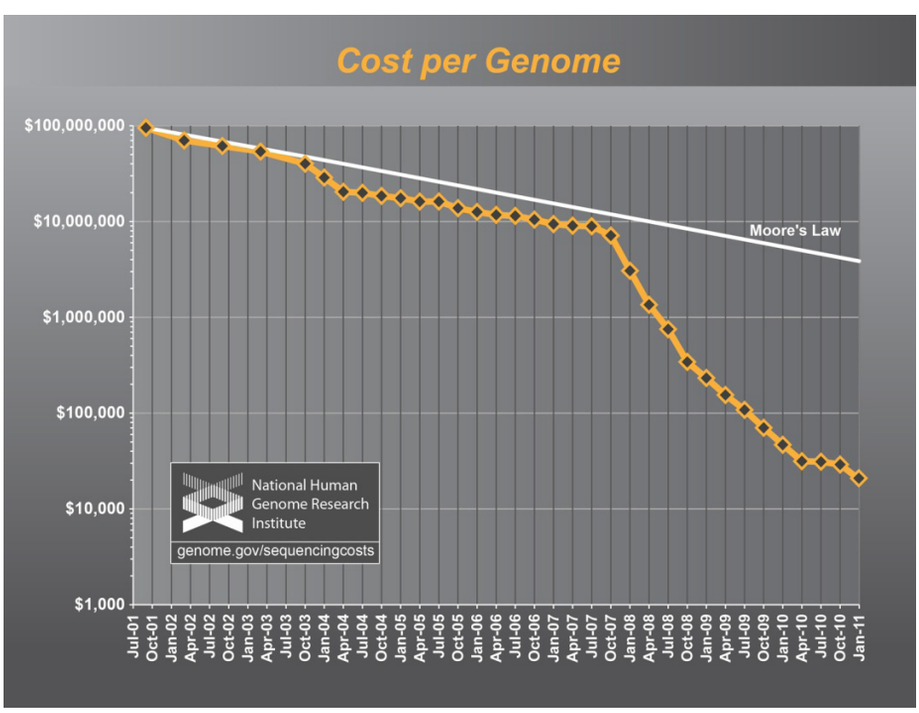


Figure 9.2 Cost (in US$) per human genome of DNA sequence as determined by the National Human Genome Research Institute.



9.3 Economic evaluation of genetic testing

Although the literature regarding the economic evaluation of WGS itself is sparse, we can learn about issues related to WGS by looking at examples related to genetic testing. Several systematic reviews of economic evaluations of genetic testing published recently have shown that studies published in this field have more often opted for CEA and have been primarily used for screening technologies rather than diagnostics[340-343]. Although the use of QALYs was found in the literature, it was not as widespread as in other areas of health economic evaluation, perhaps due to the inability of the QALY to explicitly incorporate non-health related outcomes (and therefore a limitation of the CUA method). Information provided from genetic testing is itself a key benefit/dis-benefit, both for the individual and their family[341]. Such 'process utility' is difficult to unbundle from outcomes derived from the sequelae of treatment and other decisions following testing. Also, QALYs are usually designed for the patients in question who actually receive the intervention (hedonistic utility) and not for those whose quality of life may also be affected. Therefore QALYs may not be able to capture the 'externalities' or benefits falling outside the main players in a doctor-patient contract. For example, not only does the genetic testing of an inherited mutation in cancer provide a diagnosis for the individual with subsequent treatment and prophylactic surgery as outcomes, but it also has implications for family members in potential screening, diagnosis and treatment which may or may not incur additional costs, benefits or dis-benefits.

Wordsworth *et al*.[344] have investigated the cost-effectiveness of using array-based comparative genomic hybridisation (aCGH) versus standard cytogenetic analysis (including karyotyping and multi-telomere FISH) for diagnosing idiopathic learning disability in the NHS. The average cost of a test using aCGH was £442 per single (patient) sample versus the £117 per sample using karyotyping. The increase in cost was largely accounted for by the array cost. Despite this increased average cost of an individual test using aCGH, the average overall cost of karyotyping (including a single multi-telomere FISH assay) was £4,957 compared to £3,118 for aCGH which yields around 10% more diagnoses in a hypothetical cohort of 100 idiopathic learning disability children. As can be seen, when the costs of follow-up tests and the differing diagnostic yields are incorporated into the analysis it can be concluded that using this technology is likely to be more cost-effective per diagnosis. However, the full magnitude of costs are unknown for additional tests following initial karyotyping and although the costs of such testing would likely escalate, many diagnoses achievable by aCGH would still be missed[344].

Regier *et al*.[345] have estimated the willingness-to-pay (WTP) for a diagnostic test to find the genetic cause of idiopathic developmental disability from families with an affected child. They used a discrete choice experiment to obtain WTP values and found that these families were willing to pay CDN$1118 (95% CI CDN$498-1788) for the expected benefit of twice as many diagnoses using aCGH and a reduction in waiting time of 1 week when compared to conventional cytogenetic analysis. This WTP value compares favourably to CDN$710 per aCGH test that the British Columbia Medical Services Commission paid. Again however, the full costings of additional tests would need investigating although downstream savings would be expected by avoiding unnecessary tests once a diagnosis is found. Regier *et al*.[345] suggest that the use of CBA would allow all aspects of diagnostic testing to be incorporated into an economic evaluation in order to allow a decision on whether a new technology provides good value for money.

Grosse *et al*.[346] also argue that CBA, especially using discrete choice experiments (DCEs) to estimate WTP, are necessary in order to estimate the full value of genetic testing as many important outcomes or attributes of genetic testing do not easily fit within the traditional measure of health used in CUA. WTP can allow monetary estimation of both health and non-health outcomes and as such place a value on the information received via a genetic test. DCEs are a form of analysis whereby respondents are asked to choose an intervention from a set of scenarios that describe the healthcare intervention that differ with respect to certain predefined attributes allowing these individual attributes to be valued in some measure. However, few studies have used DCEs to examine the value of information from a diagnostic test [346].

Yang *et al*.[347] have recently used a decision analytic model to show that prophylactic surgery in women with Lynch syndrome/hereditary non-polyposis colorectal cancer mutations is cost-effective when compared to surveillance (either gynaecologic surveillance or annual examinations alone) from a societal healthcare cost perspective. Although the surgical and surveillance arms of the model had a similar effectiveness, the surveillance group had a greater burden of cancer diagnosis during the patients' lifetime. This leads to increased costs as cancer treatment and care are very expensive. Although the model has limitations when compared to a prospective randomised control trial incorporating an economic evaluation, use of explicit assumptions as well as capturing the most important costs and outcomes of clinical scenarios can allow decision makers a quantitative tool for making decisions in the face of uncertainty[347].

9.3.1 Alternative methods of decision making

Veenstra *et al*.[348] have developed a framework for assessing the health related utility of genomic tests. A decision model is used to conduct an evaluation of the risks and benefits of a genomic test. This is followed by defining a range of health related utility outcomes of genomic tests as improvements that are measurable as: the proportion of patients with a reclassified risk status (*e.g.* a patient with a strong family history of breast cancer may be placed in a high risk category based on family history alone but if genetic testing finds that they are not a mutation carrier then they may move from high risk to normal risk), the proportion of patients receiving an alternative treatment strategy (*e.g.* a different drug to avoid adverse drug reactions), the proportion of patients likely to choose an alternative testing or treatment strategy, by clinical event rates (both benefits and harms), life expectancy and QALYs. Finally, a policy matrix is used to display results and facilitate the interpretation and implementation of these results. Interpretation could lead to three outcomes: a recommendation of the technology; a rejection of the technology; or a 'wait and see' approach based on further data collection on both costs and outcomes.

9.3.2 Multi-criteria decision analysis

The use of multi-criteria decision analysis (MCDA) may also facilitate practical evaluation and translation of genetic technologies as several goals can be incorporated into the decision making process, not just those related to test performance such as accuracy (sensitivity, specificity and predictive value). For example, the importance of informational benefits of genetic testing to both individuals and families, equity and acceptability to patients may represent other issues, albeit carrying different importance or weight. MCDA is a useful technique for both deciding on resource use between programmes as well as within them. It is widely applied by economists for resource allocation decisions and priority setting but can equally be used for inter-programme resource allocation where greater technical efficiency of delivering a particular programme (say genetic testing) is required.

9.4 Factors to consider in the economic evaluation of NGS

9.4.1 Substitution versus addition

Economic evaluation relies on assessment of incremental costs and incremental benefits. When conducting an economic evaluation involving NGS, it can be argued that any costs that would be the same regardless of whether NGS or an existing genetic test is used, can be ignored as long as the levels of use are similar (*e.g.* use of a genetic counsellor and number of sessions). For conditions where clinical pathways use existing genetic tests, the use of NGS is primarily as a replacement technology and as such should not require a full economic evaluation. It could also be argued that if relative effectiveness is similar for current genetic testing technology and NGS then a cost analysis evaluation alone may be sufficient. If at this stage we only consider the costs of the test the situation seems straight forward. However, a change to an investigation can also lead to a change in the treatment pathway with subsequent changes in treatment costs and even outcomes. In this case a full economic evaluation may be needed. The decision problem can be thought of as a 3x3 table where a new (genetic) test is worse, the same or better than the current test. Costs too can be higher, the same or lower than of the existing test. We can then get some indication of potential cost-effectiveness as indicated in Table 9.1.

Table 9.1 Cost-effectiveness acceptability matrix

Costs	*Outcomes*		
	Worse	**Same**	**Better**
More expensive	Reject	Reject	Consider?
Same	Reject	Consider?	Accept+
Less expensive	Reject	Accept+	Accept++

Assuming that new genetic test is being evaluated because it is "better" in some way, the first column under worse outcomes can largely be ignored. In this situation the existing technology would always dominate. The most promising scenarios for direct substitution are shown in the grey boxes but even here only one of these is clearly dominant (better outcome, better or lower costs). The two other possibilities in this group would need to be subjected to full economic evaluation depending on the ratio of incremental costs and benefits and what cost-effective threshold is acceptable. Technology assessments have typically used CEA as this is more amenable for use alongside clinical studies, where one outcome or measure of process is seen as the end-point (*e.g.* cases detected; deaths avoided) and also for reasons mentioned above. Cost-effectiveness in mutation testing and familial cancers for example has depended on the frequency of mutations in the population tested, the uptake of effective treatment, the effectiveness of treatment, the age at which testing was undertaken (and associated discount rates used) and the alternative test with which the test is being compared[342].

However, this view is over simplistic, because in most situations NGS is not a like-for-like technology replacement (for example, it enables more genes to be examined and thus increases the diagnostic rate). NGS may lead to a shift in the current clinical care pathway. In these situations, costs following on from an initial diagnosis can also be causally attached to this initial diagnostic investigation.

For example, WGS has the potential to identify all genomic alterations in a cancer sample in one process. For a cancer patient this can lead to a streamlined diagnosis and personalised management pathway which could result in cost-savings by identifying drug non-responders as well as improving health outcomes by minimising potential drug use side-effects. Reducing both costs and improving health outcomes would lead to improvements in cost-effectiveness. However, effects on the wider healthcare system are uncertain and are likely to be highly variable from individual to individual. Full costs emanate from resource use not only in testing and diagnosis but also those resources consumed in the whole pathway including future treatments.

With potential costs savings from WGS likely to be realised in the future a further question is raised how to encourage the initial outlay spent now to justify these future savings as this major investment is required at a time of financial restraint in the NHS. *Inter alia* this depends on the discount rate used. A discount rate converts future costs and benefits to a present value to take account of opportunity costs of use of resources and preferences for the timing of both costs and benefits.

9.4.2 Sensitivity analysis

Sensitivity analysis is required in a good economic evalution. It is a means of varying estimates (one at a time or in combination) in a given parameter and seeing the impact on results to determine how robust the model is to changes in parameter estimates. Given the uncertainty around these estimates sensitivity analysis should be included in any framework that evaluates the use of NGS. The following factors are likely to affect the uptake of NGS and include both quality issues as well as those that would influence efficiency (outcomes):

- Mutation penetrance (how often does a positive genetic test accurately reflect future disease and its management)
- Speed of test result (including throughput, capacity, runtime, and ability to multiplex)
- Test accuracy (including sensitivity, specificity, predictive value, quality of base calls, read depth, read length)
- What does a negative or test of unclear clinical significance mean? (what is done next/what is the clinical pathway)
- How many tests are conducted and on which platform (*e.g.* need a critical mass to allow economies of scale)
- What costs are incurred by testing (including machine and reagent costs, staff costs, treatment costs) and what is avoided (*e.g.* drugs not used/additional referrals and testing avoided)
- What is the capital investment required (*e.g.* machine costs), who bears these costs and whether they should impact on an economic evaluation of a clinical care pathway
- What are the running costs that will be needed (*e.g.* overheads and reagent costs)
- Discount rate applied (and time frame of when the costs and outcomes/benefits occur)

9.4.3 Which costs to measure

Regardless of which method of economic evaluation is undertaken, an appropriately robust costing methodology should be a common feature. Costs can be thought of as the value of resources required to produce a service or good. There are three key steps in a costing analysis. First, the identification of costs in terms of which resources might be affected by the programme or treatment. Second, the measurement of costs identified as important. (A further question arising here is how to monitor the levels of resource use.) Third, the valuation attached to each of the resources. The problem for NGS is that very little 'real' information is available on the actual costs for NGS from the NHS perspective and the NHS Department of Health Reference Costs Database and PSSRU, where standard NHS costings are listed, are generally not helpful (S Wordsworth, personal communication).

Costs can also be categorised into tangible costs and intangible costs. Tangible costs can be further broken down into direct medical costs, non-direct medical costs, and indirect costs. Direct medical costs would include costs associated with the test itself (such as capital costs of sequencing machines and cost of reagents), costs of clinic visits and any resulting hospitalisation, costs of obtaining results, costs of confirming results (possibly using a different method), cost of genetic counselling, costs of any resulting follow-up tests and also costs of any treatment used or costs saved from avoided adverse drug reactions. Capital expenditure and implementation costs should be incorporated into an investment appraisal using annualised costs in order to determine the actual cost of the test. However, these costs are sometimes wrongly treated as 'sunk costs' and not factored into an economic evaluation of the clinical pathways. Non-direct costs include any costs that are incurred as a result of the programme or treatment but are not directly related to the medical care itself. Examples include the costs incurred by the patients (for example the cost of travelling to the clinic), administration costs incurred by the programme and utility (*e.g.* electricity), costs incurred during the programme or treatment and overheads. The direct and non-direct medical costs can be grouped together.

Indirect costs can be thought of as losses in productivity or resources foregone by the patient or a carer as a result of participating in the programme or treatment. Examples include reduced productivity as a direct result of the condition itself or the side effects of treatment and time lost in participating in the programme or undertaking the treatment.

Intangible costs can be thought of as the emotional costs associated with anxiety, pain and suffering as a result of information or the side effects of the treatment itself. These costs are

often difficult to quantify and value and so are often just excluded in economic evaluations. In genetic testing the potential value of genotypic information to individuals includes better understanding of their own prognosis as well as the risk to family members regardless of whether that knowledge affects clinical management or not[349]. Reduced anxiety as a result of reassuring a relative that they do not carry the disease mutation is a potential benefit of genetic testing but is difficult to quantify due to the variation in utility gain for different individuals. Grosse *et al*.[346] suggest that in using CBA, it is possible to use the WTP approach to estimate the value of information received from genetic testing results. Other costs to also factor in relate to costs of tracing, counselling and testing living relatives with potential treatment costs as a result of testing.

9.4.4 Costs involved in sequencing

In their costing of DNA sequencing for both 'cost per genome' and 'cost per Mb of DNA' the National Human Genome Research Institute (NHGRI) distinguish between 'production costs' and 'non-production costs'[77]. Production costs are included in the calculations that produce Figures 9.1 and 9.2. In their costings, the NHGRI attempt to appropriately account for costs that are covered by significant subsidies to sequencing centres (*e.g.* a grant institution providing funds to purchase large equipment). Production costs include:

- Labour, administration, management, utilities, reagents, and consumables
- Sequencing instruments and other large equipment (amortized over three years)
- Informatics activities directly related to the production of sequence (*e.g.* laboratory information management systems and initial data processing)
- Shotgun library construction (required for preparing DNA to be sequenced)
- Submission of data to a public database
- Indirect costs as they relate to the above items

The costs associated with non-production activities that are not reflected in Figures 9.1 and 9.2 are listed below and these represent substantial additional costs and human resources to be considered:

- Quality assessment/control for sequencing projects
- Technology development to improve sequencing pipelines
- Development of bioinformatics/computational tools to improve sequencing pipelines or to improve downstream sequence analysis
- Management of individual sequencing projects
- Informatics equipment
- Data analysis downstream of initial data processing (*e.g.* sequence assembly, sequence alignments, identifying variants, and interpretation of results)

For reference, an example of current diagnostic sequencing test costs offered by Genetic Testing Network laboratories for colorectal cancer (OMIM 114500) along with reporting times are listed below in Table 9.2. This highlights the disparity across laboratories within the UK for a single test. In a research setting the cost of sequencing an exome has already dropped below $1,000.

9.4.5 Measuring outcomes

Outcomes that flow directly from the diagnostic test itself should always be included in any economic analysis as well as the outcomes related to any subsequent treatment. If a current genetic testing clinical care pathway exists and downstream outcomes are similar from the existing test and WGS then these can potentially be omitted with the evaluation only including test and treatment costs. CBA measure monetary outcomes, CUA use QALYs and CEA use natural units. Decision models can be developed for a sufficiently long time horizon in order to cover an individual's lifetime if for example natural unit outcomes such as number of life years gained are chosen. However, longer-term models are generally less reliable due to greater uncertainty surrounding estimates.

Generally, NICE require the use of QALYs as a quality of life measure using the *EuroQol EQ-5D* descriptive system (EQ-5D) as a reference case. The EQ-5D is a generic measure of health related quality of life questionnaire mapping respondents onto a health state that is defined by five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and each of the five dimensions has 3 levels of severity (level 1 = no problems, level 2 = some problems, and level 3 = extreme problems). Although the last two dimensions (pain/discomfort and anxiety/depression) may be relevant for treatment coupled with (or decided by) a genetic test (for example chemotherapy for cancer) and for the information associated with the genetic test itself, the first three dimensions are not very specific to genetic testing. Also, these dimensions measure health related quality of life for the individual and do not measure improvements in quality of life of relatives for example. Because of the lack of sensitivity for measuring the psychological impacts of genetic testing, it has been suggested that specific instruments should be developed[346]. However, these may not necessarily have the properties required to measure personal utility.

Table 9.2: Cost of genetic testing for sequencing the colorectal cancer gene *MLH1* in three UKGTN NHS diagnostic testing labs (further details available at http://www.ukgtn.nhs.uk/).

UK GTN Diagnostic member lab	Reporting time (in days)	Cost
West Midlands Regional Genetics Laboratory. Birmingham Women's Hospital NHS Trust, Birmingham.	40	£539
Sheffield Diagnostic Genetics Service. Sheffield Children's NHS Foundation Trust, Sheffield.	56	£430
Yorkshire RGC. Yorkshire Regional DNA Laboratory, St James's University Hospital, Leeds.	80	£1050

The gene tested was mutL homolog 1, colon cancer, nonpolyposis type 2 (*E. coli*); *MLH1* (OMIM 120436) and the test itself involved sequencing of the entire coding region of a gene – determining the nature and order of the nucleotide bases within the coding region of the gene being tested including the exon and intron boundaries.

In order to overcome these issues, Veenstra *et al.*[348] suggest economic analyses should report a multitude of health outcomes including QALYs, clinical events (both harms and benefits), life expectancy *etc*, known as comparative effectiveness review (CER). Clinical events could also be included such as the number of diagnoses, number of life years saved and adverse drug reactions avoided by use of personalised treatment. CER of WGS is an area that needs further work from a health economic viewpoint.

9.5 Broader economic issues and impact on related industrial sectors

With so many expensive resources required in R&D, NGS cannot be divorced from its macroeconomic implications and impacts. The scientific endeavour although now spread across many countries was first established in the 1950's in the UK with much of the industry infrastructure and subsequent spin-offs still embedded in both private and public sector initiatives. The for-profit sector has expanded and mainly serves niche diagnostics markets, whilst the public sector is predominately university based and deals mainly in primary research, albeit with a practical focus. Unlike the mainly overseas based and much larger pharmaceutical industry, the potential for the industry to migrate completely abroad is perhaps less due to these mutual private-public relationships that enhance the research potential for both parties. However, as diagnostics applications expand and separate off, global markets could become more attractive and it is possible that domestic NGS diagnostics industry could migrate internationally to seek wider, more profitable markets and take advantage of the most favourable economic condition for its production.

Lodging an early patent has ensured that some genes can be protected for use by their discoverers. Some companies have thrived on their property rights and patents over certain genes and obtained virtual monopolies (*e.g.* the discoverers of *BRCA1* and *BRCA2)*. Technological efficiency and productivity has now made it possible to produce affordable and reliable tests not only for the research market but also for healthcare and even (increasingly) private individuals with the disposal income to buy them. A main attraction of outsourcing to these companies for the health industry is the avoidance of expensive investment and manpower locally, although other quality control and regulatory issues may be raised.

Funding of these technologies, normally associated with government research bodies in advanced countries, has often taken place outside of government or university laboratories. The prospects of commercial gain have attracted levels of private sector investment hard to replicate from the public purse. This has had a fundamental influence on the structure of the genetic testing industry which is more competitive and restricted than other areas of research and technology and has helped to drive down costs. It is predicted that the price of sequencing a complete human genome will drop to US$1,000 in 2014 and significantly below this figure in the following years using existing DNA nanoarray technology[25;350]. However, the investment required not only in capital costs but also in resources and human skills is still considerable both on the part of technological input, end-users such as medical professionals, development of necessary bioinformatics and clinical decision making, and a wide range of commissioners and providers to embed technologies in clinical pathways.

The sustained take-off and use of WGS depends not only on the comparative clinical advantage to conventional genetic tests, but also on the economic viability, especially the long-term unit cost (*i.e.* being less costly). The future pattern of use of WGS will depend on the validity, reliability and reproducibility of tests being developed and the underlying economic rationale that can be made for its wider adoption. The principle economic condition for wider adoption of these technologies is that both cost and health outcome advantages exist. However, integral to this is developing a viable and sufficiently large demand for this service and hence the development of economies of scale. Therefore the technological dispersion and user-demand are likely to operate in a dynamic equilibrium. Factors that influence the diffusion of technology include: relative advantages over existing technology; compatibility with existing frameworks of technology use; and the ease/difficulty of understanding and applying the new technology. People with "knowledge" of the technology are key drivers of the diffusion process with researchers using a scientific justification for the implementation of WGS, alongside clinicians, patient groups and possibly commissioners arguing for WGS on the basis of medical need.

Mukoyama[351] has suggested a diffusion model more adapted to high-technology products that suggests a step-wise adoption with a lagged variation in speed depending predominantly on skill distribution of users (with early adopters being more highly skilled than late adopters). Technology diffusion in this model is a non-trivial process and in general takes place over a substantial period of time. The model incorporates feedback between users and producers and postulates this as the main driver of quality improvement and innovation. Thereafter such improvement leads to step-wise changes in the numbers adopting the technology and prices fall. The model initially assumes a skilled workforce, similar to geneticists, bioinformaticians and clinicians who whilst in the minority would be the first wave of early adopters, helping to drive the development and uptake of NGS technology. In this model costs actually increase in the early stages as end-users demand better machines and producers engage in more R&D. Ultimately however, costs of production do start to fall as late adopters start to use the technology.

Although plausible, this model is based on supply-led assumptions and the presence of demand should also be taken in account. A steadily falling price is an important stimulus to increasing wider demand as the law of demand predicts *('demand is inversely related to price'*). However, the elasticity of demand (a measure of how responsive demand of a good or service is to a change in its price) determines the exact response of quantity demanded to price and depends on perceptions of the value obtained (utility) against changes in price. In practice, consideration of non-price attributes of any good is also an important aspect of the willingness-to-pay and so is captured by price. Exogenous tastes, preferences and incomes change over time and may also influence the speed of change-over to new technologies.

Economic theory predicts that when relative prices fall between two similar goods or close substitutes (*i.e.* health outcomes) the cheaper good, *ceteris paribus* (all else being equal), will be substituted for the more expensive one in an effort to maintain the consumer's total utility when they face a fixed budget. But this depends on how close a 'substitute' it is in practice. The vagaries of the substitution effect are certainly seen in many areas of medicine when for example, an older well tested drug comes off patent, becomes cheaper and preferentially prescribed by many physicians. However, the new drug's manufacturer will want to emphasise the lack of substitutability through emphasising its non-price benefits, in order to gain market share at this artificially high price. Non-price competition relies on some intrinsic product characteristics of the good such as convenience, speed, and patient acceptability. The net substitution effect will therefore be influenced by both price relativities and the non-price characteristics. Adoption of innovation usually depends on social as well as economic determinants.

Although unquestionably essential, it is not viable for a new technology just to be 'supply-led' by inventors. There must also be so-called 'demand pull'. There is a growing expectation on the part of the patient that technological improvements must lead to improvements in treatment received and therefore improvement in health outcomes. Policy push/pull may also come from:

- Scientists (scientific justification)
- Clinicians (clinical need)
- Patient groups (patient pressure)
- Government (pride and status, global dominance of markets, need for self-sufficiency, perceived requirement to develop good sciences, economic need to support a thriving biotech industry, need to create/support employment)
- Public health advocates because of new opportunities for population health
- Biotechnology industry (profit, supports strategic direction)
- Society (*e.g.* demand for DTC testing – but societal disquiet might also act as a brake)

9.6 Conclusion

There is a very thin evidence base for the economic evaluation of NGS technology along with a lack of studies linking diagnostic test results to final patient outcomes (clinical utility). Economic modelling can be used to evaluate the extra information gained against the savings or additional costs incurred in cases of both substitution or complementation of standard tests by new generation genetic tests or whole genome sequencing. For focused clinical testing where there is no change to the current existing testing pathway, NGS is a replacement technology (substitution, even if initially both tests are run in parallel and thereby exerting extra costs). If this is the case and NGS is 'at least as good' as existing genotyping platforms then it only needs to become cheaper than the existing test cost for adoption to take place.

If no genetic test currently exists (perhaps due to excessive cost) and the introduction of NGS introduces a new test then a new clinical care pathway will be needed. To implement these tests without having the necessary infrastructure and means of dealing with the consequences would be unwise and as yet remains economically unjustified. In theory, the greater the number of single tests currently used the more efficient it could become to substitute them all with whole genome sequencing (perhaps at birth or at sentinel stages of life). However, the economic rationale and conditions for this to occur are as yet unclear. For example, across the NHS, costs of the single tests are not standardised with a lack of a nationally agreed price list [352](see also Table 9.2). Criteria for testing may need to be altered if more mutations are tested in 'one go' by WGS instead of the current sequential method of testing.

The economics of WGS is still unfolding and both the micro- and the macroeconomics implications need to be assessed as more widespread use occurs. At some stage sufficient data would allow alternative clinical pathways to be compared with and without WGS in terms of the costs and benefits. All new studies should consider these economic dimensions in order to accumulate a body of evidence as rapidly as possible and allow more sophisticated economic assessments to be made.

Meanwhile there is a strong need for further health economic evaluations looking just at aspects of the potential cost-savings that can be made by genetic testing via WGS (*e.g.* inherited cancers) as well as others looking at health related utility measures and outcomes derived from this technology.

10 Laboratory service delivery models

10.1 NHS pathology services

10.1.1 Genetics laboratories

NHS genetics services include a range of clinical, laboratory and screening services, which are delivered from a network of 23 Regional Genetics Services (RGS) across the UK. They provide an effective, coordinated service to patients and families with inherited diseases. Thus, most molecular genetic and cytogenetic testing still takes place within the context of integrated RGSs.

Most RGSs:

- Serve a population of 2-5 million
- Have a hub and spoke arrangement with peripheral clinics to facilitate access to the service
- Are multidisciplinary, employing a range of scientific, medical, counselling and data handling personnel
- Are typically associated with university departments
- Provide the full range of clinical genetics services (diagnosis, genetic risk assessment and counselling, follow up of family members)
- Incorporate both molecular genetic and cytogenetic laboratory services, which also receive referrals from clinicians outside the RGS

There are many advantages to this type of arrangement. By enabling a close working relationship between laboratory scientists and clinicians, the management of patients and families with inherited disease is enhanced. The results of tests may be highly predictive of disease and have implications for family members making it desirable to provide testing within a counselling context.

In addition, the rapid developments in both knowledge and technology that have always been common to genetics mean that proximity to academia facilitates technology transfer into the NHS as well as enhancing opportunities for teaching and training across other specialties. As knowledge of the genetic contribution to disease aetiology has grown, the laboratories of RGSs increasingly provide a source of genetic tests and specialist interpretation for clinicians in other specialties. Though most testing for inherited diseases takes place in RGSs, some is done in a variety of other labs *e.g.* specialist biochemistry labs, newborn screening labs or haematology labs. With the exception of the latter, however, these tests are usually not DNA-based. It is unclear whether this model will continue to be the best arrangement once WGS becomes widespread; due to the enormous informatics and IT support required to process and interpret WGS derived information.

In the wake of the Human Genome Project, there was considerable speculation about the impact that advances in genetic technology would have on healthcare. Therefore, the Department of Health (DH) established a working group[353] to review laboratory genetic services in the NHS and to identify future developments and their potential implications, which published its report in August 2000. It recommended that there should not be any change at that time to the configuration of genetic laboratories, but in recognising that the nature of the service may change as technology developed and a wider range of tests became available, the situation should be kept under review. The members of the working group foresaw that a time would come when developments in pharmacogenetics and other areas would shift the emphasis of testing away from the management of families and into mainstream medicine. At this point radical re-configuration of genetic laboratory services would need to be considered.

The work of the group revealed considerable variation across the country in many aspects of the service and highlighted the need to coordinate the evaluation, prioritisation, commissioning and funding of genetic tests. The group made a number of recommendations to address these issues, including the establishment of:

1. An expert national body to give a strategic steer to commissioning and coordinate at national level important service elements such as the assessment and evaluation of new genetic tests
2. A UK-wide genetic testing network to address rare single gene disorders and data harmonisation for planning purposes
3. One or more centres of excellence to evaluate new genetic technologies

The recommendations in the report subsequently informed initiatives set out in the Government White Paper, *Our Inheritance, Our Future* – realising the potential of genetics in the NHS (2003)[354] and led to the establishment of the Genetic Commissioning Advisory Group (GenCAG), the UK Genetic Testing Network (UKGTN) and two National Genetics Reference Laboratories (NGRL) in Manchester and Wessex.

10.1.2 GenCAG

GenCAG was set up by the DH to take a strategic national overview of genetics in healthcare delivery. It aimed to give advice to commissioners of genetic services, to enable them to provide appropriate services for NHS patients and their families. The membership of the group was drawn from relevant professional bodies and groups: the medical Royal Colleges, the patient umbrella group Genetic Alliance UK, and members of the specialised commissioning groups. Following the recent review of public bodies, GenCAG has now been disbanded and its functions will be taken over by the new NHS Commissioning Board.

10.1.3 UKGTN

The UKGTN Steering Group was set up in 2002 as an advisory group and sub-group of GenCAG. In 2003, the molecular genetic labs in the UK (and selected specialist labs) were invited to apply for membership of a genetic testing network if they fulfilled a defined set of criteria, including

Clinical Pathology Accreditation (CPA); in later years, cytogenetic labs were also able to join. Thus it has evolved to become a collaborative group of 53 genetic testing laboratories, clinicians and commissioners, and patient support groups. It advises the NHS across the UK on genetic testing and aims to ensure the provision of high quality, equitable testing services.

To further these aims, the UKGTN has made significant progress since its establishment. Amongst its most notable achievements are that it has:

- Ensured NHS services are delivered by labs which meet minimum quality standards and have been specifically designated to provide tests
- Developed and implemented the Gene Dossier process to evaluate new genetic tests before they are recommended for NHS funding
- Developed and implemented testing criteria to promote appropriate referrals for genetic testing from clinicians in other specialties
- Developed the NHS Directory of Molecular Genetic Testing, which lists all genetic tests approved for NHS funding to aid commissioners
- Developed an online database which shows the labs providing each of the tests in the Directory and the testing methods used
- Provided data to inform the development of tariffs
- Provided a horizon scanning function and commissioned evaluation of emerging new technologies such as array CGH

The Gene Dossier process

The gene dossier process was developed by the UKGTN as a mechanism for the evaluation of genetic tests to ensure they are appropriate for inclusion on the NHS Directory of Genetic Testing. Only tests which have been agreed to be appropriate for clinical use are placed on the Directory. The 'dossier' itself provides a standard format for the presentation of the information required for the evaluation of the test and the process ensures that the decision making process is explicit, transparent and evidence-based.

The gene dossier evaluation framework is based on the ACCE framework of **A**nalytical Validity; **C**linical Validity; **C**linical Utility and **E**thical, legal and social implications. Applying laboratories are required to submit comprehensive information to inform the evaluation, including information on the disease to be tested for; technical aspects of the proposed test and its clinical characteristics in terms of sensitivity and specificity *etc*.; proposed use of the test and its clinical utility; and test costs.

The last few years have seen a marked increase in the proportion of genetic tests that are not ordered by clinicians within the specialist genetic services, but by clinicians in a range of other specialities. In order to help those wishing to order a test to ensure that the test requested is appropriate for the intended purpose, submitting laboratories are now required to develop 'testing criteria' as part of the gene dossier process. However the gene dossier process was originally designed for closed tests. The UKGTN has only just moved in the direction of open tests by evaluating array CGH technology for developmental delay.

The work done by UKGTN to develop a robust framework in which a high quality genetic testing service can be provided within the NHS is widely admired and has been acknowledged by both the House of Lords[355] and the recent White Paper – E*quity and excellence: Liberating the NHS*[356].

10.1.4 National Genetics Reference Laboratories

The National Genetics Reference Laboratories (NGRLs) in Wessex and Manchester were established in 2002 by the Department of Health to support the UK genetic laboratory services. Though based in England, the laboratories work closely with genetic service providers in all of the UK home countries.

The laboratories work in close conjunction with the RGLs which provide NHS diagnostic services to the local population in the areas in which they are based. The NGRLs are thus well placed to evaluate technologies and systems that are close to service or in service and assess their applicability to the NHS.

The NGRLs aim to work with the genetics community to:

- Assess, validate, compare and develop new technologies and devices
- Develop quality assurance
- Provide an evidence base for strategic decisions
- Help increase the capacity, efficiency and flexibility of genetic testing in the NHS
- Act as a source of expert advice and knowledge

In furthering a broad, common remit the two laboratories have complementary work programmes which are overseen by a steering group.

The NGRLs are currently funded until March 2012. How their work streams will be taken forward after this date, when central funding ceases, is not yet clear.

NGRL Wessex

The work programme of the NGRL in Wessex during the 2007-2011 funding period focused on four specific areas: array CGH, next generation sequencing, non-invasive prenatal diagnosis (NIPD) and quality assurance. Their principal focus in 2011/2012 will be next generation sequencing for mutation scanning, copy number analysis and non-invasive prenatal diagnosis.

With respect to NGS, their aim is to develop and validate procedures for targeting specific genes; resolving data from multiple patients; and to work with partners to address the bioinformatics issues relating to data analysis. Their main focus in this area for the coming year will be completion of the CMGS special interest group (SIG) NGS project, which they initiated and have been coordinating. The purpose of this study is to examine the diagnostic utility of a range of NGS pipelines using seven colorectal cancer genes as a model system. A defined set of samples has been sent to a number of service providers and technology users to perform blinded analysis using a variety of targeting and sequencing methodologies. Currently eight different protocols are underway using various combinations of targeting methodology and sequencing platform. Depending on the availability of suitable sequencing partners a further three protocols may be included. It is hoped that a full data set for analysis will have been collected by late summer 2011. The aim is to use these data to evaluate the performance characteristics of each protocol, define processes and to evaluate various analytical approaches including both commercial packages and open source software.

Other areas the laboratory currently has under investigation are the use of NGS for the analysis of copy number variation, the effective utilisation of ‘low capacity’ platforms (Roche Junior, Ion Torrent PGM and Illumina MiSeq) that are now available, together with the design of appropriate gene panels for use with the different instruments and applications.

They are currently actively exploring funding sources to enable them to continue these work themes beyond March 2012, so that the outcome of their projects can help to inform strategic decision making in relation to the implementation of NGS in the NHS.

NGRL Manchester

NGRL Manchester has developed to become a dedicated centre of expertise with health informatics and bioinformatics at its core, supporting genetic medicine in the UK. Its current work programme includes: the support and development of the Diagnostic Mutation Database, a dedicated repository for sharing genetic variants and their pathogenic status among diagnostic laboratories; support and development of SNPCheck, a software tool for quality assurance in genetic analyses using PCR; bioinformatics training for clinical scientists in collaboration with the Nowgen centre; and developing best practice and support for clinical bioinformatics, including revision of best practice guidelines for unclassified variant analysis, review and support for bioinformatic tools, and support for NGS implementation into diagnostic practice.

Alongside this work programme NGRL Manchester is a partner in EU and UK funded projects including GEN2PHEN, an EU Framework 7 project which aims to unify human and model organism genetic variation databases, an EU Committee of Experts on Rare Diseases (EUCERD) project to develop a comprehensive and coded rare disease classification, and a UK Genetic Testing Network funded project on data standards for genetic data in the NHS and development of data exchange opportunities. Its future priorities are development of sustainability for the services it provides and continued representation of genetic medicine within research and development.

10.1.5 Current NHS genetic testing services

Although some centres do use commercial testing facilities to provide specific tests or to help deal with workloads at times of pressure (*e.g.* a backlog of *BRCA* samples), the vast majority of testing in the NHS is still carried out by the genetics laboratories within the RGSs. These laboratories provide a repertoire of tests including tests recommended by UKGTN for the NHS service, and provide a service which conforms to quality standards defined by the professional bodies and quality schemes *e.g.* CPA . They receive requests from both within the RGSs and clinicians in other specialties. Tests on the NHS directory are offered on the basis of NHS funding in accordance with local commissioning arrangements. Each of these tests has been evaluated by the UKGTN through their Gene Dossier process.

Tests for some rare disorders may not be on the NHS Directory and are usually considered by laboratories on a case by case basis. Any laboratory wishing to provide a newly developed test for national provision and to have it listed on the UKGTN Directory of Genetic Testing must submit a Gene Dossier for assessment. If approved the test is recommended to NHS commissioners and added to the NHS Directory and to the UKGTN online database.

This rigorous process reassures commissioners that the funding advice they receive is robust. To aid coordination and efficiency, the SCOBEC consortium (Salisbury, Cambridge, Oxford, Bristol, Exeter and Cardiff) was set up by the laboratories to rationalise testing between labs and leverage the investment. Within this network a degree of rationalisation of work has occurred in order to avoid unnecessary duplication in the provision of less commonly requested tests.

Although all the genetic tests within the Directory are recommended for funding, final decisions on which tests are actually funded are taken by the Commissioners of individual services. Thus, though UKGTN has worked hard to ensure equity of access to tests, variations in provision do occur as a result of local funding decisions. This is the process that exists at present. Significant changes in the NHS were announced in the White Paper in 2010[356], with a move to GP commissioning and the establishment of the NHS Commissioning Board. The full implications for genetic testing services are still not fully understood. The Department of Health response to the House of Lords inquiry into genomic medicine (the inquiry took place in 2008 and the Government response was published in 2009) included the establishment of the Human Genomic Strategy Group (HGSG). The HGSG first met in 2010 and GenCAG has now ceased. The future format and functions of the UKGTN have still to be clarified.

The use of DNA based techniques has progressively increased in cytogenetics over the past few years, with the advent of techniques such as array CGH. As a result of this, the demarcation between molecular genetics and cytogenetics has become increasingly blurred. In recognition of this, there has been a move in most centres towards closer working between the two services. The pace of this change has varied across the UK. In Scotland the cytogenetics labs have now been incorporated into the genetics laboratory consortium alongside the molecular laboratories as a first step towards unification. In some centres such as Birmingham and Cambridge, they have already achieved close integration of the two and a more unified genetic laboratory service. The changes to the training of laboratory scientists under the Modernising Scientific Careers initiative will help to facilitate this development. In addition to this change, the last few years have seen an increasing proportion of the workload of both components of the service being accounted for by referrals from other specialties, such as haematology, oncology, neurology and cardiology. All these developments have argued in favour of the new proposed concept of *Genomics Laboratories* covering a wide range of services to all medical specialties[367].

10.2 Use of NGS

It is highly likely that NGS technology may eventually replace most, if not all of the current methods used in both molecular and cytogenetic laboratories. However, the process of implementing NGS technology within the NHS is likely to be a gradual one. It is not currently used as a clinically validated assay by NHS laboratories, with the exception of the Leeds laboratory, see below. In the private sector some laboratories offer diagnostic mutation testing using NGS while others are actively developing this capability. Though most NHS laboratories have access to the technology through research collaborations, standard Sanger sequencing is still used for the vast majority of testing carried out in NHS laboratories. A small number of NHS laboratories have however now acquired NGS equipment and are beginning to develop it for service use, including:

- The Leeds laboratory has an Illumina platform which they now use for diagnostic BRCA testing following recent evaluation and approval of the Gene Dossier by UKGTN
- The Great Ormond Street Hospital laboratory has a Roche Junior platform which is being developed for service use
- The Edinburgh laboratory has both a Roche Junior and an Ion Torrent platform and is developing and evaluating service use on behalf of the Scottish consortium service
- The Salisbury Genetics Reference Laboratory is evaluating the efficacy and reliability of using NGS in the service setting

So far the focus within diagnostic circles has been in the potential use of NGS to test for one or more genes of interest using targeted techniques, and using NGS to sequence panels of genes in the investigation of conditions such as inherited cardiac diseases is already faster and cheaper than using sequential Sanger sequencing reactions. Most currently involved in service provision do not appear to see whole genome sequencing as realistic or appropriate in diagnostic testing for most applications within the foreseeable future. However, as previously discussed, the time may come when cost considerations drive a change of approach, and informatics targeting is used on whole genome sequences.

10.2.1 Onco-pathology/molecular diagnostic laboratories

Knowledge about the molecular basis of the aetiology of many acquired diseases such as cancer has expanded considerably in recent years and continues to grow. With this has come the development of a rapidly increasing number of molecular tests used for a variety of diagnostic, prognostic and monitoring purposes. This trend is set to continue at an even greater pace as the knowledge base grows and new pharmaceutical interventions appear and the development of new technologies such as NGS make molecular tests easier, faster and less costly to perform. At the present time the use of molecular tests for acquired disease is probably greatest in haematology and oncology,

where the availability of targeted therapies is rapidly increasing. However, their use is expanding in the other pathology disciplines too, and it is reasonable to predict that the next few years will see the widespread use of molecular techniques across all the pathology services. Thus, the Royal College of Pathologists (RCPath)[207] and others have highlighted the need to plan now for the coordination of developments as molecular diagnostic services grow. This will be vital to prevent *ad hoc* developments which are inefficient, wasteful and compromise quality.

10.2.2 Current use of molecular diagnostic tests (outside of clinical genetics)

The types of tests in use and the purposes, for which they are used, are changing all the time in this evolving field. Some of the more established DNA-based tests and technologies include:

- Karyotypes and micro-array-based analysis of structural variation in malignancies
- FISH for *HER2* to guide therapy for breast and gastric cancers, or for prognostic classification of neuroblastomas
- Real-time reverse PCR for monitoring minimal residual disease in leukaemias
- Molecular testing for *EGFR, KRAS* and *BRAF* to guide therapy in melanoma, lung or colorectal cancers
- Molecular tests for infectious micro-organisms (outside the scope of this report)

In relation to molecular diagnostic tests, there is neither a list of appropriately evaluated tests, equivalent to the UKGTN Directory of tests, nor a single, formal structure tasked with evaluating and approving such tests. This lack of consensus information for commissioners on the value and appropriateness of tests, together with the lack of agreed funding mechanisms, have been highlighted by the RCPath as major contributors to the wide variation in the provision of molecular diagnostic tests seen at present. However, the UKGTN does now consider gene dossiers for sequence based molecular tests submitted by laboratories in the network.

In addition, NICE has now established a formal process to consider and approve companion diagnostics. Active liaison between the two organisations has been established and links are being put in place which will aid coordination of activity in this area. Once these new arrangements are bedded in, this should begin to address the concerns about appropriate evaluation in this evolving field.

10.2.3 Laboratories providing molecular diagnostic tests

There is currently considerable variation across the country both in the tests which are offered and in where they are performed. Haematology has long led the field in relation to both the integration of molecular testing into diagnostic pathways and in acquiring the skills and facilities to provide some cytogenetic and molecular tests within their own laboratories. The use of molecular tests has been more recent in relation to solid tumour management and though some tests are provided from within histopathology and oncology departments, most are currently performed in molecular genetic laboratories. As the number of available tests has increased, cytogenetic, molecular and sequence based tests for acquired disease performed for other specialties have come to account for a significant (and increasing) proportion of the work done in the laboratories of RGSs.

In theory any clinician can request a DNA-based test, regardless of the laboratory in which it is done. However, currently haematologists and medical oncologists are the groups most likely to order such tests. This may be because there is currently a greater range of tests that are relevant for the management of patients for these specialties, or it may be a reflection of a lack of awareness of available tests by clinicians in other fields. As the range of useful tests grows, it will be vital that medical education is adapted to ensure that those working in all specialties have the knowledge to recognise the value and availability of molecular tests, to be able to request appropriate tests within the patient pathway and to interpret the results in dialogue with laboratory scientists.

As others have already pointed out (in the genetics White Paper, *The Review of Genetics in relation to Healthcare in Scotland*)[354;357] to underpin this, it is necessary to ensure that the education and training of healthcare professionals equips them with enough knowledge about genetics and the contribution genetic factors play in disease causation.

As discussed above, there is currently very limited use of NGS in the provision of routinely requested molecular tests. Currently available methodology is used for most of the testing done for either inherited or acquired disease. Where NGS technology is used, it is used primarily to increase testing efficiency and lower the costs of testing for one or more candidate genes of interest. The situation will no doubt change if and when tumour profiling is conclusively demonstrated to have benefits for patient management.

10.3 Private sequencing providers

In addition to the NHS, there is an increasingly significant private sector presence developing. A wide range of molecular diagnostic tests are provided by either independent, standalone laboratory organisations such as The Doctors Laboratory (TDL) or privatised laboratory services provided (in partnership with other organisations) from within NHS organisations, such as that of Guys, St Thomas' and Kings College hospitals (GSTS) with SERCO. Thus, there is currently a complex web of laboratory provision for molecular diagnostic testing in the UK. The need for significant rationalisation and coordination of pathology services within the NHS in the interests of efficiency, cost effectiveness and quality standards has already been highlighted in the first and second Carter reports[358;359]. Moves are already under way to implement their key recommendations. The more recent report from the RCPath[207], explores the case for urgently addressing the coordination of developments in relation to molecular diagnostic testing, and their recommendations are in line with those made in the Carter reports.

The advent of expensive new technology such as NGS, make rationalisation and coordination even more important so that the technology can be exploited to its full potential in the most effective fashion. The second Carter report[359] drew attention to the need to address short comings in the QA systems and IT connectivity of pathology services. The type of rationalisation and coordination of services now called for will make this more important and challenging. It will also have significant implications for the training of laboratory scientists. The current policy environment and encouragement towards further use of the private sector will add a further layer of complexity to addressing these issues.

10.4 Research sequencing centres

There are two major players in the provision of large scale sequencing in the UK – the Medical Research Council (MRC) and the Wellcome Trust – though neither currently offers a clinically accredited service and both are currently focused on large-scale research projects.

10.4.1 MRC sequencing hubs

In response to demand from the scientific community, the MRC launched a multimillion pound initiative to strengthen UK-wide capability and expertise in high-throughput sequencing and in excess of £9m was committed to fund the establishment of 4 regional sequencing hubs (located in the East of England, Oxford, the North of England and Scotland). The funding was to be used to provide high-throughput sequencing facilities for a range of biomedical researchers in the local community, building on existing sequencing resources and expertise in collaborative regional consortia in the four centres chosen. The investment also provided funding for technical support and bioinformatics expertise to underpin the research activities of the hubs. All hubs were funded for an initial period of three years with the possibility of this being extended for a further two years. The expectation was that the hubs would become self-sustaining facilities operating on a cost recovery model.

- The **East of England** hub is based at Addenbrookes' Hospital in Cambridge and represents a collaboration between the University of Cambridge, the European Bioinformatics Institute (EBI), and National Institutes of Health (NIHR) Cambridge Biomedical Research Centre. Its research development and strategic aim is to apply NGS technology to routine medical diagnostic uses, particularly in HLA-typing for transplantation and cord blood stem cells, prenatal diagnosis and resequencing of multiple genes associated with similar phenotypes *e.g.* inherited cardiac conditions. The hub works in collaboration with the National Blood Service and the local NHS Regional Clinical Genetics Service.
- The **Oxford** hub is based at the Wellcome Trust Centre for Human Genetics, building on existing resources and aiming to widen access to high-throughput sequencing facilities for the research community. The University of Oxford, NIHR Oxford Biomedical Research Centre, Oxford Radcliffe Hospital and five Oxford based MRC units are supported in a wide programme of research, which includes human and mouse genetics, pathogen genetics, translational research and technological development.
- The **North of England** hub is based in Liverpool and expands the capacity of the Advanced Genomics Facility. It enhances access to facilities and expertise to support research groups across the north of England in the Universities of Liverpool, Sheffield, Manchester and Lancaster.
- The hub in **Scotland** is based at GenePool, the University of Edinburgh's NGS and genomics facility. MRC funding expanded the capacity of GenePool and enhanced the service it could provide to research groups across Scotland in the Universities of Edinburgh, Glasgow, Dundee and Aberdeen. The research portfolio of the Scottish groups addresses a wide range of issues including the genetics of psychiatric disorders, of cancers and of cardiovascular disease as well as the genetics of early development and aspects of the pathology of infectious diseases.

The primary aim of this MRC initiative was to strengthen the research base. Without doubt however, the substantial body of work it underpins will lead to discoveries which can be directly translated into improved patient care and pave the way for the eventual integration of NGS technologies into laboratory service provision in the UK healthcare setting.

10.4.2 The Wellcome Trust Sanger Institute

Unlike the MRC Hubs, which were primarily intended to support small and medium sized projects, the Sanger Institute was established specifically as a facility for large scale research. The then Sanger Centre, a joint Wellcome Trust/ MRC venture was set up in 1993 as a new research centre designed to play a role in mapping, sequencing and decoding the human genome and the genomes of other organisms. The Institute has evolved from primarily a sequencing centre to become a leading biomedical research facility. The Wellcome Trust Sanger Institute is now a non-profit organisation primarily funded by the Wellcome Trust and run by the charity Genome Research Ltd.

Though still leaders in the development and exploitation of sequencing technologies, including NGS, the current research programme is more broadly based and focused on understanding the role of genetics in health and disease, including as a partner in the 1000 Genomes Project, the aim of which is to create a public reference database of human genetic variation. Although it has numerous collaborations with clinicians both within and outside of the UK, and there have been suggestions that the Sanger's considerable facilities could be used to strengthen capacity for NHS testing, the Institute remains focused on research and lacks diagnostic laboratory accreditation. Nevertheless, the Sanger Institute remains a powerhouse of innovation and technological development and the work done there will help to inform the development of NGS services in the NHS and the complex bioinformatics systems that will be required to support them.

10.5 International providers

In addition to these UK sequencing centres, there are a growing number of international providers with the potential to provide services to UK clients, including:

- **GATC**, a privately owned company based in Germany. It provides a wide range of sequencing services using both capillary sequencing and NGS for individual DNA samples through to complete genome projects. It also offers a comprehensive bioinformatics service for sequence evaluation. Primarily aimed at present at the industry and academic research sectors, the company clearly has the potential to widen its sphere of interest into the healthcare sector. It currently has a broad customer base in over 40 countries and subsidiary companies in the UK, France and Sweden.
- **Complete Genomics**, a US company established in 2005 with the specific aim of providing a comprehensive human DNA service for pharmaceutical and academic research. Using high-throughput NGS employing DNA nanoball arrays and combinational probe anchor ligation reads, they provide complete human genome sequencing together with full analysis of the data for use in complete human genome studies. Though specifically aimed at the research market at present, they also have the potential to move into the provision of complete genome sequencing for individuals, or the private and public healthcare sectors should the demand develop.
- **BGI** (formally Beijing Genomics Institute), established in Beijing in 1999 and now the first citizen-managed, non-profit research institution in China. It has built up an impressive capability in large-scale new generation genome sequencing (in human, animal and plant DNA), supported by powerful bioinformatics analyses. It has produced the first Asian Diploid Genome sequence, worked as a collaborating partner in international projects such as the Human Genome and HapMap projects and is currently involved in the 1000 Genomes Project. Though currently focused on research, the organisation has both the capability and potential to move into the healthcare market and has established offices in both the US and Europe.
- **DNA Vision**, a research and development company with specific expertise in genetic and genomic analyses, founded in 2004 and based in Belgium. It offers services in the fields of pharmacogenetics, pharmacogenomics, diagnostics, biobanking, plant and animal genetics and sequencing, utilising a variety of technologies. In relation to NGS, it offers complete genome sequencing, targeted resequencing, *de novo* sequencing and transcriptome sequencing. Their laboratories have both ISO/IEC17025 accreditation and CAP certification and their QA systems have been designed to ensure compliance with the QA standards of relevant regulatory bodies.

10.6 Commercial providers to the NHS

In comparison with many other types of clinical testing, DNA-based molecular tests can be complex, expensive and require equipment which becomes obsolete quickly as the technology advances. This makes them prime candidates for commercial providers to consider. With easier access to investment capital and lower overhead costs than NHS services, such providers are well placed to offer molecular tests at significantly lower costs than their NHS counterparts. Although there are a number of established UK-based (and international) organisations with the capability to provide sequencing services to the NHS, the RCPath and others have observed that so far the sector has not expanded as quickly as might be supposed and NHS take up of available services to date is low. In the UK, the best established private providers of clinical molecular testing include:

- **TDL (The Doctors Laboratory)**, a large, independent pathology organisation based in London, providing a diverse range of molecular tests for a range of clients from both the private sector and the NHS. The genetics laboratories are CPA accredited. It is a clinically led organisation and its genetics section is headed by a senior clinical geneticist with close ties to NHS services.

- **GSTS Pathology,** a large, independent provider of pathology services. It is a public private partnership between Guy's and St Thomas' and Kings College Hospitals NHS Trusts and Serco plc. The molecular and cytogenetics laboratories based at Guy's and St Thomas' are an integral part of the RGS alongside the clinical genetics service at this hospital. The haematology laboratory at Kings and the laboratories at Guy's have CPA accreditation and are designated by the UKGTN as diagnostic providers. They offer a comprehensive range of tests to both NHS and private sector clients. The DNA laboratory at Guy's also offers preimplantation genetic diagnosis for an increasing number of disorders. In the cytogenetic laboratory, array CGH is offered as the first line test for constitutional chromosomal imbalance, almost completely replacing FISH and conventional karyotyping (except where balanced rearrangements are suspected). A dedicated cytogenetic and molecular oncology service offers a range of tests and interpretative expertise for clinical oncologists and haematologists. In addition, GSTS has an NGS platform and are currently developing NGS for service use.
- **NewGene (Next Generation Diagnostics),** a newly established commercial partnership between Newcastle Hospital NHS Foundation Trust and Newcastle University based in the International Centre for Life in Newcastle. As its badging implies, its focus is on molecular laboratory testing and it was set up to meet a perceived need for additional capacity in this area. The organisation has a close relationship with the RGS, and both its Director and Assistant Medical Director are senior clinical geneticists within the NHS. The centre offers DNA and RNA based testing services to the NHS, private sector, research community and pharmaceutical industry, including NGS technologies.
- **LGC,** founded in 1996 following the privatisation of the Laboratory of the Government Chemist. The company has grown significantly since 1996 and now has 29 laboratories and centres across Europe as well as a presence in the US, Brazil, China and India. Still retaining its statutory role as Government Chemist, the organisation has expanded its activities from an original measurement standards focus to be a major player in a wide range of laboratory services in forensic science (being the major provider of paternity testing in the UK), food science, life science and genomics. Amongst a broad spectrum of genomic services they offer DNA sequencing services which are used by the pharmaceutical industry, research communities and medical institutions. Their NGS platforms (Illumina H. Seq 2000, Roche GSFLX Titanium and ABI3730xl) allow them to offer a range of NGS services, including full genome sequencing of organisms. Unlike the other UK players above, LGC does not have such close clinical ties to the NHS genetics services, however, they would appear to be poised to provide testing services for the NHS should opportunities present themselves.

In the absence of published data, it is difficult to assess exactly how much molecular genetic testing of NHS patients is currently being performed by commercial laboratories, either within the UK or abroad. What is clear from the above, however, is that there is a growing capacity in the private sector which would be able to take on this work should those commissioning services decide to go down this route. The organisations clearly have (or can acquire) state of the art equipment and high levels of bioinformatics expertise, and will have greater ability than the NHS to acquire new technology as it appears.

Within the healthcare diagnostic arena NGS technology is being used primarily for targeted sequencing of one or more candidate genes, mainly *BRCA* at present. Nevertheless, even in this area the use of NGS enables not only faster, but larger volume testing (parallel patient testing) and the economies of scale that this allows can have a significant effect on costs when compared with Sanger sequencing. This will no doubt become an important factor for commissioners, particularly if NGS technology does not become readily available within NHS services and private providers are seen to be more competitive. Any private laboratories can apply for UKGTN laboratory membership. In order to achieve membership the laboratory must adhere to defined criteria and quality standards. UKGTN offers two types of membership 1) diagnostic provider and 2) technical service provider. The technical service laboratories provide an analytical service for UKGTN diagnostic service laboratories. They do not provide the clinical interpretation but report directly to the referring Network laboratory. It is within this context that technical service laboratories are UKGTN members. These laboratories will comply with standards such as ISO 17025.

Though a move to use private providers may be seen to have advantages in terms of costs, it would bring with it challenges in other ways which should not be forgotten. Shortcomings in the laboratory accreditation system, the need for agreed QA standards against which the Care Quality Commission can inspect laboratories and the need for greatly improved IT connectivity in NHS pathology services are all issues which have been highlighted in the second Carter report[359]. It is reasonable to accept that patients are entitled to the same standard of service regardless of whether commissioners purchase them on their behalf from NHS or from private providers. This will require coordination and standardisation of laboratory accreditation and compliance with standards such as ISO 17025, (the main international standard used by testing and calibration laboratories) as well as comparable regulation arrangements (see Chapter 8). The involvement of different organisations in providing services will add to the complexity of ensuring appropriate IT connectivity especially in this sensitive field. These problems will be further compounded by the involvement of providers from outside the UK.

Once costs come down sufficiently, it may become time for a reappraisal of the current approach because it may prove cheaper to sequence the whole genome rather than to target a small number of genes in the diagnostic setting. If and when NGS is used to do whole genome sequencing in place of targeted testing, another whole level of complexity will be added to these issues and to the commissioning process itself. The quantity of data will be much greater, and bring with it IT design issues together with all the problems of data access and storage discussed in previous chapters of the report. There is also likely to be a significant amount of information that is not pertinent to the clinical question (so-called incidental findings). Who decides what happens to these findings, and who has access to them? What of this information is given to the patient and by whom? This is all still uncharted territory for the NHS, so there are no ground rules which can be just adapted for use with the private sector. All this will need to be addressed in the contract with the service provider.

10.7 Direct-to-consumer testing

10.7.1 What is direct-to-consumer testing?

Medical tests that are both marketed and sold directly to the public, without the supervision of a healthcare professional, are classed as direct-to-consumer (DTC) tests. The DTC market currently consists primarily of self-test kits for various blood or urine analytes, such as glucose or cholesterol, many of which are available over-the-counter at high street chemists. More recently, this market has expanded to include DNA testing services that are available over the internet. In this case, instead of purchasing a testing device, the customer sends a sample of tissue away for testing and their results are made available online. Although some companies offer (or require) the involvement of a qualified healthcare professional before and/or after testing, the ethos is broadly one in which the consumer has direct access to his or her own genome so that they can take charge of their own health. The development of this 'consumer genomics' industry has fuelled international debate about the implications of widespread, medically unsupervised access to genetic information.

There are numerous players in the field offering a multitude of different tests and services. A regularly updated list of tests and providers is available online from the US Genetics and Public Policy Centre, and a Report published in 2010 from the UK Human Genetics Commission (HGC) identified 11 different categories of DTC genetic tests (see Table 10.1)[360]. Whilst specialist providers exist that offer just a single analysis (*e.g. APOE* genotyping, paternity testing, *etc.*), because of the nature of genomic analysis and the development of affordable high-throughput technologies, increasingly a company may offer multiple tests on one sample as a single service. Although the majority of companies are not currently using NGS technologies, the entire industry is becoming increasingly geared towards a genome-wide approach to analysis; ultimately, numerous different tests and categories of tests could be combined into a single service. This breadth of analysis makes questions around the evaluation and regulation of these services much more complicated than those associated with individual medical tests, whether available DTC or otherwise.

10.7.2 Applications and key players

Sequencing currently represents a very small portion of the DTC genomics market, and the major players have instead focused on offering cheaper DNA arrays to genotype around a million common polymorphisms (SNPs) across the genome. By using results of numerous genome-wide association studies for hundreds of complex diseases, companies such as 23andMe, deCODEme and Navigenics are able to estimate an individual's risk of multiple different diseases by combining an estimated population absolute risk of the disease (*i.e.* incidence) with the individual's relative genetic risk (see Chapter 7 on risk prediction). In addition, a carrier testing array for multiple Mendelian recessive diseases is offered by several services (such as Counsyl).

Knome was the first company to offer WGS to consumers. It launched in November 2008, just two months after 23andMe and deCODEme, offering wealthy customers the sequence of their genomes for US$350,000. Since then, as sequencing technology has improved, the retail cost of complete genome sequences has plummeted (though still remains at least an order of magnitude more expensive than genotyping arrays). In April 2009, Knome auctioned a genome sequence to an anonymous individual on eBay for US$68,000, and subsequently established this as its retail price point – a five-fold drop over just 12 months. Two months later, genetic technology company Illumina launched its own personal genome sequencing service for US$48,000, and a year later dropped its price to US$19,500. Illumina's EveryGenome service is provided through a Clinical Laboratories Improvement Amendments (CLIA)-certified and CAP-accredited laboratory and is now available (through healthcare professionals) for as little as US$9,500 per genome for medically indicated cases. It is likely that, as the price of sequencing and data processing for WGS continues to drop; more companies will offer WGS directly to consumers at an affordable price, either as raw data or with various forms of interpretation.

Aside from WGS of individuals, other applications using NGS could potentially be more successful in the DTC market. For example, the use of NGS for non-invasive prenatal testing would doubtless be very popular if it became available DTC prior to being offered by the NHS, particularly for relatively common conditions such as Down syndrome[361]. Similarly, targeted NGS-based approach to offering universal carrier screening may also be popular and is likely to be available DTC before being provided universally by the NHS. Applications in individual cancer monitoring or screening might also prove to be possible in the longer term.

Table 10.1: Types of tests covered by the UK Human Genetics Commission Common Framework of Principles for Direct-to-Consumer Genetic Testing Services (*adapted and used with permission from the HGC*)

#	Test type	Description
1	**Diagnostic**	Tests intended to diagnose a medical condition in a person with symptoms and/or signs
2	**Presymptomatic/ predictive**	Tests intended to predict with a high probability that an asymptomatic person will develop a condition (*e.g. BRCA* testing for breast cancer, mutation testing for monogenic conditions)
3	**Carrier**	Tests intended to show that a person is a carrier of a recessive condition, so that although they are not themselves affected, there is a risk they may have affected children
4	**Prenatal**	Tests intended to identify medical information about a fetus, or to establish fetal sex or paternity, during pregnancy
5	**Susceptibility/ predisposition**	Tests intended to provide an indication of the absolute lifetime risk and/ or relative risk of an individual developing a condition compared with the general population (*e.g. APOE* testing)
6	**Pharmacogenetic**	Tests intended to predict the response profile of an individual to a drug or course of therapy
7	**Nutrigenetic**	Tests intended to provide information about an individual's responsiveness to a particular nutrient or diet and how this affects metabolism, health status and risk of disease
8	**Lifestyle/ behavioural**	Tests intended to provide information about an individual's behavioural propensities, performance capacities (physical or cognitive) or response to certain environmental conditions, which are designed to assist the individual to modify the outcomes of any of these by elective changes in behaviour (excluding the administration of prescribed medicines)
9	**Phenotype**	Tests intended to provide information about how an individual's phenotype is conditioned by their genotype (*e.g.* height, eye colour)
10	**Genetic relatedness**	Tests intended to determine/or provide information about a genetic relationship, including paternity tests
11	**Ancestry**	Tests intended to provide information about relatedness to a certain ancestor or ancestral group and/or how much of an individual's genome is likely to have been inherited from ancestors from particular geographical areas or ethnic groups

10.7.3 Issues

In addition to all the issues associated with WGS in a clinical setting, numerous commentators have raised various concerns specifically in relation to DTC personal genomics[362-364]:

1. *Information provision* – in the absence of a healthcare professional to guide an individual through the testing process, there are particular concerns about the provision of accurate and transparent information to ensure that individuals are able, firstly, to make an informed decision about whether or not to have a test and, secondly, to correctly interpret the results. This issue is much more pertinent in the context of WGS relative to genotyping, as the likelihood of any result containing clinically or personally significant findings is much higher.

2. *Analytical validity* – tests may be performed outside of a clinically accredited laboratory, which may result in inaccurate raw data. Although many companies now use accredited laboratories, there is currently no internationally agreed quality assessment process to ensure clinical grade data from either DTC laboratories or NGS technologies.

3. *Clinical validity* – there is no system to ensure that only clinically valid interpretations are offered. Although many of the leading companies make every effort to include only robustly identified variants in their algorithms, the clinical evidence required to correctly interpret most genomic variants in the context of an individual remains woefully inadequate. In the absence of clinical symptoms, interpretation of WGS data is extremely challenging.

4. *Utility* – although proponents of DTC genomics argue that 'personal' (rather than clinical) utility is sufficient for consumers, there is scant evidence to date that genomic risk assessment is either beneficial or harmful to healthy individuals (*i.e.* outside of a clinical context)[292]. If WGS is more widely adopted, it is possible that both the benefits and potential harms will increase as rarer variants are uncovered.

5. *Privacy and confidentiality* – although many DTC companies use password protected logins to access genomic information, the potential for either accidental or intentional security breaches is increased simply by virtue of the data being online. The risk of these events can be dramatically mitigated through careful data security policies, and most well-known DTC genomics companies appear to have adopted reasonable protections against these threats.

6. *Testing without consent* – it may be almost impossible to mitigate the risk that individuals may have their genome sequenced without their consent. Some genome scan companies (*e.g.* 23andMe) have argued that the risks of covert testing are reduced by their sample collection method, which requires 2ml of saliva; in addition, individuals are asked to sign to confirm that the sample belongs to them (or that they have gained consent from the individual to whom it belongs). However, neither of these methods will have any effect on the possibility of sequencing DNA from children, which is a particularly contentious issue within DTC genomics.

7. *Knock-on effect* – there is a concern that confused and distraught consumers of DTC genomics will turn to their health service to offer advice and support. Although this is likely to be a small minority of individuals initially, if WGS were to become popular amongst consumers, clinical validation, follow-up testing and interpretation of findings could become a crippling burden for the NHS.

These issues are extremely pertinent to the regulation of such services, including the extent to which there is a need to protect unwary consumers from harm. Currently DTC genomics companies are poorly regulated, as they offer a service, rather than simply a kit like most other DTC offerings, and claim not to provide medical advice. Various organisations are engaged in developing codes of conduct and best practice guidelines, but different jurisdictions have already taken different approaches to the problem, ranging from banning DTC access to genetic testing outright, to allowing the market to develop entirely unchecked[365].

Although it remains unclear which of the various regulatory options will become dominant for DTC genomics companies, it is likely that regulators could have a major and potentially stifling impact on this fledgling market. To date, the market for DTC genomics is small – perhaps only around 100,000 individuals in total have purchased a service, the majority of which are US-based. The potential future impact of consumer genomics on the NHS, or the health of the UK population, is therefore very difficult to judge and could remain minimal.

10.8 Future service models

The current configuration of molecular genetic laboratories has undoubtedly served the NHS well to date. It has enabled RGSs to deliver high quality, coordinated laboratory services for patients and families with inherited disease across the UK. However, it has always been recognised that this may not hold true forever and that the situation would need to be reviewed in the wake of new knowledge and technologies. The 2003 White Paper[354] highlighted the need for genetic laboratories to work more closely with those in the other pathology disciplines. Funding was made available for pilot sites to trial models of more integrated working, in line with the Modernising Pathology Services[358] agenda. The expansion of molecular techniques in cytogenetics is blurring the boundaries with molecular genetics. The use of both services by other specialities (particularly haematology and oncology) is increasing significantly and will continue to do so as more targeted treatments are developed. As a result, an increasing proportion of the work of genetic laboratories is accounted for by tests provided outside of a family based context appropriate for inherited disease.

The advent of NGS technologies could potentially replace almost all of the methods currently used in both molecular and cytogenetic laboratories. In addition, the capacity of NGS equipment is huge, and it works most efficiently and cost-effectively when working at full capacity. The Clinical Molecular Genetics Society (CMGS) produces an annual audit of the genetic testing activity undertaken by the laboratories of its members. The CMGS audit for 2009/2010 indicated that around 210,000 samples for DNA based tests were received by the molecular genetics laboratories of its members during that 12 month period, a 16% increase on the previous year. The laboratories represented include nearly all of the UK NHS molecular genetics services plus some specialist services; therefore, even though many of these tests would not require sequencing, the figure is a reasonable measure of molecular genetics laboratory activity in the UK.

Although NGS might have many more applications and increase the demand substantially, it is unlikely that the work of the genetics service alone will justify the establishment of NGS technology within each RGS. The capacity of the NGS platforms could only be used efficiently and cost-effectively if the equipment was used collaboratively with other services, either with other genetics services or with other laboratory services co-located within the same institution. In view of all of these factors, the time would seem ripe for a reappraisal of the way that genetics laboratory services are positioned and delivered.

Whilst these changes in technology and working practice are occurring in genetics laboratories, the use of molecular techniques has been increasing significantly in most other pathology disciplines. The need to coordinate molecular diagnostic services as they develop across pathology has been discussed above. The NHS is currently considering what the future configuration of pathology services should be in order to meet the recommendations in the Carter report for centralisation, and to achieve the significant cost savings identified. Many think that the opportunity should be taken not only to increase the integration of pathology disciplines more closely, but that genetics laboratories should be included within the integration to sit alongside the other laboratories. The imminent arrival of NGS with its large capacity and capability to provide massively parallel testing make it even more appropriate to consider such integration as a matter of urgency. As the RCPath report[207] points out, in relation to molecular testing, co-locating departments on one site would allow the sharing of equipment, expertise and workforce. It would also enable the coordination of tests done on a single sample and prevent unnecessary duplication of tests.

If the arguments for integration of services are accepted, the question is at what level this integration should take place. Closer working and dialogue between clinicians and laboratory and informatics experts is desirable in the interests of patient care, especially with the use of complex tests. Much support has been given to the use of a clinical service model in which patient management decisions are made within the context of a multidisciplinary team, and where molecular tests are ordered in a defined sequence within a pre-agreed diagnostic protocol. These factors would argue for the local provision of an integrated service which would allow for easier dialogue between clinicians and laboratory scientists. However, in relation to molecular testing, the need to concentrate testing in order to maximise the use of expensive equipment and capitalise on the expertise in the current genetics laboratories weighs against this and favour a

regional model. The introduction of NGS would strengthen this argument further, not only in terms of equipment capacity, but because of the sophisticated bioinformatics infrastructure that would be required to handle the large amount of data and its interpretation. Both the Carter report[358] and the RCPath report[207] argue strongly in favour of regional delivery of specialist pathology services, which would include molecular testing for both acquired and inherited disease.

There has been growing acceptance that this is the direction in which services should move. Many areas in the UK are already working towards the goal of greater coordination and appropriate reconfiguration of services. The pace at which this is occurring and the models being adopted do however vary:

- In **Northern Ireland**, molecular testing in the various pathology disciplines has been integrated in Belfast into a virtual molecular pathology service, which includes the NHS genetics laboratory. They have adopted a four hub functionally based model.
- In **Scotland** proposals have been put forward to coordinate the molecular pathology diagnostic testing in the four centres and use a consortium type of approach, similar to the genetics laboratory consortium model, for coordination and rationalisation of testing. There would be close ties between the consortia and an overarching group providing strategic direction.
- In **Wales**, all genetic testing for both inherited and acquired disease is done in laboratories within the University Hospital of Wales (UHW) in Cardiff. In relation to molecular pathology diagnostic testing, some of the tumour diagnostic testing is done in the laboratories of the All Wales Medical Genetics service, which provide the service for inherited disease. A range of other molecular tests are provided by other pathology or haematology laboratories on the UHW site. A review is currently underway and will be assessing options for future models of service with closer coordination of molecular pathology testing.
- In **England**, the recommendations in the Carter reports[358;359] have driven the move towards integration of pathology services, including molecular testing. Again models and progress are varied. In some areas partial integration has been achieved so far to form Institutes of Pathology such as that at Bart's and the London. In others, more integrated services have been achieved as in Cambridge. Here, all the molecular testing for all the pathology disciplines together with that for the genetic service is delivered by an integrated service with co-located laboratories. The service operates on a model of three overlapping hubs that share technology and expertise, whilst retaining independence of reporting. This allows equipment to be used most efficiently and expertise to be capitalised on to streamline service delivery – for example DNA extraction from all blood samples received by the service is carried out in a single hub. Skill mix and training of laboratory personnel will allow cross working between hubs, which provides greater efficiency of service delivery and resilience especially when there is capacity surge in any area.

Without doubt the types of service models described above are those envisaged by both the Carter report and the RCPath. They would also ensure that a service would be well placed to readily adopt NGS technology when it enters service use, and to use it with maximum efficiency. The true capacity of NGS has however not yet been fully tested in the service setting. Until this happens, it will be difficult to predict how many centres or services should be equipped with the technology to provide an efficient and cost-effective service for NHS patients across the country.

How fast and how widespread this type of service reconfiguration will be is uncertain. Before the last general election, the then Strategic Health Authorities (SHAs) were directed to submit plans to the Department of Health outlining how they intended to realise the substantial cost savings they were required to make in the wake of the second Carter report[359]; the clear implication being that this would require significant reconfiguration of pathology services. However, the new Government announced in its 2010 NHS White Paper[356] that SHAs would cease to exist under the new NHS arrangements. In the absence of SHAs it is unclear who will drive the service redesign envisaged. It is presumed that the new NHS Commissioning Board will play a part in its role of providing a strategic direction for commissioning of services, and the Human Genomics Strategy Group (HGSG) will advise on strategic planning for genetics and genomics within the NHS.

Until such rationalisation of services occurs, costs of molecular testing will remain higher than they might be under a more centralised system. In line with the policy direction and the Carter report, there is considerable encouragement to consider a broader range of providers, and the existence of commercial sector providers with considerable capacity and the potential to offer testing at competitive prices will inevitably become apparent. It is highly likely that costs will come down substantially in the private sector when NGS is more fully employed. This will likely lead to an increasing number of molecular tests being purchased from laboratories outside the NHS. The most likely scenario for the future is therefore one of a mixed economy, with testing services being provided by both the public and private sectors, with all the challenges that this will bring.

The discussion above is based on the assumption that the majority of routine genetic testing will be done in laboratories using large, high capacity sequencing platforms. This appears to be the prevailing current view. However, it is important to remember that technological development in this field is fast moving and it can be expected that the capability and performance of bench top sequencing platforms will become much more sophisticated in the near future, as will their design *e.g.* the development of modular formats. Together with development of streamlined analysis tools, this could enable a configuration of testing services which are aligned to clinical need, rather than dictated by the available technology.

In such a model, smaller platforms could be used to provide testing in situations where results are needed rapidly to guide therapeutic decisions, whilst the majority of testing, which does not require such a rapid turnaround time, is done in bigger centres on larger platforms. Diagnostic genetics is already complex in the broad spectrum of conditions tested and the types of test required, and this will only increase with the introduction of such a disruptive technology. It is unlikely that a single platform or testing methodology would suit all situations: different technologies will suit different strategies depending on attributes such as capacity, turnaround, read length and the ability to detect different types of genetic anomaly (*e.g.* copy number variation, methylation). Strategically it would be ill-advised to entirely de-skill existing regional services on the basis of the exciting capabilities of second generation sequencing when the full potential of third generation technologies is, as yet, unclear. It is important that the existing regional clinical and testing services are intimately involved in the discussions and that any service reconfiguration is led by clinical need rather than technological capability.

11 Analysis and Policy Development

11.1 Introduction

Policy development operates in a broad context. In order to inform how genome sequencing might be adopted within the NHS, three alternative scenarios for the use of genome sequencing were discussed and developed at a two-day invited expert workshop. This chapter describes these alternatives, clinical, research and commercial (direct-to-consumer) which are characterised by the primary purpose of genome sequencing. Using these hypothetical scenarios, we explore the differences in rights and responsibilities of the main stakeholders, including the genome 'donor', genetic relatives of the donors, service providers, and, where relevant, commissioners and funders.

The workshop identified three specific issues for further consideration within the clinical scenario that are discussed in more detail: the nature of the results that should be communicated to patients and 'at-risk' relatives and the processes involved; whether or not individual genome sequences should be stored for future use, and the extent to which these technologies might be used for genome-wide screening. These three issues reflect the prevalent subject of debate. However participants had a range of views on how these issues might be resolved in practice.

Points of agreement following the workshop discussions are summarised in the next chapter and form the basis of policy recommendations.

Finally, a more detailed analysis of the clinical scenario, which is the subject of the report, is presented including the drivers of whole genome sequencing, barriers to implementation and related opportunities offered by the implementation of WGS within the NHS.

11.2 Scenarios for whole genome sequencing

There are broadly three contexts in which an individual could access genetic testing, or have their whole genome sequenced. These can be categorised either by purpose or by who pays for the assay and analysis (see Table 11.1):

1. **CLINICAL**: Testing is for the purpose of clinical diagnostic testing and/or population screening. Tests are offered to individuals or families deemed to be at increased risk of a specific condition. They aim to answer a specific medical question in order to improve diagnosis, prognosis, management and surveillance. The assay itself may be provided by a state-funded or private accredited laboratory; in either case, it conforms to agreed standards of quality assurance and validation. The test is funded by the state (or health insurance).

2. **RESEARCH**: The primary purpose of testing in a research context is to further biomedical research. Typically assays are offered to families, groups or sub-populations, with the aim of improving scientific understanding of a particular disease through genomics and are funded by research funding bodies (such as the Wellcome Trust or Medical Research Council in the UK) or commercial companies. The assays are often performed outside of an accredited laboratory and consequently would need to be validated to be of sufficient accuracy for individual diagnosis.
3. **PERSONAL (direct-to-consumer, DTC)**: Such tests vary widely in purpose, and individuals may order them either with a specific medical concern or driven by curiosity about genomics, disease risk or personal history (such as ancestry or relatedness). The accuracy and validity of such tests is highly variable. They are usually paid for by individuals wishing to access personal genomic information.

These contexts are not mutually exclusive, and the boundaries between them are often blurred. However, the three different scenarios entail substantially differing rights and responsibilities on the part of both the genome 'donor' (patient, research participant or customer) and the service provider and other relevant parties (see Table 11.1).

Table 11.1 Differences between the three different scenarios in which an individual can have their genome analysed.

SCENARIO	CLINICAL	RESEARCH	PERSONAL (DTC)
Purpose	Improved health for individual or family member, or for reproductive choice	Scientific knowledge	Mixed individual aims
Funding	State (or health insurance)	Research body or company	Individual /or commercial company
Recipient/subject of testing (genome 'donor')	Patient or target at-risk population	Research participant	Consumer
Rights[1]	NHS constitution within the UK	Participant information sheet, informed consent and on-going dialogue with researchers	Consumer protection legislation
Responsibilities[1]	Professional practice, medical ethics and professional guidance	Research ethics bodies	Trading and advertising standards
Feedback of results	Clinically relevant results communicated to patient through defined clinical pathway	Often nothing fed back to individuals; anonymised aggregate results usually disseminated to participants and society through publications and shared databases	Varies from complete raw sequence data to targeted interpretation of specific findings, based on price and consumer preference
Service provider	Clinical laboratory (or accredited commercial laboratory)	Research laboratory	Commercial laboratory (which may or may not be accredited)

[1] These sources may not be comprehensive.

It should be noted that, in general, the amount and type of information received by the genome 'donor' is decided by the funder and the corresponding normative framework of the service provider. For example, while a DTC genomics company could reasonably be expected to provide an individual customer with their raw genomic data if that was advertised and specified in the sale, a researcher has no obligation to provide feedback to individual research participants regarding their own data (even if feedback of medically actionable results could prevent avoidable harm to the participant) and this is usually explicitly stated in the consent procedure to enter the study. Importantly, the rights and responsibilities associated with one scenario cannot be directly transported into another, as the agreement entered into, whether implicitly or explicitly, between the genome 'donor' and the service provider is different in each scenario.

11.3 Issues for further consideration

Three main issues were identified for further consideration if WGS were to be used in a clinical context; these relate to the feedback of results, storage of data and the use of NGS technologies for genome-wide screening, and are discussed below.

11.3.1 Feedback of results

Within any context, there are a variety of options regarding what to feedback to the genome 'donor'. In the clinical context the key question is what *results* to provide to the patients, *results* implying the interpretation of sequence data in a particular context to answer a specific clinical question (*e.g.* to either diagnose or rule out a condition).

We note three broad categories of results that may be communicated to the patient or family:

1. Results directly pertaining to the clinical question
2. Results that are incidental to the clinical question, but which are an unavoidable side-effect of the investigation undertaken ('incidental' findings)
3. Results of deliberate extended interrogation of WGS data

It is currently unclear both what results patients want to receive, and what health professionals want, or are obliged, to disclose. However, within a publicly funded health care system, physicians are not obliged to provide any test or analysis simply because a patient requests it, but must also weigh its relative risks and benefits both to the patient and wider community. Particular difficulties arise over variants of unknown significance (VUS) that have a potential clinical impact, *e.g.* a novel variant in a gene known to be associated with cancer predisposition. In such cases, even if the gene is well characterised, the effect of a previously unseen variant can be very difficult to predict. Although discovery of these variants can be minimised by confining the analysis to variants of known clinical significance, they will always exist, and can only be resolved through databases that relate genomic variation to phenotype or subsequent natural history.

Results pertaining to the clinical question

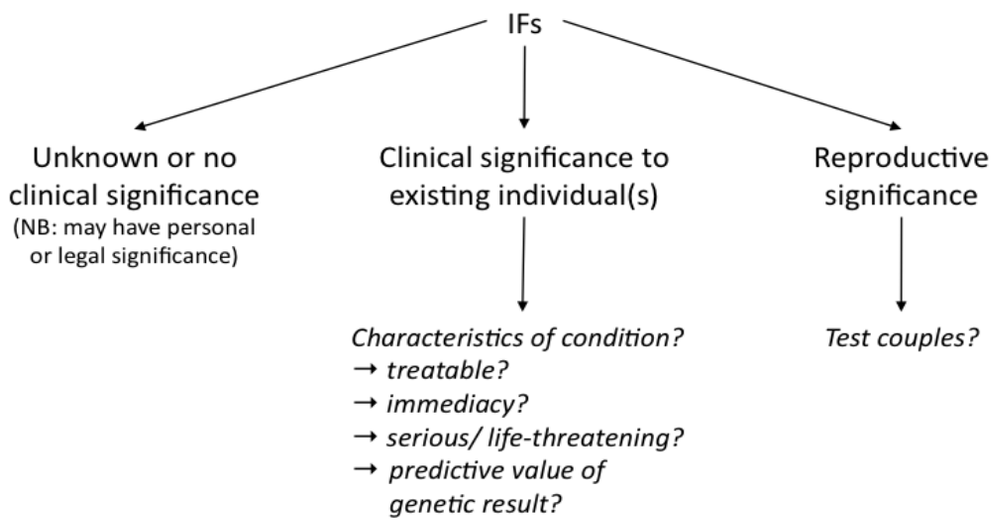
Within a clinical context, it would be expected that pertinent results relating to the clinical question (based on symptoms, family history, phenotype or a risk-category, such as a particular ethnic group) would be communicated to patients (and possibly 'at-risk' relatives) for diagnosis, prognosis, preventive management or treatment and to allow autonomous decision-making and reproductive choice. However, even in this situation, case law has established that patients are not necessarily entitled to receive all the results of a test and the professional is entitled to withhold clinically relevant results if it is believed they may do more harm than good.

Results that are incidental to the clinical question

While it may be possible to mask the majority of incidental findings by using filters to analyse the genome sequence and target analyses towards only clinically relevant variants, it will be impossible to entirely eliminate them, particularly in cases where the molecular diagnosis is unknown and genome-wide analysis is required. Findings such as variants linked to different diseases that happen to occur in the same region of the genome being investigated will be inevitable, as will non-medical findings such as misattributed parentage or incest. The number of such 'incidental' findings necessarily increases with the volume of data analysed and becomes a much greater issue in the context of WGS. The interpretation of these findings as pathological or otherwise will also change over the years as scientific understanding evolves.

As discussed in Chapter 8, the idea of grouping incidental findings into categories has been proposed previously, so that a patient or physician could decide which categories of findings to share (see Figure 11.1 for simplified example). However, defining appropriate categories, the boundaries between them and what should go into each category has not been adequately addressed to date.

Figure 11.1 Initial simple categorisation of incidental findings (IFs) and potential follow-on considerations



Although good clinical practice in dealing with incidental findings that arise within clinical genetic testing is currently not formally clarified, and findings are usually not fed-back from genetics research studies, many clinicians feel a moral duty to inform patients about any serious, clinically actionable results. However, professionals should not communicate results that are outside their own clinical expertise, but should refer patients to a competent individual.

11.3.2 Results of deliberate extended investigation of genomic sequence data

Pro-actively searching a patient's genome for other specific variants that are not integral to the clinical question, but may be otherwise useful, would constitute *opportunistic screening*, rather than diagnostic testing. Importantly, in this case the results should no longer be classed as incidental, as they would be relevant to the purpose of the extended panel of tests (see Section 11.3.4).

In the UK, screening tests require consideration against a range of screening criteria (related to the performance of the test, for example the positive predictive value of a particular set of variants in asymptomatic individuals) and a wide ranging robust population-based evaluation of outcomes and cost-effectiveness, as well as consideration of organisational, ethical, legal and social impact. Although some tests might meet some of these criteria – for example, determining carrier status

for sickle cell anaemia through targeted testing for one specific variant – many will not, and the effect of even well characterised variants (such as *BRCA* mutations) in an asymptomatic individual with no family history is currently unknown. Again the question of competence arises, and the importance of ensuring that professionals do not give advice beyond their own clinical expertise.

11.3.3 Data storage

It is useful to make the distinction between an assay (*i.e.* measurement of a biomarker) and a test (*i.e.* interpretation of that assay for a particular disease, in a particular population or person, for a particular purpose). The performance of the test relates not only to the accuracy of the assay, but also to the purpose for which it is used, *e.g.* screening versus diagnosis. In the context of WGS, this distinction is of great practical importance as the two can be separated entirely and, in practice, may be undertaken by different providers. Performing the assay involves sequencing a genome; storage of assay results could include either storing whole or minimal genome-wide data; in contrast, the test describes the analysis of any particular variant(s) in an individual compared with a reference genome relating to a specific condition for a particular purpose. There are thus an infinite number of discrete *tests* that could be performed on a single genome *assay*, ranging from highly predictive diagnostic tests to non-clinical analyses. Importantly, storage of just the result of a test, or multiple tests, requires much less data storage capacity and generates fewer data confidentiality issues than storing the result of a genomic assay.

There is continuing debate about the storage of genome assay information gathered in a clinical context. Concerns fall into two broad categories:

1. *Practical* (storage capacity, networks speeds for transfer, integration with electronic health records, access and requirement for further analysis or reanalysis of data, *etc.*)
2. *Ethical* (consent, confidentiality, privacy, access, and possible misuse, for example in ways that may be stigmatising or discriminatory, *etc.*)

There are two related but separate purposes for storing genome sequences:

1. Storage of *individual* linked genomic data of relevance to diagnosis, prognosis and management of disease(s) in that individual
2. Storage of *aggregated* anonymised genomic data to create an evidence base for interpretation of genomic results in future

If the practical issues can be overcome and the ethical issues managed, storing individual genomic assay information has the obvious advantage that the data would be instantly available if needed in the future (*e.g.* for pharmacogenetic analysis, or additional diagnoses) and that different analyses or tests could be offered to an individual based on their age, further clinical circumstances and/or personal preferences. It could be argued that *not* storing individual genomic data and reanalysing it in this way would present an enormous missed opportunity to improve both individual and population health. However, storing entire or minimal genome sequences for individual patients would require the use of electronic health records (at least in part), which has major practical and ethical implications. In addition, future technological developments may result in a substantial improvement in the quality of sequencing and genome assembly, and thus make resequencing an individual (as required) a better option.

Instead, just the results relevant to that specific clinical investigation (one or more tests) could be stored, and the rest discarded from the medical record. Some commentators have argued that once the price of sequencing drops sufficiently, there will be no need to store individual genome assay information for medical purposes, and specific test results should be stored in the medical record. If genomic data is required at a later date, it would be cheaper, easier and far more secure to simply resequence an individual's genome. In addition to side-stepping many of the issues associated with data storage outlined above (and see Chapter 4), this strategy has the advantage that the assay could be more closely tailored to the clinical question, (*for example,* relevant somatic genomes, metagenomes and/or epigenomes could be sequenced simultaneously).

Although for individual patients the balance is still uncertain between the benefits of storing comprehensive genome assay results versus only storing relevant test results and resequencing genomes when the need arises, it is clear that storage of *aggregated* anonymised genomic assays derived from clinical testing will be invaluable for developing a comprehensive database of variation with which to interpret variants of unknown clinical significance. This activity should not wait for resolution of the question of whether to store individual genomes linked to medical records. Resolving the issue of individual data storage will depend upon the cost of sequencing, utility of referring to an individual genome multiple times, NHS IT and network provision, and public opinion. In addition, the legal status of electronic health records remains unclear in many jurisdictions: the norm is that patients have the right to view personal data relating to them (subject to the caveats discussed above) but the records are owned by a combination of the creator of the record and owner of the system on which it is held. Given the proliferation of electronic health record systems, this is an area that is evolving very quickly.

Whilst electronic health records and the development of genomic databases offer novel methods for interrogating genomic data for clinical and research purposes, the simultaneous development of social networks has created opportunities to access and share genomic data in unprecedented ways. These reflect a general trend in which individuals may wish to take more responsibility for decisions which impact upon their health. These developments are likely to have wider societal implications, including changing our understanding of privacy and confidentiality.

11.3.4 Genome-wide screening

Analysis of a genome can either be targeted to be relevant to a specific clinical condition (where the approach is to concentrate the analysis solely on specific variants of interest for a particular diagnostic question, or to inform treatment or management), or wider to include variants of importance in the prediction and prevention of disease(s) either in the patient or potential offspring. Where wider use is more speculative and is not linked to a specific clinical history or symptoms, it can be regarded as screening, for which robust evidence of clinical validity and utility is needed to ensure the benefits of testing outweigh the potential harms caused by over-diagnosis and unnecessary treatment. If such testing were undertaken it should fulfil criteria for screening rather than for diagnostic or clinically based testing. In addition, the use of genome-wide technologies for screening requires specific consideration of a broad set of areas, including which conditions and variants might be tested and who should decide on the appropriate panel and keep it updated with emerging scientific and medical evidence.

It remains unclear how to determine long term effectiveness and cost-effectiveness of these technologies at individual and population levels, as well as how to ensure equitable and timely access to genome screening programmes throughout an individual's lifespan. Other issues to be considered include the commissioning of such wider genomic examination and funding of follow-up queries and counselling. It is essential to evaluate public perception of such screening, including public attitudes towards receiving information about their genomes whilst respecting an individual's right not to know. Further research is needed to assess the overall effect on social, psychological well-being of the population of offering a genomic screening programme versus individual testing in a clinical context.

11.4 Drivers, barriers and opportunities

The focus of this report is on the first scenario described above – medical testing of individuals, families, populations or sub-populations by clinical services for the purpose of improving health. Participants at the workshop were asked to consider and discuss the drivers of clinical whole genome sequencing, and barriers and related opportunities offered by the implementation of WGS within the NHS. The summarised results of this discussion are provided in Table 11.2. The main issues raised are presented below under broad headings of five specific drivers of genomics and future planning within the health service.

11.4.1 Potential for genomics and whole genome sequencing to improve health

The pursuit of scientific knowledge and the belief that this can be used to improve human health provide the main driving force behind the development of better, quicker and cheaper sequencing. The complexity of realising these benefits through the many steps of translation and implementation presents, on the one hand, a series of barriers and, on the other a set of opportunities that could have effects beyond WGS as a single new 'health technology'.

As outlined in Chapters 5-7, a number of specific clinical applications of NGS and WGS are becoming available particularly for diseases with a strong heritable component, where they replace and augment current genetic testing technologies, and for diagnosis and management of cancer. For inherited disease, the lack of robust evaluation that covers the various domains of clinical utility and relevant health economics, organisational and population impact is likely to impede the integration of these tests into routine patient pathways.

A fundamental problem, particularly in the area of rare, heterogeneous disorders is the lack of the necessary evidence, for example variant-phenotype association, or evidence on natural history of disease, changes to health outcomes or health service utilisation following testing. For cancer management, implementation of WGS has to be preceded by greater understanding of tumour biology and its association with the genetic background. First steps would include setting up databases of cancer patients with associated information on histological tumour types, in addition to clinical and genetic information.

The opportunity therefore exists to create an evidence base related to genomic testing. This should be linked with and expand current work related to evaluation of diagnostic testing in general (the NICE Diagnostics Assessment Programme) and the work of the UKGTN on Gene Dossiers. It must include: methods, resources and incentives for collecting and making available the underlying data; appropriate ways of grouping tests (*e.g.* all variants for a particular condition versus each variant separately); what should be covered in all the various domains of evaluation (*e.g.* ACCE framework); and how complexities should be dealt with.

As proof of principle, the initial focus should be on evaluation of some primary applications where there is high expectation for proven health benefits for specific groups of patients. For example, where WGS can be used as a replacement approach for specific tests (such as testing multiple genes that may carry known causative mutations) or, in cancer, where it offers a novel method for quantifying disease burden to inform treatment and predict recurrence.

The slow and detailed work to develop, evaluate and implement WGS in health systems will take place in an environment of conflicting expectations, concerns and a wide range of other priorities. On the one hand there may be inflated expectations around the power of genomics to 'revolutionise medicine' and, on the other, a distrust of the use of personal genomic information, perceived potential for unfair discrimination and concern about the apparent appropriation of the human genome for medical purposes ('medicalisation of the genome').

The apparent need to develop particular solutions for the use of genomic information in health decision-making must also be balanced by resistance of genetic exceptionalism – in which genomic information is treated as being inherently different from other information. Finally, the integration of genomic tests into clinical medicine must take place against a backdrop of acknowledged low genomics expertise in clinical specialities and primary care. The anticipated rise in breadth and volume of use of WGS should be harnessed to provide a further incentive to leaders in education and organisations such as NGEDC to emphasise the development of the necessary genetics competences throughout the curricula.

The use of WGS to answer specific clinical questions potentially provides a larger opportunity to benefit future health, for the individual and, collectively for the population by widening the filters and looking for other variants that may be predictive of other unrelated health concerns. As well as the problem of how to deal with incidental findings, this approach poses a larger set of questions about how proactive and systematic the health service may decide to be and how it ensures an acceptable benefit-harm ratio.

The complexity of work to take this forward is daunting – however, a step-by-step approach, which deals first with how wider/incidental findings are managed in other open-ended tests (such as whole body CT/MRI scans or other forms of genetic testing such as array-CGH) would provide a useful starting point. An exploration of good clinical practice, the confluence of understanding and expectations of the relationship between health care professional and patient in test provision, result communication and on-going advice and management would be an important component. And finally, the opportunity should be taken to engage the wider public health community and organisations such as the National Screening Committee to consider the role that identification of genetic variants might make in personalising prevention programmes. Thus, again the complexity of issues and requirement to take action in the case of WGS provides both a barrier to use and also an opportunity to take forward wider important questions about the use of genomics in health care and disease prevention.

There is a need and opportunity for clinicians and researchers to work together to develop validated clinical analyses to allow targeted filtering of the genome for specific clinical purposes. Clinicians can then use these filters to focus their tests on clinically relevant regions of the genome, minimising incidental findings. Different filters would be required for each different clinical question but each individual genome sequence could be used for either single or multiple tests depending upon the situation. Filters could be removed if further testing were required at a later stage. The need to develop these clinical filters adds to the difficulty of implementing WGS, and will require the development of increased clinical bioinformatics capacity with necessary interfaces with clinicians and laboratories. This, in itself is a barrier, but, once in place, the existence and maintenance of clinical filters linked to decision-support systems for clinicians who currently do not have the necessary skills to interpret sequence data (for example, cardiologists, renal physicians) provides a wider opportunity for the NHS and will help to ensure that the value of genomics is reached throughout clinical medicine.

11.4.2 Decreasing cost of DNA sequencing

One of the major drivers of WGS in clinical practice is the rapidly decreasing cost (and increasing speed) of DNA sequencing using next generation technologies. This, in itself, represents an opportunity, and yet it is important that it does not become a burden. The costs of providing sequence information rarely reflect the additional costs involved with analysis and interpretation, or the relatively high investment required to set up a WGS service and on-going costs to maintain the service. These costs include purchase and updating of equipment, training of staff, and development of appropriate analysis pipelines. The sheer quantity and complexity of genome-wide data mean that data analysis and interpretation are much more time consuming and challenging, an issue which is further compounded by the lack of bioinformatics expertise within the NHS. Costs may also need to take account of the impact on genetics services of dealing with and resolving uncertain results, possibly from tests originating elsewhere. Finally, for laboratories, there is currently a lack of standard operating procedures and quality assurance processes either for the sequencing assay or the analysis processes.

All of these barriers could be addressed with sufficient investment in research and development and strategic planning. However, without a well-coordinated and resourced approach there is a danger that piecemeal investment in equipment will take place, with different laboratories piloting the technologies locally and a plethora of resulting different practices, standards and outcomes.

With decreasing costs of sequencing, there is also likelihood that private accredited laboratories will seize the opportunity to provide sequence information to the NHS. Again, this represents both an opportunity for the NHS and a possible threat. It may reduce the initial need for capital

investment in NGS technologies, which is an important factor while the technologies are still in a state of flux, with new platforms becoming available regularly. However, increasing the proportion of commercial sequence provision may mean that NHS laboratories do not gain the necessary expertise, find it more difficult to ensure a close link between laboratory and clinician, and have little leverage against the charges made by commercial providers. Moreover, it will be important to use bioinformatics expertise within the NHS to develop and curate a genomic knowledge-base, based on collection of sequence information from a wide range of patients and conditions in order to inform future clinical decision systems. This will be dependent on the NHS owning this data – a requirement that will be lost if commercial companies become the primary providers of sequence information.

There is also an opportunity to provide a more comprehensive genetic testing service – there are a huge number of genes with known association with disease for which tests are currently unavailable. The NHS presents a wonderful opportunity to implement WGS in a way that is evidence-based, systematic, and efficient and can collect evidence prospectively. Thus an opportunity exists for the UK to become world leaders in this endeavour.

11.4.3 Move towards personalised medicine

The potential for WGS to facilitate a more rational approach to personalised or stratified medicine is clear, particularly through pharmacogenetics and tumour profiling. The hope is that this will lead to more effective treatments, with a reduction of harm from side effects, and less wastage. The use of companion diagnostics to select specific molecularly-targeted treatments is already used in cancer. By selecting the most appropriate intervention for an individual based on a tumour profile, for example, the use of unnecessary, ineffective and harmful alternative treatments can be minimised.

For many potential applications, such as genomic risk prediction for complex diseases, the reality is a lack of evidence of clinical validity and utility to support personal genomics as applied to the individual within the NHS. However, despite this lack of evidence, opportunities currently exist to influence decision-making directed at populations. For example, modelling has shown that the use of genetic variants to stratify women according to risk for breast cancer could be used to fine-tune the mammography programme, starting those at higher risk at a younger age, and resulting in a more cost-effective programme with reduced harm from false positives[366].

Nonetheless, the continuing drive towards personalised medicine has led to unrealistic expectations about providing genuinely personalised disease prevention programmes tailored to an individual's genome. The increasing belief in the power of personal genomics and predictive medicine could increase the burden on the NHS from the 'worried well', who have undertaken personal genome sequencing and request further advice and/or numerous follow-up tests and treatments. Such expectations may be influenced in a variety of ways. For example, the rise of 'citizen power', which may be fuelled by the ability of large numbers of individuals with a common cause or interest to communicate easily over the internet, could influence the agenda of personalised medicine significantly. It will be important not to ignore such developments, but to acknowledge them and harness their power to improve health and optimise healthcare use.

To date, affordable consumer genomics products (<£1,000) have been restricted to genotyping of known variants, often without robust disease associations, and thus have had limited impact and are often dismissed by the medical establishment. However, the importance of this industry could substantially increase once a whole genome sequence can be purchased at an affordable price. Rather than being primarily aimed at the 'worried well', low cost genome sequencing could provide an alternative diagnostic route for families with an undiagnosed inherited/ heritable disease, or individuals struggling to choose between alternative cancer treatments or trials. This route could become the fastest available to patients if commercial genome sequencing providers, coupled with not-for-profit DIY online interpretive services, outpace NHS adoption of WGS. The publicity via patient group websites that would inevitably follow even a single success story from anywhere in the world could be enough to turn consumer genomics into a major driver for the medical application of WGS.

11.4.4 Economic downturn and widespread public spending cuts

The current economic climate and resultant widespread cuts in NHS spending is a major driver that cannot be ignored when planning for the implementation of new medical technologies. There will be less money available for investment and strong pressure to make immediate and short-term savings. The suggestion of investing now to save in the future by improving health and reducing future health service costs will have little leverage.

The difficult situation for genetic testing is that NHS budget 'silos' have even shorter-term costs and savings resulting from genetic testing are very hard to determine, particularly as the main savings are often outside the clinical genetics budget (for example, array-CGH testing replacing a range of other paediatric investigations in learning disability). This, together with the lack of accurate cost information from laboratories and no standard way of costing tests within the NHS, makes economic analysis difficult. The intangible benefits of making a molecular diagnosis, such as the opportunity for parents to avoid having another affected child, to prevent other costly medical investigations or ensure more ready access to appropriate support services, are extremely difficult to quantify, value and compare with other health care interventions.

However, this is an opportunity both to review and modify existing NHS budgets and to recognise the savings made through the integration of laboratories and a merging agenda for the common goal of genetic provision. New economic models should therefore be developed that take into account added value and savings throughout the entire patient pathway and potential long-term health benefits.

11.4.5 NHS re-organisation

Although there is widespread agreement on the need for re-organisation of certain services within the NHS to improve efficiency, at the time of writing the details of specific changes remain uncertain. This uncertainty is itself a barrier to strategic planning, capacity building and the implementation of any new technologies within the NHS. Commissioning pathways and pathology services are likely to undergo major revision within the NHS over the next few years. This period of change, coupled with the expected reduction in NHS managers, may make decision-making and strategic investment more difficult.

However, the planned reorganisation of NHS services also presents a substantial opportunity for establishing commissioning structures and contractual arrangements that are supportive of the new potential for genetic test provision within the contexts of specialised genetics services and wider clinical specialties. The 2010 White Paper *'Liberating the NHS'* has provided a chance to review commissioning arrangements for genetics services as part of the transition from Specialised Services National Definitions Sets (SSNDS) to the NHS Commissioning Board. This work has been led by UKGTN and has recommended the principle that the NHS Commissioning Board should retain responsibility for commissioning clinical and laboratory services provided by the regional genetics centres, whilst GP Consortia should commission genetic tests requested by non-clinical geneticists for services (within established care pathways) that are not funded by the NHS Commissioning Board. With the development also of a national tariff for genetic tests this will enable diagnostic costs to be aligned with the appropriate speciality within the patient care pathway. Following evaluation by either NICE (for common conditions) or UKGTN (for rare conditions) tests would be priced and included in a service directory to promote equitable access through commissioning processes[367].

As part of this review of commissioning, it is also explicitly acknowledged that technological developments and whole genome sequencing in particular are driving organisational change in laboratories. This change is already occurring in health services throughout the world, and the rate at which other countries adopt WGS could act as a driver for implementation within the NHS. Opportunities to achieve the most effective and efficient use of WGS will be central within the pathology modernisation agenda (driving the response to the development and integration of genomic technologies in other pathology disciplines) and the Modernising Scientific Careers programme within which the appropriate amalgamation of separate disciplines such as molecular and cytogenetics may take place. The new proposed concept of *Genomics Laboratories*, covering a

wide range of services to all medical specialties, has been considered by the Service Development Group of the HGSG when considering the way forward and may be an important opportunity to securely embed WGS within pathology service[367].

Table 11.2 Summary of drivers, barriers and opportunities for implementation of WGS in the UK NHS

DRIVERS	BARRIERS	OPPORTUNITIES
Potential for genomics and whole genome sequencing to improve health	• Rise of genetic determinism and genetic exceptionalism • Translation gap and lack of evidence of clinical validity/utility • 'Enthusiasm'-based decision-making • Conservatism within clinical genetics community • Lack of genetics expertise amongst GPs and other medical specialties • Uncertainty about what to do with incidental findings and cost of following them up • 'Medicalisation' of the genome • Fear of genetic discrimination and loss of privacy	• Use of WGS in situations with proven clinical utility to improve/refine diagnosis and management of disease • Meet unmet health needs • Development of clinical 'windows' into the genome to act as packages of clinical tests • Evaluation body to ensure validity of variants included in clinical test packages and to ensure evidence-based decision-making • Development of competencies and improved medical training • Learn from other open-ended tests, *e.g.* whole body CT/MRI scans
Decreasing cost of DNA sequencing	• High investment costs of setting up and maintaining WGS service, *e.g.* equipment and training • Quantity and complexity of data and interpretation • Lack of bioinformatics expertise within NHS • Lack of SOPs and agreed QA processes for NGS technologies	• Private (or charitably-funded) DNA sequencing providers • Development and curation of genomics knowledge-base, *e.g.* database of variants • Increased understanding of genomics and biology in future • Increased understanding of how genetics and environment together influence phenotype
Move towards personalised medicine	• Hype and unrealistic expectations obscuring true benefits • Potential increased burden from the 'worried well' on NHS • Lack of evidence of clinical validity/ utility	• Public and professional education about genomics • Potential cost savings from stratified medicines • Personalised cancer diagnostics/ prognostics • Increased individual autonomy • Consumer-driven diagnostic use of low cost genome sequencing • Harnessing 'citizen power'
Economic downturn and widespread public spending cuts	• Lack of money for investment • NHS budget silos make it hard to evaluate savings of genetic testing outside of clinical genetics	• Development of new economic models to evaluate added value and savings throughout entire patient pathway including reduction in diagnostic odyssey, and fund services appropriately
NHS re-organisation	• Lack of NHS managers • Uncertainty about future organisation of commissioning pathways and genetics/ pathology services • Unequal service provision leading to health inequalities • Lack of planning and capacity building	• Development of national specialist commissioning • Rational decision-making and planning within the NHS • National organisation and provision of networks • Learn from implementation of WGS in other countries

12 Policy recommendations

12.1 Points of agreement

A number of aspects of the implementation of WGS within the NHS are uncontroversial and were agreed by our expert working group:

- **APPLICATION:** The NHS will use WGS to answer clinical or health related questions about genetic variation, though the timescale is still uncertain. It is very likely that NGS technologies will be used routinely in the near future, as a replacement technology for specific genetic analyses where the cost of doing the test is lower using NGS than other existing techniques (*e.g.* multigene panels), but the technology will probably first be used mainly for targeted gene sequencing before the use of WGS becomes widely implemented. In other applications, (such as in cancer genomics) the use of NGS technologies and WGS offers novel opportunities to improve patient management and treatment.

- **STORAGE:** If individual genomes were to be stored, rather than be disposed of, retention of a 'minimum' genome (~30MB) would be feasible with respect to data storage capacity. There is enormous additional value in storing aggregated anonymised data from patients and populations who are tested, to build up a database of variants to allow better understanding and more effective clinical use in future.

- **ETHICS:** The use of WGS raises no *fundamentally new* ethical questions. However, the extreme complexity, scale and breadth of impact make it important that some vital issues are addressed urgently and with a wide constituency. This will ensure a firm basis and for consistent practice. Areas raised as priorities include the scope of informed consent, dealing with information from WGS that is additional to the immediate clinical question and testing of minors.

- **ECONOMICS:** Existing health economic models are inappropriate for evaluating and comparing WGS against other healthcare interventions. In addition, realising cost savings of genomic analyses within the NHS is hampered by existing commissioning pathways and budget silos.

- **PUBLIC PERCEPTION**: Genetic determinism should be guarded against, both amongst professionals and the public. More research is needed on people's need and want for personal/family genetic information (such as communicating additional findings beyond the immediate clinical question) and the effect it might have on individuals, families and society.

12.2 Context for implementation

One of the core aims of this project is to develop recommendations to facilitate the adoption of WGS into the NHS where it improves patient care, which should only occur where it satisfies evidence-based criteria for test validity and utility. We suggest that the following specific applications of WGS are likely to be used within clinical practice in the short to medium term:

- **Targeted analysis of known genes or variants to improve the diagnosis and management of recognised diseases with a strong heritable component.** This is likely to be the first and main use of NGS technology, and applies equally whether the sequencing assay itself is targeted (*e.g.* exome sequencing) or genome-wide, as the sequence data will be analysed in a targeted way to test only known variants of relevance to a particular condition or set of conditions. This targeted analysis approach is highly flexible and allows either a single or multiple analyses to be performed on an individual genome sequence. It also minimises the occurrence of incidental findings as any targeted analysis necessarily has to be tailored to a specific purpose or disease. (Note that, as with any medical test, it is impossible to avoid incidental findings entirely, which here may be caused by variants/results associated with multiple conditions.) Although, in principle, this 'black box' approach would facilitate the use of genomics throughout mainstream medicine, it places a substantial burden on collecting, maintaining and assessing robust evidence for gene/variant-disease association upon which the targeted analyses are based.

  (It should be noted that the validity of this targeted analysis approach is neither dependent upon the number of variants analysed nor the number of diseases tested for, but on the evidence associated with each variant. Therefore the approach neither requires nor precludes wider testing of genomes for predictive purposes, such as part of a screening programme.)

- **Genome-wide analysis to identify the cause of undiagnosed diseases with a strong heritable component, where the genetic aetiology is currently unknown**. Such analyses have been carried out for decades using increasingly high resolution technologies – from simple karyotypes to DNA microarrays – in cases where the precise cause of the disorder is unknown but a primarily genetic basis is strongly suspected, *e.g.* children with severe developmental disorders. This application of WGS is likely to remain within the purview of specialist clinical genetics services and clinical researchers, who are able to deal with the wealth and complexity of genomic data. Again, rather than being an open-ended test, the analysis should be targeted towards variants of likely pathogenicity based on prevalence, novelty, inheritance, size, gene function, haploinsufficiency, *etc*.

- **Genome-wide analysis of tumour genomes in order to offer stratified interventions.** Either through the development of personalised biomarkers based on circulating cell-free tumour DNA, or the analysis of biopsy samples to determine the genomic characteristics of the tumour, it will be possible to molecularly characterise an individual's cancer and thus tailor treatment, management and surveillance options accordingly. NGS technologies are already better than standard techniques for assaying heterogeneous samples and detecting genetic variation present in only a minority of cells. If coupled with appropriate treatments, this molecular approach could have a major impact on patient care as well as cancer mortality and morbidity.

- **Targeted analysis of genes associated with either drug response or the immune system to allow treatment stratification or tissue matching.** This approach could be widely applied to different treatments throughout medicine, and would need evidence of accuracy and clinical utility before replacing current standard practice.

The UK National Health Service offers an infrastructure that would allow implementation of WGS in a way that is evidence-based, systematic and efficient, and allows for the prospective collection of evidence. This creates an opportunity for the UK to become a world leader in this area, but to do so will require a range of concerted activities.

12.3 Recommendations for implementation

We have identified **ten** recommendations that would facilitate the responsible and effective implementation of genome-wide sequencing within the UK NHS.

Throughout these recommendations, we make no particular distinction between the use of next generation sequencing (NGS) technologies and whole genome sequencing (WGS), and assume that cost will be the key driver for switching between a targeted sequencing assay and sequencing the whole genome. Instead, we make a distinction between region-specific *targeted analysis* (whether this be through sequence capture prior to sequencing, or downstream bioinformatics analysis following sequencing) and *genome-wide analysis*.

Recommendation 1: NHS use of genome sequencing

We recommend that next generation sequencing technology should be implemented within the UK NHS in the short to medium term for applications where it offers clear clinical or cost benefits over existing tests – specifically, for the diagnosis of diseases with a strong heritable component and the management of cancer.

Evidence from our review suggests that the likely benefits from NGS technologies outweigh their potential harms if they are introduced responsibly and systematically for specific applications. Whether targeted, exome or whole genome sequencing is used should depend simply on the cost, accuracy and technical implications of each method, rather than the clinical question. The NHS presents an excellent opportunity to implement NGS (including WGS) in an efficient, equitable and evidence-based manner. The UK has an opportunity to become a world leader in using these scientific and technological advances in medicine.

Recommendation 2: Clinically targeted analysis

The analytical approach for clinical interrogation of genome-wide sequence data should be clinically directed, such that only variants of relevance to the specific condition are analysed and shared with patients. At this time, we do not recommend interrogating genomic data more extensively for preventive purposes in the absence of a clinical indication.

There is a shared concern among health professionals about how to best use whole genome data for the benefit of patients in the short term, and a consensus is emerging amongst clinicians that targeted testing should be used to enable specific clinical questions to be answered. By treating the genome sequence as an *assay* upon which specific targeted *tests* are performed, genome-wide sequencing would fit into existing models of care without altering the current relationship between the patient and health care professional or unduly distorting their respective rights and responsibilities. Systems need to be put in place to develop, agree, maintain and update a method for filtering and prioritising variants, which should draw on appropriate expertise from laboratory scientists, informaticians and clinicians.

In the case where a specific diagnostic question can be asked – such as the presence of known variants associated with an inherited cardiac condition, or specific drug receptors in a tumour genome - the ‘test’ would constitute a targeted analysis of the genome sequence for known, clinically validated pathogenic variant(s). Where the diagnostic question cannot be framed so specifically – such as determining the cause of an undiagnosed developmental disorder, or cataloguing the aetiologically important differences between germline and somatic genomes – the ‘test’ would constitute a genome-wide analysis of the genome sequence for rare or novel variants based on criteria for predicting likely pathogenicity. Multiple different analyses could be performed on an individual genome sequence based on clinical diagnostic or preventive need, and the filtering process could be easily expanded, adjusted or removed as biological knowledge increases.

Recommendation 3: Biomedical informatics

The NHS should urgently seek to develop clinical bioinformatics expertise and infrastructure to ensure the availability of sufficient technical support to allow clinical interrogation of genomic sequence data. This may best be achieved through the establishment of a National Biomedical Informatics Institute, in addition to employing bioinformaticians embedded in local clinical services.

Bioinformatics expertise is a necessary prerequisite for the implementation of NGS technology, and there is an urgent need for bioinformaticians within the NHS to:

- Develop a method to facilitate clinically targeted genomic analyses to answer specific clinical questions (see Recommendation 2)
- Develop and curate a database of variants (with associated clinical phenotypes where possible) to facilitate interpretation of rare variants in individuals (see Recommendation 4)
- Develop and curate a standardised method for collating and comparing phenotype data
- Develop and support standardised data exchange protocols, file formats and interfaces with the NHS
- Develop clinical decision support mechanisms for interfacing with a range of clinicians (geneticists and others) to allow genomic data to inform diagnosis of inherited disease and other stratified medicines initiatives
- Support local clinical teams (health care professionals and laboratories) in the interpretation of WGS data

These activities could be coordinated initially by a centralised Biomedical Informatics Institute where bioinformaticians work closely with clinicians to ensure activities are directly tailored to clinical need. The Institute would undertake development, curation and provision of bioinformatics tools for the NHS, as well as providing training and professional governance for bioinformaticians embedded within local services and appropriate educational support for clinicians to understand and use those bioinformatics resources.

Although there are currently a handful of bioinformaticians employed within the NHS, they have no professional accreditation body (equivalent to the ACC or CGMS for laboratory scientists) and thus may lack specific required competencies and professional codes of conduct. The regulation and governance of this emerging discipline will need to be clarified, including defining the nature of the responsibilities owed by bioinformaticians and demarcation of liability for negligence.

Recommendation 4: Developing the evidence base

An evidence base is needed to allow clinical interpretation of genome-wide sequence data, with standardised databases of normal and pathogenic genomic variation at its core, and linked analytical tools to facilitate clinical use. Construction and maintenance of this evidence base should be an urgent and ongoing task of a National Biomedical Informatics Institute.

> ***Sub-recommendation 4a:*** **Anonymised and/or linked-anonymised genomic data should be stored centrally to act as an evidence base to facilitate interpretation of individual variants in future.**
>
> ***Sub-recommendation 4b:*** **Population and clinical research projects should receive sufficient infrastructural and financial support to ensure that normal and pathogenic genetic variation can be accurately distinguished and interpreted.**
>
> ***Sub-recommendation 4c:*** **Formal evaluation systems should be set up to collect data on the clinical validity and utility of genetic tests and genomic analyses, to inform and validate available clinical analyses and set national standards of practice.**

Development of an evidence base for clinical interpretation of genomic data requires both

integration of existing resources (such as DECIPHER, DMuDB and numerous clinical locus-specific databases), and development of additional tools to allow data generation and gathering, as well as analysis and interpretation. Constructing the evidence base with a layered structure (databases of genomic variation that can be interrogated by multiple interpretive tools and accessed via numerous portals) would allow for development, evaluation and access at different levels.

There is currently no formal system to record and share genomic information amongst clinicians. A database should therefore be set up to allow storage and sharing of information especially on variants of unknown clinical significance; anonymised genomic data should be aggregated for population use and to provide a database of variants to facilitate interpretation. The extent of data storage, access and disposal should be considered both within and between different jurisdictions, and appropriate data protection safeguards will need to be put in place to ensure individual confidentiality, and public perception will need to be carefully managed.

At the same time, methods need to be developed to improve the analysis and annotation of genomes based on the growth of such databases and increased scientific understanding. Data on the extent of normal genomic variation from research studies (such as the 1000 Genomes Project), as well as from large control cohorts used in disease association studies is invaluable for interpreting variants in patients and predicting their likely pathogenicity. This work must be complemented by clinical research into the genetic basis of disease, as well as translational research to evaluate the validity and utility of gene-disease associations, in order to facilitate the prompt clinical implementation of findings and products.

In order to facilitate clinically targeted analysis (see Recommendation 2), a formal evaluation mechanism will be needed to select and validate the variants that should be included within directed genomic tests for specific conditions. There are currently insufficient data to support robust evaluation of NGS technologies and genomic analyses, and a lack of an established system to generate and collect this type of data. To date, this role has been played in part by the UKGTN predominantly for molecular genetic tests within the clinical genetics service. If expanded and adequately supported, workshop participants were broadly in favour of the UKGTN (supported by and NICE for higher volume testing) acting as a gatekeeper for evaluating diagnostic genomic analyses.

Recommendation 5: Policy development

Policy research is needed to define the evolving relationship between the health service and patient, their respective rights and responsibilities in the context of genome-wide analysis, and to develop professional guidance for clinical use of WGS.

In the short to medium term WGS will largely be used as a replacement technology for currently available genetic tests, and thus current professional guidance relating to the rights and responsibilities of health services and patients will apply. However, the potential availability of extra data and information that might be gained from WGS, the potential use of the technology by a wider group of health professionals for a wider range of conditions, and the likely expectations of patients lead us to recommend that some further development of good practice and formalisation of information exchange prior to testing would help to avoid a mismatch between expectations and use. Such guidance should clarify professional practice with respect to:

- The information that should be communicated to patients to secure meaningful consent (including the probability and significance of true and false positive/ negative findings, and incidental findings)
- The elements of analysis that will, and will not, be included
- The results that will and will not be reported back to the patient (including variants of unknown significance and incidental findings)
- Any continuing obligation to re-examine this data (in the light of developing scientific knowledge), and the possibilities and practicalities of recalling the patient

- Whether patients have the right of access to their own genomic data for subsequent personal, clinical or family uses and under what circumstances
- Dealing with customers of consumer genomics companies who have identified potentially clinically important variants

Whilst a tailored approach would decrease the number of incidental findings, by targeting analyses only at relevant genes/variants, it will still be necessary to explore the implications for securing informed consent (bearing in mind the greater potential for incidental and unanticipated findings, increased need for data sharing, and the possibility that lapses in confidentiality might lead to discrimination or stigmatisation). The development of such guidance is a substantial area of work that should involve clinical geneticists and a wide range of other specialists alongside a range of policy experts, service providers, public and patient groups and other stakeholders.

Recommendation 6: Competences and best practice guidelines

Competences and best practice guidelines should be developed for health care professionals to facilitate the responsible and equitable translation of WGS into the NHS.

Developments in NGS technologies, and the expected wide use of WGS across a range of clinical specialities, reinforce and emphasise the need to ensure inclusion of genomics in medical education and training, from undergraduate to post-graduate and post-qualification levels. We would not recommend that WGS be treated as a special case, either with respect to education, or requirements that professionals would need to be accredited in some way to use this knowledge. However, health care professionals at all levels need to be educated about WGS so they can understand its limitations as well as benefits.

Evidence-based adoption of WGS within specific clinical pathways should be supported by formal guidelines and processes (for example, patient leaflets) with a commitment to on-going development. We recommend that the lead for such development should be taken within clinical specialties supported by experts in clinical and laboratory genetics and bioinformatics, possibly supported by dedicated funding from the Service Delivery and Organisation programme of NIHR.

Recommendation 7: Service provision

A modular approach to service provision should be taken, with a small number of sequencing laboratories (or providers) acting as regional hubs to provide national coverage to the NHS, in addition to maintaining local expertise for interpretation.

There are several separable components involved in providing a NGS service: data generation (*i.e.* the sequencing assay), read mapping, variant calling and annotation, and interpretation. The first stages would benefit from centralisation and might best be located in sequencing hubs with the requisite technical capacity, quality assurance processes, informatics power and professional expertise. However, the latter stages could remain distributed throughout regional centres and involve the patient's clinician directly in interpretation of the results.

Service organisation changes have already started under the pathology modernisation agenda. Some centralisation of services is desirable to use NGS equipment at full capacity to maximise efficiency, and ensure optimum bioinformatics support and genetics expertise to interpret the results. Although it is too early to decide the exact number of laboratories with genome sequencing capacity and associated bioinformatics expertise needed by the NHS, there is general consensus that two or more central sequencing centres will be required, and certainly every local hospital will not need to be involved in sequencing. However, where rapid turnaround times are required, testing could be provided locally by smaller capacity bench top sequencers.

Concentrating expertise in the 23 existing regional genetics services may be a possible solution, though fewer super-regional sequencing hubs may suffice. The picture for cancer services is currently less well defined than for clinical genetics, but the same principle is likely to apply for tumour sequencing for the foreseeable future. This configuration and the levers through which it could be achieved are actively under discussion by the Service Development sub-group of HGSG.

Sequencing providers could be established through a range of models, including public-private partnerships or collaborative partnerships with academic centres, with agreed SOPs and QA standards put in place. Workshop participants were broadly of the opinion that there was no fundamental objection to outsourcing the assay to private providers (*i.e.* purchasing genome sequence data outside of the NHS) provided that appropriate quality assurance, security measures and accreditation were in place, and this might be the most cost-effective option in the short to medium term. However, the analysis and expertise required for interpretation would need to rest within the NHS.

Recommendation 8: Health economics

Outcomes evidence and health economic modelling of the impact of genomic analyses within the NHS is urgently needed to identify costs and savings both for diagnostic applications and in order to identify where it can be used to stratify interventions.

Where NGS is used as a replacement technology for an existing assay, the economic issue is simply whether this can be provided at lower cost. However, this direct substitution is likely to be rapidly superseded by expansion of variants tested or of clinical criteria for testing, such that a more complex cost-effectiveness analysis will be required. The current sparse evidence base for economic assessment of WGS and lack of studies linking genetic diagnostic tests to patient outcomes will then become a problem. Many methodological difficulties exist for health economics analyses in this context. Standard health economics analysis based on quality adjusted life years (QALYs) are often unworkable within a clinical genetics setting, as the outcomes, related to aspects such as reproductive choice or benefits to family members or reducing future disease risk are qualitatively different to standard approaches. These need to be tackled with urgency if evidence-based commissioning and provision of WGS is to take place. Although existing health economic models may be more appropriate for evaluating the impact of genomic analyses to enable stratified interventions (*e.g.* tumour profiling), in practice there is still a lack of empirical data adequately evaluating the test itself rather that the intervention. In addition, research is required on quantitative and qualitative differences in benefits and risks for patients and families, and suitable methods to address issues of complexity in health economic assessment.

Recommendation 9: Commissioning

Rational, clear and transparent commissioning pathways need to be developed and agreed between all relevant stakeholders to enable NGS technologies to be accessed and delivered effectively and equitably.

Where NGS can be regarded as a substitute technology, the decision to use it is simply a provider investment issue. However, where it is considered to be a novel technology, it needs to be embedded within commissioning processes for all the relevant services, along with other forms of genetic and genomic testing. Commissioning of specialised laboratory and clinical genetics services represents the more straightforward element, where it is now agreed that the NHS Commissioning Board will have responsibility in England. This will require development and adoption of a service specification that would include a set of service standards. Similarly, structures for commissioning cancer specific tests to guide stratified treatment strategies already exist.

More problematic will be the use of specific genomic technologies such as NGS in patient pathways for inherited disorders that are delivered outside the clinical genetics services (such as services for FH, haemophilia and sickle cell disease) and NGS that is used for non-inherited disease conditions. These will be commissioned by GP consortia within established care pathways. Such commissioning of companion diagnostics would, in theory be evaluated first by NICE. However, it is not clear what capacity NICE will have across a broad range of uses. In practice it seems likely that GP consortia may make a variety of different decisions influenced by local experts and pressure, funding and different priorities.

Particular questions for NGS will include:

- How will commissioners be provided with the necessary evidence for decision-making and can this be developed and coordinated at a national level?
- How will commissioners prioritise particularly when it may be necessary to invest early in order to achieve savings later?
- What (if any) influence may commissioners be able to exert over the configuration of test providers (for example the rationalisation of laboratories or the use of private testing companies)?

Recommendation 10: Genomic screening

Policy research is needed to consider under what circumstances wider screening of the genome might be offered, how relevant health policy should be developed and what issues might arise.

These technologies could ultimately be used in more radical ways to seek and prevent a wider range of disease in those tested for clinical reasons (opportunistic screening) or even proactively directed towards a population of asymptomatic individuals. Using WGS applications in this way is, however, fundamentally different from the clinical context and should be considered against criteria applied to potential screening programmes, particularly:

- The scientific and clinical validity and utility of these tests in the relevant populations
- Robust evidence of clinical validity and utility to ensure the benefits of testing outweigh the potential harms caused by over-diagnosis and unnecessary treatment
- The capacity of health services to respond effectively to those testing positive

Multiple predictive genome-wide analyses have the potential to revolutionise medical practice and to catalyse a paradigm shift from diagnosis to prevention. Systematic use of these technologies on a population basis either for disease prevention or to support reproductive choice does, however, raise profound questions about the nature of society that we find ourselves in and the preferred direction of travel. On-going research on the public desire for and acceptability of these technologies, and the likely societal effects, is therefore of utmost importance.

13 References

1. Ahn SM, Kim TH, Lee S *et al.*. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res* 2009;19:1622-1629.

2. Bentley DR, Balasubramanian S, Swerdlow HP *et al.*. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456:53-59.

3. Drmanac R, Sparks AB, Callow MJ *et al.*. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;327:78-81.

4. Kim JI, Ju YS, Park H *et al.*. A highly annotated whole-genome sequence of a Korean individual. *Nature* 2009;460:1011-1015.

5. Levy S, Sutton G, Ng PC *et al.*. The Diploid Genome Sequence of an Individual Human. *PLoS Biol* 2007;5:e254.

6. McKernan KJ, Peckham HE, Costa GL *et al.*. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* 2009;19:1527-1541.

7. Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 2009;27:847-852.

8. Wang J, Wang W, Li R *et al.*. The diploid genome sequence of an Asian individual. *Nature* 2008;456:60-65.

9. Wheeler DA, Srinivasan M, Egholm M *et al.*. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;452:872-876.

10. Pleasance ED, Stephens PJ, O'Meara S *et al.*. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010;463:184-190.

11. Pleasance ED, Cheetham RK, Stephens PJ *et al.*. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010;463:191-196.

12. Lander ES, Linton LM, Birren B *et al.*. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.

13. Venter JC, Adams MD, Myers EW *et al.*. The sequence of the human genome. *Science* 2001;291:1304-1351.

14. Waterston RH, Lander ES, Sulston JE. On the sequencing of the human genome. *Proc Natl Acad Sci USA* 2002;99:3712-3716.Copyright (2002) National Academy of Sciences, U.S.A.

15. Clamp M, Fry B, Kamal M *et al.*. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci USA* 2007;104:19428-19433.

16. Iourov IY, Vorsanova SG, Yurov YB. Somatic genome variations in health and disease. *Curr Genomics* 2010;11:387-396.

17. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol* 2010;28:1057-1068.

18. Li Y, Willer C, Sanna S, Abecasis G. Genotype Imputation. *Annual Review of Genomics and Human Genetics* 2009;10:387-406.

19. Rosa-Rosa JM, Gracia-Aznarez FJ, Hodges E *et al.*. Deep sequencing of target linkage assay-identified regions in familial breast cancer: methods, analysis pipeline and troubleshooting. *PLoS One* 2010;5:e9976.

20. Vissers LE, de Ligt J, Gilissen C *et al.*. A *de novo* paradigm for mental retardation. *Nat Genet* 2010;42:1109-1112.

21. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-1073.

22. Lupski JR, Reid JG, Gonzaga-Jauregui C *et al.*. Whole-Genome Sequencing in a Patient with Charcot-Marie-Tooth Neuropathy. *N Engl J Med* 2010;362:1181-1191.

23. Roach JC, Glusman G, Smit AF *et al.*. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010;328:636-639.

24. Jones S, Laskin J, Li Y *et al.*. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biology* 2010;11:R82.

25. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010;11:31-46.

26. Tucker T, Marra M, Friedman JM. Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. *The American Journal of Human Genetics* 2009;85:142-154.

27. Voelkerding KV, Dames SA, Durtschi JD. Next-Generation Sequencing: From Basic Research to Diagnostics. *Clin Chem* 2009;55:641-658.

28. Mardis ER. A decade/'s perspective on DNA sequencing technology. *Nature* 2011;470:198-203.

29. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. *Genomics* 2009;93:105-111.

30. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009;458:719-724.

31. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74:5463-5467.

32. Smith LM, Sanders JZ, Kaiser RJ *et al.*. Fluorescence detection in automated DNA sequence analysis. *Nature* 1986;321:674-679.

33. Luckey JA, Drossman H, Kostichka AJ *et al.*. High speed DNA sequencing by capillary electrophoresis. *Nucl Acids Res* 1990;18:4417-4421.

34. Margulies M, Egholm M, Altman WE *et al.*. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376-380.

35. Shendure J, Porreca GJ, Reppas NB *et al.*. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005;309:1728-1732.

36. Harris TD, Buzby PR, Babcock H *et al.*. Single-molecule DNA sequencing of a viral genome. *Science* 2008;320:106-109.

37. Bowers J, Mitchell J, Beer E *et al.*. Virtual terminator nucleotides for next-generation DNA sequencing. *Nat Methods* 2009;6:593-595.

38. Eid J, Fehr A, Gray J *et al.*. Real-time DNA sequencing from single polymerase molecules. *Science* 2009;323:133-138.

39. Karow J. Life Tech Details Real-Time Single-Molecule Tech at AGBT; Combines Qdots with FRET-Based Detection. *GenomeWeb In Sequence* 2010.

40. Karow J. Ion Torrent Systems Presents $50,000 Electronic Sequencer at AGBT. *GenomeWeb In Sequence* 2010.

41. Rothberg JM, Hinz W, Rearick TM *et al*.. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011;475:348-352.

42. Aksimentiev A, Heng JB, Timp G, Schulten K. Microscopic Kinetics of DNA Translocation through Synthetic Nanopores. *Biophysical Journal* 2004;87:2086-2097.

43. Garaj S, Hubbard W, Reina A, Kong J, Branton D, Golovchenko JA. Graphene as a subnanometre trans-electrode membrane. *Nature* 2010;467:190-193.

44. Howorka S, Cheley S, Bayley H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat Biotech* 2001;19:636-639.

45. Branton D, Deamer DW, Marziali A *et al*.. The potential and challenges of nanopore sequencing. *Nat Biotech* 2008;26:1146-1153.

46. Kircher M, Kelso J. High-throughput DNA sequencing--concepts and limitations. *Bioessays* 2010;32:524-536.

47. Astier Y, Braha O, Bayley H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J Am Chem Soc* 2006;128:1705-1710.

48. Wu HC, Astier Y, Maglia G, Mikhailova E, Bayley H. Protein nanopores with covalently attached molecular adapters. *J Am Chem Soc* 2007;129:16142-16148.

49. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 2009;4:265-270.

50. Krivanek OL, Chisholm MF, Nicolosi V *et al*.. Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature* 2010;464:571-574.

51. Linnarsson S. Recent advances in DNA sequencing methods - general principles of sample preparation. *Experimental Cell Research* 2010;316:1339-1343.

52. Medvedev P, Fiume M, Dzamba M, Smith T, Brudno M. Detecting copy number variation with mated short reads. *Genome Res* 2010;20(11):1613-22.

53. Summerer D. Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* 2009;94:363-368.

54. ten Bosch JR, Grody WW. Keeping Up With the Next Generation: Massively Parallel Sequencing in Clinical Diagnostics. *J Mol Diagn* 2008;10:484-492.

55. Mamanova L, Coffey AJ, Scott CE *et al*.. Target-enrichment strategies for next-generation sequencing. *Nat Methods* 2010;7:111-118.

56. Saiki RK, Scharf S, Faloona F *et al*.. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985;230:1350-1354.

57. Fredriksson S, Baner J, Dahl F *et al*.. Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. *Nucleic Acids Res* 2007;35:e47.

58. Varley KE, Mitra RD. Nested Patch PCR enables highly multiplexed mutation discovery in candidate genes. *Genome Res* 2008;18:1844-1850.

59. Meuzelaar LS, Lancaster O, Pasche JP, Kopal G, Brookes AJ. MegaPlex PCR: a strategy for multiplex amplification. *Nat Meth* 2007;4:835-837.

60. Tewhey R, Warner JB, Nakano M *et al*.. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol* 2009;27:1025-1031.

61. Nilsson M, Dahl F, Larsson C, Gullberg M, Stenberg J. Analyzing genes using closing and replicating circles. *Trends Biotechnol* 2006;24:83-88.

62. Akhras MS, Unemo M, Thiyagarajan S *et al*.. Connector inversion probe technology: a powerful one-primer multiplex DNA amplification system for numerous scientific applications. *PLoS One* 2007;2:e915.

63. Dahl F, Stenberg J, Fredriksson S *et al*.. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc Natl Acad Sci USA* 2007;104:9387-9392.

64. Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M. Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. *Nucl Acids Res* 2005;33:e71.

65. Krishnakumar S, Zheng J, Wilhelmy J, Faham M, Mindrinos M, Davis R. A comprehensive assay for targeted multiplex amplification of human DNA sequences. *Proc Natl Acad Sci USA* 2008;105:9296-9301.

66. Li JB, Gao Y, Aach J *et al*.. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res* 2009;19:1606-1615.

67. Albert TJ, Molla MN, Muzny DM *et al*.. Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 2007;4:903-905.

68. Gnirke A, Melnikov A, Maguire J *et al*.. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009;27:182-189.

69. Chou LS, Liu CS, Boese B, Zhang X, Mao R. DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. *Clin Chem* 2010;56:62-72.

70. Porreca GJ, Zhang K, Li JB *et al*.. Multiplex amplification of large sets of human exons. *Nat Methods* 2007;4:931-936.

71. Hodges E, Xuan Z, Balija V *et al*.. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007;39:1522-1527.

72. Brockman W, Alvarez P, Young S *et al*.. Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* 2008;18:763-770.

73. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* 2008;36:e105.

74. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Meth* 2008;5:16-18.

75. Binladen J, Gilbert MT, Bollback JP *et al*.. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2007;2:e197.

76. Meyer M, Stenzel U, Myles S, Prufer K, Hofreiter M. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucl Acids Res* 2007;35:e97.

77. Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Programme. Available at: www.genome.gov/sequencingcosts [date of access 01.03.2011]

78. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998;8:175-185.

79. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998;8:186-194.

80. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008;18:1851-1858.

81. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52-58.

82. He D, Choi A, Pipatsrisawat K, Darwiche A, Eskin E. Optimal algorithms for haplotype assembly from whole-genome sequence data. *Bioinformatics* 2010;26:i183-i190.

83. Campbell PJ, Stephens PJ, Pleasance ED *et al.*. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 2008;40:722-729.

84. Flicek P, Birney E. Sense from sequence reads: methods for alignment and assembly. *Nat Methods* 2009;6:S6-S12.

85. Magi A, Benelli M, Gozzini A, Girolami F, Torricelli F, Brandi M. Bioinformatics for Next Generation Sequencing Data. *Genes* 2010;1:294-307.

86. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: Accurate indel calls from short-read data. *Genome Res* 2010; 21:961-73

87. Ding L, Wendl MC, Koboldt DC, Mardis ER. Analysis of Next Generation Genomic Data in Cancer: Accomplishments and Challenges. *Human Molecular Genetics* 2010; 19:188-96.

88. Stein L. Genome annotation: from sequence to biology. *Nat Rev Genet* 2001;2:493-503.

89. Kent WJ. BLAT-The BLAST-Like Alignment Tool. *Genome Res* 2002;12:656-664.

90. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;447:799-816.

91. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008;118:1590-1605.

92. Harrow J, Denoeud F, Frankish A *et al.*. GENCODE: producing a reference annotation for ENCODE. *Genome Biology* 2006;7:S4.

93. Flicek P, Amode MR, Barrell D *et al.*. Ensembl 2011. *Nucl Acids Res* 2011; 39:D800-6.

94. Kent WJ, Sugnet CW, Furey TS *et al.*. The Human Genome Browser at UCSC. *Genome Res* 2002;12:996-1006.

95. Kuhlenbaumer G, Hullmann J, Appenzellerm S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat* 2010; Nov 18. [Epub ahead of print.

96. Tranchevent LC, Capdevila FB, Nitsch D, De MB, De CP, Moreau Y. A guide to web tools to prioritize candidate genes. *Brief Bioinform* 2011;12:22-32.

97. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. *Human Molecular Genetics* 2010;19:R125-R130.

98. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010;20:110-121.

99. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010;6:e1001154.

100. Adzhubei IA, Schmidt S, Peshkin L *et al.*. A method and server for predicting damaging missense mutations. *Nat Meth* 2010;7:248-249.

101. Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function. *Annual Review of Genomics and Human Genetics* 2006;7:61-80.

102. Vitkup D, Sander C, Church G. The amino-acid mutational spectrum of human genetic disease. *Genome Biology* 2003;4:R72.

103. Kuntzer J, Eggle D, Klostermann S, Burtscher H. Human variation databases. *Database* 2010; Jul 17;2010:baq015. Print 2010. Review

104. Eyre TA, Ducluzeau F, Sneddon TP, Povey S, Bruford EA, Lush MJ. The HUGO Gene Nomenclature Database, 2006 updates. *Nucl Acids Res* 2006;34:D319-D321.

105. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000;25:25-29.

106. Sintchenko V, Coiera E. Developing Decision Support Systems in Clinical Bioinformatics. *Meth Mol Med* 2008;141:331-351.

107. Kawamoto K, Lobach D, Willard H, Ginsburg G. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Medical Informatics and Decision Making* 2009;9:17.

108. Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP. Computational solutions to large-scale data management and analysis. *Nat Rev Genet* 2010;11:647-657.

109. Schatz MC, Langmead B, Salzberg SL. Cloud computing and the DNA data race. *Nat Biotech* 2010;28:691-693.

110. Connecting for Health. Information Governance Toolkit. *https://www igt connectingforhealth nhs uk/* [Accessed 9/12/2010.]

111. Care Quality Commission. Guidance about Compliance Essential standards of quality and safety. 2010.

112. Care Record Development Board. Report of the Care Record Development Board Working Group on Secondary Uses of Patient Information. 2007.

113. Cabinet Office. Making a Difference: Safe and Secure Data Sharing Between Health and Adult Social Care Staff. *Better Regulation Executive Report* 2006.

114. BMA Statement on Connecting for Health. British Medical Association. 2006.

115. Greenhalgh T, Stramer K, Bratan T *et al.*. The Devil's in the Detail. Final report of the independent evaluation of the Summary Care Record and Healthspace programmes. 2010.

116. Greenhalgh T, Stramer K, Bratan T, Byrne E, Russell J, Potts HWW. Adoption and non-adoption of a shared electronic summary record in England: a mixed-method case study. *BMJ* 2010;340:c3111.

117. Liberating the NHS: An Information Revolution. A consultation on proposals. Department of Health. 2010.

118. Firth H and Hurst JA. Oxford- Desk Reference - Clinical Genetics. Oxford University Press 2005.

119. Conrad DF, Keebler JE, Depristo MA *et al.*. Variation in genome-wide mutation rates within and between human families. *Nat Genet* 2011.

120. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), [Accessed 7th Feb 2011].

121. Baird D, Walker J, Thomson AM. The causes and prevention of stillbirths and first week deaths. III. A classification of deaths by clinical cause; the effect of age, parity and length of gestation on death rates by cause. *J Obstet Gynaecol Br Emp* 1954;61:433-448.

122. Donaldson D. On the state of public health: Annual report of the Chief Medical Officer 2009. 2010. Department of Health.

123. Burton H, Alberg C, Stewart A. Heart to Heart: Inherited Cardiovascular Conditions Services - A Needs Assessment and Service Review. *PHG Foundation* 2009;ISBN 978-1-907198-01-4.

124. Moore T, Burton H. Genetic Ophthalmology in focus. *PHG Foundation* 2008.

125. Rahi JS, Cable N. Severe visual impairment and blindness in children in the UK. *Lancet* 2003;362:1359-1365.

126. Ford D, Easton DF, Stratton M *et al.*. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* 1998;62:676-689.

127. Cancer Research UK. CancerStats: Genes and cancer. 2007.

128. Bisgaard ML, Fenger K, Bulow S, Niebuhr E, Mohr J. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. *Hum Mutat* 1994;3:121-125.

129. Roeleveld N, Zielhuis GA, Gabreels F. The prevalence of mental retardation: a critical review of recent literature. *Dev Med Child Neurol* 1997;39:125-132.

130. Hagberg B, Kyllerman M. Epidemiology of mental retardation--a Swedish survey. *Brain Dev* 1983;5:441-449.

131. Department of Health. Valuing people: a new strategy for learning disability for the 21st century. 2001. London: The Stationary Office, 2001.

132. Deafness Research UK. Factsheet: Genetics and Deafness. 2007.

133. Keats BJB. Genetic Hearing Loss. In: Allan IB, Akimichi K, Gordon MS *et al.*., eds. *The Senses: A Comprehensive Reference*. New York: Academic Press; 2008;139-148.

134. Van Camp G, Smith RJH. Hereditary Hearing Loss Homepage. *URL: http://hereditaryhearingloss.org*

135. Simpson JL. Preimplantation genetic diagnosis at 20 years. *Prenat Diagn* 2010;30:682-695.

136. Norbury G, Norbury CJ. DNA analysis: what and when to request? *Archives of Disease in Childhood* 2006;91:357-360.

137. Tarpey PS, Smith R, Pleasance E *et al.*. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* 2009;41:535-543.

138. Daiger SP, Sullivan LS, Bowne SJ *et al.*. Targeted High-Throughput DNA Sequencing for Gene Discovery in Retinitis Pigmentosa. In: Anderson RE, Hollyfield JG, LaVail MM, eds. *Retinal Degenerative Diseases*. 664 ed. Springer New York; 2010;325-331.

139. Shearer AE, DeLuca AP, Hildebrand MS *et al.*. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc Natl Acad Sci USA* 2010; 107:21104-9.

140. Morgan JE, Carr IM, Sheridan E *et al.*. Genetic diagnosis of familial breast cancer using clonal sequencing. *Hum Mutat* 2010;31:484-491.

141. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Hum Mol Genet* 2010;19:R119-R124.

142. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003;33:228-237.

143. Ng SB, Turner EH, Robertson PD *et al.*. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272-276.

144. Ng SB, Bigham AW, Buckingham KJ *et al.*. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790-793.

145. Johnston JJ, Teer JK, Cherukuri PF *et al.*. Massively Parallel Sequencing of Exons on the X Chromosome Identifies RBM10 as the Gene that Causes a Syndromic Form of Cleft Palate. *The American Journal of Human Genetics* 2010;86:743-748.

146. Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet* 2010;42:13-14.

147. Worthey E, Mayer A, Syverson G *et al.*. Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* 2011; 13:255-62.

148. Mujezinovic F, Alfirevic Z. Procedure-related complications of amniocentesis and chorionic villous sampling: a systematic review. *Obstet Gynecol* 2007;110:687-694.

149. Lo YMD, Corbetta N, Chamberlain PF *et al.*. Presence of fetal DNA in maternal plasma and serum. *The Lancet* 1997;350:485-487.

150. Illanes S, Denbow M, Kailasam C, Finning K, Soothill PW. Early detection of cell-free fetal DNA in maternal plasma. *Early Hum Dev* 2007;83:563-566.

151. Lo YMD, Rossa CWK. Prenatal diagnosis: progress through plasma nucleic acids. *Nat Rev Genet* 2007;8:71-77.

152. Lun FMF, Chiu RWK, Allen Chan KC, Yeung Leung T, Kin Lau T, Dennis Lo YM. Microfluidics Digital PCR Reveals a Higher than Expected Fraction of Fetal DNA in Maternal Plasma. *Clin Chem* 2008;54:1664-1672.

153. Lo YMD, Chan KCA, Sun H *et al.*. Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. *Science Translational Medicine* 2010;2:61ra91.

154. Wright CF. Cell-free fetal nucleic acids for non-invasive prenatal diagnosis, Report of the UK expert working group. *PHG Foundation* 2009.

155. Finning KM, Chitty LS. Non-invasive fetal sex determination: Impact on clinical practice. *Sem Fetal & Neonatal Med* 2008;13:69-75.

156. Norbury G, Norbury CJ. Non-invasive prenatal diagnosis of single gene disorders: How close are we. *Sem Fetal & Neonatal Med* 2008;13:76-83.

157. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet* 2010;19:R145-R151.

158. Firth HV, Richards SM, Bevan AP *et al.*. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 2009;84:524-533.

159. Human Genetics Commission. Increasing options, informing choice: A report on preconception genetic testing and screening . 6-4-2011.

160. Bittles AH. Consanguinity and its relevance to clinical genetics. *Clin Genet* 2001;60:89-98.

161. Bittles AH, Black ML. Consanguinity, human evolution, and complex diseases. *Proc Natl Acad Sci USA* 2010;107:1779-1786.

162. Anand KS, Alan HB. Consanguinity and child health. *Paediatrics and Child Health* 2008;18:244-249.

163. Bittles AH, Black ML. The impact of consanguinity on neonatal and infant health. *Early Hum Dev* 2010;86:737-741.

164. Cousens NE, Gaff CL, Metcalfe SA, Delatycki MB. Carrier screening for Beta-thalassaemia: a review of international practice. *Eur J Hum Genet* 2010;18:1077-1083.

165. Samavat A, Modell B. Iranian national thalassaemia screening programme. *BMJ* 2004;329:1134-1137.

166. Kaback M. Population-based genetic screening for reproductive counseling: the Tay-Sachs disease model. *Eur J Paediatr* 2000;159:192S-195.

167. Burton H, Levene S, Alberg C, Stewart A. Tay Sachs Disease carrier screening in the Ashkenazi Jewish population: A needs assessment and review of current services. *PHG Foundation* 2009;ISBN 978-1-907198-00-7.

168. Srinivasan BS, Flannick J, Patterson S *et al.*. A Universal Carrier Test for the Long Tail of Mendelian Disease. *Nature Precedings* 2010; 21:537-51.

169. Dinwiddie D, Bell C, Miller N *et al.*. Carrier Screening of Recessive Genetic Disorders by Exon Capture and Next-Generation Sequencing. *Journal of Biomolecular Techniques* 2010;21:S26.

170. Bell CJ, Dinwiddie DL, Miller NA *et al.*. Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing. *Science Translational Medicine* 2011;3:65ra4.

171. Wald NJ, Rodeck CH, Hackshaw AK, Walters J, Chitty L, Mackinson AM. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and ultrasound Screening Study. *Health Technol Assess* 2003;7.

172. Morris J. Trends in Down's syndrome live births and antenatal diagnoses in England and Wales from 1989 to 2008: analysis of data from the National Down Syndrome Cytogenetic Register. *BMJ* 2009;339.

173. National Screening Committee. NHS Fetal Anomaly Screening Programme - Screening for Down's syndrome: UK NSC Policy recommendations 2007-2010: Model of Best Practice. 2010;Gateway Reference 9674.

174. Chiu RWK, Chan KCA, Gao Y *et al.*. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci USA* 2008;105:20458-20463.

175. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci USA* 2008;105:16266-16271.

176. Chiu RWK, Akolekar R, Zheng YWL *et al.*. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ* 2011;342.

177. Watson M, Lloyd-Purvear M, Mann M, Rinaldo P, Howell R. Newborn Screening: Toward a Uniform Screening Panel and System. *Genet Med* 2006;8:S1-S252.

178. Burton H, Moorthie S. Expanded newborn screening: A review of the evidence. *PHG Foundation* 2010;ISBN 978-1-907198-03-8.

179. Human Genetics Commission. Profiling the newborn: a prospective gene technology? A report from a Joint Working Group of the Human Genetics Commission and the UK National Screening Committee. 2005.

180. WHO. WHO Cancer Factsheet No297. 2011.

181. WHO. Projections of mortality and burden of disease 2004-2030. 2011.

182. Office for National Statistics. Cancer and mortality in UK 2005-2007.

183. Maddams J, Brewster D, Gavin A *et al.*. Cancer prevalence in the United Kingdom: estimates for 2008. *Br J Cancer* 2009;101:541-547.

184. Olivier M, Hussain SP, Caron de Fromentel C, Hainaut P, Harris CC. TP53 mutation spectra and load: a tool for generating hypotheses on the etiology of cancer. *IARC Sci Publ* 2004;247-270.

185. Weir B, Zhao X, Meyerson M. Somatic alterations in the human cancer genome. *Cancer Cell* 2004;6:433-438.

186. Stewart A, Brice P, Burton H, Pharoah PD, Sanderson S, Zimmern RL. Genetics science and technology. *Genetics, Health Care and Public Policy*. Cambridge University Press; 2007;23-64.

187. Strachan T and Read AP. Cancer genetics in Human Molecular Genetics.

188. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. *Briefings in Functional Genomics & Proteomics* 2009;8:174-183.

189. Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. *Nat Rev Cancer* 2006;6:924-935.

190. Mc Dermott U. Genomics and the Continuum of Cancer care. NEJM . 2011. 364:340-350

191. McBride DJ, Orpana AK, Sotiriou C *et al.*. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosom Cancer* 2010;49:1062-1069.

192. Leary RJ, Kinde I, Diehl F *et al.*. Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing. *Science Translational Medicine* 2010;2:20ra14.

193. Balmain A, Barrett JC, Moses H, Renan MJ. How many mutations are required for tumorigenesis? implications from human cancer data. *Mol Carcinog* 1993;7:139-146.

194. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998;396:643-649.

195. Prindle M, Fox E, Loeb L. The Mutator Phenotype in Cancer: Molecular Mechanisms and Targeting Strategies. *Current Drug Targets* 2010;11:1296-1303.

196. Bodmer W. Genetic Instability Is Not a Requirement for Tumor Development. *Cancer Research* 2008;68:3558-3561.

197. Stewart SA, Weinberg RA. Telomeres: Cancer to Human Aging. *Annual Review of Cell and Developmental Biology* 2006;22:531-557.

198. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010;11:685-696.

199. Robison K. Application of second-generation sequencing to cancer genomics. *Briefings in Bioinformatics* 2010;11:524-534.

200. Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Briefings in Bioinformatics* 2010;11:484-498.

201. Cowin PA, Anglesio M, Etemadmoghadam D, Bowtell DDL. Profiling the Cancer Genome. *Annual Review of Genomics and Human Genetics* 2010;11:133-159.

202. Suda K, Tomizawa K, Mitsudomi T. Biological and clinical significance of KRAS mutations in lung cancer: an oncogenic driver that contrasts with EGFR mutation. *Cancer and Metastasis Reviews* 2010;29:49-60.

203. Mardis ER, Ding L, Dooling DJ *et al.*. Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. *New England Journal of Medicine* 2009;361:1058-1066.

204. van der Velden VHJ, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJM. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* 0 AD;17:1013-1034.

205. Stephens PJ, McBride DJ, Lin ML *et al.*. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 2009;462:1005-1010.

206. Department of Health. Improving outcomes: a strategy for cancer. 2011.

207. The Royal College of Pathologists. The future provision of molecular diagnostic services for acquired diseases in the UK. 2010.

208. Humphries SE, Cooper JA, Talmud PJ, Miller GJ. Candidate gene genotypes, along with conventional risk factor assessment, improve estimation of coronary heart disease risk in healthy UK men. *Clin Chem* 2007;53:8-16.

209. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. *Nature* 2009;461:724-726.

210. Agarwal A, Williams GH, Fisher ND. Genetics of human hypertension. *Trends Endocrinol Metab* 2005;16:127-133.

211. Williams RR, Hunt SC, Hopkins PN *et al.*. Genetic basis of familial dyslipidemia and hypertension: 15-year results from Utah. *Am J Hypertens* 1993;6:319S-327S.

212. Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr. *et al.*. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* 2004;291:2204-2211.

213. Wang Q. Molecular genetics of coronary artery disease. *Curr Opin Cardiol* 2005;20:182-188.

214. Talmud PJ, Cooper JA, Palmen J *et al.*. Chromosome 9p21.3 Coronary Heart Disease Locus Genotype and Prospective Risk of CHD in Healthy Middle-Aged Men. *Clin Chem* 2008;54:467-474.

215. Drenos F, Whittaker JC, Humphries SE. The use of meta-analysis risk estimates for candidate genes in combination to predict coronary heart disease risk. *Ann Hum Genet* 2007;71:611-619.

216. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 2008;358:2796-2803.

217. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. *Nat Rev Genet* 2003;4:937-947.

218. Dishy V, Sofowora GG, Xie HG *et al.*. The Effect of Common Polymorphisms of the ß2-Adrenergic Receptor on Agonist-Mediated Vascular Desensitization. *New England Journal of Medicine* 2001;345:1030-1035.

219. Kaye DM. Beta-adrenoceptor genotype influences the response to carvedilol in patients with congestive heart failure. *Pharmacogenetics*. 2003; 13:379-82.

220. Pirmohamed M. *et al.*. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ*. 2004;329:15-9.

221. Adams JU. Pharmacogenomics and personalised medicine. *Nature Education* . 2008.

222. McDonald MG, Rieder MJ, Nakano M, Hsia CK, Rettie AE. CYP4F2 Is a Vitamin K1 Oxidase: An Explanation for Altered Warfarin Dose in Carriers of the V433M Variant. *Molecular Pharmacology* 2009;75:1337-1346.

223. Wang L, McLeod HL, Weinshilboum RM. Genomics and Drug Response. *New England Journal of Medicine* 2011;364:1144-1153.

224. Epstein RS, Moyer TP, Aubert RE *et al.*. Warfarin Genotyping Reduces Hospitalization Rates: Results From the MM-WES (Medco-Mayo Warfarin Effectiveness Study). *J Am Coll Cardiol* 2010.

225. Russmann S, Kaye JA, Jick SS, Jick H. Risk of cholestatic liver disease associated with flucloxacillin and flucloxacillin prescribing habits in the UK: Cohort study using data from the UK General Practice Research Database. *British Journal of Clinical Pharmacology* 2005;60:76-82.

226. Daly AK, Donaldson PT, Bhatnagar P *et al.*. HLA-B[ast]5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat Genet* 2009;41:816-819.

227. Mallal S, Phillips E, Carosi G *et al.*. HLA-B*5701 Screening for Hypersensitivity to Abacavir. *New England Journal of Medicine* 2008;358:568-579.

228. Hughes AR, Spreen WR, Mosteller M *et al.*. Pharmacogenetics of hypersensitivity to abacavir: from PGx hypothesis to confirmation to clinical utility. *Pharmacogenomics J* 2008;8:365-374.

229. Callahan R, Hurvitz S. Human epidermal growth factor receptor-2-positive breast cancer: Current management of early, advanced, and recurrent disease. *Curr Opin Obstet Gynecol* 2011;23:37-43.

230. Wagner MJ. Pharmacogenetics and personal genomes. *Per Med.* 2009. 6:643-652.

231. Swen JJ, Nijenhuis M, de Boer A *et al.*. Pharmacogenetics: From Bench to Byte[mdash] An Update of Guidelines. *Clin Pharmacol Ther* 2011;89:662-673.

232. Van de Velde H, De Rycke M, De Man C *et al.*. The experience of two European preimplantation genetic diagnosis centres on human leukocyte antigen typing. *Human Reproduction* 2009;24:732-740.

233. Lee SJ, Klein J, Haagenson M *et al.*. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood* 2007;110:4576-4583.

234. Howell WM, Carter V, Clark B. The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. *Journal of Clinical Pathology* 2010;63:387-390.

235. Gabriel C, Danzer M, Hackl C *et al.*. Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. *Human Immunology* 2009;70:960-964.

236. Bentley G, Higuchi R, Hoglund B *et al.*. High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens* 2009;74:393-403.

237. Lind C, Ferriola D, Mackiewicz K *et al.*. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. *Human Immunology* 2010;71:1033-1042.

238. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SGE. The IMGT/HLA database. *Nucl Acids Res* 2010; 39:D1171-6.

239. Harris SR, Feil EJ, Holden MT *et al.*. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010;327:469-474.

240. Turner M. Microbe outbreak panics Europe. *Nature* 2011;474:137.

241. Gardy JL, Johnston JC, Ho Sui SJ *et al.*. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011;364:730-739.

242. McGuire A, Fisher R, Cusenza P, Henley D, *et al.*. Confidentiality, privacy and security of genetic and genomic test information in electronic health records: points to consider. *Genetics in Medicine* 2008;10:495-499.

243. Lowrance WW, Collins FS. Ethics. Identifiability in genomic research. *Science* 2007;317:600-602.

244. Lin Z, Owen AS, Altman RB. Genomic research and human subject privacy. *Science* 2004;305:183-184.

245. Pearson TA, Manolio TA. How to Interpret a Genome-wide Association Study. *JAMA* 2008;299:1335-1344.

246. Genetic Information Nondiscrimination Act. *http://www.eeoc.gov/laws/statutes/gina cfm* 2008.

247. Equal Employment Opportunity Commission. Regulations Under the Genetic Information Nondiscrimination Act of 2008. *http://www.federalregister. gov/articles/2010/11/09/2010-28011/regulations-under-the-genetic-information-nondiscrimination-act-of-2008.*

248. Belmont J, McGuire AL. The futility of genomic counselling: essential role of electronic health records. *Genome Medicine* 2009;1:1-3.

249. Genetic Alliance, Private Access Inc. Private Access and Genetic Alliance Announce Strategic Partnership to Accelerate Medical Research While Protecting Privacy. *Press release http:// www geneticalliance org/pr partnership* 2008;August 18.

250. Data Protection Act. *Chapter 29* 1998.

251. Durant v. Financial Services Authority. *EWCA Civ* 1746 2003.

252. Curren L, Boddington P, Gowans H, Hawkins N, Kanellopoulou N, Kaye J. Identifiability, Genomics and UK Data Protection Law. *European Journal of Health Law* 2010;17:329-344.

253. Information Commissioner' Office. Data Protection Technical Guidance Determining what is personal data. *v1.0* 2007.

254. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. *Nature Reviews Genetics* 2008;9:406-411.

255. P3G Consortium, Church G, Heeney C *et al.*. Public Access to Genome-Wide Data: Five Views on Balancing Research with Privacy and Protection. *Plos Genetics* 2009;5:1-4.

256. Knoppers BM, Abdul-Rahman M, Bédard K. Genomic Databases and International Collaboration. *King's Law Journal* 2007;18:291-312.

257. Lakeman P. Preconceptual carrier couple screening for cystic fibrosis and hemoglobinopathies. An ancestry-based offer in a multiethnic society. *Vrije Universiteit* 2008.

258. Lemke T. "A slap in the face". An exploratory study of genetic discrimination in Germany. *Genomics, Society and Policy* 2009;5:22-39.

259. Hudson KL. Genomics, Health Care, and Society. *New England Journal of Medicine* 2011;365:1033-1041.

260. HM Government, Association of British Insurers. Concordat and Moratorium on Genetics and Insurance. *Department of Health* 2005.

261. U.S.Equal Opportunity Commission. Questions and Answers for Small Businesses: EEOC Final Rule on Title II of the Genetic Information Nondiscrimination Act of 2008. *U S Equal Opportunity Commission* 2010.

262. Information Commissioners Office. The Employment Practices Code: Supplemental Advice. *Good Practice Recommendations - Part 4 Information from Genetic Testing* 2004;Paragraph 4.5.

263. Casserley C. Discrimination and genetic pre-disposition - the legal framework. *Written for a meeting at the Human Genetics Commission Jan 2010* 10 A.D.

264. *Sidaway v Board of Governors of the Bethlem Royal Hospital*.1. *All ER* 643. 1985.

265. *Bolam v Frien Hospital Management Committee*. 2. *All ER* 118. 1957.

266. Bredenoord AL, Kroes HY, Cuppen E, Parker M, van Delden J. Disclosure of individual genetic data to research participants: the debate considered. *Trends in Genetics* 2011;1-7.

267. McGuire AL, Caulfield T, Cho MK. Research ethics and the challenge of whole-genome sequencing. *Nature Reviews Genetics* 2008;9:152-156.

268. Report of a working party of the British Society of Human Genetics. Genetic Testing of Children. *British Society of Human Genetics* 2010.

269. Mayer AN, Dimmock DP, Arca MJ *et al.*. A timely arrival for genomic medicine. *Genets Med* 2011; 13:195-6.

270. Choi M, Scholl UI, Ji W *et al.*. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009;106:19096-19101.

271. Henderson M. Family become first to have DNA sequenced for non-medical reasons. *The Times* 2010; http://www.timesonline.co.uk/tol/news/science/genetics/article7100159.ece.

272. Laurie G. Genetic privacy. A Challenge to Medico-Legal Norms. *Cambridge University Press* 2002.

273. Lucassen A, Parker M. Confidentiality and sharing genetic information with relatives. *The Lancet* 2010;375:1507-1509.

274. Human Genetics Commission. Inside Information: Balancing interests in the use of personal genetic data. *A report by the Human Genetics Commission* 2002.

275. Tabor HK, Cho MK. Ethical implications of array comparative genomic hybridization in complex phenotypes: points to consider in research. *Genet Med* 2007;9:626-631.

276. Rios J, Stein E, Hobbs H, Cohen J. Identification by whole-genome resequencing of gene defect responsible for severe hypercholesterolemia. *Human Molecular Genetics* 2010;1-6.

277. *Hardman v Amin*. *P N L R 11* 2001.

278. *Farraj v King's Healthcare NHS Trust . 1 W L R 2139 (CA(Civ Div))* 2010.

279. Pelak K, Shianna K, Ge D, Mala J, *et al.*. The Characterization of Twenty Sequenced Human Genomes. *PloS Genetics* 2010;6:e1001111-10.

280. Organisation for Economic Co-operation and Development. OECD Guidelines on Human Biobanks and Genetic Research Databases. *European Journal of Health Law* 2010;17:191-204.

281. Lawrenz F, Sobotka S. Empirical Analysis of Current Approaches to Incidental Findings. *J Law Med Ethics* 2008; 249-255.

282. Wolf S, Lawrenz F, Nelson C *et al.*. Managing Incidental Findings in Human Subjects Research: Analysis and Recommendations. *J Law Med Ethics* 2008;36:219-211.

283. Pichert G, Mohammed S, Ahn J, Ogilvie C, Izatt L. Unexpected findings in cancer predisposition genes detected by array comparative genomic hybridisation: what are the issues? *J Med Genet* 2011.

284. Booth TC, Jackson A, Wardlaw JM, Taylor SA, Waldman AD. Incidental findings found in "healthy" volunteers during imaging performed for research: current legal and ethical implications. *The British Journal of Radiology* 2010;83:456-465.

285. Stjernschantz Forsberg J, Hansson AG, Eriksson S. Changing perspectives in biobank research: from individual rights to concerns about public health regarding the return of results. *EJHG* 2009;17:1544-1549.

286. Knoppers BM, Joly Y, Simard J, Durocher F. The emergence of an ethical duty to disclose genetic research results: international perspectives. *EJHG* 2006;14:1170-1178.

287. McGuire AL, Lupski JR. Personal genome research: what should the participant be told? *Trends in Genetics* 2010;199-201.

288. Bredenoord A, Onland-Moret C, VanDelden J. Feedback of individual genetic results to research participants: in favor of a qualified disclosure policy. *Human Mutation* 2011;uncorrected accepted article.

289. Berg J, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: Meeting the challenge one bin at a time. *Genetics in Medicine* 2011;13:499-504.

290. Lo D, Chan A, Sun H *et al*.. Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. *Science Translational Medicine* 2011;2:1-13.

291. McGuire A, Diaz C, Wang T, Hilsenbeck S. Social Networkers's Attitudes Toward Direct-to-Consumer Personal Genome Testing. *Am J Bioeth* 2009;9:3-10.

292. Bloss CS, Schork NJ, Topol EJ. Effect of direct-to-consumer genomewide profiling to assess disease risk. *N Engl J Med* 2011;364:524-534.

293. Fuller C, Middendorf L, Benner S *et al*.. The challenges of sequencing by synthesis. *Nature Biotechnology* 2009;27:1013-1023.

294. Holt R, Jones S. The new paradigm of flow cell sequencing. *Genome Res* 2011;18:839-846.

295. The European Parliament, The Council of the European Union. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on *in vitro* diagnostic medical devices. *Official Journal of the European Communities* 1998;L331/1.

296. British Society for Human Genetics. Revision of Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. Public consultation (closure data 15th September 2010). *1 September* 2010.

297. Kaye J. Building a Foundation for Biobanking: The 2009 OECD Guidelines on Human Biobanks and Genetic Research Databases (HBGRDs). *European Journal of Health Law* 2010;17:187-190.

298. Organisation for Economic Co-operation and Development. Guidelines for Quality Assurance in Molecular Genetic Testing. *OECD* 2007;http://www.oecd.org/dataoecd/43/6/38839788.pdf.

299. Mattocks C, Morris M, Matthijs G *et al*.. A standardized framework for the validation and verification of clinical molecular genetic tests. *EJHG* 2010;1-13.

300. Hawkins N. Human Gene Patents and Genetic Testing in Europe: A Reappraisal. *Scripted* 2010;7:453-473.

301. Cook-Deegan R, Heeney C. Patents in Genomics and Human Genetics. *Annals Rev Genom Hum Genet* 2010;11:383-425.

302. United States District Court Southern District of New York. *Association for Molecular Pathology, et. al. v. U.S. Patent and Trademark Office, et. al. Case 1:09-cv-04515-RWS* 2010.

303. Vorhaus D, Conley J. Pigs Fly: Federal Court Invalidates Myriad's Patent Claims. *Genomics Law Report* 2010;30.3.10.

304. Evans JP. Restricting Gene Patents: A Pro-market Agenda. *Genomics Law Report* 2010;30.11.10.

305. Gene Patents and Licensing Practices and Their Impact on Patient Access to Genetic Tests: Report of the Secretary's Advisory Commmitee on Genetics, Health and Society. *Secretary's Advisory Committee on Genetics, Health and Society* 2010.

306. *Monsanto Technology LLC v Cefetra BV* (Case C-428/08). *[2011] F S R 6 Court of European Justice* 2010.

307. Aymé S, Matthijs G, Soini S, ESHG Working Party on Patenting and Licensing. Patenting and licensing in genetic testing. Recommendations of the European Society of Human Genetics. *EJHG* 2008;16:405-411.

308. Soini S, Aymé S, Matthijs G, members of the Public and Professional Policy Committee (PPPC) and Patenting and Licensing Committee (PLC) on behalf of the ESHG. Patenting and licensing in genetic testing: ethical, legal and social issues. *EJHG* 2008;16:S10-S50.

309. Chandrasekharan S, Heaney C, James T, Conover C, Cook-Degan R. Impact of gene patents and licensing practices on access to genetic testing for cystic fibrosis. *Genet Med* 2010;12:S194-S211.

310. Khoury MJ, Evans JP, Burke W. A reality check for personalized medicine. *Nature* 2011;464.

311. Bayer R. Public Health and the AIDS Epidemic. An End to HIV Exceptionalism? *New England Journal of Medicine* 1991;324:1500-1504.

312. Ross LF. Genetic Exceptionalism vs. Paradigm Shift: Lessons from HIV. *Journal of Law, Medicine and Ethics* 2001;29:141-148.

313. Gostin LO, Hodge Jr JC. Genetic Privacy and the Law: An End to Genetics Exceptionalism. *Jurimetrics* 1999;40:21-58.

314. Lazzarini Z. What Lessons Can We Learn from the Exceptionalism Debate (Finally)? *Journal of Law, Medicine and Ethics* 2001;29:149-151.

315. Green J, Hewison J, Bekker H, Bryant L, Cuckle H. Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review. *Health Technology Assessment* 2004;8.

316. Marteau T, French D, Griffin S *et al.*. Effects of communicating DNA-based disease risk estimates on risk-reducing behaviours (Review). *The Cochrane Collaboration* 2010.

317. Collins R, Wright A, Marteau T. Impact of communicating personalized genetic risk information on perceived control over the risk: A systematic review. *Genetics in Medicine* 2011;13:273-277.

318. Bethune M. Time to consider our approach to echogenic intracardiac focus and choroid plexus cysts. *Australian and New Zealand Journal of Obstetrics and Gynaecology* 2008;48:137-141.

319. Getz L, Kirkengen A. Ultrasound screening in pregnancy: advancing technology, soft markers for fetal chromosomal aberrations, and unacknowledged ethical dilemmas. *Social Science and Medicine* 2003;56:2045-2057.

320. Fisher J. First-trimester screening: dealing with the fall-out. *Prenat Diagn* 2011;31:46-49.

321. Kai J, Ulph F, Cullinan T, Qureshi N. Communication of carrier status information following universal newborn screening for sickle cell disorders and cystic fibrosis: qualitative sutdy of experience and practice. *Health Technology Assessment* 2009;13.

322. Ahmed S, Green J, Hewison J. Attitudes towards prenatal diagnosis and termination of pregnancy for thalassaemia in pregnant Pakistani women in the North of England. *Prenat Diagn* 2006;26:248-257.

323. Ahmed S, Atkin K, Hewison J, Green J. The influence of faith and religion and the role of religious and community leaders in prenatal decisions for sickle cell disorders and thalassaemia major. *Prenat Diagn* 2006;26:801-809.

324. Hewison J, Green J, Ahmed S *et al.*. Attitudes to prenatal testing and termination of pregnancy for fetal abnormality: a comparison of white and Pakistani women in the UK. *Prenat Diagn* 2007;27:419-430.

325. President's Council on Bioethics. The changing moral focus of newborn screening. *President's Council on Bioethics* 2008.

326. Tarini B, Singer D, Clark S, Davis M. Parents' Interest in Predictive Genetic Testing for Their Children When a Disease Has No Treatment. *Pediatrics* 2009;124:438.

327. Johnsson L, Helgesson G, Rafnar T *et al.*. Hypothetical and factual willingness to participate in biobank research. *EJHG* 2010;18:1261-1264.

328. Green M, Botkin J. "Genetic Exceptionalism" in Medicine: Clarifying the Differences between Genetic and Nongenetic Tests. *Ann Intern Med* 2003;138:571-575.

329. Caulfield T. Direct-to-consumer testing: if consumers are not anxious, why are policymakers? *Hum Genet* 2011.

330. Heshka J, Palleschi C, Howley H, Wilson B, Wells P. A systematic review of perceived risks, psychological and behavioural impacts of genetic testing. *Genetics in Medicine* 2008;10:19-32.

331. Bloss C, Ornowski L, Silver E *et al.*. Consumer perceptions of direct-to-consumer personalized genomic risk assessments. *Genetics in Medicine* 2010;12:556-566.

332. Genomes Unzipped http://www.genomesunzipped.org

333. Evans E, Meslin E, Marteau T, Caulfield T. Deflating the Genomic Bubble. *Science* 2011;331:861-862.

334. Nicholson D. The Year: NHS Chief Executive's annual report 2008/2009. *available at: http://www dh gov uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_099700 pdf* 2009.

335. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Third Edition. *Oxford University Press,*

Oxford, UK 2005.

336. Raiffa H. Decision analysis: introductory lectures on choices under uncertainty. *Addison-Wesley, Reading, MA, USA* 1968.

337. Briggs A, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. *Oxford University Press, Oxford, UK* 2006.

338. Claxton K, Sculpher M, McCabe C *et al.*. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ* 2005;14:339-347.

339. Drummond M, Griffin A, Tarricone R. Economic evaluation for devices and drugs--same or different? *Value Health* 2009;12:402-404.

340. Carlson JJ, Henrikson NB, Veenstra DL, Ramsey SD. Economic analyses of human genetics services: a systematic review. *Genet Med* 2005;7:519-523.

341. Jarrett J, Mugford M. Genetic health technology and economic evaluation: a critical review. *Appl Health Econ Health Policy* 2006;5:27-35.

342. Rogowski W. Genetic screening by DNA technology: a systematic review of health economic evidence. *Int J Technol Assess Health Care* 2006;22:327-337.

343. Rogowski W. Current impact of gene technology on healthcare. A map of economic assessments. *Health Policy* 2007;80:340-357.

344. Wordsworth S, Buchanan J, Regan R *et al.*. Diagnosing idiopathic learning disability: a cost-effectiveness analysis of microarray technology in the National Health Service of the United Kingdom. *Genomic Med* 2007;1:35-45.

345. Regier DA, Friedman JM, Makela N, Ryan M, Marra CA. Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children. *Clin Genet* 2009;75:514-521.

346. Grosse SD, Wordsworth S, Payne K. Economic methods for valuing the outcomes of genetic testing: beyond cost-effectiveness analysis. *Genet Med* 2008;10:648-654.

347. Yang KY, Caughey AB, Little SE, Cheung MK, Chen LM. A cost-effectiveness analysis of prophylactic surgery versus gynecologic surveillance for women from hereditary non-polyposis colorectal cancer (HNPCC) Families. *Fam Cancer* 2011.

348. Veenstra DL, Roth JA, Garrison LP, Jr., Ramsey SD, Burke W. A formal risk-benefit framework for genomic tests: facilitating the appropriate translation of genomics into clinical practice. *Genet Med* 2010;12:686-693.

349. Grosse SD, Khoury MJ. What is the clinical utility of genetic testing? *Genet Med* 2006;8:448-450.

350. Drmanac R. The advent of personal genome sequencing. *Genet Med* 2011;13:188-190.

351. Mukoyama T. A theory of technology diffusion. *PhD Dissertation, University of Rochester, USA* 2003.

352. Payne K. Fish and chips all round? Regulation of DNA-based genetic diagnostics. *Health Econ* 2009;18:1233-1236.

353. Laboratory Services for Genetics - Report of an expert working group to the NHS Executive and the Human Genetics Commision. Department of Health, editor. 2000.

354. Department of Health. Our Inheritance, Our Future - realising the potential of genetics in the NHS. The Stationary Office. 2003.

355. A report by the House of Lords Select Committee on Science Technology. Genomic Medicine. The Stationery Office. 2009.

356. Department of Health. Equity and Excellence, Liberating the NHS. The Stationery Office. 2010.

357. Scottish Executive Edinburgh. Review of Genetics in relation to healthcare in Scotland. 2006.

358. Chaired by Lord Carter of Coles. Independent review of NHS pathology services. 2006. Department of Health.

359. Chaired by Lord Carter of Coles. Report of the second phase of the review of NHS pathology services in England. 2008. Department of Health.

360. Human Genetics Commission. A Common Framework of Principles for direct-to-consumer genetic testing services. *July* 2010.

361. Chiu RWK, Akolekar R, Zheng YWL *et al.*. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ; 1-11-2011.

362. Gurwitz D, Bregman-Eschet Y. Personal genomics services: whose genomes? *Eur J Hum Genet* 2009;17:883-889.

363. Patch C, Sequeiros J, Cornel MC. Genetic horoscopes: is it all in the genes? Points for regulatory control of direct-to-consumer genetic testing. *Eur J Hum Genet* 2009;17:857-859.

364. Hogarth S, Javitt G, Melzer D. The current landscape for direct-to-consumer genetic testing: legal, ethical, and policy issues. *Annu Rev Genomics Hum Genet* 2008;9:161-182.

365. Wright CF, Gregory-Jones S. Size of the direct-to-consumer genomic testing market. *Genet Med* 2010;12:594.

366. Wang X, Pankratz VS, Fredericksen Z *et al.*. Common variants associated with breast cancer in genome-wide association studies are modifiers of breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Human Molecular Genetics* 2010;19:2886-2897.

367. UK Genetic Testing Network. Review of commissioning arrangements for genetic services and strategic recommendations. *March* 2011. www.ukgtn.nhs.uk

14 Appendices

Appendix 1: Steering group

Dr James Brenton	Group Leader and Honorary Consultant in Medical Oncology, Cambridge Research UK, Cambridge Research Institute, Cambridge
Professor Sir John Burn	Professor of Clinical Genetics, Institute of Human Genetics, Newcastle University
Dr Peter Campbell	Group Leader, Cancer Genome Project, Wellcome Trust Sanger Institute, Cambridge
Mr Andrew Devereau	Director, National Genetics Reference Laboratories, St Mary's Hospital, Manchester
Dr Paul Flicek	Team Leader, Vertebrate Genomics, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge
Dr Timothy Hubbard	Head of Informatics, Wellcome Trust Sanger Institute, Cambridge
Dr Kathleen Liddell	Senior Lecturer, Faculty of Law, University of Cambridge
Professor Anneke Lucassen	Professor of Clinical Genetics, Wessex Clinical Genetics Service & University of Southampton, Princess Anne Hospital, Southampton
Mr Chris Mattocks	Senior Clinical Research Scientist, National Genetics Reference Laboratory, Wessex)
Dr Jo Whittaker	Director, Genetics Laboratories & Co-Director Investigative Sciences Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge

PHG Foundation members	
Chair: Dr Hilary Burton	Director
Ms Alison Hall	Project Manager (Law)
Dr Anna Pokorska-Bocci	Project Coordinator
Dr Caroline Wright	Programme Associate (formerly Head of Science 2008-2010)

Appendix 2: Workshop 1 participants

Dr Stephen Abbs	Head of DNA Laboratory, GSTS Pathology, Guy's Hospital, London
Dr James Brenton	Group Leader and Honorary Consultant in Medical Oncology, Cambridge Research UK, Cambridge Research Institute, Cambridge
Professor Anthony Brookes	Chair of Genomics & Bioinformatics, Department of Genetics, Leicester University
Professor Sir John Burn	Professor of Clinical Genetics, Institute of Human Genetics, Newcastle University
Dr Rachel Butler	Consultant Clinical Scientist, All Wales Molecular Genetics Laboratory, University Hospital of Wales, Cardiff
Dr Hilary Burton	Director, PHG Foundation, Cambridge
Dr Peter Campbell	Group Leader, Cancer Genome Project, Wellcome Trust Sanger Institute, Cambridge
Professor Nick Cross	Professor of Human Genetics, Wessex Regional Genetics Laboratory, Southampton University
Mr Andrew Devereau	Director, National Genetics Reference Laboratories, St Mary's Hospital, Manchester
Dr Wybo Dondorp	Senior Researcher & Ethicist, Department of Health Ethics & Society, Maastricht University, The Netherlands
Dr Helen Firth	Consultant Clinical Geneticist, Department of Medical Genetics, Addenbrooke's Hospital, Cambridge
Dr Paul Flicek	Team Leader, Vertebrate Genomics, EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge
Dr Richard Fordham	Senior Lecturer, Medical School, University of East Anglia, Norwich, Norfolk
Ms Alison Hall	Project Manager (Law & Policy), PHG Foundation, Cambridge
Dr Fiona Hemsley	Head of Strategic Projects, Cancer Research UK, London
Professor Jenny Hewison	Professor of the Psychology of Healthcare, Leeds Institute of Health Sciences, Leeds University
Dr Timothy Hubbard	Head of Informatics, Wellcome Trust Sanger Institute, Cambridge
Dr Mark Kroese	Consultant in Public Health Medicine, Peterborough Primary Care Trust, Peterborough, Cambridgeshire

Professor Anneke Lucassen	Professor of Clinical Genetics, Wessex Clinical Genetics Service & University of Southampton, Princess Anne Hospital, Southampton
Dr Anna Middleton	Genetic Counsellor, & ethics researcher, Wellcome Trust Sanger Institute, Cambridge
Dr Sowmiya Moorthie	Project Coordinator, PHG Foundation, Cambridge
Mrs Gail Norbury	Commissioning & Governance Director for Genetics Laboratories, Guy's & St Thomas' NHS Foundation Trust, Guy's Hospital, London
Dr Christine Patch	Consultant Genetic Counsellor/Manager, Guys & St Thomas' NHS Foundation Trust, Guys' Hospital, London
Dr Anna Pokorska-Bocci	Project Coordinator, PHG Foundation, Cambridge
Dr Imran Rafi	GP and Senior Lecturer in Primary Care Education, Department of Community Health Care Sciences, St George's University of London
Dr Hilary Robinson	Consultant Paediatrician with Special Expertise in Cardiology, Royal United Hospital, Bath NHS Trust
Dr Simon Sanderson	Consultant in Public Health Medicine & Associate, PHG Foundation, Cambridge
Dr Rosalind Skinner	Senior fellow, PHG Foundation, Cambridge
Dr William Spooner	Founder and Technical Director, Eagle Genomics, Cambridge
Dr Steve Sturdy	Deputy Director and Senior Lecturer, ESRC Genomics Policy and Research Forum, University of Edinburgh
Dr Jenny Taylor	Programme Director, Oxford Biomedical Research Centre Genetics Theme, Wellcome Trust Centre for Human Genetics, Oxford
Dr Susan Wallace	Lecturer, Department of Health Sciences, University of Leicester
Dr Sarah Wordsworth	Senior Research Officer, Health Economics Research Centre, Department of Public Health, Oxford University
Dr Jo Whittaker	Director, Genetics Laboratories & Co-Director Investigative Sciences Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge
Dr Caroline Wright	Project Associate, PHG Foundation, Cambridge, and Senior Scientific Manager, Wellcome Trust Sanger Institute, Cambridge
Dr Ron Zimmern	Chairman, PHG Foundation, Cambridge

Appendix 3: Workshop 2 Participants

Professor Tim Aitman	Professor of Clinical and Molecular Genetics, Institute of Clinical Science, Imperial College, London
Dr Graham Bell	Lead Specialist, Stratified Medicine, Technology Strategy Board
Dr Ewan Birney	Senior Scientist, European Bioinformatics Institute
Professor Sir Tom Blundell	Professor of Biochemistry, University of Cambridge
Dr Helen Bodmer	Head of Innovation, Office for Life Sciences, Department for Business, Innovation & Skills
Dr Philippa Brice	Head of Knowledge and Communications, PHG Foundation, Cambridge
Dr Hilary Burton	Director, PHG Foundation, Cambridge
Dr Nick Crabb	Associate Director, Diagnostics Assessment Programme, Centre for Health Technology Evaluation, NICE
Professor Peter Farndon	Professor of Clinical Genetics, Director of NHS National Genetics Education and Development Centre
Dr Robert Frost	FORUM Manager, Academy of Medical Sciences
Professor Peter Furness	President, The Royal College of Pathologists
Ms Alison Hall	Project Manager (Law & Policy), PHG Foundation, Cambridge
Dr Fiona Hemsley	Head of Strategic Projects, Cancer Research UK
Dr Richard Henfrey	Director Marketing, Europe Illumina UK Ltd
Dr Timothy Hubbard	Head of Informatics, Wellcome Trust Sanger Institute, Cambridge
Dr Dragana Josifova	Consultant Clinical Geneticist, Guy's and St Thomas' Hospital, London
Mr Alastair Kent	Director, Genetic Alliance
Professor Anneke Lucassen	Professor of Clinical Genetics, Wessex Clinical Genetics Service & University of Southampton, Princess Anne Hospital
Dr David Lynn	Head of Strategic Planning and Policy, Wellcome Trust
Mr Chris Mattocks	Senior Clinical Research Scientist, National Genetics Reference Laboratory, Wessex)
Professor Jonathan Montgomery	Professor of Health Care Law, University of Southampton

Dr Sowmiya Moorthie	Project Coordinator, PHG Foundation, Cambridge
Dr Anna Pokorska-Bocci	Project Coordinator, PHG Foundation, Cambridge
Dr Lucy Raymond	Reader in Neurogenetics, Department of Medical Genetics, University of Cambridge
Dr Beverly Searle	CEO, Unique Rare Chromosome Disorder Support Group
Dr Rosalind Skinner	Senior Fellow , PHG Foundation, Cambridge
Ms Beth Thompson	Policy Officer, Wellcome Trust
Dr Caroline Thureau	Regional Market Manager UK and South of Europe, Diagnostic Illumina UK Ltd
Professor John Todd	Professor in Department of Medical Genetics, University of Cambridge
Dr Chris Tyler	Director, Centre for Science and Policy, University of Cambridge
Dr Caroline Wright	Project Associate, PHG Foundation, Cambridge, and Senior Scientific Manager, Wellcome Trust Sanger Institute, Cambridge
Dr Ron Zimmern	Chairman, PHG Foundation, Cambridge

About the PHG Foundation

The PHG Foundation is an independent, non-profit organisation based in Cambridge, UK, with the mission making science work for health. We identify the best opportunities for 21st century genomic and biomedical science to improve health and tackle disease in ways that are rapid and effective, equitable and responsible. This entails work to promote the prompt translation of scientific innovation into medical and public health policy and practice.

We provide knowledge, evidence and ideas to stimulate and direct well-informed debate on the potential and pitfalls of key genomic and biomedical developments, and to inform and educate stakeholders – policy makers, health professionals, patients and the public. We also provide expert research, analysis, health services planning and consultancy services for governments, health systems, and other non-profit organisations.

PHG Foundation
2 Worts Causeway
Cambridge CB1 8RN

Tel: +44 (0) 1223 740200
Fax: +44 (0) 1223 740892

ISBN 978-1-907198-08-3

phg foundation
making science work for health

www.phgfoundation.org