Data sharing to support UK clinical genetics and genomics services

Workshop report

December 2015

Authors:

Sobia Raza, Alison Hall, Chris Rands, Sandi Deans, Dominic McMullan, Mark Kroese

Acknowledgements:

NB: URLs in this report were correct as at 30 November 2015

This report can be downloaded from our websites:
www.phgfoundation.org
www.acgs.uk.com

Comments to:
sobia.raza@phgfoundation.org
bshg@bshg.org.uk
#PHGACGS

How to reference this report:
Data sharing to support UK clinical genetics & genomics services
PHG Foundation (2015)
ISBN 978-1-907198-20-5

Contents

Executive summary

Introduction

Access to high quality data is fundamental to delivering clinical genetics / genomics diagnostic services

Diagnosing a patient with a rare genetic disease is often complex and time-consuming as it involves accumulating data from multiple sources to make a robust assessment of the cause of a patient's disease. This entails a careful assessment of the patient's genetic / genomic data, in combination with phenotypic and family history information. It also requires access to as much pre-existing genetic / genomic and phenotypic data relevant to the clinical case as possible; in rare diseases this includes data from unrelated patients with the same or similar disorders. This aggregated data enables expert interpretations of the clinical significance of genomic variants which are essential for delivering high quality and safe clinical genetics / genomics testing.

There are challenges to sharing and accessing the patient data required by NHS clinical genetics / genomics laboratories to deliver safe and effective diagnostic services

Sharing patient information - particularly in the context of genetic and genomic data - is complex, since some of this data *e.g.* very rare genetic variants, could be seen as potentially identifiable information and thus subject to additional data protection. Variations between NHS Trusts also arise due to the various laws and regulations relevant to the sharing of genetic / genomic data being applied in different ways, through varying interpretations being made by local Caldicott Guardians. These variations result in a lack of consistent data sharing practices across laboratories, as highlighted by our survey (**see** Page 12**). In addition to these legal and regulatory challenges, there are logistical barriers and other disincentives to sharing: the curation of data requires time and resources, and currently there is no specific designated database to meet the needs of the NHS clinical genetics / genomics services.

Joint PHG Foundation and ACGS workshop on data sharing

A multidisciplinary approach is needed to consider the challenges to data sharing and to identify priorities for policy development

PHG Foundation, an independent health policy organisation, and the Association for Clinical Genetic Science (ACGS), the professional group for clinical genetics scientists, jointly organised and hosted a workshop on 23rd June 2015 to discuss the most pressing challenges around data sharing. Resulting from the workshop and our supplementary analysis is a set of recommendations to inform and facilitate improvements to data sharing, particularly within the NHS.

The meeting convened 60 stakeholders including clinicians, laboratory scientists, bioinformaticians, regulators and policy makers. Six invited speakers, including the National Data Guardian for health and care, provided clinical, laboratory, and legal / regulatory perspectives on the subject. Three parallel multidisciplinary focus groups deliberated the clinical need and case for data sharing, the technical and practical routes for enabling data sharing and the legal avenues to enable responsible data sharing. As part of our analysis we surveyed clinical genetics laboratories in advance of the workshop to better understand current data sharing practices and the perceived impediments to sharing.

Outcomes and key recommendations

Current arrangements for sharing genetic / genomic data within the NHS are unsatisfactory

The workshop discussions and pre-workshop survey underscored that laboratories require access to data generated elsewhere for both the direct clinical care of their patients and for improving the quality and safety of their services. Legal uncertainty, lack of practical support for data curation, and the absence of a designated infrastructure into which NHS laboratories can deposit data, are all impeding the sharing of data. Inconsistent sharing practices are causing significant differences in patient care and are compromising quality and safety. The most serious consequences of sub-optimal data sharing are, variations in the quality of testing services, potential misdiagnosis resulting in inappropriate patient care, and delays in patient diagnosis.

There is an urgent need for national agreement and leadership

There is no single solution to achieving greater data sharing. Improvements will require a multi-layered approach which strikes a balance between facilitating data sharing whilst securing proportionate safeguards against data breach or other types of harm to patients.

Amongst the most crucial developments needed to improve, optimise and transform existing practice are:

» **Strong leadership** by the multiple responsible health organisations to demonstrate the benefits associated with data sharing, as well as the burdens and risks associated with sharing and not sharing, and to fully exploit new opportunities for building genomic capacity and services in the UK.

» **National agreement to optimise sharing** of (genetic / genomic) data within the NHS. All those involved need to develop a common understanding of the legitimacy of data sharing. We support a responsible and proportionate approach that takes account of a set of common principles to demonstrate trustworthiness. These include transparency about the purpose, risks, benefits and safeguards involved.

» **Standardised operational processes** to achieve robust, effective and consistent sharing practices. There needs to be a designated sustainable database or mechanism for sharing data across NHS clinical genetics / genomics services with clear governance, oversight, standards and safeguards. A nationally accessible resource is integral to improving clinical outcomes and to support the effective delivery of clinical genetics / genomics services and as such should be long-term and sustainably resourced.

Improvements in data sharing can be achieved by building and capitalising on the momentum for harnessing advances in data for health

In addition to national agreement and leadership, further momentum for achieving greater data sharing can be created by: building public and patient trust (Recommendation 5, Recommendation 10); communicating criteria for national data sharing systems (Recommendation 1, Recommendation 2, Recommendation 3); future proofing systems and ensuring their sustainability and suitability for the intended purpose (Recommendation 4, Recommendation 6, Recommendation 7); encouraging compliance with data sharing through mandates and monitoring processes (Recommendation 8); through guidance for staff (Recommendation 5); and developing a balanced assessment of the risks / benefits of sharing and not sharing, through a collaborative multi-agency approach (Recommendation 11).

A number of relevant policy development initiatives are underway. Moreover there is an ongoing national strategic ambition to use better data to improve health, transform quality, and reduce costs of health services. Together these developments present an opportunity to achieve greater genetic / genomic data sharing and in doing so enable NHS clinical genetics services to deliver the best possible care for patients, and for the UK to position itself at the forefront of genomic medicine.

1. Introduction

1.1 Joint PHG Foundation / ACGS workshop on data sharing to support clinical genetics / genomics services

On the 23 June 2015 the PHG Foundation and the Association for Clinical Genetic Science (ACGS) jointly hosted a workshop to discuss the most pressing challenges around sharing data to support NHS clinical genetics and genomics services. This report sets out the background to this meeting, summarises the discussions, and draws out the key recommendations.

Experts and stakeholders at the meeting came from diverse disciplines related to genetics / genomics and data sharing, including clinical, laboratory, bioinformatics and managerial representatives from 19 of the 23 regional clinical genetics laboratories across England, Scotland and Wales. Also in attendance were four representatives from the specialist laboratories. This was in addition to legal and health policy experts and government officials. Meeting delegates and their affiliations are listed in Appendix 8.1.

Invited speakers highlighted the challenges to data sharing before the meeting broke into focus group discussions. Presentations were delivered by:

» **Dr Ann Dalton** - Chair ACGS and Head of the Sheffield Diagnostic Genetics Service

» **Dr Helen Firth** - Consultant Clinical Geneticist, Cambridge University Hospitals NHS Foundation Trust and Honorary Faculty Member Wellcome Trust Sanger Institute

» **Professor Sian Ellard** - Head of Molecular Genetics and Consultant Clinical Scientist, Royal Devon & Exeter NHS Foundation Trust and Professor of Molecular Genetics and Genomic Medicine, University of Exeter Medical School

» **Dominic McMullan** - Chair ACGS Science-Subcommittee and Consultant Clinical Scientist, West Midlands Regional Genetics Laboratory

» **Dr Jon Fistein** - Ethics and Policy Lead on Clinical Data and Research for the Medical Research Council (MRC)

» **Dame Fiona Caldicott** - National Data Guardian for health and care, Independent Information Governance Oversight Panel

Three parallel focus group discussions were held to explore ways to improve data sharing to support the effective development and delivery of genetic and genomic services. Overall representation within the focus groups was weighted towards either clinical genetics experience, or laboratory and bioinformatics experience, or legal and regulatory experience, but the groups were intentionally kept multidisciplinary in order to inform discussions. An overlapping series of questions were posed to the three groups to obtain broad professional perspectives on the considerations to data sharing (Figure 1). These discussions are summarised in section's 2.5, 2.6 and 3.6 of the report and covered:

» The clinical need and case for data sharing

» The technical and practical routes for enabling data sharing

» The legal and regulatory avenues to enable responsible data sharing

Figure 1: The themes of questions considered by the focus groups. A complete list of the questions is given in Appendix 8.3

Group 1 (Clinical focus)	**Group 2** (Laboratory / technical focus)	**Group 3** (Legal / regulatory focus)
Data types to share	Data types to share	Data types to share
Consent & related matters	Curation & deposition	Consent & related matters
Safeguards	Safeguards	Safeguards
Data access	Data access	Data access
DB management	DB management	—
Facilitating data sharing	Facilitating data sharing	Facilitating data sharing

1.2   The multiple and complex considerations to data sharing

The analysis of genomic and genetic data is becoming an increasingly important part of NHS clinical services, and is a key component for the diagnosis and management of patients with rare diseases. However, in most cases, the clinical utility of genetic and genomic data only materialises after the aggregation of data, when individual patient data is interpreted with data from other patients and the wider population. Combining genomic, phenotypic and contextual information from a variety of individuals, facilitates the identification and characterisation of genetic variants, allowing the determination of the underlying causes of genetic diseases and enabling genetic diagnoses [1].

While the themes for data sharing in a clinical genetics and genomics context have been examined before, often in the framework of international initiatives *e.g.* the Global Alliance for Genomics and Health [2], or the Human Variome Project [3], there remain unresolved practical, technical and legal considerations, particularly in the context of health services. This is because individual health services operate within different jurisdictions and have varying levels of infrastructure and resources to share data. The novelty of this work was to analyse and evaluate these considerations within the context of the NHS, using a deliberately multidisciplinary approach which brought together clinicians, laboratory scientists, regulators and policy makers to identify priorities for future policy development. This workshop and our supplementary in-house analysis formed the basis for developing the recommendations set out in this report. It is expected that this report will stimulate further relevant work and it will be used to help draft ACGS best practice guidelines on the sharing of data to support NHS clinical genetics and genomics services.

What are the benefits of data sharing?

In general, the sharing of relevant genetic / genomic and associated data amongst clinicians and laboratory scientists within the NHS (with appropriate safeguards) has the following potential benefits:

» **Increased, improved and faster diagnosis for patients**

  Confident genetic diagnoses can usually only be offered to patients once their condition has been observed in other patients and the genetic variants cross-validated

» **Improved and more tailored treatments for patients**

  A new genetic diagnosis may suggest novel treatment options for patients based on functional information about the genetic element that the disease appears to disrupt [4]

» **Cost and efficiency savings for the NHS**

  By connecting genetic and genomic data across NHS laboratories, clinical insights obtained in one laboratory can be accessed by another, thereby saving time and resource in investigating findings

Why are current data sharing practices variable?

Despite the clinical benefits, data sharing is not universally conducted nor standardised across clinical genetics laboratories for several reasons:

» **Uncertainty about the legality of data sharing**

Data should be shared in compliance with the law, but the regulatory landscape surrounding the sharing of patient data is complex, multi-layered, and ambiguous

» **Lack of suitable and sustainable infrastructure for sharing data**

Generally current databases have not been set-up specifically with the needs of NHS clinical genetics services in mind, and there are logistical barriers to data sharing (time taken; challenges with curation). Moreover, the long term sustainability of existing infrastructure is not necessarily guaranteed

» **Disincentives to share data**

Competition within the health system can lead to dis-incentives or incentives to share data

Addressing the data sharing challenge

To realise the potential benefits of sharing data for clinical genetics and genomics services, and to overcome the issues surrounding the sharing of data, there is a need to establish:

» **What specific types of data, if shared now, can enable clinical benefits to be realised?**

There are a range of data that could be shared to improve and support genetic and genomic services, including: genetic data (*i.e.* different categories of genetic variants and complete sequence data), accompanying phenotypic and clinical data and other metadata (*i.e.* quality metrics or methodological information). It is important to establish exactly which elements of data are critical to share for the effective delivery of services

» **How should this data be shared to optimise ease of use for laboratory scientists and clinicians?**

Data could be shared within existing infrastructures, such as the NHS portal of the DECIPHER database that stores genetic variants implicated in rare diseases and developmental disorders (section 2.2). Alternatively, existing databases could be modified or new tailored infrastructure constructed. The relative merits and complexities of the various options require assessment

» **How can data be shared in a legally permissible and ethical manner?**

To ensure data is shared in accordance with the Data Protection Act and other relevant laws, there needs to be clarity on the types of data and the extent and purpose of the sharing. Issues to consider include data anonymisation and patient consent

1.3   The current landscape of data sharing by laboratories

To better understand the current landscape of genetic / genomic and related data sharing, NHS clinical genetics laboratories were surveyed in advance of the meeting. Thirteen of the 23 NHS regional clinical genetics laboratories and two of the eight specialist laboratories completed the survey, with representation from across England, Scotland and Wales. Whilst it is accepted that the results of this survey are not comprehensive and may not be representative of laboratories who did not participate in the survey, the responses did reveal some broad trends:

» 93% of all surveyed laboratories (14 / 15) refer to data generated elsewhere to deliver their clinical genetics services. The remaining one laboratory refers to such data for a sizeable minority of cases

» 87% (13 / 15) of laboratories surveyed currently deposit some data to databases of genetic variants

» The 13 surveyed laboratories who do deposit data were asked questions on how, where, and what data are shared:

Where is variant data deposited?

• 85% (11 / 13) have deposited data into the DECIPHER (DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources) database, often in addition to other clinical, locus-specific, and disorder-specific variant databases

Figure 2: Databases that the 13 laboratories reported depositing data into

Into which databases are data deposited by your laboratory?
(tick any that apply)
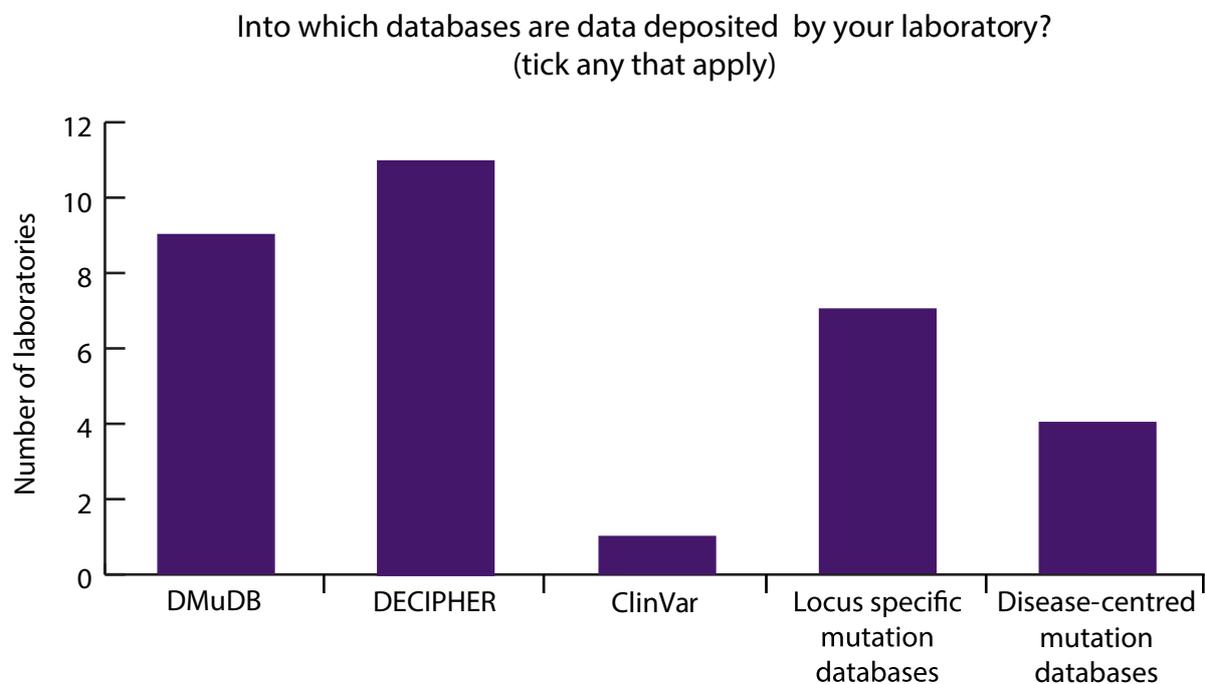
What types of variant data are deposited?

- 100% (13 / 13) deposit pathogenic variants and variants of uncertain clinical significance (VUS) and 31% (4 / 13) deposit non-pathogenic variants

- 69% (9 / 13) deposit copy number variants (CNVs) and 39% (5 / 13) deposit single nucleotide variants (SNVs)

How is the variant data deposited?

- 92% (12 / 13) deposit some data manually and 54% (7 / 13) deposit some data using automated scripts or via bulk uploads

- 46% (6 / 13) do not seek specific patient consent prior to depositing genetic variant data into databases, while the remainder seek consent in at least some instances

» Survey participants were asked to consider which factors were impeding the deposition of data into databases and the extent to which this was the case. Fourteen of the fifteen laboratories surveyed considered resource and technical limitations to be a strongly or moderately impeding factor and ten laboratories considered uncertainty around the legal considerations as a strongly or moderately impeding factor (Figure 3)

Figure 3: Perceived impediments to data sharing from representatives across 15 different NHS clinical genetics laboratories

What factors, if any, are impeding the deposition of data?
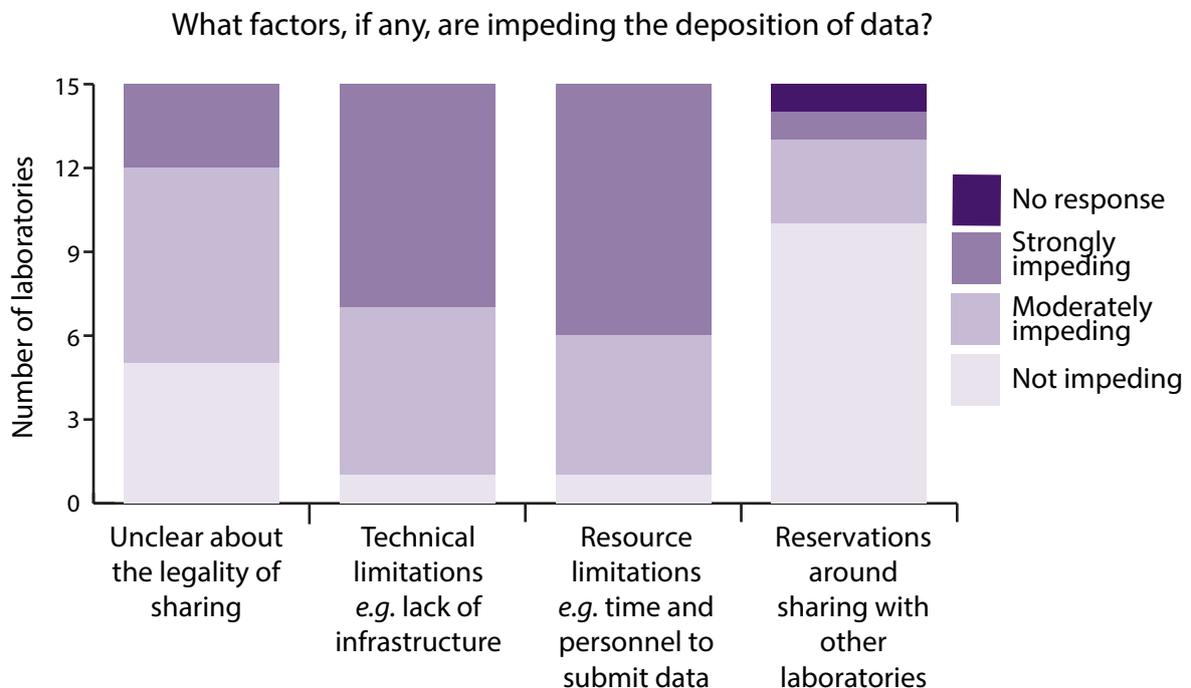
2. Facilitating data sharing: clinical necessity and technical routes

2.1 Introduction: the clinical case for data sharing and systems for enabling sharing

Accurate determination of the genetic cause of a rare disease depends on knowledge of the variants in a patient's genome that could underlie their condition. This necessitates the comparison of a patient's genetic / genomic data with existing genetic / genomic and clinical data on the wider population and other patients with the same or similar illnesses. It is therefore essential for clinicians and scientists to have routine access to all existing information on the genetic variants and associated clinical characteristics relevant to their patient's disorder, in order to make the most informed interpretation and diagnosis.

Accurate diagnosis also depends on the quality and completeness of the underlying knowledge base. Every human has millions of variants in their genome, ranging from those that are common in the population (and unlikely to be the cause of a rare genetic disease), those that are very rare and found in just a small subset of individuals and sometimes those that are unique. When interpreting their clinical significance, variants are categorised into a gradient of groups from those that are definitively disease causing (*i.e.* pathogenic) to those that are not disease causing (*i.e.* benign). Despite efforts to standardise this classification of variants, the interpretation of the importance of the same variant by multiple clinical laboratories can differ, resulting in some variants being classified as pathogenic in one database and benign / not pathogenic in another [5]. Since at least one interpretation must be wrong, this disparity in interpretation could lead to an incorrect diagnosis and therefore an inappropriate medical intervention. Pooling information on variants is integral to refining variant interpretations, and improving the robustness of diagnostic decisions [6], particularly given that the clinical significance of most rare variants is unknown.

Data sharing – what, how, with whom and where?

The types of data that are valuable to share for interpretative and diagnostic purposes range from the genetic data itself to associated clinical and phenotypic and metadata. From a clinical perspective the greater the availability of these data, the more informed the interpretation. However, there is a chance that these data, in combination and sometimes arguably even in isolation (if the data is exceptionally unique), could be used to uncover the identity of the patient from whom the data are derived. So a careful consideration of the clinical justification for sharing these different data types within or beyond the NHS is warranted, to balance competing interests which regard individual privacy claims as absolute.

Although a number of databases for sharing and storing genetic data exist [1, 7, 8, 9], many emerge from research endeavours and currently there is database resource dedicated to the sharing of data across NHS clinical genetics services. In 2005 the Diagnostic Mutation Database (DMuDB) was specifically established as a practical and secure repository of diagnostic variant data for this purpose, with funding from the Department of Health through the National Genome Reference Laboratories. However funding has ceased (in March 2012) and since 2014 the database can no longer accept new data submissions. In the absence of a sustainable strategy for aggregating and sharing data, the logistical barriers to sharing data (time, resources, and insufficient infrastructure) can, for many laboratories, outweigh the benefits, given the risk that time and resource expended in uploading data may be in vain.

2.2 Presentation: DECIPHER - enabling diagnosis and discovery through responsible data sharing

Presenter: Dr Helen Firth, Consultant Clinical Geneticist, Cambridge University Hospitals NHS Foundation Trust and Honorary Faculty Member Wellcome Trust Sanger Institute

Synopsis: This presentation described the operation and functionality of the DECIPHER database [10] and its approach to enabling flexible data sharing at public or restricted-access levels as a means of facilitating genetic diagnosis whilst respecting patient preferences on data access.

Background: The DECIPHER (DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources) is an online database of genetic variation with associated phenotypes, designed to facilitate the identification and interpretation of phenotype-linked pathogenic genetic variation in rare disorders. The DECIPHER project was initiated in 2004 to help with the interpretation of novel genetic variants of uncertain clinical significance because of the difficulties in determining if these are pathogenic or benign when analysed in isolation.

Key features and functionality of DECIPHER: An expansive and dynamic database for clinical genetic variants and phenotypic data

DECIPHER is web-based, international and dynamic in that it provides 'real time' representation of genomic variation, *i.e.* with the most current genomic context and interpretation. It is designed to be comprehensive and scalable, and can be accessed via a flexible online browser. The database covers the nuclear and mitochondrial human genomes, cataloguing single nucleotide and copy number variation linked to phenotypic data. The latter are recorded using standardised nomenclature, the Human Phenotype Ontology (HPO) terms, to facilitate association analyses between phenotypes and genomic variants.

DECIPHER includes data from ~14,000 patients and their parents recruited from 23 UK NHS Regional Genetic Services as part of the Deciphering Developmental Disorders (DDD) project (www.ddduk.org), a study of children with severe undiagnosed developmental disorders. Sharing of DDD data within DECIPHER has enabled the identification of unrelated patients with overlapping clinical features, or identical genetic variants, thereby strengthening analytical confidence on the significance of variants, and providing a new diagnosis for some patients.

DECIPHER has an interactive and customisable user-interface offering different options for viewing rare patient variation in both the phenotypic or genotypic context. Relevant information is provided on each variant in order to support analysis and interpretation, including information on the consequence of the variant on the protein sequence, and a genome browser interface showing the precise location of the variant within a gene and chromosome. Functionality for displaying the conservation of genomic sequences across different species was added in July 2015.

A multi-tiered model for sharing genetic variants based around patient consent

DECIPHER enables the flexible sharing of data through a tiered approach. It is possible for submitting centres to retain their patient data as private (*i.e.* only visible to authorised users from the same centre), or to share - but only with preferred partners for example within a consortia (*e.g.* the DDD project consortium or a recent embryonic NHS group). Patient consent is not sought for data shared within the private network, on the basis that the data is (pseudo-)anonymised and only users with legitimate and privileged access can share or access this information. However explicit written consent is taken for any data shared publically through DECIPHER. A key principle of DECIPHER is proportionality, whereby only the minimum necessary amount of data is shared. In general, patients are very keen to share their data knowing that this could potentially aid their diagnosis.

Security features for protecting patient identity

DECIPHER is designed with privacy and security features:

» To provide a high level of anonymity, the global location of the submitting laboratory or patient is not revealed to non-authorised users, although the data submitters can be contacted via a request process

» DECIPHER is regularly tested with penetration assessments to ensure that it is robust against hacking

» Unique patient identifier and codes are used for added security so that it is not possible for those external to DECIPHER to identify patients, as doing so would require breaching the link between the patient identifiers and personal patient information held securely elsewhere

Collaborating partners and future planned developments

The project has been successful in catalysing the discovery of rare diseases with over 700 publications in the last four years using DECIPHER data. It is also a key partner with the Matchmaker Exchange programme of the Global Alliance initiative (GA4GH) which is working towards a federated platform to facilitate the matching of cases with similar phenotypic and genotypic profiles even across data held in different databases. DECIPHER is collaborating with ClinVar, a USA based National Institutes of Health (NIH) funded resource for genomic variant data, and will in the future integrate with data gathered by Genomics England as part of the 100,000 Genomes Project. DECIPHER plans to release an automated application programming interface for submitting data, which would increase the speed and efficiency of data deposition.

2.3 Presentation: towards automatic variant submission at the point of clinical reporting

Speaker: Professor Sian Ellard, Consultant Clinical Scientist, Royal Devon & Exeter NHS Foundation Trust and Professor of Molecular Genetics and Genomic Medicine, University of Exeter Medical School

Synopsis: This presentation described a laboratory's perspective on the importance of access to variant interpretations and the practical steps that could be taken to facilitate data sharing (*e.g.* through automated deposition into databases) and to improve the reliability of data that is shared.

Background: Sharing genetic variants can facilitate classification of pathogenicity and disease relevance, which in turn can aid genetic diagnostics. However, it is often difficult to accurately assign variants to different categories based on their functional and disease effects. Furthermore, there are technical challenges to overcome in order to aid the submission of data into infrastructure such as the DECIPHER database.

High quality genetic testing requires access to expert variant interpretations

A recent publication on the ClinGen Clinical Genome Resource by Rehm and colleagues [11] illustrates the challenges and risks inherent in relying on an incomplete and often inaccurate knowledge base. In the example, genetic testing identified a likely pathogenic variant in a patient who on post mortem was found to have suffered from hypertrophic cardiomyopathy leading to sudden cardiac death. The patient's relatives were tested for the variant and informed whether or not they were at risk of cardiomyopathy and required monitoring. Five years later a cardiologist queried a database and discovered that the variant was now classified as 'likely benign'. The family were re-tested, and a different variant found to be pathogenic. A family member who had previously tested negative was now found to be positive for the new variant and received an implantable cardioverter defibrillator. Had the cardiologist not queried the variant, there might have been a grave outcome for the patient with the false negative diagnosis.

Given the pace at which genomic knowledge is developing there will be similar cases of patients with an incorrect genetic diagnosis because at the time they were tested there was insufficient evidence. High quality clinical genomic testing requires expert interpretation of variants at the point of clinical reporting which incorporates:

» A systematic approach and best practice guidelines [12] to interpreting and classifying variants

» Variant interpretation to take account of clinical phenotypes

» Access (via databases) to expert interpretation for known variants

» Notification systems within databases to flag variants for which new information or an updated interpretation is available

InSiGHT: An example of expert interpretation of variants through sharing

The International Society for Gastrointestinal Hereditary Tumours (InSiGHT) is a multidisciplinary and international scientific organisation with a focus on examining variants in genes involved in hereditary bowel (and related) cancers and associated Lynch syndrome. Their aim is to better understand the clinical significance of variants and classify them according to pathogenicity from Class 5 (pathogenic) to Class 1 (benign). Many existing classifications are inaccurate and a collaborative approach which requires sharing of information has been integral to improving the quality of the underlying knowledge base. For example, using a standardised model to classify variants, the InSiGHT team reduced the proportion of variants previously classified as 'of uncertain significance' from a half to a third. However, over 20% of over 2,000 variants originally classified as 'known pathogenic' variants were reclassified to a benign status and over 10% of considered benign variants were reclassified as pathogenic [6].

The impact of variant interpretation: examples from diabetes testing

Examples from monogenic diabetes testing illustrate the significance and potential consequences for a patient's wellbeing and safety of an informed or misinformed genetic test and interpretation of data. A confident correct genetic diagnosis enabled a patient who had been treated for 27 years with insulin injections to be able to safely switch to an oral treatment as the test indicated the condition could be managed using tablets. However, there are also cases where a misguided genetic test or interpretation could lead to an incorrect genetic diagnosis and inappropriate patient care with potentially harmful consequences for the patient. As genetic testing becomes more mainstream and specialities beyond genetics refer patients for testing, there will be increasing demands on genomic laboratories to provide clear reports that are interpretable beyond clinical genetics services and indicate where evidence is weak or requires further investigation.

Facilitating the sharing of genetic variant data: the Exeter laboratory's experience

The Exeter laboratory has been performing molecular testing since 1996 and originally used Access databases for storing variant data. Since 2000 monogenic diabetes variant data have been shared through a series of Mutation Updates published in Human Mutation [13]. During 2013 these data and variants identified through testing for >75 rare diseases were formatted for submission into DMuDB. However, since DMuDB no longer accepts submissions, there has been work this year to transfer the data from DMuDB into the NHS portal within the DECIPHER database.

The manual curation and submission of data is labour intensive and the laboratory has been working towards streamlining this process through digital management systems for automating the deposition of genetic variant data into databases. In 2010 the laboratory adopted a laboratory information management system (StarLIMS) and configured this to bulk export data into the DMuDB database. The StarLIMS system was then configured to submit data to the NHS portal of DECIPHER. The data upload occurs monthly and is currently semi-automated, although full automation is the target. From the laboratories perspective the process of data curation and submission needs to be resourced and a sustainable database solution is essential in order to make the effort and investment in (automated) data submission worthwhile.

Sharing of variant data should ultimately be with a global intent

Ultimately high quality genomic testing requires global data sharing of expertly interpreted variants. Although widespread sharing across the NHS would be a positive start, specialised genetic testing services (such as the tests offered by the Exeter laboratory) that are not available widely across the NHS will require data from further afield to achieve the greatest clinical utility.

2.4   Presentation: sharing structural variation - from large to small and back again

Presenter: Dominic McMullan, Chair ACGS Science-Subcommittee, and Consultant Clinical Scientist, West Midlands Regional Genetics Laboratory

Synopsis: This presentation described how structural variants are associated with various different rare diseases and cancers, but classifying and interpreting these variants in their phenotypic context is challenging. Mechanisms to share structural variants through databases, such as DECIPHER, are required to improve genetic diagnostics.

Background: Structural variants occur when large sections of DNA sequences are duplicated, deleted, translocated or inverted in the genome. These mutations are less common than smaller single nucleotide variants but can be an important source of genetic variation that contributes towards diseases and disorders. It is important that an approach to variant sharing considers the different types of variation.

Chromosomal microarrays (CMAs) are currently the frontline test for all postnatal referrals for neurodevelopmental delay and / or congenital anomalies (birth defects), and increasingly the first choice test for invasive prenatal diagnoses following detection of a structural anomaly during an ultrasound scan. These tests constitute more than 40,000 patients assays per year in the UK, probably making CMA one of the most common rare disease tests. Hence sharing data on structural variants (in addition to other types of genetic variation) is essential for improving the diagnosis of genetic disorders. Previous to implementation of CMA, lower resolution conventional cytogenetic test results were traditionally shared via the Association of Clinical Cytogenetics Chromosome Abnormality Database, which contained approximately 150,000 cytogenetic patient records and the European Cytogeneticists Association Register of Unbalanced Chromosome Aberration (ECARUCA) database. Neither initiative sought consent for data sharing.

There are a number of databases for storing CNVs, including dbVar [14] for population level data on CNVs, and DECIPHER and ClinVar for clinical data. CNVs are classified into five different classes based on their pathogenicity. The West Midlands Regional Genetics Laboratory (WMRGL), like most NHS diagnostic laboratories, deposits variants into DECIPHER that fall within the pathogenic, likely pathogenic and uncertain categories (Classes 5, 4 and 3 respectively). Data is entered at, or shortly after, clinical reporting, manually if immediate clinical interpretation is needed, or later as part of a batch submission.

There is a wealth of existing CNV data and the volume is increasing. By 2009 WMRGL had deposited 70 rare CNVs into DECIPHER; by 2015 this number had risen to over 1,000 from both clinical genetics and paediatric referrals.

Some of these variants are associated with the diverse ethnicity of the West Midlands population; in particular the Pakistani population which constitutes ~15% of the Birmingham demographic. There is a paucity of control data for this population and pooling more data would further improve and facilitate the classification of these variants, particularly those variants of uncertain significance, the reporting of which to patients can cause anxiety.

Patient consent and the deposition of structural variants

Consent is sought for public data sharing within DECIPHER: following the genetic test a patient consent form for public data sharing is issued and emailed to the clinical geneticist along with the laboratory's report. It was noted that consent is more likely to be obtained with Class 4 or 5 variants than Class 3 – variants of uncertain significance. Of the ~1000 CNVs of clinical interest identified by WMRGL laboratory, only 22% (208) have been consented for public data sharing, while the remaining 78% (748) are visible privately to the WMRGL teams. The retrospective consent process is probably impeding wider sharing of data and is hindering increased accuracy in classification of variants of uncertain significance.

The view from WMRGL is that the NHS consortium data sharing model outlined in Dr Helen Firth's presentation, where data is visible within the consortium without requiring wider patient consent, would substantially aid the sharing (and therefore the classification of) variants of uncertain significance and those variants that lie within genes of uncertain significance. Governance may be required if the consortium model is to be taken forward, perhaps via an oversight panel or steering group to determine how the database should be used most effectively. There is also the need for designated laboratory or clinical leads to be responsible for data curation.

Illustrating the complexities of interpreting structural variants

Genetic variation has patterns that correspond, in some instances, to different ethnic populations, and this can present certain challenges. An example that was given demonstrated the complexities of rare disorders coming from (probably closely related) individuals from a small region in Pakistan. Among these large families there was common occurrence of infantile seizures (sometimes resulting in premature death), neurodevelopmental delay, and microcephaly. After years of metabolic and genetic testing, an implicated homozygous genetic variant in the *WWOX* gene was found in 2014 via a SNP array analysis. Heterozygous variants are relatively common in this region in healthy individuals, so there is a need for further investigation to understand the genotype to phenotype link and establish a clear disease pathway.

The necessity to share all types of genetic variant that have potential clinical relevance

Clinical data sharing is needed to make use of the large NHS data sets that are still largely untapped. CNVs need to be combined in the same database with other types of genetic variation, particularly for cases where the mutational effects may be additive. This will become increasingly important as testing coalesces to a single assay for detection of all variants (CNV, SNV, INDEL and larger structural variation). There is an urgent need to better capture phenotypic information. In the view of Dominic McMullan, DECIPHER currently represents an immediate option for heterogeneous disease data collection.

2.5 Small group discussion on the clinical rationale for sharing different data types

Each of the three focus groups was asked about the data types that are necessary to share in the short term and desirable to share in the future and how widely they should be shared to enable consistent high quality service delivery within the UK. The full list of the questions is provided in Appendix 8.3.

Data types to share

What data is it necessary to share in the short term for the clinical benefits to be realised?

Genetic variant data

» **Pathogenic variants**

There was a consensus across the three groups that sharing pathogenic and likely pathogenic variants (Class 5 and 4 variants respectively based on ACGS variant classification guidelines [12]) is essential, as these can be used by clinicians across different genetic services to directly inform the diagnosis of their patients

» **Variants of uncertain clinical significance (VUS)**

Equally there was consensus that sharing VUS (Class 3 and 2 variants) is essential, as pooling information can help facilitate their interpretation as either pathogenic or benign

» **Benign variants**

Benign (Class 1) variants do not contribute to disease, so sharing them can help filter down the number of variants observed in a patient to those that are most likely to be disease causing. There was not a consensus across the focus groups whether it was necessary to share these variants in the short term or desirable in the medium term. The robustness and certainty with which the variants are assigned as 'benign' is significant as an incorrect classification can misinform data interpretation and diagnosis. As described in Professor Sian Ellard's presentation over 10% of variants considered benign were reclassified as pathogenic by InSiGHT

Clinical and phenotypic data

» **Patient diagnosis and clinical and phenotypic data**

There was a consensus across the three groups that the patient's putative diagnosis, along with some basic standardised phenotypic information, (*e.g.* using the Human Phenotype Ontology (HPO) terms [15]) can provide the critical genotype-phenotype link necessary to interpret variants, and is thus essential to share. Standardised descriptors can enable the direct identification of traits, comparison and interpretation of data, and therefore may reduce the likelihood of variant of misinterpretation. Beyond this minimal data, more detailed clinical and phenotypic information may not be necessary or practical to share in the short term, although in the medium term this information could be useful in helping further define syndromes and their symptoms if appropriately standardised nomenclature is used

Patient history

» **Family history and pedigree data**

Information provided by family history in the context of the diagnosis may facilitate variant interpretation and the understanding of disease inheritance. However, the legal / regulatory focus group noted that such sharing is legally complex because consent has not usually been sought to share data about other family members

» **Patient ethnicity**

Ethnicity can be important for contextualising variants, for instance, a variant may appear rare in some populations but not others and so it is important to understand this genetic background when interpreting pathogenicity, since common variants are unlikely to be pathogenic. However, ethnicity is an imperfect proxy for designating a genetic population and, increasingly with whole genome sequence (WGS) data, the genetic architecture and other patient information (*e.g.* consanguinity) may be best defined from the data. However, in the short-medium term, whilst patient WGS data is not routine, the ethnicity of the patient would be useful to share

Further contextual data

» **Name of testing laboratory or contact details**

Further details of a variant may be provided on request and therefore the ability to contact a testing laboratory was deemed essential by the three groups. Further details could be provided via a submitter form, as used by the DECIPHER database, so the laboratory remains anonymised

» **Metadata**

The legal / regulatory group thought that contextual and methodological data would also be essential to share in order to help build an evidence base to support classification, as well as aid transparency. However, the clinical need group felt that this was not essential in the short term, and that it would be sufficient to share data without such information providing the data had fulfilled defined quality control standards

What data would it be desirable to share in the longer term?

» **Clinical management and outcomes**

All groups agreed that information on clinical management and outcomes would be highly desirable for applications beyond diagnostics, such as permitting improved and more tailored treatments, *e.g.* stratified-care in the cancer context

» **Sequence data**

There was a consensus across the three groups that assembled sequence data (and the raw data from which this is derived) would not generally be necessary for clinical care but would, if shared in the future, help improve the robustness of definitions of variants and inform clinical research

Table 1: Summary of outcomes of discussions on data sharing by data type, across the three focus groups

Data category	Data type
Consensus that it is necessary to share now	
Genetic variants	Pathogenic / likely pathogenic variants (Classes 5, 4)
Genetic variants	Variants of uncertain clinical significance (Classes 3, 2)
Clinical and phenotypic data	Patient diagnosis and basic phenotypic data
Further contextual data	Contact details for testing laboratory
Either necessary or desirable and requiring more policy development	
Genetic variants	Classified benign variants (Class 1)
Clinical and phenotypic data	Detailed clinical and phenotypic data
Patient history	Family history and pedigree data
Patient history	Patient ethnicity
Further contextual data	Metadata and contextual information
Consensus that it is desirable to share in the future	
Clinical and phenotypic data	Clinical management and outcomes
Sequence data	Assembled genome sequences
Sequence data	Raw genome sequences

Data access

How widely should the data be shared?

In general, it was agreed that the sharing of relevant genomic data amongst relevant clinicians and scientists within the NHS, possibly via a consortium model as is used by the DECIPHER project, is necessary for the clinical benefits to be realised. The specifications for an NHS Consortium data sharing model should be developed and clearly defined. These should include the community operating the system; their knowledge and expertise; data sharing infrastructure; form, content and nature of data shared; purpose of sharing; and the safeguards in place. All these elements contribute to trustworthiness.

Sharing a sub-set of this data more widely beyond the NHS, such as public global sharing, would have clinical and research benefits, particularly for very rare diseases that may only be seen once or a few times within the NHS. The wider the data is shared, the more potential health utility can be reaped from it, but also the greater the risk of privacy breaches and data misuse.

The legal / regulatory group discussed some specific data types and felt that some known pathogenic variants could be shared in the public domain through disorder / disease-specific or locus-specific variant databases. The legal group also agreed that the NHS number (or equivalent in the devolved nations) is a direct identifier and access should be limited to NHS genetic laboratories or to laboratories commissioned to provide services for NHS patients, therefore including private providers. Increasingly complex encryption methods are being developed based on the NHS number, which allow data relating to the same individual to be linked without attributing that data to a particular individual.

2.6 Small group discussions on the technical and practical mechanisms for facilitating data sharing

The laboratory / technical group were asked about the practicalities of sharing different data types. All groups were asked to broadly consider factors, or principles, that may facilitate and expedite data sharing and who should be responsible for any database in place specifically to support NHS clinical genetics / genomics services.

Data curation, deposition and data management

Technical feasibility of sharing data

The laboratory representatives were asked to consider whether it was 'technically feasible' to share the types of data identified as necessary to share now for clinical care. 'Technical feasibility' encompassed the ability to deposit, store and access data. Although it is technically possible to share the necessary data types, curation and deposition of data are chiefly manual processes and therefore demand resources, time and personnel, which are limited. Delegates noted that the DECIPHER database has much of the functionality for enabling the collation of and access to the critical data types. However, it is not currently practical to share a number of the data types identified as highly desirable within existing infrastructure, namely genomic sequence data, partly due to their size and volume and because the generation of exome or genome level data is a relatively recent undertaking in the NHS setting. It was noted that Genomics England is developing infrastructure for storing whole genome sequence data generated as part of the 100,000 Genomes Project. The aim of this project is to create a lasting legacy for patients, the NHS and the UK economy. Since this workshop took place, the possibility of using the DECIPHER NHS consortium model to help inform the analysis of data arising from the 100,000 Genomes Project is being explored.

Who should 'own' any database for specifically supporting NHS clinical genetics / genomics services?

This question was posed to all three small groups and it was emphasised that patient data should be contained within a resource assigned for NHS use only. The management and maintenance of the physical database infrastructure may well be contracted out, as long as robust processes and adequate agreements are in place to secure and manage the database and proprietary interests are not exerted over the data. There was consensus that any database provider should be trustworthy (in the public's perception); the database should be sustainably funded and processes must meet Information Governance (IG) standards. It is vital to have a resourced and sustainable infrastructure with robust processes for management.

Mechanisms for facilitating data sharing

Sustainable infrastructure and mechanisms / support to facilitate data curation and upload

There was consensus that data sharing would be facilitated by a designated and sustainable infrastructure in which to deposit data and processes to support data curation. Support could range from dedicated personnel to curate and upload data, to modifications to LIMS to enable data deposition to occur at the point of clinical reporting. Ideally the process should be fully automated to increase accuracy and throughput and to reduce overheads. The costs of curating and developing systems to deposit data could potentially be built into budgets for providing diagnostic services.

Incentivisation to encourage data sharing

A variety of potential mechanisms to incentivise data sharing were discussed. The challenge is in finding an effective balance of measures. Data sharing could be mandated through service commissioning processes, although strategies for dealing with non-compliance would need to be considered. For the laboratory / technical focus group, the key incentive to share data is to facilitate patient diagnosis. Data sharing processes should be simplified as much as possible with appropriate safeguards and with clear national agreement on the legality, then virtually all laboratories would comply. Best practice guidance would facilitate sharing to the standards required for service delivery. 'Reward' mechanisms could include micro-attribution, *i.e.* giving contributors credit for submitting data. One suggestion was to rate laboratories based on the proportion of variants that they share, *e.g.* by capturing metrics in a (clinical reference group) quality dashboard.

Clarity on where to share data

National agreement on both the legitimacy of sharing and NHS operational agreement on the use of an approved database, whether DECIPHER or another resource, will improve the consistency of data sharing practices by providing clarity on data sharing at a local (Trust) level and give laboratories the authority to share data. Although the initial collection of data should be focused into a principal, robust database in which users have confidence of the consistency and quality of data, from a clinical and laboratory perspective there is also value in sharing data into different database systems (*e.g.* disease centred or locus specific databases).

It was noted that as the volume and complexity of genomic data increases, laboratories are seeking advanced solutions that enable them to manage, store and analyse data. For example, one laboratory commented that they have undergone an IG process for archiving genetic data on the Amazon cloud. Other laboratories are seeking IG approval in order to use commercially developed analytic software systems that rely on the transfer of data to servers outside the E.U.

3. Securing responsible data sharing through proportionate law and regulation

3.1   Introduction to legal / regulatory issues

The law and regulation relating to data sharing attempts to strike a balance between competing interests: on the one hand it seeks to protect individuals from foreseeable harms such as privacy breaches by restricting the extent of data sharing, but on the other, to enable prompt, safe care and effective management of health services, by increasing data sharing. This tension, between the individual and wider population based concerns, is a feature of many areas of medicine and in practice it is sometimes difficult to resolve, especially in areas such as public health and genomics.

Finding the right balance between policies that prioritise individual protection and allow individuals to dictate the extent to which their data is shared (*i.e.* a restrictive model), and those where data is shared more widely to promote the public good, is particularly difficult in the context of genetic and genomic data, where:

» Very rare genomic variants could be considered to be unique (and therefore potentially identifying)

» Genomic variants from a patient may also be found in their family members

» The harms that arise from privacy breaches might not be imminent, or they might be speculative in nature

Due to the way the existing laws and regulations are structured, applying these to genetic and genomic data is challenging:

» **The extent to which genetic / genomic data is identifiable depends on context**

  Whole genome sequence data is unique to an individual and therefore potentially identifiable if linked to other identifiable data such as phenotypic data. At the other end of the spectrum, common polymorphisms are not identifiable or identifying

» **There is not always a clear distinction between research and clinical activities in genomics**

  The law distinguishes between data sharing that is necessary for clinical care and sharing for a secondary use (including medical research, clinical audit): these are not easily distinguished in practice

» **There is disagreement about whether specific consent is required to share genomic data to support clinical services and, if so, how extensive this should be**

  As a result there is variation in practice between laboratories and countries despite initiatives to standardise practice

3.2   Presentation: data sharing from the diagnostic laboratories perspective: challenges and opportunities

Presenter: Dr Ann Dalton, Chair of ACGS and Director of Sheffield Diagnostic Genetics Service

Synopsis: This presentation described the arguments for and against sharing data into databases, and presented the laboratory perspective on considerations for achieving a proportionate balance between these competing interests.

Background: There are multiple reasons to share genetic variant data; in short, increased information facilitates more accurate diagnosis, prognosis and treatment and management of patients with a genetic condition. Whilst the majority of variant data represent common occurrences in the population, some changes are unique to an individual. The local Caldicott Guardian has expressed the view that such 'unique' variant data cannot be shared if a subset of individuals can identify themselves from that data, even if they are not identifiable by any other person. This opinion has been upheld in a letter from the national Council of Caldicott Guardians, which stated that if an individual could identify themselves on any database through identifying a unique variant, then prospective and retrospective consent is required for data deposition.

Why is it necessary to share genetic / genomic and associated phenotypic data and what are the potential benefits and risks?

The greater the available clinical information on a given variant, the better the ability to interpret relevant genetic finding(s) and return a robust report to the referring clinician. Therefore, from a laboratory perspective, the ideal would be to access as much genetic / genomic and phenotypic data relevant to the clinical case as is possible. However, deposition of data into databases carries a risk of patient identification. The magnitude of this risk varies from case to case as there is a gradation of anonymisation associated with genetic variants, depending on whether they are common, rare, or novel, and whether they occur in dominant or recessive disorders. Arguably for some conditions the variant and related clinical information are not particularly identifiable *e.g.* autosomal dominant hereditary breast and ovarian cancer. However, for rarer conditions where the population incidence of each variant is considerably smaller, the data becomes increasingly unique to the patient. For example, the potential for compound heterozygosity in autosomal recessive diseases, such as Wilson Disease, means that a combination of variants, together with relevant phenotypic information, could be unique to an individual and might be potentially identifiable by a patient. Moreover, some types of clinical data associated with disease could be regarded as particularly sensitive, such as gynecomastia, impotence or erectile dysfunction, neuromuscular features in spinobulbar muscular atrophy, cognitive deterioration or Parkinsonism in Wilson disease. Some patients (or family members) may object to recording such data on the basis that it might be shared, or if asked would refuse their consent.

The role of consent

Evidence of the consent process (including pre-test information provision, consent obtained and return of results) is typically poor in patient records; it is not clear whether this failure reflects poor consent procedures or poor documentation. In a published audit of neurology tests, 26 / 136 documented consent in the notes, and only 50 / 132 had return of those results to the patient recorded. Current reliance on paper records, compounded by poor handwriting, increases the potential for inaccuracy of data deposition.

Achieving a proportionate balance

Not only is the quality of consent in places highly variable, but data stores are also under increasing security attack through hacking and also through inadvertent data breaches and data loss. Data will become more identifiable in the future as the scale and nature of data acquisition and integration increases exponentially, yet submission of well curated phenotypic data is crucial for good quality diagnostic genetics services. So whilst data sharing is vital for diagnosis, prognosis and treatment, it can result in loss of privacy and potential harm to patients and thus potential litigation. Ensuring adequate consent procedures, and appropriate regulation and legislation is one way of achieving a proportionate balance.

From the laboratory perspective, a number of factors need to be taken into consideration:

» High quality data will require investment if benefits are to be realised

» Extraction and transfer of data must be efficient, robust and secure

» Laboratories cannot be held responsible for the nature or quality of consent; valid consent must be assumed or implied

» Regulation and legislation on data sharing must be clear and explicit to avoid the NHS being vulnerable to litigation

» The NHS is increasingly using private sector providers across all parts of the patient pathway, which operate under different constraints and have different drivers

3.3   Presentation: genomic data sharing and the law

Presenter: Dr Jon Fistein, Barrister, Ethics and Policy Lead on Clinical Data Research for the Medical Research Council (MRC)

Synopsis: This presentation described the legal framework for sharing genetic / genomic data including confidentiality, privacy and data protection law and relevant contractual, professional and legal obligations. The limitations of data sharing using this legal framework were discussed and implications for confidentiality and consent practices were highlighted.

The legal framework: confidentiality, privacy and data protection law

Confidentiality, privacy and data protection law impacts upon the provision and use of genomic information in multiple conflicting ways. The common law duty of confidentiality protects information that has the "*necessary quality of confidence about it*" [16]. The Human Genetics Commission (HCG) suggested in *Inside Information* [17] that *"private personal genetic information should generally be treated as being of a confidential nature and should not be communicated to others without consent except for the weightiest of reasons"*.

A right to privacy allows individuals to operate without interference, a right conferred by Article 8 of the Human Rights Act 1998, "*everyone has the right to respect for his private and family life, his home and his correspondence*". This includes medical records [18]. The HGC suggested that "in the absence of justification based on overwhelming moral considerations, a person should generally not be obliged to disclose information about his or her genetic characteristics" [1]. However, privacy rights are not absolute.

The Data Protection Act 1998 (DPA) enacts the EU Data Protection Directive (95/46/EC) and establishes eight principles to ensure the 'fair and lawful' processing of personal information from living individuals. These include requirements to: collect data for a specific purpose; ensure that data is relevant and up-to-date; and hold as much data as is needed for only as long as is necessary. Schedules 2 and 3 of the DPA contain conditions for processing 'sensitive personal data' including the processing of (some) medical / genomic data, with consent; where necessary for 'medical purposes' (prevention, diagnosis, research, care, treatment and the management of health services); and by order of the Secretary of State.

Contractual and ethical / professional obligations

Contractual obligations and professional guidelines, including the NHS Confidentiality Policy, the Code of Practice on the Use of Confidential Information (Health and Social Care Information Centre (HSCIC)) and Confidentiality Guidance (General Medical Council), provide additional safeguards. All these requirements must be balanced against the obligation for appropriate information sharing, which is essential to the efficient provision of safe, effective care, both for the individual patient and for the wider community of patients [19]. Establishing trustworthiness is also a key determinant for effective data sharing. Although confidentiality is central to trust, trust does not depend on confidentiality.

How do these laws, professional and legal obligations impact upon the sharing of genetic / genomic information?

Confidentiality, privacy and data protection are not absolute obligations, confidentiality only applies if the information is 'confidential in nature' (*e.g.* relating to an individual) and there is an expectation of confidentiality. Privacy laws only apply if infringement of privacy is threatened. Whilst some suggest that sharing a genome may intrinsically breach privacy [20], these risks can be mitigated by limiting access to genomic data or sharing only part of the genome. The DPA only applies to personal data not anonymised data.

Relevant questions include:

» Is genomic data intrinsically identifiable / personal / confidential / private? Does it uniquely identify an individual?

» Even if not, can identities be inferred (*e.g.* through jigsaw identification)?

» What can be shared legally (*e.g.* through relying on consent or using anonymised data)?

» How can risks be minimised (*e.g.* using automation, with appropriate contractual safeguards, or through sharing part of the genome)?

Sharing of confidential information may be allowed: if the patient consents; if there is a public interest justification (*e.g.* to prevent serious harm to a named person); if there is a legal requirement or alternative ground enabling disclosure (*e.g.* the use has been sanctioned under s.251 of the Health and Social Care Act). The scope of these exceptions is limited because public interest exceptions are generally framed narrowly, implicit for care but expressly for other purposes, and s.251 sets aside confidentiality but not data protection obligations.

In the past, obtaining consent to sharing was seen as the gold standard. But many questions remain:

» What information should be given to patients? What would a reasonable patient want to know? [21] How might this differ for research and clinical use?

» What expectations do patients have around data sharing? (How? With whom? Sharing to multiple databases? Implications of sharing with family members?)

» Are broad or open consents permissible in law?

It remains to be seen whether new legislation will be developed (unlikely) or fresh approaches adopted.

3.4   Presentation: information governance and data sharing in support of genomics in clinical practice

Presenter: Dame Fiona Caldicott, National Data Guardian for health and care, Chair of the Independent Information Governance Oversight Panel

Synopsis: This presentation described recent developments in Information Governance (IG), the impact that IG is having on genomic data sharing, and suggested measures to facilitate data sharing.

Background: There is increasing awareness of issues of privacy and confidentiality, indeed a key finding of the Information Governance Review (2013) was the importance of ensuring the right balance between sharing and protecting information [19]. This report concluded that there was a need to address the culture of anxiety around data sharing and achieve more proactive sharing: the Panel identified a new 7th Caldicott Principle establishing that 'the duty to share information can be as important as the duty to protect patient confidentiality' and that 'health and social care professionals should have the confidence to share information in the best interests of their patients within the framework set out by these principles, and should be supported by the policies of their employers, regulators and professional bodies'. Therefore, there is a need to be more aware of the harms associated with not sharing relevant information, and appropriate education and training of professionals will be required to reinforce this. The important doctrine of 'no surprises' should guide Health and Social Care organisations in ensuring that patients and service users know what information is being recorded, and how it is used and shared. This transparency is important in safeguarding trustworthiness and to ensure that appropriate lessons are learned from the care.data programme.

Recent developments: establishment of the Independent Information Governance Oversight Panel and National Information Board

The Independent Information Governance Oversight Panel was established in September 2013; this has independent oversight of information governance in the Health and Social Care systems and responsibility for reporting directly to the Secretary of State for Health on progress in implementing the Caldicott Recommendations. Developing IG knowledge and facilitating and managing consent are important priorities, particularly since the *Montgomery* [21] case and increasingly are acknowledged as such within Government. The National Information Board was established in 2014, and has developed nine work streams designed to improve data and technology use and access in order to transform outcomes for patients and citizens [22]. Work stream four is very relevant: its focus is to build and sustain trust, including giving the role of the National Data Guardian a statutory basis and creating a roadmap to consent based information sharing – with the assurance of safeguards - whilst ensuring that health and social care data is utilised as a rich resource for UK citizens, and optimising access to the newest and most effective treatments, including earlier diagnosis *e.g.* the 100,000 Genomes Project. Creating a permanent post of National Data Guardian will also help to build trustworthiness and to safeguard the public interest in relation to data. A consultation on how this can be achieved was underway in autumn 2015.

The impact of information governance on sharing genomic data

The IG principles for providing direct care for patients with genetic conditions are the same as for the direct care of any patient: there must be a legitimate (legal) relationship and 'to care appropriately, you must share appropriately'. Information governance has become a barrier to data sharing and a culture of anxiety amongst staff must be changed to one of confidence. The established presumption is that explicit consent is required for sharing personal confidential data with staff that are not part of the team providing direct care, or if the use is other than that for which the information was originally collected: implied consent is only acceptable in the context of direct care. However, the effect of the recent *Montgomery* judgement might be to move practice from paternalism ('doctor knows best') to increased transparency and choice, giving a greater role to patient preferences.

What distinguishes genetics / genomic data from other types of data?

For genetic diseases, the shared nature of familial inheritance makes it more difficult to characterise and assess the legitimacy of the relationship between the geneticist and family members; additional complexity derives from the blurred interface between research and clinical care. The lack of maturity for children to make autonomous decisions about their data is another relevant factor. Establishing the purpose of the test is critical, as is clarifying the motivation for sharing information. However when the potential risks and burdens are described, people do not tend to decline data sharing.

Suggestions for ways forward

There is no 'silver bullet' likely to achieve greater data sharing: creating sufficient momentum by building trust, strengthening infrastructure (through mandating stronger oversights for information governance at board level), and improving education for staff and better information for patients and service users will help to dispel the culture of anxiety. These developments, together with increased political commitment as exemplified by the creation of the position of National Data Guardian, suggests that engagement with policy makers (including the National Information Board and Genomics England) would be productive. Identifying empirically what preferences patients have over the use of legacy data through focus group research might also be useful to aid policy development.

3.5   Legal / regulatory plenary discussion

These talks highlighted the fine balance that needs to be struck in order to facilitate data sharing whilst securing proportionate safeguards against data breach or other types of harm to patients. They generated discussions about how best these competing tensions could be reconciled:

What features might a database take?

There was general agreement that a secure data sharing structure within the NHS firewall was both necessary and desirable to support clinical genetics and genomics services. The optimal form of data repository needs to be clarified: rare disease variant data could be held as part of an existing database (such as a subset of the 100,000 Genomes Project dataset, or part of the DECIPHER database) or operate as a secure federated system between NHS laboratories. The provenance for a federated system was the National Genome Informatics Network (NGIN) proposed by the Human Genomics Strategy Group. Both possibilities should be investigated. There are benefits in being able to link data about a single individual from different sources: this could be done using pseudonymised NHS number or equivalent as an identifier.

Future proofing

There are advantages in limiting the scope of the database to rare disease variants. However this limits the utility of the database: variant pathogenicity may be determined by surrounding exons or pseudogenes which requires that longer sequences of genome are shared; understanding of common complex diseases relies upon more extensive knowledge of other risk factors (such as lifestyle or environmental data) to yield significant health gains.

Trustworthiness

Numerous strategies have been explored in the past to facilitate effective data sharing. One proposal was to identify accredited safe havens for data. More recently, focus has switched to 'trustworthiness'. This principle is useful, because it can take account of multiple dimensions of data sharing – the extent of identifiability of the data, purpose of data sharing, the nature of the users – to determine a proportional approach. Features of trustworthiness include probity; safeguarding; security; describing clear sanctions for data breaches: in short, demonstrating the standards that the public would expect of an accountable system.

Challenges

The most effective strategy is to focus on how data sharing can be improved for newly acquired data rather than existing legacy data. Optimising newly acquired data sharing is the greatest priority. Solutions that require minimal data curation prior to sharing, or that utilise existing data flows are likely to be most effective (*e.g.* National Registration systems for cancers and for rare diseases). However, this might limit functionality because of insufficient specificity to a single gene. A dynamic representation of the genome within a browser type system would be ideal.

There was some discussion about how far a patient's confidentiality might be breached if they (*e.g.* by inference) learn something from records relating to themselves, or were able to self-identify their own variant on a database. Currently, variants are often reported without accompanying frequency data. Without this, inferring the data source is almost impossible. Instances of this type are very unlikely to constitute a breach of confidentiality.

Scoping

Better use could be made of existing data: delegates felt that if this was pooled the probability of identification would be negligible. Scoping would help to ascertain the potential power of this data; the likelihood that individuals would be identified, and risks of resultant harm.

3.6   Small group discussion on legal and regulatory aspects

As described in section 1.1 one group focused predominantly on legal and regulatory aspects of data sharing. These discussions explored two aspects: the scope of any consent used and processes that could be adopted; and the safeguards or protections that could be adopted to mitigate the risks associated with data sharing. High-level questions on consent and safeguards were also posed in the other two focus groups.

Consent and related matters

a)   The practicality of seeking explicit consent for sharing genomic variant data

Responsibility for seeking and recording consent lies with the clinician who has a relationship with the patient rather than the laboratories. Whilst it might be good practice to seek consent for sharing genomic variant data, clinicians have a hierarchy of issues to discuss with their patients, and prioritising data sharing over matters that are of immediate clinical relevance could distort clinical encounters. A detailed discussion may be impractical. Establishing consistent routine practice for consent discussions would be desirable if genetic variant data is retained on databases after initial sharing of variant data for direct clinical care.

There was consensus that limiting sharing to a national 'NHS consortium model' with a federated open structure would be beneficial and reduce foreseeable risks and harms associated with variant data sharing. This would help to address worries that laboratories have regarding the legitimacy of sharing and could also promote greater public trust.

Sharing variant data outside this consortium would offer additional benefits. Sole reliance on a closed system risks insularity, but also lost opportunities to participate in global data sharing initiatives such as the Global Alliance's BRCA Challenge.

b)   Is routine consent sought for deposition of genetic / genomic variants in databases?

Laboratories receive samples from a diverse group of clinicians with different clinical interests. As a consequence, the Trusts do not routinely seek consent to data sharing from patients and some delegates felt it would be very difficult to do so. It is necessary to distinguish between seeking consent prospectively for future sharing, and seeking retrospective consent for existing variants already held in databases. Both pose significant, but different, ethical and practical challenges.

Factors to be addressed include:

» The extent of sharing – whether restricted to a 'trustworthy environment' such as an NHS consortium or more widely

» The blurring between clinical care and research – sometimes it is difficult to differentiate direct care (such as clarifying whether a variant of uncertain significance is pathogenic or not) and research (understanding the impact of that variant in a wider population)

» Demonstrating patient (rather than public) benefit - consent is required for sharing personal identifiable data if the purpose of sharing is not for the direct care of patients

» Formal consent discussions could be supplemented by other options for informing patients about data sharing practices (leaflets, posters *etc.*)

» Opt out – the extent (prospective and / or retrospective), purpose and impact of different systems of opt-out need to be considered. Offering an opt-out for wider sharing could be a pragmatic approach. Complete opt-out can, paradoxically, result in intrusive data collection every time a patient has contact with the health care system which duplicates effort and could cause distress

» Public expectations – many patients expect their data to be shared within the NHS. Evidence from national registration systems suggest that over 99% of respondents are in favour of national sharing. This suggests that there would be widespread support for sharing genetic variants within managed systems such as the NHS, and seeking explicit and specific consent from every patient might not be necessary

c) What consent processes should be adopted?

Consent is a transactional process which should be dynamic and flexible. The fact that obtaining consent is impractical or unfeasible does not negate the legal / regulatory requirement to seek consent for sharing personal identifiable data outside of direct care (or to seek consent for sharing of confidential data without alternative legal authority). If a consortium model is adopted, which is singly commissioned, there must therefore be clarity about purpose *i.e.* whether sharing data goes beyond clinical care, and the extent of sharing and linkage.

d) Could anonymisation / pseudonymisation be used as an alternative?

Genetic variants (and genotypes) vary in identifiability: if not previously reported, there is a possibility that they could be unique and that individuals could self-identify themselves from this data. However, such self-identification is unlikely to represent a breach of confidentiality. Many other genetic variants are non-identifiable and therefore fall outside Data Protection legislation. The risks associated with sharing data can be mitigated by de-identification of genetic variants but methods vary in effectiveness and any claims should be accurate (in order to safeguard public trust). The measure of effective anonymisation would be differently construed if data sharing were limited to within the NHS, as compared to wider data sharing.

e)   Establishing a secure laboratory network with proportionate safeguards

A proportionate approach is needed which balances the very small - arguably hypothetical - risks of breach and harm through wider data sharing, with the potential benefits of variant sharing. The majority felt that professionals, clinicians and laboratory scientists would be failing in their duty of care to patients by not sharing data outside individual hospital trusts. Indeed litigation could result from the failure to share data as much as from the wrongful sharing of data [23]. Data sharing may also be beneficial in other respects: within a secure network it can also act as a quality standard, particularly if frequency data is also shared. The greater danger is not that the risks will be overstated but rather that there is insufficient effort to explain the potential benefits of data sharing.

4. Conclusions - responsible and proportionate sharing of data for patient benefit

The workshop discussions and pre-workshop survey clearly underscore that clinical genetics laboratories require access to data generated by other laboratories or elsewhere for both the direct clinical care of their patients and for improving the quality and safety of their services. The value of data sharing is exemplified by the success of the DDD project in diagnosing rare diseases by collating and sharing data within a structured framework [1]. Critically, the current situation whereby laboratories do not have the confidence to share due to legal uncertainty, nor the practical support or designated infrastructure in which to do so, is creating inequalities in services and compromising patient safety and must be addressed urgently at a number of levels. Although the exact quantity is unknown, valuable patient data already exists and continues to accrue without being shared across clinical genetics services. Whilst they are segregated, the maximal clinical utility of these data is not being realised.

Laboratories emphasised that they are willing to share data they have generated with other NHS clinical laboratories (and beyond) for wider patient benefit. Indeed, since this meeting on data sharing took place (on 23 June 2015) there has been a 120% increase in the number of cases shared in the NHS Consortium DECIPHER project, rising from 6,662 cases (before 23rd June) to 14,618 (correct as of October 2015) and an increase in the centres involved from six to ten. Although the precise motivation for laboratories depositing this data following this meeting is unclear, knowledge of the consortium, awareness around how other laboratories are sharing and increased confidence in sharing constituting part of the laboratory's duty of care, could all be possible incentives. However, in the absence of clear national guidance, laboratories will continue to receive inconsistent advice from their local NHS Trust and the consequent inconsistencies in practice will continue. Given that access to genetic and genomic variant databases is required for interpreting patient data in a diagnostic context, improving service quality and reducing harm to patients, data sharing should be regarded as necessary for routine NHS care and service delivery. As such, arguably consent to sharing within the NHS provider (or NHS consortium) system can be implied provided that there is no breach of confidentiality. Ideally an NHS consortium or provider system specifically for the purpose of sharing between NHS clinical genetics / genomics services and providers should be established.

Sharing some data beyond a closed NHS only consortium, and globally, can also be applicable for the direct clinical care of patients, particularly those with very rare disorders, and for improving service quality by collaborating with international experts in a disease area as demonstrated by the work of the international InSiGHT committee. Clearly public level sharing has different implications for privacy and confidentiality than sharing within a closed NHS framework. Therefore the types of data that might be placed in the public domain, and the consent mechanisms required, would be different. The recommendations consider significant elements around sharing within a closed NHS framework and beyond, as well as the broader principles which apply to both scenarios.

5. Recommendations

Enabling consistent and responsible data sharing

Recommendation 1

Sharing genetic / genomic variants is a necessary part of clinical care and NHS service delivery. Current arrangements for sharing genetic / genomic data within the NHS are unsatisfactory: inconsistent practices are causing significant differences in patient care and are compromising quality and safety. Three elements are needed to improve, optimise and transform existing practice:

» There is a need for **strong leadership** by the multiple responsible health organisations to demonstrate the benefits associated with data sharing, as well as the burdens and risks associated with sharing and not sharing, and to fully exploit new opportunities for building genomic capacity and services in the UK

» **National agreement** is urgently needed in order to optimise sharing of genetic / genomic data within the NHS. All those involved need to develop a common understanding of the **legitimacy of data sharing**. We support a responsible and proportionate approach that takes account of a set of common principles to demonstrate trustworthiness. These include transparency about the purpose, risks, benefits and safeguards involved

» **Standardised operational processes** need to be developed to achieve robust, effective and consistent sharing practices. There needs to be a **designated sustainable database or mechanism for sharing data across NHS clinical genetics / genomics services with clear governance, oversight, standards and safeguards**. A nationally accessible resource is integral to improving clinical outcomes and to supporting the effective delivery of clinical genetics / genomics services and as such should be long-term and sustainably resourced

Data sharing within the NHS

Where to share?

Recommendation 2

A variety of approaches might be suitable for a national data sharing system: **end-users should engage early with stakeholders who are already developing systems and applications** to ensure that optimal mechanisms are developed for clinical use, including (where appropriate) interoperability with existing relevant infrastructure. If the most effective strategy to enable and promote data sharing is to build on existing knowledge and systems such as DECIPHER and adapt this for the NHS, then the principles from Recommendation 1 must still apply.

What to share?

Recommendation 3

As a minimum all disease causing or potentially disease causing genetic variants identified through clinical investigations should be accessible and shared. This should be supplemented by diagnostic and phenotypic data and metadata. This data should be accessible to authorised users within the NHS (and some elements to other users subject to appropriate safeguards). (Further detail in section 2.5).

How to share

Recommendation 4

Data deposition, sharing and curation are essential elements of laboratory service delivery. These must be safeguarded by secure, efficient, robust and sustainable mechanisms and processes. A variety of tools and approaches are likely to be required, ranging from dedicated resources for data curation and sharing, to developing a collaborative framework for laboratories to share tools, experience and best practice for facilitating data sharing.

Recommendation 5

Comprehensive, **peer reviewed, best practice guidelines for data sharing** for clinical genetics and genomics practice should be developed by laboratory scientists and relevant clinical professionals to provide consensus, establish minimum quality standards and therefore promote greater data sharing.

Future proofing

Recommendation 6

Mechanisms ought to be established for sharing and managing whole genome and exome sequences and raw genomic reads, in ways that promote the future development and improvement of clinical genomic services. These mechanisms should take account of advanced or innovative methods of data storage, processing and analytics (such as cloud computing) as well as novel approaches to consent (such as dynamic or machine readable consent).

Data sharing beyond the NHS

Recommendation 7

Systems and legal processes need to be put in place to allow the contents of the NHS Database system (whether a consortium or dedicated database) to be shared more widely outside the NHS. In order to address proposed legislative changes, the optimal method of establishing a firm legal basis for sharing identifiable patient data beyond the clinical care of the patient would be to seek routine appropriate consent. This will contribute to building public trust.

Wider considerations – achieving consistent and supported data sharing

Recommendation 8

Once a robust data sharing model has been established, through identifying a database or securing access to variant data via a federated system, **data sharing should be made more robust by mandating requirements for sharing genetic variants and monitoring compliance** through commissioning (as part of the service specifications for clinical genetics laboratory provision). These approaches should be supported by relevant commissioning bodies and professional clinical guidelines.

Recommendation 9

In order to facilitate public engagement and public trust, **there needs to be a concerted effort by NHS providers and other relevant stakeholders to inform patients about how their data are used** both to support routine clinical care and for other secondary purposes (such as research, commissioning, and education and training). This could be achieved through a variety of processes including information leaflets, posters or by verbal communication.

Recommendation 10

Relevant regulatory agencies and professional groups should work together to ensure that there are harmonised appropriate approaches to evaluating and assessing health and care data use by organisations and individuals, and that sanctions are applied in a robust and consistent manner.

Recommendation 11

In developing a national consensus on data sharing the risks and harms of sharing ought to be balanced with the risks, harms and opportunity costs of not sharing data firstly within the NHS, and then also beyond NHS services. **A collaborative, multi-agency assessment of the probability of the risks of privacy breaches occurring and the impact they may have on patients, as well as the risks and impact on patients of not sharing data, is warranted.**

6. Relevant initiatives

Since the workshop was held, there have been a number of initiatives which are relevant to policy development in this area:

» The National Information Board has begun a comprehensive programme of policy development to take forward the recommendations from Delivering the Five Year Forward View: Personalised Health and Care 2020 [24]. This now comprises nine work streams that target various aspects of the report with the aim of harnessing advances in data and technologies to transform health and care services and deliver greater quality and efficiency. Four of these work streams contain initiatives which are relevant to genomic data sharing: work stream 1.1 - to provide for genomic data to be included in future clinical decision support; work stream 2.2 - to create a core genomics minimum dataset for secondary use; work stream 4 - to build and sustain trust; and work stream 5 – to support innovation and growth

» The Health and Social Care (Safety and Quality) Act 2015 (in force on 1.10.15) enshrines an explicit duty to share information internally within a health or adult social care provider or commissioner, for purposes of health or social care

» The Information Governance Alliance will generate guidance on this legislative change (and other legal concepts such as the implied consent) for practitioners and Caldicott Guardians.

» A working group has been set up to facilitate capability within the NHS to support the future agenda on genomics and molecular pathology datasets (chaired by Dr Mark Bale and Professor Sue Hill) and maximise benefits from the 100,000 Genomes Project

7. References

URLs in this report were correct as at 30 November 2015

1. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015 Mar 9; 519(7542): 223-8.

2. Callaway E. Global genomic data-sharing effort kicks off. Nature. 2014 Mar 3; doi. 10.1038/nature.2014.14826.

3. The Human Variome Project [Internet]. Avaiable from: www.humanvariomeproject.org/

4. Soden SE, Saunders CJ, Willig LK *et al.* Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. Sci Transl Med. 2014; 6(265): 265ra168.

5. Observed variation in classification of pathogenicity of sequence variants: UK NEQAS* for Molecular Genetics Pathogenicity of Sequence Variants pilot EQA runs 2012, 2013, 2014 and 2015. Dr Sandi Deans personnel communication.

6. Thompson BA, Spurdle AB, Plazzer JP *et al.* Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet. 2014. 46(2): 107-15.

7. ClinVar [Internet]. Available from: www.ncbi.nlm.nih.gov/clinvar/

8. Leiden Open Variation Database [Internet]. Available from: www.lovd.nl/

9. Human Genome Mutation Database [Internet]. Available from: www.hgmd.org

10. DECIPHER database [Internet]. Available from: decipher.sanger.ac.uk

11. Rehm HL, Berg JS, Brooks LD *et al.* ClinGen – The Clinical Genome Resource. N Engl J Med; 2015: 372:2235-2242.

12. Wallis Y, Payne S, McAnulty C. *et al.* Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics. ACGS/VKLR. 2013.

13. Colclough K, Bellane-Chantelot C, Saint-Martin C *et al.* Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha and 4 alpha in maturity-onset diabetes of the young and hyperinsulinemic hypoglycemia. Hum Mutat 2013; 34(5): 669-85.

14.   dbVar. [Internet]. Available from: www.ncbi.nlm.nih.gov/dbvar/

15. Groza T, Kohler S, Moldenhauer D *et al.* The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease. Am J Hum Genet. 2015 July 2; 97(1):111-24.

16. *Saltman Engineering Co. Ltd. v Campbell Engineering Co. Ltd.* [1948] 65 RPC 203.

17. Human Genetics Commission. Inside Information: Balancing interests in the use of personal genetic data. Human Genetics Commission, p.14.

18. *Z v Finland* 1997 ECtHR.

19. Independent Information Governance Oversight Panel. Information: To share or not to share? The Information Governance Review. Department of Health (UK), 2013. Available from www.gov.uk/government/publications/the-information-governance-review

20. Gymrek M, McGuire AL, Golan D *et al.* Identifying Personal Genomes by Surname Inference. Science. 2013 Jan 18; 339 (6117): 321-324.

21. *Montgomery (Appellant) v Lanarkshire Health Board (Respondent) (Scotland)* [2015] UKSC 11.

22. National Information Board. Personalised Health and Care 2020, Using Data and Technology to Transform Outcomes for Patients and Citizens, A Framework for Action. Crown Copyright; 2014.

23. *ABC v St George's Healthcare NHS Trust and others* [2015] EWHC 1394 (QB).

24. Delivering the Five Year Forward View: Personalised Health and Care 2020 [Internet]. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/437067/nib-delivering.pdf

8. Appendix

8.1 Workshop attendees

Delegate	Title
Dr Joo Wook Ahn	ACGS - Bioinformatics Representative, and Bioinformatics Lead, Guy's and St Thomas' Hospital, London
Dr Anwar Alhaq	Project Director, King's College Hospital NHS Foundation Trust, London
Dr Pavlos Antiniou	High Throughput Bioinformatician, University of Oxford, Oxford Molecular Diagnostics Centre, John Radcliffe Hospital, Oxford
Dr Thalia Antoniadi	Consultant Clinical Scientist, West Midlands Regional Genetics Laboratory, Birmingham Women's Hospital NHS Foundation Trust, Birmingham
Maria Athanasopoulou	Bioinformatician, Institute of Neurology, University College London, Queens Square, London
Dr Mark Bale	Deputy Director, Health Science and Bioethics Division, Department of Health, London
Dr David Baty	Head of Laboratory, NHS Tayside Genetics, Ninewells Hospital, Dundee
Jennie Bell	Consultant Clinical Scientist / Deputy Director, West Midlands Regional Genetic Laboratories Birmingham Women's NHS Trust, Birmingham
Cliff Billing	Clinical Scientist, Leeds Genetics Laboratory, St James's Hospital, Leeds
Dr David Bourn	Clinical Scientist, Northern Genetics Service, Institute of Genetic Medicine, Newcastle upon Tyne
Dr Christopher Boustred	Bioinformatician, Great Ormond Street Hospital, London
Professor Sir John Burn	Professor of Clinical Genetics Institute of Genetic Medicine, Central Parkway, Newcastle upon Tyne
Rachel Butler	Head of Genetic Laboratory, All Wales Genetic Laboratory, Institute of Medical Genetics, University Hospital of Wales, Heath Park, Cardiff
Dame Fiona Caldicott	National Data Guardian, Independent Information Governance Oversight Panel, Leeds

Delegate	Title
Carolyn Campbell	Consultant Clinical Cytogeneticist, Oxford University Hospitals NHS Trust, The Churchill Hospital, Oxford
Dr David Cockburn	Deputy Director (Molecular Genetics), Leeds Genetics Laboratory, St James's Hospital, Leeds
Lara Cresswell	Head of Cytogenetics Service, University Hospitals of Leicester, Leicester
Dr Ann Dalton	Chair of ACGS and Head of the Sheffield Diagnostic Genetic Service, Sheffield
Val Davison	National Laboratory Lead Genomics NHSE and Genomics Scientific Adviser HEE, Birmingham
Dr Sandi Deans	Chair of ACGS Quality Committee, Scheme Director, UK NEQAS / ACGS, Lab Medicine, Royal Infirmary of Edinburgh
Andrew Devereau	Head of Data Modelling, Genomics England Ltd, Queen Mary University of London
Mrs Victoria Donnelly	Infrastructure Programme Manager, The National Congenital Anomaly and Rare Disease Registration Service, Public Health England, Skipton House, London
Professor Sian Ellard	Head of Molecular Genetics & Professor of Human Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter
Thomas Finnegan	Legal / Regulatory Policy Analyst, PHG Foundation, Cambridge
Dr Helen Firth	Consultant Clinical Geneticist, Addenbrooke's Hospital, Cambridge and Hon Faculty Member Wellcome Trust Sanger Institute
Dr Jon Fistein	Ethics and Policy Lead on Clinical Data and Research for the MRC
Professor Frances Flinter	Consultant in Clinical Genetics and Caldicott Guardian, Genetics Department, Guy's & St Thomas' NHS Foundation Trust, London
Dr Lorraine Gaunt	Director, Genetics Diagnostics Laboratory, Manchester Centre for Genomic Medicine, Manchester
Dr Vanessa Gibbons	Clinical Scientist / Quality Manager, Institute of Neurology, London
Alison Hall	Head of Humanities, PHG Foundation, Cambridge
Carol Hardy	Clinical Scientist, Birmingham Children's Hospital NHS Foundation Trust, Birmingham

Delegate	Title
Dr Susan Kenwrick	Principal Genetic Counsellor, Clinical Genetics, Addenbrooke's Hospital, Cambridge
Will King	Quality and Service Delivery Manager, SW Thames Regional Genetics Laboratory, St George's University London, London
Dr Mark Kroese	Deputy Director, PHG Foundation, Cambridge
Sam Loughlin	Acting Head of Service (Molecular Genetics), Great Ormond Street Foundation Trust, Regional Genetics, London
Gordon Lowther	Head of Service, Genetics, NHS Greater Glasgow and Clyde, Southern General Hospital, Glasgow
Dr Fiona Macdonald	UKGTN Scientific Advisor, West Midlands Regional Genetics Laboratory, Birmingham
Eddy Maher	Laboratory Director, South East Scotland Cytogenetics Service, Western General Hospital, Edinburgh
Dr Joanne Mason	Lead Scientist - NGS Core Facility, BRC/NHS Oxford Molecular Diagnostics Centre, John Radcliffe Hospital, Oxford
Una Maye	Deputy Cytogenetics Department, Liverpool Women's Hospital, Cheshire and Merseyside Regional Genetics Laboratory, Liverpool Women's Hospital, Liverpool
Michelle McConnachie	Senior Clinical Scientist, NHS Tayside Genetics, Ninewells Hospital, Dundee
Dominic McMullan	ACGS Chair – Science Subcommittee and Consultant Clinical Scientist, West Midlands Regional Genetic Laboratories, Birmingham Women's Hospital, Birmingham
Dr Shehla Mohammed	Consultant Geneticist and Head of Service, Guys and St Thomas' NHS Foundation Trust, London
Andrew Parrish	Trainee Healthcare Scientist (Bioinformatics), Royal Devon and Exeter Foundation Trust, Exeter
Mary Anne Preece	Consultant Clinical Scientist, Department of Newborn Screening & Biochemical Genetics, Birmingham Children's Hospital NHS Trust, Birmingham
Dr Simon Ramsden	Central Manchester University Hospitals NHS Foundation Trust, Genomic Diagnostics Laboratory, Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Manchester

Delegate	Title
Dr Chris Rands	Project Manager (Science), PHG Foundation, Cambridge
Dr Sobia Raza	Interim Programme Lead (Science), PHG Foundation, Cambridge
Eileen Roberts	Head of Department, Bristol Genetics Laboratory, North Bristol NHS Trust, Bristol
Dr Anneke Seller	Director of Regional Genetics Laboratories, Oxford University Hospitals NHS Trust, The Churchill Hospital, Oxford
Dr Abid Sharif	Head of Department, East Midlands Regional Molecular Genetics Service, Nottingham University Hospitals, Nottingham
Dr Joanne Staines	Clinical Scientist, Addenbrooke's Hospital Genetics Laboratories, Cambridge University Hospitals NHS Foundation Trust, Cambridge
Rohan Taylor	Consultant Clinical Scientist Lead for Genetics, SW Thames Regional Genetics Laboratory, St George's University London, London
Lorraine Warne	Laboratory Support Manager, Bristol Genetics Laboratory, Southmead Hospital, Bristol
Dr Christine Waterman	Head of Service, Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury
Dr Paul Westwood	Deputy Head, Laboratory Genetics, NHS Greater Glasgow and Clyde, Southern General Hospital, Glasgow
Dr Joanne Whittaker	Fellow, PHG Foundation, Cambridge
Dr Michael Yau	Research & Development Lead, Genetics, VIAPATH, Guys Hospital, London

8.2   Meeting programme

Tuesday 23 June 2015			
		Timings	**Talk overview**
09.30	**Arrival and registration**	**30 mins**	
10.00	**Introduction to the day** Dom McMullan (ACGS) Dr Sobia Raza (PHG Foundation)	10 mins	Day summary
10.10	**Dr Ann Dalton** Chair of ACGS and Head of the Sheffield Diagnostic Genetic Service	15 mins + 5min questions	Data sharing from the diagnostic laboratories perspective: challenges and opportunities
10.30	**Dr Helen Firth** Consultant Clinical Geneticist, Cambridge University Hospital Trust	15 mins + 5min questions	DECIPHER - enabling diagnosis and discovery through responsible data sharing
10.50	**Professor Sian Ellard** Head of Molecular Genetics and Consultant Clinical Scientist, Royal Devon & Exeter NHS Foundation Trust	15 mins	Towards automatic variant submission at the point of clinical reporting
	Dominic McMullan Chair ACGS Science-Subcommittee and Consultant Clinical Scientist, West Midlands Regional Genetics Laboratory	15 mins + 10 min shared questions	Sharing structural variation - from large to small and back again
11.30	**Coffee**	**(15mins)**	
11.45	**Dr Jon Fistein** Ethics and Policy Lead on Clinical Data and Research for the MRC	20 mins + 5 min questions	Genomic data sharing and the law
12.10	**Dame Fiona Caldicott** National Data Guardian for health and care. Chair of the Independent Information Governance Oversight Panel	20 mins + 5 min questions	Information governance and data sharing in support of genomics in clinical practice
12.35	Panel discussions	25 mins	
13:00	**Lunch**	**(45 mins)**	
13.45	Focus group discussions	1h 30 mins	
15.15	**Coffee**	**(15 mins)**	
15.30	Feedback from groups and follow on discussion (15mins / group)	45 mins	(10 mins feedback and 5 mins discussion / group)
16.15	Plenary discussion and way forward	15 mins	
16.30	**Closing remarks**		

8.3   Focus group questions

Different subsets of these questions were asked to the three different groups.

» **Data types to share**

a)   Which types of data is it absolutely necessary to share now (in the short term), in order to enable the delivery of clinical genetic and genomic services and facilitate the clinical care of patients?

b)   Which types of data would be highly desirable to share, going forward, if the appropriate support and structures were in place for facilitating clinical care and wider applications in order to improve and enhance the quality of clinical genetics and genomics services?

Could you please indicate those broad data categories that you consider appropriate in groups (a) or (b)?

(a) Necessary now	(b) Highly desirable (future)	
☐	☐	Pathogenic variants
☐	☐	Variants of uncertain clinical significance
☐	☐	All other categories of variants, including non-pathogenic variants
☐	☐	(Some) clinical and phenotypic data
☐	☐	Diagnosis
☐	☐	Clinical management / outcomes
☐	☐	Family history / pedigrees
☐	☐	Patient ethnicity
☐	☐	Location / name of testing laboratory
☐	☐	Sequence data (assembles / aligned sequences)
☐	☐	Sequence data (raw)
☐	☐	Other data types of metadata (please state)

» **Consent and related issues**

a)   In your view, is it practicable for clinicians to seek or obtain an explicit patient consent for sharing genomic variant data with other databases or initiatives (on which laboratories will rely)?

b)   Does your Trust routinely seek consent to the deposition of genetic and / or genomic variants in databases?

c) How could / should consent systems be flexible enough to accommodate

   i. Withdrawal of patient consent

   ii. The addition of new databases

d) Alternatively if considering data anonymisation then:

   i. Thinking of the types of data points from the first question (on data types), which categories of data can be anonymised or pseudo-anonymised?

   ii. Could laboratories use and share anonymised or pseudo-anonymised information rather than patient identifiable data?

   iii. If not, why do you think this might limit current or future activity or be impracticable or impossible in some other way?

» **Data curation, deposition and database management**

a) Is it currently technically feasible to share the above sets of data identified as necessary (now)?

b) Are there existing databases and initiatives which could be used or adapted for this purpose, or is there the need for a bespoke 'NHS' resource?

c) Equally, does adequate infrastructure exist to share (transfer / store / access) the data types identified as highly desirable (future)?

d) Who should 'own' any database for specifically supporting NHS clinical genetics / genomics services?

» **Safeguards for sharing the data**

a) Do we agree that data should be shared between laboratories within a secure framework with safeguards?
If yes – what are the safeguards that should be put in place?
If no – are there alternative forms of data sharing that should be adopted?

» **Data access**

a) Which categories of data (from the first question on data-types) should be only visible to other NHS genetic laboratories?

b) Which categories of data could / should be shared in the public domain? (*e.g.* through separate disease-centered or locus specific variant databases)

» **Database management**

Who should 'own' any database for specifically supporting NHS clinical genetics / genomics services?

» **Facilitating data sharing**

What could be put in place to improve the sharing of data? For example:

a) Mechanisms / support to facilitate data curation and upload?

b) Incentivisation to encourage data sharing? *e.g.* contractual and / or accreditation

c) Ensuring any database infrastructure is sustainable?

d) Mechanisms or standards to share genetic / genomic, phenotypic, and clinical data consistently? *e.g.* standard ontologies for recording phenotypes

e) Clarity as to which database(s) it is legally acceptable to share data into?

f) Clarity on what safeguards are in place for specific databases and their data policies?

g) Other considerations?

8.4   Glossary

Anonymisation/ pseudo-anonymisation	Scrambling or removing identifiable information from data, such as patient numbers, so that the remaining data is not (easily) identifiable
Autosomal dominant disease	A condition that will manifest in the next generation if a child receives one copy of the defective gene (*i.e.* from one parent)
Autosomal recessive disease	A condition that will only manifest in the next generation if a child receives two copies of the defective gene (*i.e.* from both parents)
Chromosome	Double-stranded X-like structures located in the cell nucleus that contain the nuclear DNA
Confidential	Confidential data relates to an individual patient
Consent	Where a patient offers their permission for something, such as sharing of their data. Consent may be explicit or implied and may be obtained retrospectively or prospectively
Copy Number Variant	A mutation where a stretch of DNA is duplicated, deleted or inverted in the genome
Cytogenetics	The analysis of the number and structure of chromosomes, including copy number variants
Genotype	A set of genetic sequences that together produce a particular phenotype
Indels	Mutations where two or more DNA bases are added or removed from the genome
Information Governance	The laws, policies and processes that aim to ensure the appropriate use of personal information
Microarray	A biochemical assay that can be used to identify genetic variants
Mitochondrial genome	All the DNA that is situated within the mitochondria, cellular organelles that lie outside the cell nucleus
Next Generation Sequencing	The collective term for post sanger sequencing that is high throughput and involves massively parallel DNA sequencing
Non-coding	DNA sequences that do not encode protein sequences. Some noncoding DNA has regulatory function
Nuclear genome	All the DNA that is situated within the cell nucleus
Pathogenic variant	Disease causing variant
Phenotype	Observable trait caused by some underlying genotype

Privacy	Expectation that confidential data relating to a patient is only visible to particular people
Section 251	Section 251 of the National Health Service Act allows certain patient information to be shared without consent for medical purposes
Single Nucleotide Variant	A mutation where a single nucleotide changes from one 'DNA letter' to another, *e.g.* A to C
Variant of uncertain significance	A variant whose effect and clinical importance is uncertain

8.5   Acronyms

Abbreviation	Full name
ACGS	Association for Clinical Genetic Science
BRCA	Breast Cancer; normally in the context of the *BRCA1* and *BRCA2* genes
BSGM	British Society for Genetic Medicine
CC	Conventional Cytogenetic
ClinGen	Clinical Genome Resource
CMA	Chromosomal Microarray
CNV	Copy Number Variant
CRG	Council for Responsible Genetics
DDD	Deciphering Developmental Disorders
DECIPHER	Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources
DMuDB	Diagnostic Mutation Database
ELSI	Ethical, legal and social issues
ExAC	Exome Aggregation Consortium
HGC	Human Genetics Commission
HPO	Human Phenotype Ontology
HRA	Human Rights Acts
IG	Information Governance
InSiGHT	International Society for Gastrointestinal Hereditary Tumours
JCGM	Joint Committee on Genomics in Medicine
MHRA	Medicines and Healthcare Products Regulatory Agency
MRC	Medical Research Council
NDG	National Data Guardian
NGS	Next Generation Sequencing
NHSE	National Health Service England
NIB	National Information Board
PHGF	PHG Foundation
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
UKAS	United Kingdom Accreditation Service
VUS	Variant of Uncertain (Clinical) Significance
WMRGL	West Midlands Regional Genetics Laboratory
WGS	Whole Genome Sequencing

About the PHG Foundation

The PHG Foundation is a pioneering independent think-tank with a special focus on genomics and other emerging health technologies that can provide more accurate and effective personalised medicine. Our mission is to make science work for health. Established in 1997 as the founding UK centre for public health genomics, we are now an acknowledged world leader in the effective and responsible translation and application of genomic technologies for health.

We create robust policy solutions to problems and barriers relating to implementation of science in health services, and provide knowledge, evidence and ideas to stimulate and direct well-informed discussion and debate on the potential and pitfalls of key biomedical developments, and to inform and educate stakeholders. We also provide expert research, analysis, health services planning and consultancy services for governments, health systems, and other non-profit organisations.

About the ACGS

The Association for Clinical Genetic Science (ACGS) was established in 2012 from a merger of the Association for Clinical Cytogenetics and the Clinical Molecular Genetics Society with the vision of bringing together scientists working within genetics into one professional association. It is the largest of the constituent groups of the British Society of Genetic Medicine (BSGM).

Our members are professionals working within clinical genetic science and include scientists, technologists and bioinformaticians. We aim to promote, protect and preserve the good health of the patients we serve, by the promotion, encouragement and advancement of the study and practice of clinical genetic science. We develop and promote standards in clinical genetic science to ensure best practice. We also support the advancement of education, research and innovation in clinical genetic science.
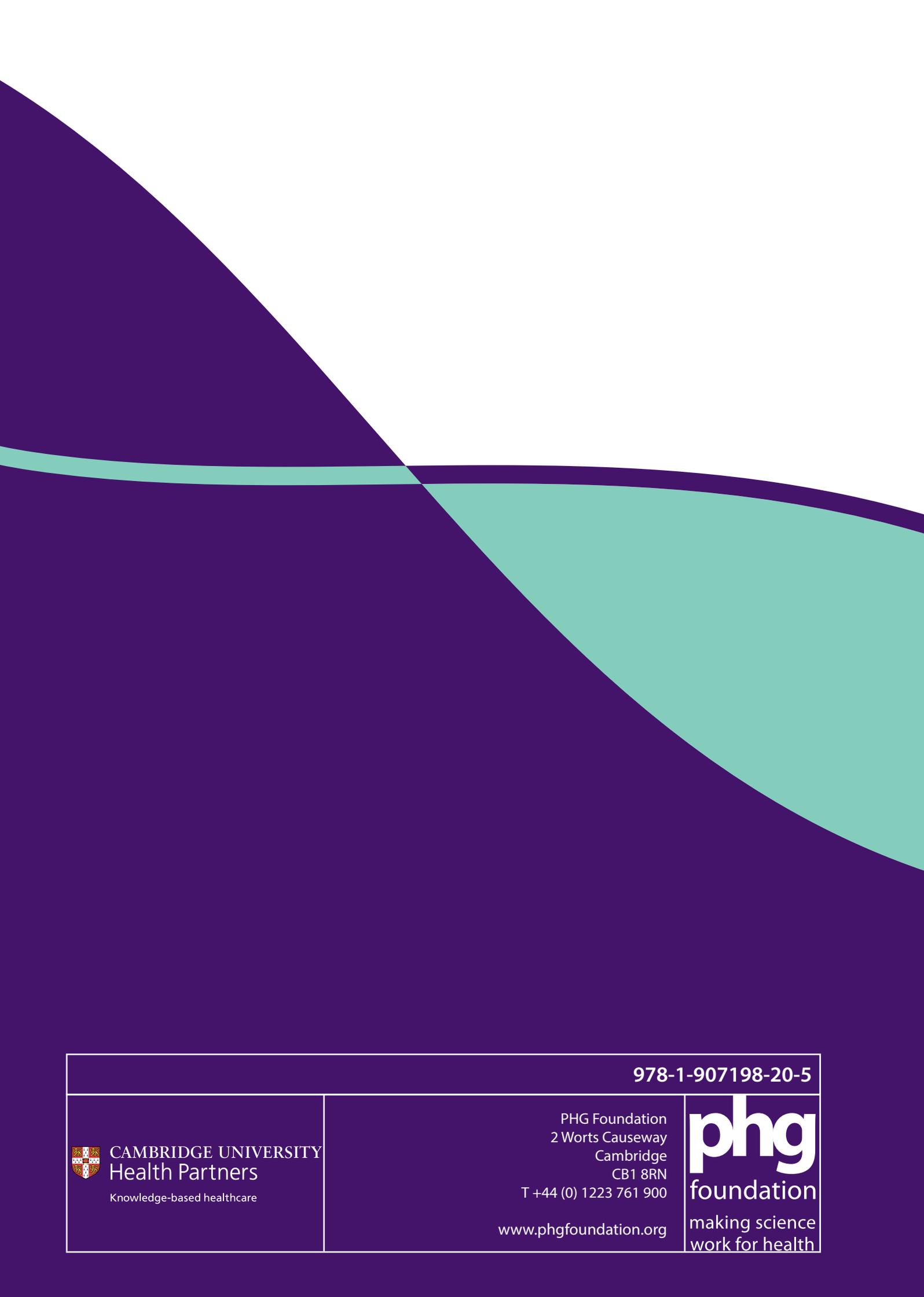
CAMBRIDGE UNIVERSITY
Health Partners

Knowledge-based healthcare

phg
foundation

making science
work for health