










OPEN LETTER

REVISED **Genomic variant sharing: a position statement [version 2; peer review: 2 approved]**

Caroline F. Wright ¹, James S. Ware ², Anneke M. Lucassen ³, Alison Hall⁴, Anna Middleton ^{5,6}, Nazneen Rahman ⁷, Sian Ellard ¹, Helen V. Firth ^{8,9}

¹Institute of Biomedical and Clinical Science, University of Exeter, Exeter, UK

²National Heart and Lung Institute, Imperial Centre for Translational and Experimental Medicine, London, UK

³Department of Clinical Ethics and Law, Faculty of Medicine, University of Southampton, Southampton, UK

⁴PHG Foundation, Cambridge, UK

⁵Faculty of Education, University of Cambridge, Cambridge, UK

⁶Connecting Science, Wellcome Genome Campus, Cambridge, UK

⁷Division of Genetics and Epidemiology, Institute of Cancer Research, UK, London, UK

⁸Department of Clinical Genetics, University of Cambridge Addenbrooke's Hospital Cambridge, Cambridge, UK

⁹Wellcome Trust Sanger Institute, Cambridge, UK

v2 **First published:** 05 Feb 2019, 4:22 (<https://doi.org/10.12688/wellcomeopenres.15090.1>)

Latest published: 04 Dec 2019, 4:22 (<https://doi.org/10.12688/wellcomeopenres.15090.2>)

Abstract





Sharing de-identified genetic variant data via custom-built online repositories is essential for the practice of genomic medicine and is demonstrably beneficial to patients. Robust genetic diagnoses that inform medical management cannot be made accurately without reference to genetic test results from other patients, population controls and correlation with clinical context and family history. Errors in this process can result in delayed, missed or erroneous diagnoses, leading to inappropriate or missed medical interventions for the patient and their family. The benefits of sharing individual genetic variants, and the harms of *not* sharing them, are numerous and well-established. Databases and mechanisms already exist to facilitate deposition and sharing of de-identified genetic variants, but clarity and transparency around best practice is needed to encourage widespread use, prevent inconsistencies between different communities, maximise individual privacy and ensure public trust. We therefore recommend that widespread sharing of a small number of genetic variants per individual, associated with limited clinical information, should become standard practice in genomic medicine. Information confirming or refuting the role of genetic variants in specific conditions is fundamental scientific knowledge from which everyone has a right to benefit, and therefore should not require consent to share. For additional case-level detail about individual patients or more extensive genomic information, which is often essential for individual clinical interpretation, it may be more appropriate to use a controlled-access model for such data sharing, with the ultimate aim of making as much information available as possible with appropriate governance.


Keywords

medical genomics, variant, data sharing, data ethics

Open Peer Review

Reviewer Status  

	Invited Reviewers	
	1	2
REVISED		
version 2	report	
published 04 Dec 2019		
version 1		
published 05 Feb 2019	report	report

1 **Gert Matthijs** , KU Leuven, Leuven, Belgium

2 **Christa L Martin**, Geisinger Health System, Danville, USA

Erin Rooney Riggs, Geisinger Medical Center, Danville, USA

Heidi Rehm, Massachusetts General Hospital, Boston, USA

Broad Institute of MIT and Harvard, Cambridge, USA

Brigham & Women's and Harvard Medical School, Boston, USA

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the [Wellcome Sanger Institute](#) gateway.



This article is included in the [Transforming Genetic Medicine Initiative \(TGMI\)](#) gateway.

Corresponding authors: Caroline F. Wright (caroline.wright@exeter.ac.uk), Sian Ellard (Sian.ellard@nhs.net), Helen V. Firth (hvf21@cam.ac.uk)

Author roles: **Wright CF:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ware JS:** Writing – Original Draft Preparation, Writing – Review & Editing; **Lucassen AM:** Writing – Original Draft Preparation, Writing – Review & Editing; **Hall A:** Writing – Review & Editing; **Middleton A:** Writing – Review & Editing; **Rahman N:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ellard S:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Firth HV:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Transforming Genomic Medicine Initiative [200990].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Wright CF *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Wright CF, Ware JS, Lucassen AM *et al.* **Genomic variant sharing: a position statement [version 2; peer review: 2 approved]** Wellcome Open Research 2019, 4:22 (<https://doi.org/10.12688/wellcomeopenres.15090.2>)

First published: 05 Feb 2019, 4:22 (<https://doi.org/10.12688/wellcomeopenres.15090.1>)

REVISED Amendments from Version 1

We have revised our manuscript in light of reviewers' comments to clarify our terminology and recommendations, describe specific variant databases in more detail, and add further references to new work. We have responded point-by-point to reviewers' comments.

Any further responses from the reviewers can be found at the end of the article

Recommendations

1. Open and widespread sharing of plausibly causal genetic variants with high-level disease or organ-level information via appropriate online databases should be routine clinical practice and should not be dependent upon consent from individual patients.
2. It is good practice to maintain a cryptic link to the laboratory or clinical service that shared the genetic data, so that clinical follow-up remains possible should knowledge of the implications of a variant change or to combine data to build evidence.
3. Disclosing case-level clinical detail, large variant sets or genome-wide data may be crucial for variant interpretation, accurate diagnosis or clinical management, but requires explicit consent to share openly.

Introduction

Making an accurate diagnosis is the cornerstone of good medical practice, essential for determining prognosis, guiding treatment and informing patient management. Across all medical specialties, the interpretation of diagnostic test results relies upon knowledge of what is 'normal' in the population versus what 'disease' looks like. This knowledge relies upon sharing test results from previous patients and population controls. Without such data, the sensitivity and specificity of the test is unknown, its clinical utility is questionable, and its continued use may be harmful.

Genomic medicine is no exception to this rule, but determining what constitutes 'normal' and 'disease' can be extremely complicated and arguably the need for ongoing pooling of data is even greater than in other branches of medicine. Increasingly, clinical testing will rely on genome-wide sequencing, rather than targeted single-gene testing, and the enormous amount of normal variation in every genome¹ means that interpreting the results from one person's genome requires knowledge of many thousands of other genomes across different populations and ancestral backgrounds. Despite ongoing efforts to sequence large cohorts²⁻⁴, every genome examined contains novel changes not previously seen. For diseases with a substantial genetic component, caused by a specific rare variant or variants in an individual's genome, determining which variants are responsible for disease—and which are simply incidental, or play a minor role—is an enormous challenge. The only way to meet that challenge is by sharing data on individual variants with associated high-level disease or organ-level information that are not uniquely identifying.

Advantages of sharing genetic variant data

The main purpose of sharing individual genetic variants is to improve the diagnostic accuracy of genetic testing; the main data processors are clinicians and clinical scientists, and the main beneficiaries are patients and publics. Within this context, there are many benefits of sharing individual genetic variants associated with specific conditions⁵:

1. **Making accurate and safe diagnoses.** Genetic testing often benefits the individual patient undergoing testing, whose diagnosis can be accurately determined and prognosis further refined. Such genetic testing is dependent on being able to compare the variant of interest to variants from thousands of other people (via a database that is accessed by the scientist or clinician doing the analysis); at a minimum, this variant comparison is necessary to characterise and usually exclude variants that are relatively common in the general population. Variants of uncertain significance are regularly generated from genome-wide testing and can most easily be resolved through being able to access and explore the context in which such variants have been observed elsewhere (see [Figure 1](#))⁶. Numerous examples exist where making a successful genetic diagnosis has only been possible as a result of being able to access variant and phenotype data from other individuals undergoing testing⁷⁻¹¹, and many new genetic causes of disease have been uncovered this way^{12,13}. While most of the published cases are clinician-led, there are an increasing number of patient-led examples of variant sharing that have also catalysed the formation of disease-specific patient support groups and created new avenues of research^{14,15}.
2. **More effective disease management and precision medicine.** In some cases, an accurate genetic diagnosis leads to specific targeted therapies that can more effectively treat disease, or, in rare cases, may even reverse or prevent disease¹⁶⁻¹⁸. As a result of variant sharing, individuals may also be recruited to clinical trials that are tailored to their specific genotype, offering the potential for therapy where none currently exists¹⁹⁻²¹. In addition, new fundamental biological insights from genetic studies may identify novel targets for future therapies. Effective data sharing facilitates research across academia, clinical practice and industry and across different diseases and specialties²².
3. **Accurate advice for family members.** Due to the shared familial nature of most genetic variants, the benefits of making a robust genetic diagnosis may be cascaded out to biological relatives and have a profound impact on both existing and future generations. Consideration needs to be given to if and when communication of relevant information to relatives needs to take place, and the means by which this might be facilitated²³⁻²⁷.
4. **Improved understanding of genetic disease.** There are also wider benefits to the community, including

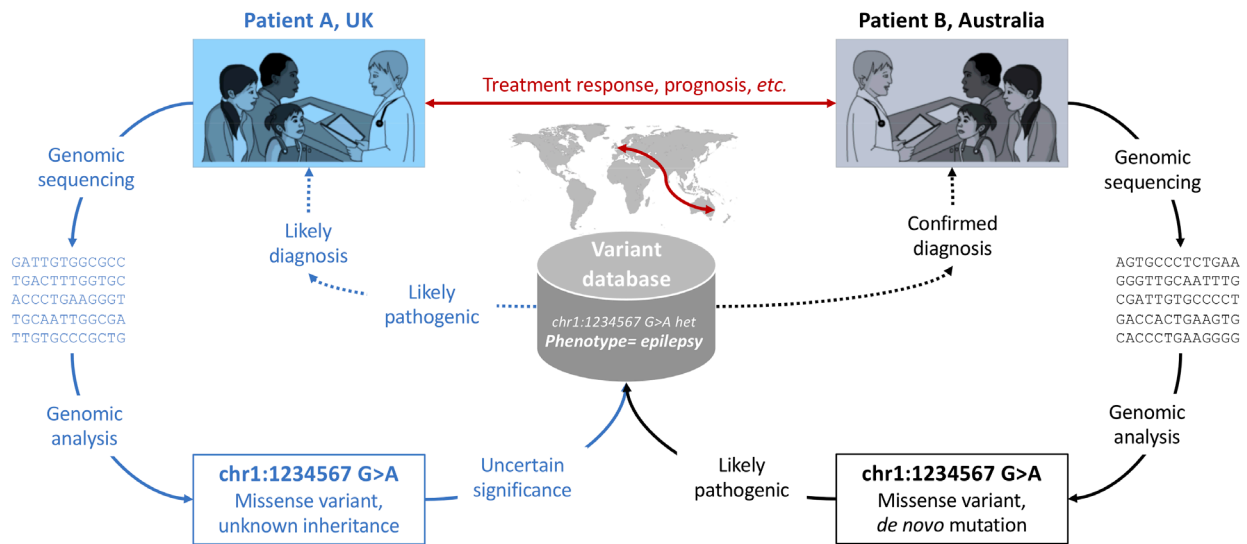


Figure 1. Global open variant sharing enables robust diagnoses to be made as quickly as possible; facilitating controlled sharing of detailed case-level information also informs clinical management and aids diagnosis in complex cases.

patients, clinicians and researchers across the globe, who are trying to understand and treat the causes of disease. Reporting new gene-disease associations, and sharing of variant-level information to discern which specific variants within each gene are pathogenic or benign or carry some degree of risk, is critical to advancing our understanding of genetic disease. Moreover, sharing variants together with phenotype, age and sex will allow an evolving understanding of incomplete penetrance and variable expressivity, improving interpretation of both diagnostic and predictive testing.

Disadvantages of not sharing genetic variant data

There is a substantial opportunity cost to not sharing clinically-oriented data that could otherwise be used to accelerate medical progress. The harms of *not* sharing individual genetic variants are well established and include delayed, missed and erroneous diagnoses, leading to inappropriate care^{28–31} and sometimes litigation^{32,33}. (See **Box 1** and **Box 2** for examples where variant sharing had a direct impact on clinical care.) Due to the familial nature of genetics, any diagnostic mistakes can easily be compounded by cascading erroneous information out to family members, thus multiplying the harms. Furthermore, without data sharing, research progress would be impeded, and the growing genomics knowledgebase—upon which the promise of personalised medicine is based—will stagnate.

Historical mistakes that exist in public variant databases³⁴ cannot be fixed without an influx of new data to allow reclassification of variants^{35,36}, without which misdiagnoses and errors in predictive algorithms will continue. Some international databases contain wrong and erroneous variant classifications³⁷, making such curation essential. Although it has been suggested that highlighting discordance in variant interpretation can be unhelpful

for clinical users^{37,38}, exposing discordant classifications allows laboratories and clinical services to work together to understand their differences, some of which may relate to incomplete penetrance of variants, and improve concordance^{28,39,40}. Organisations that actively maintain private genetic variant databases, such as commercial companies that do not share variant information for proprietary reasons⁴¹, are thus inhibiting diagnoses for other patients and undermining public health efforts in this area. Issues can arise where public databases are acquired by private companies, which despite being favourable for their survival may limit data access through prohibitive licencing fees.

Box 1. Example 1: The hazard of variant over-interpretation

In the early 2000's, a routine scan from a woman in her second trimester of pregnancy showed increased signal in the fetal bowel. This can be a sign of a chromosomal anomaly, viral infection or cystic fibrosis (CF) so an amniocentesis was offered. DNA analysis showed the fetus carried two *CFTR* variants that were said to be pathogenic. The parents were counselled that their baby would be affected by CF. They elected to continue the pregnancy.

After birth, the child was started on prophylactic antibiotics, twice daily physiotherapy, regular nebulisers and pancreatic supplements. Years later, the child was referred to the genetics clinic for review because the disease seemed unusually mild.

The clinical geneticist told the family that the status of one mutation had changed in the CFTR2 database and this combination was no longer thought to cause cystic fibrosis.

As a direct consequence of this change in variant interpretation, the child's prognosis changed from a life-limiting disorder to one of near-normal life expectancy and the day-to-day life of the child was transformed. The intensive regime of care was substantially reduced.

Box 2. Example 2: The need for population-specific variation data

A middle-aged Turkish man was referred to clinical genetics because he had colorectal cancer and numerous polyps were discovered at surgery. A homozygous variant in *MUTYH* was identified and reported to be of “unknown significance” in the diagnostic laboratory report. Biallelic *MUTYH* mutations cause *MUTYH*-associated polyposis (MAP), a recessive syndrome consistent with the diagnosis. Specific mutations are found at different frequencies in different populations.

Evaluation of available databases revealed that the variant had been identified once before in a patient with colon cancer and polyposis. Notably this second patient was also Turkish. No functional data were available and *in silico* analyses were inconclusive. The variant is extremely rare; present in only 7 individuals, all of South or East Asian origin, in the Exome Aggregation Data set of 61,486 individuals. However, no Turkish samples are listed as contributing to any of these datasets and no *MUTYH* or exome data from the general Turkish population is available.

Thus it is unclear whether this *MUTYH* variant is a pathogenic Turkish founder mutation or a non-pathogenic variant that is particularly prevalent in the Turkish population, but rare/absent in other populations. This lack of clarity presents significant clinical challenges in managing the patient and his relatives. **Sharing data generated in laboratories worldwide and across more ethnic groups would provide information to differentiate between these options and would allow clear classification of this and many other variants and reduce the potential for health disparities.**

Perceived harms of sharing genetic variant data

We have not been able to find any evidence that sharing data relating to individual genetic variants in the context of clinical applications causes harm. Nonetheless, perceived harms include re-identification of individuals across different datasets, loss of security of associated medical information (about the individual or their relatives), and the maleficent misuse of data^{42,43}. Early fears relating to genetic discrimination and the impact of genetic data on insurance premiums have not materialised in the UK and many other countries, thanks in part to genetic non-discrimination legislation and the Code on Genetic Testing and Insurance^{44,45}. Identification of an individual through knowledge of their genetic variant(s) is now perhaps the main concern. Although it is never possible to guarantee anonymity, and no data sharing system can be 100% secure, individual genetic variants—even very rare ones—are not uniquely identifying, and re-identification would require an intimate knowledge of the individual’s genotype or phenotype together with some information to trace that genotype/phenotype to a specific person. In practice, only an individual patient or their clinician would easily be able to re-identify themselves from a specific variant, neither of which would constitute a breach of confidentiality⁴⁶. A related concern is the perception that all genetic data are personal and therefore inherently sensitive, which stems from conflating genome-wide data with individual genetic variants.

Finding a balance

In our view, the definite and provable harms of *not* sharing genetic data outweigh the potential and largely hypothetical harms of sharing, a view that is corroborated by several recent litigation cases^{32,33} and supported by several large opinion surveys^{47,48}. Some empirical research has shown that patients and research participants support widespread data sharing^{48,49} and believe that the positive consequences outweigh the potential negatives⁴⁷. Clinical experience also suggests that, when the risks and benefits are explained to them and when invited to give consent, most patients are keen for their variant data and associated phenotypes to be shared. Recognising these benefits, **13 European countries have recently signed a declaration for delivering cross-border access to their genomic information**. Nonetheless, in our increasingly data-aware society, there is a perception that data sharing is inherently risky⁵⁰. A balance must therefore be struck between sharing sufficient data to reap the benefits, but only as much data as is needed to avoid the potential (perceived and actual) harms.

We have previously proposed a principle of proportionality in genetic data sharing, that balances the depth of data shared with the breadth of sharing⁵¹. With any dataset, decisions must be made about what specifically to share and how widely to share it. Many of the clinical benefits of data sharing in genetics can be realised by sharing a tiny subset of an individual’s de-identified genetic variants⁵², together with limited medical data, rather than necessarily whole genomes. This principle is in accordance with data privacy laws such as the new European General Data Protection Regulation (GDPR), which mandates that stored data are “*adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed*”⁵³.

The specifics of implementation are critical and agreeing standards for sharing variants and associated clinical data is essential. Specific data elements for sharing individual genetic variants have been outlined previously⁵⁴ and include (see [Table 1](#)):

1. a standardised genetic description of the variant(s), including Human Genome Variation Society (HGVS) nomenclature and genomic coordinates of the variant;
2. the variant classification and summary of evidence upon which that assertion was based;
3. the disease and inheritance pattern (e.g. dominant/recessive) upon which the clinical significance was asserted;
4. a standardised clinical description of the high-level disease phenotypes in the patient(s) that are included as supporting observations for the variant assertion, using appropriately controlled vocabulary/ontology; and
5. a cryptic or hidden link to the laboratory or clinical service that submitted the data, to enable further

Table 1. Example of genomic variant sharing.

	Variant 1	Variant 2
Variant	Standardised description of variant, including genomic coordinates	Standardised description of variant, including genomic coordinates
Gene	<i>e.g. MYH7</i>	<i>e.g. MYH7</i>
Genotype	Heterozygous	Heterozygous
Phenotype	Hypertrophic cardiomyopathy	Hypertrophic cardiomyopathy
ACMG/AMP variant-level evidence⁵⁵	PS1 – a different variant at the same position has previously been established to be pathogenic PM1 – occurs in the head of the protein (a functional domain with high probability pathogenicity) PM2 – absent from the general population PP3 – computational evidence suggests deleterious effect on gene product	PM1 – occurs in the head of the protein (a functional domain with high probability pathogenicity) PM2 – absent from the general population PP3 – computational evidence suggests deleterious effect on gene product
Interpretation (based on public data)	Likely pathogenic	Variant of uncertain significance
Aggregated case-level evidence	Observed in 1/10,000 individuals referred with diagnosis of HCM	Lab A – variant observed in 2/3,000 total cardiomyopathy patients sequenced Lab B – 2/4,000 Lab C – 1/3,000 Lab D – 1/1,000 patients
Interpretation (with variant sharing)	Likely pathogenic	Likely pathogenic

information to be requested and avoid data duplication but obscure the precise geographical location.

We recommend that openly sharing variant-level data, such as that included in Table 1, should be routine practice. No personal identifiers should be openly shared (e.g. name, hospital IDs, address, etc), and only the minimal genetic and clinical information required (as outlined in the five points above) to assist with interpreting a similar variant should be included. We recommend a cryptic link to the individual case-level data is maintained in a de-identified fashion via the laboratory or clinical service that submitted the data, that may obscure its precise geographical origin by deposition via another platform, to enable clinical follow-up if needed. Linking basic clinical information with information about genetic variation is crucial for supporting variant interpretation and aiding diagnoses. However, as with more extensive genome-wide data, or genomic risk scores, different levels of clinical detail will require different modes of sharing, i.e. open versus controlled access. Controlled sharing of more detailed phenotypes allows for more accurate diagnosis by enabling an independent evaluation of the clinical fit; if a diagnosis is simply stated in association with a variant, the validity of that association cannot be evaluated. Including this detailed clinical information with a genetic test result also avoids potential attrition, where individual clinicians need to go back to the original data generator to obtain sufficient information with which to make a diagnosis in their patient.

A flexible platform with broad international sharing of variant data together with national/local sharing of more granular phenotypic data would enable both needs to be addressed. Numerous databases already exist for collating and sharing genetic information, which may have differing requirements

for data deposition and thus offer different advantages and disadvantages. For example, US-based ClinVar^{56,57} is one of the largest genetic variant deposition databases, with >600,000 open access variants assayed primarily through laboratory genetic testing services, of which 60% of the >170,000 pathogenic/likely pathogenic variants have at least some supporting evidence, either as a written evidence summary and/or PubMed citations. UK-based DECIPHER^{10,58,59} is a global platform containing detailed case-level clinical data associated with >65,000 variants, of which 90% of pathogenic/likely pathogenic variants have associated phenotypes. DECIPHER uses a tiered access model whereby around half the cases are open access and half are accessible to members of closed groups to enable data-sharing that is compliant with local or national governance requirements. DECIPHER and many other variant databases internationally are now part of Matchmaker Exchange (MME), which was created to address the issue of data siloes by establishing “a federated network connecting databases of genomic and phenotypic data using a common application programming interface”^{7,8}. MME has facilitated gene discoveries that would not have been possible were the data from individual rare disease patients siloed in individual databases (see <https://www.matchmakerexchange.org/statistics.html>).

Establishing good practice

Uncertainty about what are permissible types of genetic variant sharing and when explicit consent is required means that current data sharing practices across regional genetics centres are highly variable⁴⁶. The inclusion of genetic data within Article 9 of the European GDPR, “*Processing of special categories of personal data*”, has created further confusion about the legality of sharing individual variants. There is therefore a need to establish and agree best practice⁶⁰ for data sharing within genomic medicine, to avoid inconsistent

practices across different regions, communities and jurisdictions, and ensure transparency and consistency when speaking to patients. Genetic variant data of the sort described above does not meet a recently proposed Data Sharing Privacy Test⁶¹, as the data is neither inherently sensitive nor uniquely identifying. Within the UK, the National Data Guardian has stated that “*the duty to share information can be as important as the duty to protect patient confidentiality*”⁶², a principle that applies to all data generated across the UK National Health Service. The American College of Medical Genetics and Genomics recently published a position statement in 2017 that “*laboratory and clinical genomic data sharing is crucial to improving genetic health care*”⁶³. However, genomic medicine is inherently a global enterprise, so more countries need to follow suit⁶⁴. The approach to data sharing espoused by the Global Alliance for Genomics and Health^{65,66} is rooted in international human rights legislation, focussing on our ‘solidarity rights’ to genomic information^{67,68} and emphasising the social good that can derive from appropriate data sharing. The handful of patients with the same rare diagnosis may be scattered across different countries, and are therefore best served when data are shared as openly and as widely as possible. Patients across the globe currently benefit from shared data and derived knowledge in databases such as ClinVar, DECIPHER and the Leiden Open Variation Database (LOVD)⁶⁹. Services that are not currently sharing their clinical data owe a substantial data debt and risk perpetuating current data biases.

Explicit consent should not be required for individual variant sharing

A recent analysis of the ethical principles that should guide genomic medicine services suggested that the “*use of genomic data for the advancement of medical knowledge should be permitted without explicit consent*”⁷⁰. In addition to variants from current and future patients, in whom the benefits of sharing vastly outweigh the potential harms, enormous swathes of legacy data exist from decades of patients who have undergone genetic testing. Some of these individuals are no longer alive and most are no longer in touch with their clinicians, making obtaining consent for data sharing impossible. Sharing variants from these tests could potentially benefit many thousands of patients without posing any risk of harm to the data subjects.

Although considering ownership of data has often been used as a route to determine what can be done with it, examining who controls access to the data is perhaps a more useful way forward than entering into ownership debates which, even if resolved, would not answer the question of what can legitimately be done with the data⁷¹. Individuals have a right to control access to data relating to them, but when it is not uniquely identifying and can benefit others without harming the individual—as is the case for genetic variants—rights of veto should be limited to the most unusual situations. A link between a particular genetic variant and associated disease is not personal information any more than the link between high blood cholesterol and heart disease, for example.

We therefore propose that patient consent should *not* be required in order to share variant-level data on individual genetic

variants, with minimal disease information⁵⁴. Agreeing this principle of “clinical variant-level sharing”⁵⁴ would remove the onus from data generators to ensure that they have the appropriate consents and permissions in place, and replace it with an unambiguous policy that is clear and transparent for both data generators and data subjects. In addition, we suggest that more detailed case-specific information generated within a particular healthcare system should initially remain within that healthcare system, sensitive to the quirks of each individual regulatory regime, but with the aim of eventual open data sharing following discussion with the patient and subject to their explicit consent.

Conclusions

All interpretation of genetic data is fundamentally dependent upon data sharing, since it is rarely possible to robustly demonstrate an association between a particular genetic change and a disease with an “N-of-one”. Therefore, sharing genetic variant data—albeit aggregated at some level and de-identified as far as possible—is inseparable from the practice of genomic medicine. Clinicians cannot treat patients appropriately if they cannot compare their patient’s data with data from healthy populations and other patients to establish a safe genetic diagnosis. It is therefore beholden upon those who generate and interpret genetic test results to allow access to relevant data as widely and as openly as possible, by depositing the data into appropriate databases and making it available to others to access whilst remaining compliant with local and national legislation and data governance. Numerous databases exist with aggregated genetic information, and although they differ in their deposition requirements and governance structures, ensuring interoperability between them through initiatives such as Matchmaker Exchange will prevent information silos and ensure longer-term sustainability.

Despite the overwhelming benefits of genetic variant sharing, and paucity of proven harms, there remain anxieties around deposition of individual genetic variants to open access databases. We propose that consent should not be required for widespread, open sharing of individual de-identified genetic variants linked with high-level phenotypes (i.e. associated disease or organ-level information), and that sharing such data should become standard practice in genomic medicine. We also recommend that richer case-level phenotypic detail (such as individual phenotype terms with age and other case-specific information) is shared within healthcare systems to facilitate robust diagnosis and that consent is routinely sought at the time of diagnosis to share such data openly. Ultimately, both the promise and the safety of genomic medicine will depend on our ability and willingness to share.

Data availability

No data are associated with this article

Acknowledgments

The authors wish to thank Fiona Cunningham, Ewan Birney, Matthew Hurles, David FitzPatrick, Graeme Black and Patrick Chinnery for helpful comments and input on this manuscript.

References

1. 1000 Genomes Project Consortium, Auton A, Brooks LD, *et al.*: **A global reference for human genetic variation.** *Nature.* 2015; **526**(7571): 68–74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Turnbull C, Scott RH, Thomas E, *et al.*: **The The 100 000 Genomes Project: bringing whole genome sequencing to the NHS.** *BMJ.* 2018; **361**: k1687.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Sudlow C, Gallacher J, Allen N, *et al.*: **UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.** *PLoS Med.* 2015; **12**(3): e1001779.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Lek M, Karczewski KJ, Minikel EV, *et al.*: **Analysis of protein-coding genetic variation in 60,706 humans.** *Nature.* 2016; **536**(7616): 285–291.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Raza S, Hall A: **Genomic medicine and data sharing.** *Br Med Bull.* 2017; **123**(1): 35–45.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Vears DF, Sénécal K, Clarke AJ, *et al.*: **Points to consider for laboratories reporting results from diagnostic genomic sequencing.** *Eur J Hum Genet.* 2018; **26**(1): 36–43.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Philippakis AA, Azzariti DR, Beltran S, *et al.*: **The Matchmaker Exchange: a platform for rare disease gene discovery.** *Hum Mutat.* 2015; **36**(10): 915–921.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Sobreira NLM, Arachchi H, Buske OJ, *et al.*: **Matchmaker Exchange.** *Curr Protoc Hum Genet.* 2017; **95**: 9.31.1–9.31.15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Boycott KM, Rath A, Chong JX, *et al.*: **International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases.** *Am J Hum Genet.* 2017; **100**(5): 695–705.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Chatzimichali EA, Brent S, Hutton B, *et al.*: **Facilitating collaboration in rare genetic disorders through effective matchmaking in DECIPHER.** *Hum Mutat.* 2015; **36**(10): 941–949.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Rehm H: **Rapid communication of efforts to resolve differences or update variant interpretations in ClinVar through case-level data sharing.** *Cold Spring Harb Mol Case Stud.* 2018; **4**(5): pii: a003467.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Deciphering Developmental Disorders Study: **Prevalence and architecture of de novo mutations in developmental disorders.** *Nature.* 2017; **542**(7642): 433–438.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. 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**(10): 681–691.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Might M, Wilsey M: **The shifting model in clinical diagnostics: how next-generation sequencing and families are altering the way rare diseases are discovered, studied, and treated.** *Genet Med.* 2014; **16**(10): 736–737.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Lambertson KF, Damiani SA, Might M, *et al.*: **Participant-driven matchmaking in the genomic era.** *Hum Mutat.* 2015; **36**(10): 965–973.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Finkel RS, Chiriboga CA, Vajsar J, *et al.*: **Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study.** *Lancet.* 2016; **388**(10063): 3017–3026.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Desmond A, Kurian AW, Gabree M, *et al.*: **Clinical Actionability of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Risk Assessment.** *JAMA Oncol.* 2015; **1**(7): 943–951.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Camp KM, Parisi MA, Acosta PB, *et al.*: **Phenylketonuria Scientific Review Conference: state of the science and future research needs.** *Mol Genet Metab.* 2014; **112**(2): 87–122.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Vat LE, Ryan D, Etchegary H: **Recruiting patients as partners in health research: a qualitative descriptive study.** *Res Involv Engagem.* 2017; **3**: 15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Grill JD: **Recruiting to preclinical Alzheimer's disease clinical trials through registries.** *Alzheimers Dement (N Y).* 2017; **3**(2): 205–212.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Thompson R, Johnston L, Taruscio D, *et al.*: **RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research.** *J Gen Intern Med.* 2014; **29**(Suppl 3): S780–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Conroy M, Sellors J, Effingham M, *et al.*: **The advantages of UK Biobank's open-access strategy for health research.** *J Intern Med.* 2019; **286**(4): 389–397.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Bredenoord AL, Kroes HY, Cuppen E: **Disclosure of individual genetic data to research participants: the debate reconsidered.** *Trends Genet.* 2011; **27**(2): 41–47.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Beskow LM, Burke W: **Offering individual genetic research results: context matters.** *Sci Transl Med.* 2010; **2**(38): 38cm20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Porsdam Mann S, Savulescu J, Sahakian BJ: **Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue.** *Philos Trans A Math Phys Eng Sci.* 2016; **374**(2083): pii: 20160130.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Royal College of Physicians, Royal College of Pathologists and British Society for Genetic Medicine: **Consent and confidentiality in genomic medicine.** (Royal College of Physicians and Royal College of Pathologists), 2019.
[Reference Source](#)
27. Parker M, Lucassen A: **Using a genetic test result in the care of family members: how does the duty of confidentiality apply?** *Eur J Hum Genet.* 2018; **26**(7): 955–959.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Harrison SM, Dolinsky JS, Knight Johnson AE, *et al.*: **Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar.** *Genet Med.* 2017; **19**(10): 1096–1104.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Shah N, Hou YC, Yu HC, *et al.*: **Identification of Misclassified ClinVar Variants via Disease Population Prevalence.** *Am J Hum Genet.* 2018; **102**(4): 609–619.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Moynihan R, Doust J, Henry D: **Preventing overdiagnosis: how to stop harming the healthy.** *BMJ.* 2012; **344**: e3502.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Manrai AK, Funke BH, Rehm HL, *et al.*: **Genetic Misdiagnoses and the Potential for Health Disparities.** *N Engl J Med.* 2016; **375**(7): 655–665.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Lucassen A, Gilbar R: **Alerting relatives about heritable risks: the limits of confidentiality.** *BMJ.* 2018; **361**: k1409.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Deborah L: **Lawsuit raises questions about variant interpretation and communication: Ambiguity of lab and clinician roles could be at issue if case proceeds.** *Am J Med Genet A.* 2017; **173**(4): 838–839.
[PubMed Abstract](#) | [Publisher Full Text](#)
34. Biesecker LG: **Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project.** *Genet Med.* 2012; **14**(4): 393–398.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Eccles DM, Mitchell G, Monteiro AM, *et al.*: **BRCA1 and BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance.** *Ann Oncol.* 2015; **26**(10): 2057–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Piton A, Redin C, Mandel JL: **XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing.** *Am J Hum Genet.* 2013; **93**(2): 368–383.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Vail PJ, Morris B, van Kan A, *et al.*: **Comparison of locus-specific databases for BRCA1 and BRCA2 variants reveals disparity in variant classification within and among databases.** *J Community Genet.* 2015; **6**(4): 351–359.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Gradishar W, Johnson K, Brown K, *et al.*: **Clinical Variant Classification: A Comparison of Public Databases and a Commercial Testing Laboratory.** *Oncologist.* 2017; **22**(7): 797–803.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Harrison SM, Dolinsky JS, Chen W, *et al.*: **Scaling resolution of variant classification differences in ClinVar between 41 clinical laboratories through an outlier approach.** *Hum Mutat.* 2018; **39**(11): 1641–1649.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Riggs ER, Nelson T, Merz A, *et al.*: **Copy number variant discrepancy resolution using the ClinGen dosage sensitivity map results in updated clinical interpretations in ClinVar.** *Hum Mutat.* 2018; **39**(11): 1650–1659.
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Conley JM, Cook-Deegan R, Lázaro-Muñoz G: **MYRIAD AFTER MYRIAD: THE PROPRIETARY DATA DILEMMA.** *N C J Law Technol.* 2014; **15**(4): 597–637.
[PubMed Abstract](#) | [Free Full Text](#)
42. McCormack P, Kole A, Gainotti S, *et al.*: **'You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research.** *Eur J Hum Genet.* 2016; **24**(10): 1403–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Middleton A, Milne R, Thorogood A, *et al.*: **Attitudes of publics who are unwilling to donate DNA data for research.** *Eur J Med Genet.* 2019; **62**(5): 316–323.
[PubMed Abstract](#) | [Publisher Full Text](#)
44. Joly Y, Feze IN, Song L, *et al.*: **Comparative approaches to genetic discrimination: chasing shadows?** *Trends Genet.* 2017; **33**(5): 299–302.
[PubMed Abstract](#) | [Publisher Full Text](#)
45. Joly Y, Ngueng Feze I, Simard J: **Genetic discrimination and life insurance: a systematic review of the evidence.** *BMC Med.* 2013; **11**: 25.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Raza S, Hall A, Rands C, *et al.*: **Data sharing to support UK clinical genetics and**

- genomics services.** PHG Foundation, 2015.
[Reference Source](#)
47. Mello MM, Lieou V, Goodman SN: **Clinical trial participants' views of the risks and benefits of data sharing.** *N Engl J Med.* 2018; **378**(23): 2202–2211.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 48. Goodman D, Johnson CO, Bowen D, *et al.*: **De-identified genomic data sharing: the research participant perspective.** *J Community Genet.* 2017; **8**(3): 173–181.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 49. Dheensa S, Fenwick A, Lucassen A: **'Is this knowledge mine and nobody else's? I don't feel that.' Patient views about consent, confidentiality and information-sharing in genetic medicine.** *J Med Ethics.* 2016; **42**(3): 174–179.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 50. Dankar FK, Badji R: **A risk-based framework for biomedical data sharing.** *J Biomed Inform.* 2017; **66**: 231–240.
[PubMed Abstract](#) | [Publisher Full Text](#)
 51. Wright CF, Hurles ME, Firth HV: **Principle of proportionality in genomic data sharing.** *Nat Rev Genet.* 2016; **17**(1): 1–2.
[PubMed Abstract](#) | [Publisher Full Text](#)
 52. Mourby M, Mackey E, Elliot M, *et al.*: **Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK.** *Computer Law & Security Review.* 2018; **34**(2): 222–233.
[Publisher Full Text](#)
 53. **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, and repealing Directive 95/46/EC (General Data Protection Regulation).** *Official Journal of the European Union.* 2016; **L119**: 1–88.
[Reference Source](#)
 54. Azzariti DR, Riggs ER, Niehaus A, *et al.*: **Points to consider for sharing variant-level information from clinical genetic testing with ClinVar.** *Cold Spring Harb Mol Case Stud.* 2018; **4**(1): pii: a002345.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 55. 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**(5): 405–424.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 56. Landrum MJ, Lee JM, Benson M, *et al.*: **ClinVar: public archive of interpretations of clinically relevant variants.** *Nucleic Acids Res.* 2016; **44**(D1): D862–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 57. Harrison SM, Riggs ER, Maglott DR, *et al.*: **Using ClinVar as a Resource to Support Variant Interpretation.** *Curr Protoc Hum Genet.* 2016; **89**: 8.16.1–8.16.23.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 58. Bragin E, Chatzimichali EA, Wright CF, *et al.*: **DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation.** *Nucleic Acids Res.* 2014; **42**(Database issue): D993–D1000.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 59. Swaminathan GJ, Bragin E, Chatzimichali EA, *et al.*: **DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders.** *Hum Mol Genet.* 2012; **21**(R1): R37–44.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 60. Shabani M, Dyke SOM, Marelli L, *et al.*: **Variant data sharing by clinical laboratories through public databases: consent, privacy and further contact for research policies.** *Genet Med.* 2019; **21**(5): 1031–1037.
[PubMed Abstract](#) | [Publisher Full Text](#)
 61. Dyke SO, Dove ES, Knoppers BM: **Sharing health-related data: a privacy test?** *NPJ Genomic Med.* 2016; **1**(1): 160241–160246.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 62. Chan T, Di Iorio CT, De Lusignan S, *et al.*: **UK National Data Guardian for Health and Care's Review of Data Security: Trust, better security and opt-outs.** *J Innov Health Inform.* 2016; **23**(3): 627–632.
[PubMed Abstract](#) | [Publisher Full Text](#)
 63. 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; **19**(7): 721–722.
[PubMed Abstract](#) | [Publisher Full Text](#)
 64. Scollen S, Page A, Wilson J, *et al.*: **From the data on many, precision medicine for "one": the case for widespread genomic data sharing.** *Biomed Hub.* 2017; **2**(suppl 1): 481682.
[Publisher Full Text](#)
 65. Cook-Deegan R, Ankeny RA, Maxson Jones K: **Sharing Data to Build a Medical Information Commons: From Bermuda to the Global Alliance.** *Annu Rev Genomics Hum Genet.* 2017; **18**: 389–415.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 66. Global Alliance for Genomics and Health: **GENOMICS. A federated ecosystem for sharing genomic, clinical data.** *Science.* 2016; **352**(6291): 1278–80.
[PubMed Abstract](#) | [Publisher Full Text](#)
 67. Knoppers BM, Harris JR, Budin-Ljøsne I, *et al.*: **A human rights approach to an international code of conduct for genomic and clinical data sharing.** *Hum Genet.* 2014; **133**(7): 895–903.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 68. Rahimzadeh V, Dyke SO, Knoppers BM: **An International Framework for Data Sharing: Moving Forward with the Global Alliance for Genomics and Health.** *Biopreserv Biobank.* 2016; **14**(3): 256–259.
[PubMed Abstract](#) | [Publisher Full Text](#)
 69. Fokkema IF, Taschner PE, Schaafsma GC, *et al.*: **LOVD v.2.0: the next generation in gene variant databases.** *Hum Mutat.* 2011; **32**(5): 557–563.
[PubMed Abstract](#) | [Publisher Full Text](#)
 70. Johnson SB, Slade I, Giubilini A, *et al.*: **Rethinking the ethical principles of genomic medicine services.** *Eur J Hum Genet.* 2019.
[PubMed Abstract](#) | [Publisher Full Text](#)
 71. Montgomery J: **Data Sharing and the Idea of Ownership.** *New Bioeth.* 2017; **23**(1): 81–86.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 12 December 2019

<https://doi.org/10.21956/wellcomeopenres.17121.r37289>

© 2019 Matthijs G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Gert Matthijs 

Center for Human Genetics, KU Leuven, Leuven, Belgium

I thank the authors for constructively replying to my comments by clarifying all issues and complementing the manuscript where necessary. I have no further comments. As a community, we are indebted to the authors for sharing views and publishing a position statement on variant sharing.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetics, rare diseases, NGS diagnostics, societal issues of genomic medicine

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 11 March 2019

<https://doi.org/10.21956/wellcomeopenres.16463.r34792>

© 2019 Martin C et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Christa L Martin

Autism & Developmental Medicine Institute (ADMI), Geisinger Health System, Danville, PA, USA

Erin Rooney Riggs

Geisinger Medical Center, Danville, PA, USA

Heidi Rehm

¹ Massachusetts General Hospital, Boston, MA, USA

² Broad Institute of MIT and Harvard, Cambridge, MA, USA

³ Brigham & Women's and Harvard Medical School, Boston, MA, USA

Answers to the questions:

1. Is the rationale for the Open Letter provided in sufficient detail?

The authors state that the rationale for their recommendation to share genomic variants with limited clinical information is to encourage consistency and transparency among the genetics community, and to bolster the practice of genomic medicine, making it more beneficial to patients.

2. Does the article adequately reference differing views and opinions?

The article does provide a section outlining “perceived harms of sharing genetic variant data,” effectively outlining commonly proposed concerns about genomic data sharing. To be clear, our group, the Clinical Genome Resource (ClinGen), has also published articles with similar recommendations to those put forth by these authors on genomic data sharing, and we agree with the concepts presented in this manuscript. However, we are aware of at least one other “differing” opinion that was not represented here: the opinion that public data sharing highlights discordance in variant interpretation and is potentially confusing for clinical users ^{1,2}. Our group believes that exposing discordant classifications between laboratories is actually a *benefit* to data sharing, allowing laboratories to see where they differ and work together towards concordance ^{3,4,5}.

3. Are all factual statements correct, and are statements and arguments made adequately supported by citations?

In our general response, we note one place where we thought a factual statement was not completely accurate. The authors state: “...US-based ClinVar is perhaps the leading genetic variant deposition database.. but most variants have only very limited or no clinical information and no supporting evidence associated with them.” While it is true that most entries do not contain patient data, the majority (62%) of the more than 170,000 pathogenic/likely pathogenic variants in ClinVar have supporting evidence, either as written evidence summaries and/or PubMed citations.

There were also some other places in the document where additional context would be useful (for example, in order for the reader to understand the nuances between variant-level and case-level data sharing or to further explain the difference between the disease upon which a claim of variant pathogenicity was made and the phenotypic features presenting in an individual patient). These issues are noted in our general response.

4. Is the Open Letter written in accessible language?

Yes

5. Where applicable, are recommendations and next steps explained clearly for others to follow?

Providing more clarity would be helpful regarding “next steps” for readers to follow, particularly in regards to where variant information could be submitted.

General report:

Wright and colleagues have written an open letter to address genomic variant sharing. It is a thorough and excellent accounting of the rationale for this type of data sharing and we commend the authors for taking the time to thoughtfully review and provide guidance on this important topic. We have a few suggestions and some minor edits that could strengthen the article and provide additional guidance to the community.

Higher level comments and suggestions:

In the first section under “Recommendations” the first recommendation suggests only sharing “plausibly causal” genetic variants. We think this recommendation is insufficient and strongly encourage this guidance to include sharing of *all* variants that have been reviewed. The literature and databases are currently littered with false claims of causality/pathogenicity. It is critically important that we also share evidence on variants that are deemed benign or uncertain, or, at the case-level, deemed non-causal. Over three-quarters of ClinVar’s content is made up of variants classified as benign, likely benign, or variant of uncertain significance (VUS) and this data has been enormously useful to counter many of the false claims of pathogenicity from the literature.

In addition, there is a bit of conflating of the concept of variant-level versus case-level interpretation and it would be useful to better separate these concepts in the paper. We have previously defined “variant-level” information as the aggregation of all evidence and observations to define the pathogenicity of a variant (i.e., its capacity to cause disease)⁶. This may include evidence from a current case under investigation, but also takes into account all prior available data. However, whether a given variant is actually causal for the symptoms in a given patient is best called case-level interpretation and involves additional factors such as penetrance, a phenotype match with the relevant gene, and allelic information (e.g., recessive disease requires two alleles).

Related to this issue, we recommend in the section that outlines five specific data elements and references our prior publication⁶, that items 2 and 4 be swapped to start with the variant level claim and then include the patient phenotype as part of the supporting evidence. This approach is more in line with variant-level data sharing and our referenced publication, which should be distinguished from case-level sharing, which is also important, but requires additional considerations as the authors have pointed out. Similarly, the variant claim (e.g., pathogenic, benign) should be asserted against a disease, not the patient’s clinical features, which should be left to the case-level interpretation step. We have made suggested edits below to data elements 3 and 4 to better clarify these points (additions in bold, deletions indicate by strikethrough):

“Specific data elements for sharing individual genetic variants have been outlined previously⁴² and include (see Table 1):

- 1. a standardised genetic description of the variant(s), including Human Genome Variation Society (HGVS) nomenclature and genomic coordinates of the variant;*
- 2. the clinical significance and summary of evidence upon which that assertion was based;*
- 3. the **disease and inheritance pattern of the disease (e.g. dominant/recessive) upon which the clinical significance is asserted;***
- 4. a standardised clinical description **of any of the clinical features in the patient(s) that are included as supporting observations for the variant assertion, using appropriately controlled vocabulary/ontology;**and*
- 5. a cryptic link to the laboratory or clinical service that submitted the data (to enable further information to be requested and avoid data duplication).”*

Next, the authors state, “We recommend a cryptic link to the individual case-level data is maintained in a de-identified fashion via the laboratory or clinical service that submitted the data, that may obscure its geographical location by deposition via another platform, to enable clinical follow-up if needed.” We think this topic requires further consideration of the benefits and drawbacks of obscuring the submitter’s location. For most laboratories that perform a large volume of testing and receive samples from geographically diverse locations, it seems unnecessary to obscure the geographical location of the laboratory and data; indeed, the geographical location of the laboratory is easily discernible given that individual variants are attributed to specific laboratories in databases, such as ClinVar. ClinVar has operated with transparency to the submitter and their location without harm for several years now. To the contrary, it can be helpful to recognize the potential for data duplication, which is not uncommon. We would suggest a more nuanced discussion of this topic. For example, one may consider obscuring geographical location only in instances where the population is small or geographically isolated.

Finally, it would be useful if the authors gave more concrete suggestions for where laboratories should submit their classified variants today. Do the authors support that direct submission to ClinVar is one recommended option? The authors describe both ClinVar and DECIPHER, but it is unclear what their recommendation would be. Given the momentum that ClinVar has achieved, it seems important that wherever the variant classifications are initially generated and stored, that they also be easily submitted to ClinVar. DECIPHER does have the advantage of a richer connection to case-level data. If DECIPHER took on a role as an additional site of primary variant deposition (not clear if it accepts individual submitted variant interpretations), we would assume the authors would agree that it would still be important for DECIPHER to facilitate submission to ClinVar on behalf of its users, in the same way that DECIPHER is able to fully consume ClinVar data. Would this be a second recommended option? More detail around any recommendations and/or future plans would likely be useful to readers.

Minor suggestions and edits:

In the abstract the authors state: “We therefore recommend that widespread sharing of a small number of individual genetic variants associated with limited clinical information should become standard practice in genomic medicine.” We assume the authors mean a small number “per individual” but we read it as stating that each “source/laboratory” should only share a small amount of data, in general. We suggest deleting “a small number of” and saving the nuance of per individual issue for later in the paper. Alternatively, you could reword the statement to read: “We therefore recommend that widespread sharing of a small number of an individual’s genetic variants associated with limited clinical information should become standard practice in genomic medicine.” This same issue occurs in the second paragraph of the section “Finding a balance”. In the sentence “...sharing a tiny subset of..” we suggest adding “an individual’s” after “of”.

In the last sentence of the abstract the authors state “For additional case-level detail about individual patients or more extensive genomic information, which is often essential for clinical interpretation, it may be more appropriate to use a controlled-access model for data sharing.....”. We fear this could be implied as abandoning the core suggestion if one wants to also share case-level data and therefore we suggest clarifying by adding “this additional” so the sentence reads “...it may be more appropriate to use a controlled-access model for this additional data sharing.....”

For the second Recommendation “A single genetic variant is not personally identifiable information; however, it is good practice to maintain a cryptic link to the laboratory or clinical service that shared the genetic data so that clinical follow-up remains possible should knowledge of the implications of a variant change.” We suggest adding “or to combine data to build evidence”. In our experience, many

variants change classifications once labs bring their evidence/observations together, but a source for contact is needed to communicate and bring the data together.

Another good example of patient benefit, and avoidance of harm, from data sharing is Grant et al, referenced below, in case you would like to cite⁷.

The authors state: “...US-based ClinVar is perhaps the leading genetic variant deposition database.....but most variants have only very limited or no clinical information and no supporting evidence associated with them.” This statement is not completely accurate. While it is true that most entries do not contain patient data, the majority (62%) of the >170,000 pathogenic/likely pathogenic variants have supporting evidence, either as a written evidence summary and/or PubMed citations.

References

1. Gradishar W, Johnson K, Brown K, Mundt E, Manley S: Clinical Variant Classification: A Comparison of Public Databases and a Commercial Testing Laboratory. *Oncologist*. **22** (7): 797-803 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Vail PJ, Morris B, van Kan A, Burdett BC, Moyes K, Theisen A, Kerr ID, Wenstrup RJ, Egginton JM: Comparison of locus-specific databases for BRCA1 and BRCA2 variants reveals disparity in variant classification within and among databases. *J Community Genet*. 2015; **6** (4): 351-9 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Harrison SM, Dolinsky JS, Chen W, Collins CD, Das S, Deignan JL, Garber KB, Garcia J, Jarinova O, Knight Johnson AE, Koskenvuo JW, Lee H, Mao R, Mar-Heyming R, McFaddin AS, Moyer K, Nagan N, Rentas S, Santani AB, Seppälä EH, Shirts BH, Tidwell T, Topper S, Vincent LM, Vinette K, Rehm HL, ClinGen Sequence Variant Inter-Laboratory Discrepancy Resolution Working Group: Scaling resolution of variant classification differences in ClinVar between 41 clinical laboratories through an outlier approach. *Hum Mutat*. 2018; **39** (11): 1641-1649 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Harrison SM, Dolinsky JS, Knight Johnson AE, Pesaran T, Azzariti DR, Bale S, Chao EC, Das S, Vincent L, Rehm HL: Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med*. **19** (10): 1096-1104 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Riggs ER, Nelson T, Merz A, Ackley T, Bunke B, Collins CD, Collinson MN, Fan YS, Goodenberger ML, Golden DM, Haglund-Hazy L, Krgovic D, Lamb AN, Lewis Z, Li G, Liu Y, Meck J, Neufeld-Kaiser W, Runke CK, Sanmann JN, Stavropoulos DJ, Strong E, Su M, Tayeh MK, Kokalj Vokac N, Thorland EC, Andersen E, Martin CL: Copy number variant discrepancy resolution using the ClinGen dosage sensitivity map results in updated clinical interpretations in ClinVar. *Hum Mutat*. 2018; **39** (11): 1650-1659 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Azzariti DR, Riggs ER, Niehaus A, Rodriguez LL, Ramos EM, Kattman B, Landrum MJ, Martin CL, Rehm HL: Points to consider for sharing variant-level information from clinical genetic testing with ClinVar. *Cold Spring Harb Mol Case Stud*. **4** (1). [PubMed Abstract](#) | [Publisher Full Text](#)
7. Rehm H: Rapid communication of efforts to resolve differences or update variant interpretations in ClinVar through case-level data sharing. *Cold Spring Harb Mol Case Stud*. 2018; **4** (5). [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for the Open Letter provided in sufficient detail?

Yes

Does the article adequately reference differing views and opinions?

Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Partly

Is the Open Letter written in accessible language?

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow?

Partly

Competing Interests: The reviewers are investigators who receive funding from NIH/NHGRI for the Clinical Genome (ClinGen) Resource project (U41HG006834), an initiative with a focus on data sharing. In addition, several authors on the manuscript under review participate in ClinGen working groups, and ClinGen and DECIPHER co-host an annual scientific conference.

Reviewer Expertise: Genomic variant curation and clinical interpretation, broad data sharing of variant and phenotypic data

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 02 Dec 2019

Caroline Wright, University of Exeter, Exeter, UK

Wright and colleagues have written an open letter to address genomic variant sharing. It is a thorough and excellent accounting of the rationale for this type of data sharing and we commend the authors for taking the time to thoughtfully review and provide guidance on this important topic.

We thank the reviewer for their positive comments.

We are aware of at least one other “differing” opinion that was not represented here: the opinion that public data sharing highlights discordance in variant interpretation and is potentially confusing for clinical users ^{1,2}. Our group believes that exposing discordant classifications between laboratories is actually a *benefit* to data sharing, allowing laboratories to see where they differ and work together towards concordance ^{3,4,5}.

This is a good point and we agree with the reviewer. We have added a sentence about this into the manuscript with these additional references.

In the first section under “Recommendations” the first recommendation suggests only sharing “plausibly causal” genetic variants. We think this recommendation is insufficient and strongly encourage this guidance to include sharing of *all* variants that have been reviewed. The literature and databases are currently littered with false claims of causality/pathogenicity. It is critically important that we also share evidence on variants that are deemed benign or uncertain, or, at the case-level, deemed non-causal. Over three-quarters of ClinVar’s content is made up of variants classified as benign, likely benign, or variant of uncertain significance (VUS) and this data has been enormously useful to counter many of the false claims of pathogenicity from the literature.

See earlier comment about this and our concerns around linking a variant to a phenotype that it does not cause.

In addition, there is a bit of conflating of the concept of variant-level versus case-level interpretation and it would be useful to better separate these concepts in the paper. We have previously defined “variant-level” information as the aggregation of all evidence and observations to define the pathogenicity of a variant (i.e., its capacity to cause disease)⁶. This may include evidence from a current case under investigation, but also takes into account all prior available data. However, whether a given variant is actually casual for the symptoms in a given patient is best called case-level interpretation and involves additional factors such as penetrance, a phenotype match with the relevant gene, and allelic information (e.g., recessive disease requires two alleles).

We have tried to clarify this throughout the text.

Related to this issue, we recommend in the section that outlines five specific data elements and references our prior publication⁶, that items 2 and 4 be swapped to start with the variant level claim and then include the patient phenotype as part of the supporting evidence. This approach is more in line with variant-level data sharing and our referenced publication, which should be distinguished from case-level sharing, which is also important, but requires additional considerations as the authors have pointed out. Similarly, the variant claim (e.g., pathogenic, benign) should be asserted against a disease, not the patient’s clinical features, which should be left to the case-level interpretation step. We have made suggested edits below to data elements 3 and 4 to better clarify these points (additions in bold, deletions indicate by strikethrough):

“Specific data elements for sharing individual genetic variants have been outlined previously⁴² and include (see Table 1):

- 1. a standardised genetic description of the variant(s), including Human Genome Variation Society (HGVS) nomenclature and genomic coordinates of the variant;*
- 2. the clinical significance and summary of evidence upon which that assertion was based;*
- 3. the **disease and inheritance pattern of the disease (e.g. dominant/recessive) upon which the clinical significance is asserted;***
- 4. a standardised clinical description **of any of the clinical features in the patient(s) that are included as supporting observations for the variant assertion**, using appropriately controlled vocabulary/ ontology;and*
- 5. a cryptic link to the laboratory or clinical service that submitted the data (to enable further information to be requested and avoid data duplication).”*

We have made these changes.

Next, the authors state, “We recommend a cryptic link to the individual case-level data is maintained in a de-identified fashion via the laboratory or clinical service that submitted the data, that may obscure its geographical location by deposition via another platform, to enable clinical follow-up if needed.” We think this topic requires further consideration of the benefits and drawbacks of obscuring the submitter’s location. For most laboratories that perform a large volume of testing and receive samples from geographically diverse locations, it seems unnecessary to obscure the geographical location of the laboratory and data; indeed, the geographical location of the laboratory is easily discernible given that individual variants are attributed to specific laboratories in databases, such as ClinVar. ClinVar has operated with transparency to the submitter and their location without harm for several years now. To the contrary, it can be helpful to recognize the potential for data duplication, which is not uncommon. We would suggest a more nuanced discussion of this topic. For example, one may consider obscuring geographical location only in instances where the population is small or geographically isolated.

This is an interesting point, but a more nuanced discussion is outside the scope of this paper. We have changed the text to suggest that the “precise” geographic location should

be obscured, but have not discussed it in further detail as this issue will vary between countries depending upon the catchment area of the testing laboratory.

Finally, it would be useful if the authors gave more concrete suggestions for where laboratories should submit their classified variants today. Do the authors support that direct submission to ClinVar is one recommended option? The authors describe both ClinVar and DECIPHER, but it is unclear what their recommendation would be. Given the momentum that ClinVar has achieved, it seems important that wherever the variant classifications are initially generated and stored, that they also be easily submitted to ClinVar. DECIPHER does have the advantage of a richer connection to case-level data. If DECIPHER took on a role as an additional site of primary variant deposition (not clear if it accepts individual submitted variant interpretations), we would assume the authors would agree that it would still be important for DECIPHER to facilitate submission to ClinVar on behalf of its users, in the same way that DECIPHER is able to fully consume ClinVar data. Would this be a second recommended option? More detail around any recommendations and/or future plans would likely be useful to readers.

We do not wish to prescribe where users should deposit their data, as many databases offer slightly different features. Instead, we support the federation of databases using systems such as MME, to ensure that data are not siloed. We have added a sentence about MME.

Minor suggestions and edits:

In the abstract the authors state: “We therefore recommend that widespread sharing of a small number of individual genetic variants associated with limited clinical information should become standard practice in genomic medicine.” We assume the authors mean a small number “per individual” but we read it as stating that each “source/laboratory” should only share a small amount of data, in general. We suggest deleting “a small number of” and saving the nuance of per individual issue for later in the paper. Alternatively, you could reword the statement to read: “We therefore recommend that widespread sharing of a small number of an individual’s genetic variants associated with limited clinical information should become standard practice in genomic medicine.” This same issue occurs in the second paragraph of the section “Finding a balance”. In the sentence “...sharing a tiny subset of..” we suggest adding “an individual’s” after “of”.

We have made these changes.

In the last sentence of the abstract the authors state “For additional case-level detail about individual patients or more extensive genomic information, which is often essential for clinical interpretation, it may be more appropriate to use a controlled-access model for data sharing.....”. We fear this could be implied as abandoning the core suggestion if one wants to also share case-level data and therefore we suggest clarifying by adding “this additional” so the sentence reads “...it may be more appropriate to use a controlled-access model for this additional data sharing.....”

We have made this change.

For the second Recommendation “A single genetic variant is not personally identifiable information; however, it is good practice to maintain a cryptic link to the laboratory or clinical service that shared the genetic data so that clinical follow-up remains possible should knowledge of the implications of a variant change.” We suggest adding “or to combine data to build evidence”. In our experience, many variants change classifications once labs bring their evidence/observations together, but a source for contact is needed to communicate and bring the

data together.

We have made this change.

Another good example of patient benefit, and avoidance of harm, from data sharing is Grant et al, referenced below, in case you would like to cite⁷.

We have added this reference.

The authors state: “...US-based ClinVar is perhaps the leading genetic variant deposition database.....but most variants have only very limited or no clinical information and no supporting evidence associated with them.” This statement is not completely accurate. While it is true that most entries do not contain patient data, the majority (62%) of the >170,000 pathogenic/likely pathogenic variants have supporting evidence, either as a written evidence summary and/or PubMed citations.

Thank you, we have updated the text with this information.

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 March 2019

<https://doi.org/10.21956/wellcomeopenres.16463.r34794>

© 2019 Matthijs G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Gert Matthijs

Center for Human Genetics, KU Leuven, Leuven, Belgium

Correct genomic variant interpretation and classification are an important and often complex issue, not only in research, but also, and even more critically, in genetic diagnostics and clinical care. Variant interpretation relies on different criteria and different in silico tools are available.

The authors rightly argue that sharing of genomic variants is essential for improving and facilitating variant interpretation. Indeed, patients will strongly benefit from open and well-managed databases. Databases may be federated, as long as a swift and controlled exchange of information is available.

Clearly, variant sharing is more than a clinical or technical issue, it is a societal issue: citizen – and not just patients – should be informed about the necessity and value of variant sharing. They should be convinced that variant sharing is essential and safe. It is a matter of solidarity and mutual interest to share data as broadly as possible. The principle of proportionality, in relation to potential harm, is certainly rightly applicable to variant sharing.

Comment on the abstract and beyond:

- The authors use ‘de-identified’ and ‘pseudonomised’, ‘cryptic link’ and a few other descriptions. It would be good to select the best term or definition and explain it to the readership.
- It is unclear what is meant in the abstract with “a small number of ...”. Small is hard to define.
- “Information robustly linking genetic variants with specific conditions is fundamental biological knowledge.” This is a significant statement that should be explored and explained in more detail, especially if the statement adds that it “should not require consent...”.

On the recommendations:

- In recommendation 1. The 'small number' from the abstract is not well reflected (or vice versa). What are 'high level' phenotypes, and why would sharing be limited to these?
- Recommendation states no consent in 1. and explicit consent in 3. This dichotomy is not presented in the Abstract. Again, the definition of 'small' is crucial, in all instances of policy, defining a 'cut-off' is a tricky thing.
- In general, what about sharing genomic variant that are excluded from disease, i.e. definitely not linked to the disease? The best example is in trans with a known dominant, pathogenic mutation. That information is equally useful.

On the Advantages:

For 2. Clearly, individuals may be identified in data bases by the genotype, for inclusion in clinical trials. With whom shall the data be shared? Companies? What would be the conditions? Who shall be the custodian? How to warrant and permit access? It would be nice to elaborate a bit on this. It is another aspect of variant sharing, that is not covered under the umbrella of variant interpretation.

For 3. The moral duty to help has been turned into a legal obligation in France. It may be interesting to cite this, as it is an example or situation that may pop up in other countries.

For documentation of the situation in France, please visit the following sites:

<https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006072665&idArticle=LEGI>

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000027592003&categorieLien=id>

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000029921462>

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000027592025&categorieLien=id>

For 4. How shall data be linked to natural history of disease, or vice versa?

On the Disadvantages:

- Ref 29 is not tightly linked to the issue of the proposed international sharing of data. Are there other cases/references?
- There are other, early papers on re-classification, e.g. Piton et al. 2013¹.
- It is probably worthwhile to mention that some international databases are 'contaminated' i.e. contain a wrong and erroneous variant classification. So the curation is essential. Variant database should explicit how the data is collected and managed. In the diagnostic arena, there is a consensus that HGMD data should be explicitly double-checked.
- The authors also mention the issue of private databases. It is unfortunate indeed that genetic and genomic analyses that are performed in often commercial laboratories do not make it to the public databases. Several large laboratories, mostly in the US, are committed to sharing data, like for instance via ClinVar. However, bad examples do exist as well, and have been denounced early on. References could be added, e.g. Conley et al 2014².

Also of note is that several databases, that were originally open, have been acquired by commercial companies. The latter has been favourable for their survival, however, the licencing fees are often prohibiting individual laboratories to obtain access. Equally, some companies offer access to their own clinical diagnostic databases, but again, the prices are mostly prohibitive. The data in private databases, especially these that are well curated, may be considered as having a value – as a result of intellectual or other efforts to generate good data – and thus come with a price. How to deal with this?

- In parallel, the public laboratories have not been very active in submitting variants. What kind of incentive would be needed to promote data sharing?
- The statement on informed consent hints at an important shift in the policy of Decipher to request consent. This policy was reportedly very strict in the early days. The position statement pleads for a relaxed (or no) requirement for a written consent. It would be interesting to read how and why the policy has evolved so significantly.

Finding a balance:

- What is the link between the text and Table 1. Table 1 does not list all the elements that are listed in the text. It would be good to explain to the reader what the aim of Table 1 is.
- Open versus controlled access: open access shall best be promoted, given the large number of labs that will either submit or consult.
- Open access databases are not necessarily free. Are there any other incentives to urge (diagnostic or research) laboratories to share variants? The latter are invited to submit in relation to publication, the former? Linking it to reimbursement of the test would be an option for laboratories operating in a 'fee for service' (public) health system, but would not be useful for private billing.
- What about using a model of clearing houses, to offer an incentive for submission? At some moment, funding and/or a financial model for maintaining the databases will be necessary.
- Page 6: LOVD shall be mentioned, as it fulfils the criteria listed in the text.

On information silos:

- It would be good to give a brief description and view point on how silos could be broken down or avoided. What about a model of federated databases?

References

1. Piton A, Redin C, Mandel JL: XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am J Hum Genet.* 2013; **93** (2): 368-83 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Conley JM, Cook-Deegan R, Lázaro-Muñoz G: MYRIAD AFTER MYRIAD: THE PROPRIETARY DATA DILEMMA. *N C J Law Technol.* 2014; **15** (4): 597-637 [PubMed Abstract](#)

Is the rationale for the Open Letter provided in sufficient detail?

Yes

Does the article adequately reference differing views and opinions?

Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Yes

Is the Open Letter written in accessible language?

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetics, rare diseases, NGS diagnostics, societal issues of genomic medicine

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Dec 2019

Caroline Wright, University of Exeter, Exeter, UK

The authors use 'de-identified' and 'pseudonymised', 'cryptic link' and a few other descriptions. It would be good to select the best term or definition and explain it to the readership.

We have replaced pseudonymised with de-identified throughout and defined it where it first appears as being a process whereby personal identifiers are removed and replaced with linked IDs. We have kept the term cryptic link, as this has a different meaning, but have changed it to "cryptic or hidden link" and explained its purpose e.g. to geographical location.

It is unclear what is meant in the abstract with "a small number of ...". Small is hard to define.

We did not intend to define a number, as it is the combination of a few variants (of variable number) with limited clinical information that together limits the extent to which this information is identifiable.

"Information robustly linking genetic variants with specific conditions is fundamental biological knowledge." This is a significant statement that should be explored and explained in more detail, especially if the statement adds that it "should not require consent...".

This statement is not intended to apply to personal information but to scientific knowledge in general and is a philosophical assertion. According to the Human Rights Act (see Bartha Knoppers' work on this) and public health systems such as the NHS in UK, we all have a right to benefit from science. We have slightly amended the sentence to reflect these points.

On the recommendations:

In recommendation 1. The 'small number' from the abstract is not well reflected (or vice versa).

What are 'high level' phenotypes, and why would sharing be limited to these?

Since very few variants will be causal in any individual, we have not recapitulated the "small number" in the recommendation. The term "high-level" phenotypes is intended to include disease or organ-involvement, and are thus not uniquely identifying even in combination; richer case-level phenotypes may be uniquely identifying, particularly in combination, and therefore may require consent. We have amended the text to clarify this point, though left the term in the recommendation for brevity.

Recommendation states no consent in 1. and explicit consent in 3. This dichotomy is not presented in the Abstract. Again, the definition of 'small' is crucial, in all instances of policy, defining a 'cut-off' is a tricky thing.

Revised, see above.

In general, what about sharing genomic variant that are excluded from disease, i.e. definitely not

linked to the disease? The best example is in trans with a known dominant, pathogenic mutation. That information is equally useful.

We thank the reviewer for this comment, though it raises some difficult questions. Open sharing of such data linked with phenotypes can be confusing. Although there are times when sharing all clinically evaluated variants can be helpful, we feel it is better to focus on a demarcation between disease databases (containing largely pathogenic variants) and population databases (containing largely presumed benign variants).

On the Advantages:

For 2. Clearly, individuals may be identified in data bases by the genotype, for inclusion in clinical trials. With whom shall the data be shared? Companies? What would be the conditions? Who shall be the custodian? How to warrant and permit access? It would be nice to elaborate a bit on this. It is another aspect of variant sharing, that is not covered under the umbrella of variant interpretation. ***We have added a sentence to the paper about this point. We have focused primarily on data sharing with researchers, whether they are clinical, academic or commercial, potentially for any condition.***

For 3. The moral duty to help has been turned into a legal obligation in France. It may be interesting to cite this, as it is an example or situation that may pop up in other countries.

For documentation of the situation in France, please visit the following sites:

We thank the reviewer for this helpful link, and we have added a reference to it into the manuscript.

For 4. How shall data be linked to natural history of disease, or vice versa?

The aim of this article is not to provide details on how data are to be linked, but to provide a conceptual analysis of the issues to facilitate policy in this area. Electronic health records are potentially one method for linking natural history of disease with genetic data, but there are others, and we do not wish to be prescriptive on this point.

On the Disadvantages:

Ref 29 is not tightly linked to the issue of the proposed international sharing of data. Are there other cases/references?

This reference (now Ref 32) relates to litigation due to variant interpretation and communication, which is directly relevant to data sharing and the point made in this sentence. We are not aware of other better references.

There are other, early papers on re-classification, e.g. Piton et al. 2013¹.

Reference added.

It is probably worthwhile to mention that some international databases are 'contaminated' i.e. contain a wrong and erroneous variant classification. So the curation is essential. Variant database should explicit how the data is collected and managed. In the diagnostic arena, there is a consensus that HGMD data should be explicitly double-checked.

We have further emphasised this point.

The authors also mention the issue of private databases. It is unfortunate indeed that genetic and genomic analyses that are performed in often commercial laboratories do not make it to the public databases. Several large laboratories, mostly in the US, are committed to sharing data, like for

instance via ClinVar. However, bad examples do exist as well, and have been denounced early on. References could be added, e.g. Conley et al 2014².

Reference added.

Also of note is that several databases, that were originally open, have been acquired by commercial companies. The latter has been favourable for their survival, however, the licencing fees are often prohibiting individual laboratories to obtain access. Equally, some companies offer access to their own clinical diagnostic databases, but again, the prices are mostly prohibitive. The data in private databases, especially these that are well curated, may be considered as having a value – as a result of intellectual or other efforts to generate good data – and thus come with a price. How to deal with this?

We have added a sentence about this point.

In parallel, the public laboratories have not been very active in submitting variants. What kind of incentive would be needed to promote data sharing?

We agree this continues to be an issue and is part of the motivation for writing this article. We acknowledge that public resources are often limited for this sort of activity, and laboratories have historically had variable levels of activity. However, we feel this emphasises the need for better infrastructures to enable fast and efficient data sharing, rather than necessarily creating unrealistic incentives. The motivation to share data already exists, and the majority of regional genetics laboratories in the UK have now submitted variants to a shared database. Moreover, variant sharing is increasingly included in professional best practice guidelines.

The statement on informed consent hints at an important shift in the policy of Decipher to request consent. This policy was reportedly very strict in the early days. The position statement pleads for a relaxed (or no) requirement for a written consent. It would be interesting to read how and why the policy has evolved so significantly.

The field has changed substantially over the last decade as the ubiquity of genetic variation has become more apparent and the rate of diagnoses has increased through large-scale genome-wide sequencing efforts. Decipher continually reviews its data-sharing policy to remain compliant with legal and ethical standards as they evolve and change but there has been no major recent shift in our policy.

Finding a balance:

What is the link between the text and Table 1. Table 1 does not list all the elements that are listed in the text. It would be good to explain to the reader what the aim of Table 1 is.

Table 1 contains examples of specific data elements for sharing individual genetic variants. We have now explained this more clearly in the text.

Open versus controlled access: open access shall best be promoted, given the large number of labs that will either submit or consult.

We agree.

Open access databases are not necessarily free. Are there any other incentives to urge (diagnostic or research) laboratories to share variants? The latter are invited to submit in relation to publication, the former? Linking it to reimbursement of the test would be an option for laboratories operating in a 'fee for service' (public) health system, but would not be useful for private billing. What about using a model of clearing houses, to offer an incentive for submission? At some moment, funding

and/or a financial model for maintaining the databases will be necessary.

Although we agree with the reviewer on this point, finding incentives for data sharing and solving the long-term funding issues associated with databases is beyond the scope of this paper.

Page 6: LOVD shall be mentioned, as it fulfils the criteria listed in the text.

We have added this and a reference.

On information silos:

It would be good to give a brief description and view point on how silos could be broken down or avoided. What about a model of federated databases?

We have added a point about federated databases and MME to the text.

Competing Interests: No competing interests were disclosed.
