British Medical Bulletin, 2017, **123**:35–45
doi: 10.1093/bmb/ldx024
Advance Access Publication Date: 10 August 2017

Invited Review

Genomic medicine and data sharing

Sobia Raza and Alison Hall*

PHG Foundation, 2 Worts Causeway, Cambridge, CB1 8RN, UK

*Correspondence address. E-mail: alison.hall@phgfoundation.org

Editorial Decision 15 June 2017; Accepted 20 July 2017

Abstract

Introduction: Effective data sharing does not occur in the UK despite being essential for the delivery of high-quality genomic services to patients across clinical specialities and to optimize advances in genomic medicine.

Sources of data: Original papers, reviews, guidelines, policy papers and web-resources.

Areas of agreement: Data sharing for genomic medicine requires appropriate infrastructure and policies, together with acceptance by health professionals and the public of the necessity of data sharing for clinical care.

Areas of controversy: There is ongoing debate around the different technical approaches and safeguards that could be used to facilitate data sharing while minimizing the risks to individuals of identification. Lack of consensus undermines trust and confidence.

Growing points: Ongoing policy developments around genomics and health data create opportunities to ensure systems and policies are in place to support proportionate, effective and safeguarded data sharing.

Areas timely for developing research: Mechanisms to improve public trust.

Key words: genomics, data sharing, clinical care, policy

Introduction

The potential for genomics to revolutionize care and enable more targeted diagnosis and treatment has driven life sciences policy[1] over recent years and that promise is beginning to be translated into routine clinical care. Advances in DNA sequencing technologies are escalating the use of genomics to guide the diagnosis and management of a range of disorders across the lifespan of an individual.[2] Diagnostic testing for rare monogenic diseases is a

common clinical application of genetic and genomic tests; 80% of rare diseases have a genetic component and an estimated 7% of the population are affected by a rare disease at some point in their lives. Developments in high-throughput sequencing are now offering the opportunity to uncover the genetic basis of the many 1000's of rare diseases that remain undiagnosed,[3] as well as enabling more rapid and precise diagnosis. Whole exome sequencing (i.e. sequencing solely the genes that code for proteins), has contributed to a diagnostic yield of 27% in previously investigated, undiagnosed children with developmental disorders,[4] while timely use of whole genome sequencing has been demonstrated in principle to guide the rapid diagnosis and management of critically ill infants.[5,6] The clinical utility of a range of exome-based panel tests is recognized and these are now being offered by NHS genetic testing services. A pioneering initiative across the UK is underway to understand the clinical and research applications of whole genome sequencing. This is the 100 000 Genomes Project which is undertaking whole genome sequencing of NHS patients with rare diseases and cancers. Given the heterogeneity of rare diseases and cancers, including hereditary cancers, these developments in genomic testing are impacting upon care for patients across many clinical specialities. Examples include, but are not limited to, cardiology, where genomic testing can help to inform the management of patients with suspected inherited cardiac conditions; paediatrics, for the diagnosis of birth defects or developmental disorders; and oncology to target therapies, or to determine cancer predisposition in familial cancer syndromes such as hereditary breast and ovarian cancer.

As genetic tests have evolved from the interrogation of single genes to gene panels, and through to whole exomes and whole genomes, the complexity of the concomitant data analysis and interpretation has increased too. Some of this complexity can be resolved by the improved pooling and exchange of existing data and knowledge—including information on how genomic test results have been previously interpreted. Other elements of the analytical complexity—particularly pertaining to the less well understood areas of the genome—can only be resolved through further research and the timely sharing of newly generated knowledge. Ultimately, in the immediate term data sharing is essential to ensuring patients receive high-quality, accurate and timely genomic diagnoses, and in the medium to longer term is central to realizing the full clinical benefits of genomic medicine. Yet the sharing of genetic and genomic data—especially that generated within the healthcare context—is fraught with challenges, not least structural, technical, ethical, legal and regulatory considerations. As such, data sharing is at a critical juncture in the advancement of genomic medicine and its impact on patient care. As genomic testing becomes integrated into routine clinical care, data sharing is an issue which will increasingly impact upon and involve medical specialities beyond clinical genetics. In this review we will describe the analytical and clinical necessity of data sharing, the challenges to harnessing genomic data, and the potential opportunities to optimize data sharing for patient benefit in the context of the evolving policy landscape around genomics and healthcare data.

Genome analysis—a question of data aggregation

Every human genome differs from another at around 3–4 million points in their DNA sequence. These differences are known as genomic variants. Some variants are common in the population and other variants can be very rare. Moreover the frequency of a given variant can differ between populations with different genomic ancestries; i.e. what appears to be rare in one population may be common in another. Ultimately understanding the significance of genomic variation—and whether a variant is the underlying cause of disease in an individual or not—can only be done by comparison with wider population level data. Depending on the context, this may comprise of genomic data, and where relevant, phenotypic and clinical data relating to similar and divergent conditions, in related or unrelated individuals. In the context of rare genetic diseases, comparison typically involves short

segments of genomic sequence and relevant clinical details with other clinical genetics laboratories for help with interpretation and validation. In cancer care, the relevant comparison is between the cancer genome (somatic) and the person's normal genome (germline), or with progressive cancer samples over time, to illustrate responses to treatment including chemotherapy. At the point of referral, especially in the absence of digitized patient records, there is often a practical challenge in collating patient data stored in different formats (paper and/or electronic patient notes, referral letters, forms, images) from across the health system.

Data sharing and rare diseases

In the case of rare genetic diseases there are broadly two steps for which the sharing and pooling of data is critical: determining how rare a variant is in a given population and interpreting the clinical significance of rare variation. Variants with high allele frequency, i.e. common in the population are generally unlikely to be the underlying cause of a rare disease. Databases of genetic variation are therefore needed to distinguish between rare versus common variation. ExAC (Exome Aggregation Consortium) is one example of a publically accessible resource commonly used as a reference set of allele frequencies derived from exome sequence data from over 60 000 individuals of diverse ancestry.[7] Clinical scientists typically query the ExAC database to filter for candidate disease causing variants in their patients by excluding those variants that are 'common' in a similar population and focussing on those that are rare.

Having determined which variants in a patient are 'rare' in the genomic regions of interest, the next and arguably most taxing and time-consuming stage in genome analysis is in interpreting the clinical significance of any identified rare variation. Evidence to support the pathogenicity of each potentially disease causing rare variant is carefully evaluated to exclude false positives based on the patient's detailed clinical features and family history. A range of knowledge resources, in-silico tools, as well as scientific and clinical expertise in the genes and disorder in question are applied to inform the classification of a rare variant into one of five categories ranging from pathogenic (disease causing) to benign (has no impact on health).[8] Intermediate categories are classed as variants of uncertain significance (VUS), those whose clinical significance to disease is unclear. Amongst the most compelling pieces of evidence to support the interpretation of a variant as clinically significant is pre-existing information on the variant(s) of interest as well as pre-existing clinical information on patients with the same or similar disorder. In a significant number of cases—a definitive diagnosis rests on comparing the rare variation in unrelated patients with similar conditions. Given the low incidence of the diseases in question, rare variants in similar disorders might only be observed in a handful of cases nationally or even globally and any one individual laboratory cannot expect to generate all the data relevant to interpreting variants in-house. Therefore, well-curated, shared repositories of knowledge that can both be queried and contributed to by multiple users are unequivocally essential to the clinical interpretation process and fundamentally to patient care.[9–11]

Variant interpretation and the genomic testing pathway

Scaling up genomic testing so that it can become integrated into routine clinical care will require a more streamlined service: modernization plans are underway to redesign clinical genomic testing services in the UK, which currently consist of 23 NHS Regional Genetic Laboratories and several specialist testing centres. Collectively these laboratories issue an estimated 200 000 genetic test results per annum, with around 60% of test reports requested from specialties outside clinical genetics.[12] Across these services the interpretation of variants is typically undertaken by the clinical scientists, but also by the referring clinicians (e.g. clinical geneticist), and increasingly for more complex exome and genome based tests—by multidisciplinary teams comprising clinical scientists, and relevant clinicians, e.g. a cardiologist if the test is for a cardiac condition. The

routinization and greater use of genomic testing will place further emphasis on collaboration between a wider range of clinical specialities and clinical scientists in ensuring the interpretation of variants is supported through improved data sharing.

Data sharing—the importance of quality, quantity and variety

Many genetic tests are offered by most or multiple laboratories (e.g. *BRCA* gene testing is undertaken by 19 providers in the UK).[13] Differences in variant interpretation across centres can result in different clinical decisions and courses of care—often life altering and irreversible—being recommended for patients with the same rare variant. For a *BRCA* variant, these differences could range from taking no intervention where the variant is not deemed to be disease causing, to considering a prophylactic mastectomy where the same variant is interpreted as pathogenic.[14] In other scenarios the interpretation may have implications for reproductive decision making or for the care and management of family members who share the same variant. In other words, inadequate data sharing risks misdiagnosis, compromises patient safety and quality of care. This issue of conflicting and incorrect interpretations is not insignificant. One assessment of aggregated variant data in a US based database found that of the variants for which clinical interpretations have been submitted by more than one laboratory (12 895), 17% (2 229) had been interpreted differently by the submitters.[11] In another study, analysis of the ExAC database revealed that for a number of variants earlier reported to cause rare Mendelian disorders, many were subsequently found at implausibly high frequencies—incompatible with 'rare' disease.[7] Again, the possibility that these variants have been incorrectly classified as disease causing, may have profound implications in cases where the information has been used to guide diagnosis and management.

Thus data sharing across laboratories and hospitals is essential to support accurate and reliable diagnoses since this depends on (Table 1);

- The quantity of observations made (i.e. how has this variant been previously interpreted), and as these observations are rare at each individual centre they need to be aggregated nationally and internationally.
- Comparing the quality of diagnostic inferences made for rare events and resolving disparities in clinical interpretation.[15]
- Placing the individual in the correct biological context, i.e. understanding their variants in the context of similar genomic backgrounds. For example, a recent study found that African Americans undergoing genetic testing for hypertrophic cardiomyopathy were disproportionately likely to receive an incorrect diagnosis because of the historical dearth of control populations in genomic databases that include persons of diverse racial and ethnic backgrounds.[16]

Table 1 The scientific and clinical rationale for data sharing

Prevent a 'diagnostic lottery'	Whereby a patient's chances of receiving a diagnosis depends on whether the laboratory their test is referred to has access to information that could lead to their accurate diagnosis
Faster resolution of variants of uncertain significance	Consolidating information enables the more definitive interpretation of a variant as disease causing or not
Reduce risk of misdiagnosis	By identifying conflicting interpretations, and by reducing the chances of an uninformed interpretation
Keeping up with a rapidly evolving knowledgebase	Since exome and genome based approaches are revealing a greater number of novel variants, only some of which may be relevant to disease
Greater efficiency	By improving the quality and efficiency of diagnosis and reducing time spent by different testing centres trying to interpret the same variants

Harnessing data for genomic medicine—the essential elements and the key challenges

A number of key elements are necessary for effective data sharing and delivery of genomic medicine: an infrastructure for sharing variants, access to the various types of data that are needed to inform a diagnosis (including genomic, clinical and phenotypic data) and processes to integrate those data. Since data collection, sharing and integration can risk the patient's identity becoming known outside the group of people who are responsible for delivering care (risk of identification), or personal patient data being shared against the patient's will (leading to breaches of privacy, confidentiality, and the possibility of stigmatization and/or discrimination), we also consider what technical strategies and safeguards might be adopted to minimize the risks associated with sharing and integrating these data.

Infrastructure for sharing variants

Accessing information on rare variants is challenging on a practical level since data on 'rare' variants are by their nature scarce. There are a wide range of resources and initiatives that aim to make rare variant data more readily discoverable. ClinVar—supported by the US National Center for Biotechnology Information—is a publically accessible database of >158 000 interpretations of clinically relevant variants.[17] DECIPHER (DatabasE of genomiC variation and Phenotype in Humans using Ensembl) (Fig. 1)—hosted by the Sanger Institute, contains data from more than 24 000 patients who have consented to broad sharing, but also supports limited sharing via access-controlled consortia. BRCA Exchange is a global initiative to catalogue variants in the breast cancer associated genes *BRCA1* and *BRCA2*. There are numerous other disease and gene specific variant databases. Generally, these resources all rely on voluntary data submissions; predominately from either research-driven sequencing studies, or from clinical services undertaking patient genetic and genomic testing.

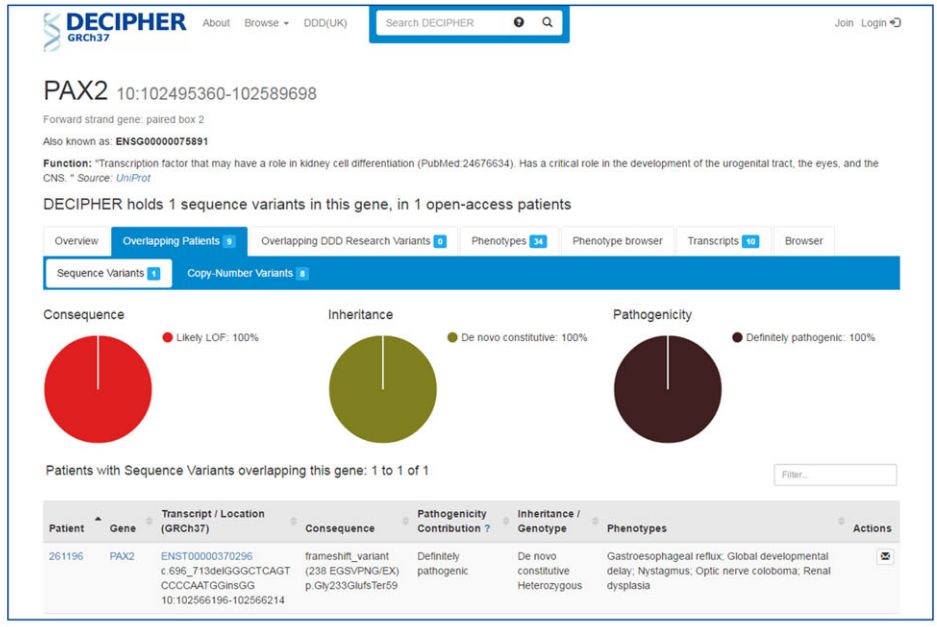
Despite the plethora of available databases and widely acknowledged necessity of collating genomic data, in practice data sharing activity by clinical genetic and genomic services is highly variable and a considerable volume of data remains siloed within individual testing centres. The reasons for this are manifold but principally include: limited technical and resource capacity to curate and upload data; concerns around the longevity and sustainability of third-party managed databases; and uncertainty around the legitimacy of sharing potentially identifiable patient data –especially into publically accessible databases.

In a research context, there is considerable variation in the extent to which genomic databases are publicly accessible. While research funders are increasingly requiring genomic data to be deposited into publicly accessible databases as a condition of funding, at the other extreme, participants of research projects such as the Personal Genomes Project have consented for their genomes to be made publicly available.[18] Moreover, in some countries such as the US, some clinical services rely heavily upon proprietary commercial databases, where access is limited to those users who purchase testing or interpretation services.

Processes for integrating data

Pooling of variant data from 'different' patients nationally and internationally is one critical element of data sharing for genomic medicine. The other fundamental requirement is the ability to integrate genomic and phenotypic data on 'individual' patients within the health system to inform the clinical interpretation of their genomic test results. Referrals to an individual laboratory for genetic and genomic tests may be received across a broad geographic spectrum. Therefore, patient data may have to be collated from across different locations. For some basic genetic tests, e.g. some single gene tests, the data collation and variant interpretation process is relatively straightforward in that it only requires a limited number of phenotypic and clinical details and the clinical expertise of just a few professionals. However increasingly, the interpretation of genomic tests for complex disorders can require more detailed phenotypic and clinical data as well as specialist input from a multidisciplinary team

(a)



(b)

DECIPHER ID	Variant	Sex	Size	Pathogenicity Contribution ?	Inheritance	Phenotype(s)
261030	2 166243484 166243484 T>A	46XY	SNV	Definitely pathogenic	De novo constitutive	Cerebral atrophy, Generalized myoclonic seizures, Generalized tonic-clonic seizures, Infantile encephalopathy, Intellectual disability, profound, Optic atrophy, Progressive microcephaly, Spasticity, Unerupted tooth
261786	2 166179687 166179690 TCCC>TCC	46XX	-1 bp	Definitely pathogenic	De novo constitutive	Autistic behavior, Bruxism, Delayed speech and language development, Postnatal microcephaly, Seizures, Short foot, Short palm, Stereotypy
264089	2 166245954 166245954 G>A	46XY	SNV	Definitely pathogenic	De novo constitutive	Edema, Global developmental delay, Hypoplasia of the corpus callosum, Progressive microcephaly, Seizures
280269	2 165798270 166304847 duplication	46XX	506.58 kb	Definitely pathogenic	Maternally inherited, constitutive in mother	Seizures
287585	2 166096116 169225805 deletion	46XY	3.13 Mb	Definitely pathogenic	De novo constitutive	Generalized myoclonic seizures, Global developmental delay
289855	2 162547702 167910910 loss	unknown	5.36 Mb	Definitely pathogenic	De novo constitutive	Generalized seizures
322847	2 163513067 168986256 deletion	46XX	5.47 Mb	Definitely pathogenic	De novo mosaic	Global developmental delay, Increased body weight, Precocious puberty in females, Seizures
259514	2 165578422 167108554 loss	46XX	1.53 Mb	Definitely pathogenic Full	De novo constitutive	Brachycephaly, Developmental regression, Generalized myoclonic seizures, Generalized tonic-clonic seizures, Intellectual disability, profound, Postnatal microcephaly, Short foot, Supernumerary nipple

Fig. 1 Example of how sequence variant data are recorded within the DECIPHER database. (**a**) An individual record of a sequence variant in the PAX2 gene. (**b**) A list of variants within genes and information related to the variant including associated phenotypes.

(MDT). The MDTs may comprise the referring clinicians, clinical geneticists, the clinical scientists, bioinformaticians, and other relevant clinical specialists—e.g. a cardiologist, or neurologist. The types and breadth of patient data that are relevant vary depending on the condition and these data may have been captured across different clinical specialities; examples include a patient's medical history and that of their relatives, the results of biochemical, physical, electrophysiological examinations, and medical images (MRI, CT scans). Appropriate infrastructure is needed to enable sharing amongst the multidisciplinary team who may be geographically distributed, and possibly also removed from the patient. Key to enabling the improved aggregation and exchange of phenotype data is the systematic, standardized and electronic capture of patient data, as well as software tools to facilitate the recording and analysis of phenotypes as well as computational tools to facilitate virtual MDT working, and to ensure that MDT deliberations are supported by appropriate and proportionate governance. While local informatics solutions and bespoke software are assisting in data capture and exchange,[19,20] arguably the lack of widespread implementation of interoperable electronic health records in the NHS remains the key technical bottleneck to patient record sharing.

Anonymization or de-identification of data

One way of reducing the potential risks associated with data sharing is to remove or obscure potentially identifying data to render the data 'de-identified' or anonymous. By disguising the source of the data, these technologies help to protect individuals from further discrimination or stigmatization. However the paradox is that 'as the utility of data increases, the privacy decreases'.[21] This is especially the case in genomics where data can be highly identifying when combined with other data, such as clinical data. Certainly the fact that genomic data can reveal information about other family members and also that it remains relatively static over time, leads some to claim, erroneously, that genomic data are inherently identifying.[22] Genomic data is not, of itself, identifying, but proportionate safeguards do need to be put in place to minimize the possibility of harms occurring. These harms could include that a person might be identified without their consent, or that the data might reveal something that they would prefer to keep private. Alternatively, the harm might be something more intangible such as people feeling that their human dignity has been infringed or that their individual or family privacy has been violated.[23,24]

A proliferation of different technologies have been developed to change data in order to prevent identification of an individual: these include removing specific identifiers such as name, date of birth, or descriptive clinical data. The disadvantage of irreversibly removing identifiers is that an individual cannot subsequently be identified, even if for good reasons, such as to act on indications that they are at increased risk of future disease. For this reason, codes are sometimes used to replace the identifiers which the process of anonymization removes. This allows linkage of data with an individual, using a key kept separately and securely, which enables relevant data to be linked with a person without identifying them. This process of 'pseudo-anonymisation' is also known as 'coding' and the data as 'coded data'.[25]

Limiting data access

Another pragmatic means of achieving proportionate genomic data sharing, is where the depth of data (what is shared) is weighted against the breadth of sharing (with whom).[26] Under this model, the most sensitive and potentially identifying data, e.g. exome/genome level data with detailed phenotypic descriptions of the patient, could be shared only with authorized users through an access-controlled system; minimal data, e.g. individual variants with a high-level phenotypic description could be shared more broadly—even publically—thereby increasing the discoverability of the rare variant but minimizing the risk of a privacy breach.

Legal and regulatory context

As well as using technical and pragmatic strategies to reduce the potential risks associated with data

sharing, more attention is being put on the regulatory and legislative framework that enables clinicians and others to share patient data safely and securely.

The extent to which data is identifiable forms the basis of how data is regulated in many countries. In the UK there are two main types of regulation that apply to genetic and genomic data: laws governing data protection which currently cover personal identifiable data but not coded or anonymized data, and the common law (or case law made by judges) which regulates how confidential data is used and shared. As genomic data is often generated as part of healthcare, the laws relating to confidentiality may often apply. Guidance is available from a number of different sources including relevant professional bodies and statutory authorities. For example, the General Medical Council guidance on confidentiality[27] sets out good practice in handling patients' information and establishes a framework for using and disclosing patient information for direct care and for secondary purposes including research. This guidance explicitly states that the familial nature of genetic data may justify disclosure of data in the public interest if failure to disclose information leaves others at risk of death or serious harm.[28]

The law on data protection is currently governed by the Data Protection Act (DPA) 1998,[29] which implemented the earlier EU Data Protection Directive.[30] The DPA regulates personal data which is defined as 'data which relate to a living individual who can be identified—(a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller' (defined as the person who determines the purposes and manner in which personal data are processed). Under this Act, genetic and genomic data are regulated to the extent that they qualify as sensitive personal data. In a medical context, processing for medical purposes is usually permissible if access is limited to health professionals or those who owe an equivalent duty of confidentiality. Data protection law will be changed by the forthcoming EU General Data Protection Regulation (GDPR)[31] which will come into force on 25 May 2018. The scope of this Regulation includes pseudonymized or coded data for the first time. In addition, special protections are required for certain types of data, including genetic and biometric data setting more onerous requirements for lawful processing of these types of data. This sets a baseline for minimum requirements, but Member States can adopt more stringent standards if they wish, suggesting that harmonization of data sharing practices across Europe will be unlikely, at least in the short term.

Within the UK, a number of different statutory authorities have responsibility for ensuring compliance with relevant legislation: the Information Commissioner's Office ensures compliance with the Data Protection laws and will be developing guidance to support the forthcoming GDPR.[32] In recognition of the importance that public trust and confidence play in the use of health data, the Government have supported the creation of a statutory National Data Guardian for Health and Social Care who advises and challenges the health and care system to help ensure that citizens' confidential information is safeguarded securely and used properly. This includes providing independent advice and support to government departments, health systems and health and social care providers to ensure that health data are stored and used lawfully and ethically and in ways that promote wider public trust. In a series of reviews, the National Data Guardian has highlighted some existing shortcomings, and has made a series of proposals designed to facilitate more effective and optimal data sharing practices. For example in the 2nd Caldicott Review,[33] Dame Fiona Caldicott emphasized how data sharing can enable better, safer care through implementing a 7th Data Sharing Principle—'the duty to share information can be as important as the duty to protect patient confidentiality' and in a more recent review has recommended adoption of revised security standards and model consent and opt-outs to streamline and legitimize use of personal identifiable data (including some genomic data) for secondary uses including research, audit and education.[34]

Empirical research on attitudes to data sharing more generally suggests that there is considerable

variation in peoples' willingness to share, and that some individuals have particular reservations about sharing some types of data (such as genetic data) for specific applications (such as use in insurance).[35]

Enabling genomic medicine in an evolving policy landscape

National developments in the U.K. and the NHS

As a result of the legislative reforms, existing data sharing practices are being examined to ensure that they comply with the more stringent requirements of the GDPR and relevant guidance. Additionally, plans are now underway to re-procure and consolidate genomic laboratory services in England, including specialist laboratory infrastructure—a process which is part of a wider Personalized Medicine Strategy for the NHS in England.[36] The reconfigured services are expected to incorporate a centralized whole genome sequencing provider, a network of 'Genomic Central Laboratory Hubs' providing routine diagnostic clinical sequencing and genome analysis, linked to local laboratories that provide services for rapid testing; all of which will be overseen by a 'National Genomics Data Centre' which will co-ordinate the sharing and collection of data between centres including variant level data.[37] Alongside these plans, the 100 000 Genomes Project, which is expected to complete patient recruitment in the next year, will provide a unique legacy through planned developments in clinical genomic services, although it is unclear whether data generated as part of this project will be accessible to the wider NHS in the future.

Realising the full potential of clinical genomics

Given this increasing momentum around genomic activity arguably genomic medicine has truly arrived in the U.K., and its impact on patient care is beginning to be realized. However the full extent of its success will be predicated on the ability of clinical services to share high-quality patient data, especially genomic variant data. The reconfiguration of genetic laboratory services presents a crucial opportunity to ensure that sustainable and robust data sharing processes are embedded within services. A fundamental element is a system that at the very least enables data sharing between the network of clinical genomic services involved in interpreting patient variant data.[38] However, wider data sharing could be justified, particularly for the interpretation of very rare variants, where the expertise of international disease consortiums might prove invaluable in informing variant pathogenicity. Although sharing data outside the UK

Genomic medicine glossary: key terms

Anonymization	The irreversible delinking of identifying information from associated data
Bioinformatician	A practitioner of the interdisciplinary field 'bioinformatics' which combines concepts and knowledge from computer science, statistics and biosciences in order to manage, mine, visualize and analyse biological and medical data
Coding/pseudonymization	The act of replacing an identifier with a code for the purpose of avoiding direct identification of the participant, except by persons holding the key linking the code and identifier
De-identification	The removal or alteration of any data that identifies an individual or could, foreseeably, identify an individual in the future
Exome	The protein coding regions of the genome (around 1–2% of the human genome)
Genome	The entire genetic material of an organism
Phenotype	The observable traits of an organism
Variant	A point or region in a sequenced genome that varies when compared to a 'reference' genome—a composite human genome sequence. Variants can be single DNA point (base) changes or larger deviations such as insertions or deletions of multiple adjoining bases

or using some types of data sharing infrastructure (such as cloud based systems) might raise additional regulatory concerns, the benefits of more systematic sharing might outweigh the potential disadvantages provided that some safeguards were adopted: for example, patients would need to be informed if their data was shared outside the UK to countries having less robust data protection regimes. Initiatives such as the Global Alliance for Genomics and Health[39] have been important in establishing exemplars for appropriate governance: this collaborative approach has generated regulatory frameworks and codes of practice as well as demonstrating technical and bioinformatics feasibility through application programming interface tools (APIs) and demonstration projects. This multidisciplinary approach has been important in developing and harmonizing policy at an international level. In general, the principles of proportionality and necessity should underpin any system which is adopted.[26]

The expected consolidation of genomic laboratories as well as the increasing demand for genomic testing will further accentuate existing challenges in sharing patient records as more referrals are received across hospital and geographic boundaries. The drive towards a paperless health system and interoperable electronic health records will be key in supporting more streamlined mechanisms for data exchange.

In addition to technical solutions and appropriate governance, in order for genomic medicine to be adopted beyond existing clinical genetics practice there must be the support and collaboration of medical professionals including those outside of clinical genetics.

Conclusion

In our view, to ensure that genomics becomes a routine component of clinical care will require three additional elements: recognition amongst mainstream clinical specialties of the utility that genomic analysis can bring for their specialisms; an increased recognition of the central role and importance of data sharing by healthcare professionals engaged in delivering care for patients; and finally, greater public understanding of the reasons why data sharing is necessary, and improved trust and confidence in the infrastructures, processes and people involved. This last component requires sustained effort from clinicians, patient groups, researchers and policy makers in order to earn public trust.

Conflict of interest statement

The authors have no potential conflicts of interest.

References

1. Life Sciences Minister George Freeman MP Leading the way for personalised medicine. http://www.newstatesman.com/2016/02/life-sciences-minister-george-freeman-mp-leading-way-personalised-medicine. in *NewStatesman* (2016).
2. Rehm HL. Evolving health care through personal genomics. *Nat Rev Genet* 2017;18:259–67.
3. Boycott KM, Vanstone MR, Bulman DE, et al. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013;14:681–91.
4. Wright CF, Fitzgerald TW, Jones WD, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* 2015;385:1305–14.
5. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med* 2015;7:100.
6. Saunders CJ, Miller NA, Soden SE, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med* 2012;4:154ra135.
7. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
8. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–24.
9. ACMG Board Of Directors. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet Med* 2017. doi:10.1038/gim.2016.196. [Epub ahead of print]

10. Chatzimichali EA, Brent S, Hutton B, et al. Facilitating collaboration in rare genetic disorders through effective matchmaking in DECIPHER. *Hum Mutat* 2015;36: 941–9.
11. Rehm HL, Berg JS, Brooks LD, et al. ClinGen—the clinical genome resource. *N Engl J Med* 2015;372: 2235–42.
12. Promoting gene testing together: fifth report of the UKGTN. (2017).
13. ACGS 2015–2016 genetic test activity audit. Association of Clinical Genomic Science. (2017).
14. Gradishar W, Johnson K, Brown K, et al. Clinical variant classification: a comparison of public databases and a commercial testing laboratory. *Oncologist* 2017. doi: 10.1634/theoncologist.2016-0431. [Epub ahead of print]
15. Harrison SM, Dolinsky JS, Knight Johnson AE, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med* 2017. doi:10.1038/gim.2017.14. [Epub ahead of print]
16. Manrai AK, Funke BH, Rehm HL, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med* 2016;375:655–65.
17. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;44:D862–8.
18. Ball MP, Thakuria JV, Zaranek AW, et al. A public resource facilitating clinical use of genomes. *Proc Natl Acad Sci U S A* 2012;109:11920–7.
19. Girdea M, Dumitriu S, Fiume M, et al. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat* 2013;34:1057–65.
20. Kohler S, Schulz MH, Krawitz P, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet* 2009;85:457–64.
21. Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLAL Rev* 2010; 57:1701–77.
22. Ayday E, Cristofaro E, Jean-Pierre H, et al. Whole genome sequencing: revolutionary medicine or privacy nightmare? *IEEE Computer* 2015;48:58–66.
23. Kaye J, Kanellopoulou N, Hawkins N, et al. Can I access my personal genome? The current legal position in the UK. *Med Law Rev* 2014;22:64–86.
24. Wright CF, Middleton A, Barrett JC, et al. Returning genome sequences to research participants: Policy and practice. *Wellcome Open Res* 2017;2:15.
25. Global Alliance for Genomics & Health Data Sharing Lexicon. Available at; http://genomicsandhealth.org/files/public/GA4GH_DataSharingLexicon_Mar15%20%282%29.pdf. The Global Alliance for Genomics & Health. (2016).
26. Wright CF, Hurles ME, Firth HV. Principle of proportionality in genomic data sharing. *Nat Rev Genet* 2016; 17:1–2.
27. Confidentiality: good practice in handling patient information. General Medical Council (2017).
28. GMC confidentiality guidance ibid. Paragraphs 73–76. General Medical Council. (2017).
29. Data Protection Act c.29. (1998).
30. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
31. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, a repealing Directive 95/46/EC (General Data Protection Regulation).
32. https://ico.org.uk/for-organisations/data-protection-reform/. The Information Commissioner's Office. Accessed on 8 May 2017.
33. Information: To share or not to share? The Information Governance Review. National Data Guardian for Health and Care. (2013).
34. Review of Data Security, Consent and Opt-Outs. National Data Guardian for Health and Care. (2016).
35. The One-Way Mirror: Public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. Ipsos MORI. (2016).
36. Improving Outcomes through Personalised Medicine. NHS England. (2016).
37. Genomic Laboratory Service Re-Design Service Specification. NHS England. Available at https://www.engage.england.nhs.uk/consultation/genomic-laboratories/. Accessed on 8 May 2017
38. Raza S, Hall A, Rands C, et al. Data sharing to support UK clinical genetics and genomics services. PHG Foundation & Association of Clinical Genetic Science. (2015).
39. http://genomicsandhealth.org/. The Global Alliance for Genomics and Health. Accessed on 8 May 2017.