nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-019-0528-2

In the format provided by the authors and unedited.

PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels

Antonio Rueda Martin[1,8], Eleanor Williams[1,8], Rebecca E. Foulger[1], Sarah Leigh[1], Louise C. Daugherty[1], Olivia Niblock[1], Ivone U. S. Leong[1], Katherine R. Smith[1,2], Oleg Gerasimenko[1], Eik Haraldsdottir[1], Ellen Thomas[1,3], Richard H. Scott[1], Emma Baple[1,4,5], Arianna Tucci[1], Helen Brittain[1], Anna de Burca[1,6], Kristina Ibañez[1], Dalia Kasperaviciute[1,2], Damian Smedley[1,2], Mark Caulfield[1,2], Augusto Rendon[1,7] and Ellen M. McDonagh[1,2]*

[1]Genomics England, London, UK. [2]William Harvey Research Institute, Queen Mary University of London, London, UK. [3]Imperial College Healthcare NHS Trust, London, UK. [4]Institute of Biomedical and Clinical Science, RILD Wellcome Wolfson Centre, University of Exeter Medical School, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK. [5]Peninsula Clinical Genetics Service, Royal Devon and Exeter Hospital, Exeter, UK. [6]Oxford Centre for Genomic Medicine, Oxford University Hospitals, Oxford, UK. [7]Department of Haematology, University of Cambridge, Cambridge, UK. [8]These authors contributed equally: Antonio Rueda Martin, Eleanor Williams. *e-mail: ellen.mcdonagh@genomicsengland.co.uk

PanelApp Crowdsources Expert Knowledge to Establish Consensus Diagnostic Gene Panels

Supplemental Information

Authors: Antonio Rueda Martin and Eleanor Williams, Rebecca E. Foulger, Sarah Leigh, Louise C. Daugherty, Olivia Niblock, Ivone U.S. Leong, Katherine R. Smith, Oleg Gerasimenko, Eik Haraldsdottir, Ellen Thomas, Richard H. Scott, Emma Baple, Arianna Tucci, Helen Brittain, Anna de Burca, Kristina Ibañez, Dalia Kasperaviciute, Damian Smedley, Mark Caulfield, Augusto Rendon, Ellen M. McDonagh

Content

Supplementary note

Method for creating virtual gene panels in PanelApp

1. Creating initial gene panels in PanelApp

There are many different genetic diagnostic laboratory services worldwide that use gene panels. To create an initial list of genes relevant to rare disease diagnosis, we selected four sources which provide high-quality gene–disease information (Supplementary Table 3). For a given rare disease, each source is manually queried with key words to identify genes that are tested for by that specific diagnostic laboratory. This query may include searching for phenotypes within the Genomics England rare disease eligibility criteria statement, or descriptions of the disorder in Orphanet and Genetics Home Reference (Supplementary Table 3). The resulting genes, along with any associated phenotype and mode of inheritance information, are collected and uploaded to PanelApp to create a new gene panel which is designated 'version 0'. At this stage, each gene in the panel is automatically given an initial confidence level rating based on the number of the four sources that the gene was included in. Different confidence levels are indicated by a traffic light system: if found in three or four of these sources the gene is Green (the highest level of confidence), two of these sources the gene is Amber (a moderate level of confidence), one of these sources the gene is Red (a lower level of confidence). Gene lists sourced from other diagnostic laboratories, research groups in particular disease areas, or other databases are also added to the panels if available, though these sources do not influence the initial automated confidence rating.

2. Crowdsourcing

Each gene on a panel can be reviewed by multiple registered reviewers, with their names and affiliations displayed for transparency and acknowledgement of their contribution. Reviewers are asked whether there is sufficient evidence for a gene to be classified as Green (a high level of evidence). Guidelines for the level of evidence required for a gene to be classified Green were based on existing ClinGen[1] and Deciphering Developmental Disorders (DDD)[2, 3] project gene guidelines to allow alignment of PanelApp datasets with other gene curation endeavours, and were adapted for classification of genes implicated in rare diseases (Figure 4 and supplementary note section below). It may be that the data for a particular gene is still inconclusive or that the gene is not relevant for the disorder; these genes should be classified Red by the reviewer. As well as rating the evidence for the gene by colour, reviewers are asked to provide information regarding the mode of inheritance, mode of pathogenicity, and whether they report variants within this gene as part of their diagnostic practice. To support their review, reviewers can add details of publications providing evidence of relevant cases, comment on the number of unpublished cases from their laboratories, and add phenotypes. Reviews are collected and displayed in real-time, with the date, time and version of the panel recorded, and a reviewer can revisit their review to edit or add new evidence.

The power of crowdsourcing

Crowdsourcing has been successfully used for other genomic curation endeavours, such as ClinVar and CIViC, to enable sharing of data and assessment of evidence for germline and somatic genetic variants[4,5]. A key advantage of crowdsourcing is that it can capture expert knowledge of variants or genes involved in rare disorders that is

not necessarily published or publicly-available. Moreover, many diagnostic laboratories have their own internal databases that record the variants and corresponding genes reported in their patients. Thus, there may be a wealth of data held in local files or private databases, and PanelApp provides a platform for these laboratories and groups to submit gene lists or reviews based on their experience and data. PanelApp is therefore a community effort and a database allowing this information to be collected, shared and unified benefits all, rather than this information remaining isolated in data silos.

Experts with different backgrounds can contribute reviews, for example clinicians, clinical scientists, researchers, bioinformaticians, scientific curators and students, encompassing experience from the clinic, academia and industry. When registering with PanelApp, we request that reviewers have expertise in one of the relevant disease areas or genes, and that they use their institutional email to register so that their credentials may be checked. Those who register to be a reviewer are further screened by the PanelApp curation team. This approach enables those involved in the 100,000 Genomes Project to provide their opinion and be involved in the genome analysis process, and expands the reach and impact of the project to the wider international scientific community. This outreach is particularly important for connecting with individuals with expertise in a given rare disease where there may only be a handful of experts, or a few published cases worldwide. Crowdsourcing of external reviews from experts in the clinical scientific community in PanelApp allows us to gather specific knowledge for each gene–disease relationship, and enables us to reach a community consensus of which genes are causative of a disease. As demonstrated in Extended Data Figure 3, there is a good consensus around gene rating; the majority of

reviews for Green and Red genes match the post-evaluation final rating. The PanelApp gene panels are publicly available, and the tools are open source, allowing clinical and research communities to access these panels and integrate them within their own genome/exome interpretation pipeline or analysis workflow.

Why be a reviewer?

Experts give their time and effort to contribute to PanelApp, and this is acknowledged by highlighting their contribution in an open access resource; reviews are displayed publicly on each panel with the name and affiliation of the registered user, allowing the reviewer to gain recognition for their contribution. Open reviews are important for initiating debate as well as acknowledging the time reviewers have taken to contribute to PanelApp (Figure 1). Reviewers can add their own publications and research findings. As more than one reviewer can review each gene, it allows experts to be part of a worldwide community, sharing different opinions and expertise, and providing opportunities for further collaboration. By contributing to this process, reviewers are helping to establish a final set of Green genes with a high level of evidence that are used for genome interpretation, and are therefore directly contributing to the diagnostic process for patients recruited in the 100,000 Genomes Project. As the panels are openly available, they can be used by the reviewers themselves, clinical laboratories within the National Health Service (NHS), or other international projects for interpretation of NGS results, and thus reviews have a wide impact that extends beyond the 100,000 Genomes Project. As the PanelApp gene panels are dynamic and have the ability to evolve as new evidence arises, we welcome further reviews on all panels to continually enrich the knowledgebase.

3. Establishing a diagnostic-grade virtual gene panel

As described in the main manuscript, to establish a final diagnostic grade set of Green genes that have a high level of evidence for use in genome analysis, Genomics England curators change the gene or genomic entity rating on a panel to reflect the overall evidence based on the PanelApp guidelines. This process takes into consideration expert reviews, published cases and information from existing publicly-available databases. Genomics England curators have a range of scientific and clinical backgrounds, and any genes/reviews that do not fit with the ruleset are discussed within the team to gain a clinical perspective and achieve a consensus. As well as evidence level, discussions also include the scope of the presenting phenotype of recruited patients and clinical utility to inform which genes are relevant to include on a panel.

4. Updates to gene panels

Curators also regularly scan key journals to identify relevant articles, look up ClinGen annotations, and OMIM[6] alerts, to add new genes or evidence to panels. An analysis of relevant Nature Genetics publications from May 2016 to May 2019 added to PanelApp rare disease panels was undertaken. Half the gene-publication annotations were added by an external reviewer and half by a member of the Genomics England curation or clinical team. Around 70% of these were added within 5 months of journal issue publication, and around 80% of the time, subsequent action (if relevant) was undertaken by a curator to update the panel within 4 months of the publication being added.

PanelApp guidelines for gene rating

Green Genes included in a Genomics England gene panel for a rare disease category should fit the criteria A-E outlined below. These guidelines were developed as a combination of the ClinGen DEFINITIVE evidence for a causal role of the gene in the disease[1], and the Developmental Disorder Genotype-Phenotype (DDG2P) CONFIRMED DD Gene evidence level[2,3] (please see the original references for full details). These help provide a guideline for expert reviewers when assessing whether a gene should be on the green or the red list of a panel.

A. There are plausible disease-causing mutations(i) within, affecting or encompassing an interpretable functional region(ii) of this gene identified in multiple (3 or more) unrelated cases/families with the phenotype(iii).
OR
B. There are plausible disease-causing mutations(i) within, affecting or encompassing cis-regulatory elements convincingly affecting the expression of a single gene identified in multiple (>3) unrelated cases/families with the phenotype(iii).
OR
C. As definitions A or B but in 2 or 3 unrelated cases/families with the phenotype, with the addition of convincing bioinformatic or functional evidence of causation e.g. known inborn error of metabolism with mutation in orthologous gene which is known to have the relevant deficient enzymatic activity in other species; existence of an animal model which recapitulates the human phenotype.
AND
D. Evidence indicates that disease-causing mutations follow a Mendelian pattern of causation appropriate for reporting in a diagnostic setting(iv).
AND
E. No convincing evidence exists or has emerged that contradicts the role of the gene in the specified phenotype.
(i)Plausible disease-causing mutations: Recurrent de novo mutations convincingly affecting gene function. Rare, fully-penetrant mutations - relevant genotype never, or very rarely, seen in controls. (ii) Interpretable functional region: ORF in protein coding genes miRNA stem or loop. (iii) Phenotype: the rare disease category, as

described in the eligibility statement. (iv) Intermediate penetrance genes should not be included.

PanelApp virtual gene panel principles

- For each rare disorder category, the panel should be a conservative "diagnostic grade" set of genes that out of the whole genome should be examined first, as variants within these genes are most likely to cause/explain the disease phenotype.
- A conservative list of genes of known clinical utility and scientific validity is required.
- We acknowledge that the diagnostic-grade Green gene list therefore will be missing genes that have been reported in association with the disease/phenotype but where the level of proof has not reached that required for it to enter use in a diagnostic setting. Variants that have passed the standard filtering criteria but are not within the Green gene panel for the relevant disease category will be assigned to a separate tier/rank.
- Genes included on the panel may have been screened in the patient previously; however, with whole genome sequencing, we may find variants of interest in regions not well covered by exome sequencing or missed by other methods. Therefore a gene panel typically contains genes listed in the prior genetic testing criteria.
- A single gene may appear in multiple gene panels.
- Genes may also be associated with other phenotypes not indicated in the gene panel.
- The gene panels will be updated as we learn from the 100,000 Genomes Project participant data and new published evidence.

An introduction to the 100,000 Genomes Project

The four main aims of the 100,000 Genomes Project are: 1) to create a transparent and ethical programme based on consent; 2) To establish a genomic medicine service for NHS England and bring benefit to patients; 3) Enable new medical insights and scientific discovery; and 4) To stimulate the development of a UK genomics industry. Thirteen Genomic Medicine Centres (GMCs) were founded across England to recruit patients, collect samples and associated clinical data required for analysis of the genomes, and validate pertinent findings prior to reporting back to participants. In addition to providing genome sequence reports to clinicians (and ultimately patients), the 100,000 Genomes Project is generating a wealth of data for research. Genomics England Clinical Interpretation Partnerships (GeCIP) domains are UK-led consortia that span a range of diseases such as 'renal', 'cardiovascular' or 'breast cancer' and cross-cutting themes including for example 'ethics' and 'population genomics'. GeCIPs bring together researchers and clinicians to interrogate and analyse the data generated within the 100,000 Genomes Project and ultimately improve the interpretation of genomes. To facilitate the translation of research into novel disease treatments, the Genomics England Discovery Forum was established. Members of the forum include biotechnology and pharmaceutical companies interested in identifying novel drug targets and therapeutic opportunities that emerge from analysis of the 100,000 Genomes Project dataset.

Specificity and sensitivity in genome analysis

Guidelines recommend the use of gene lists for whole genome analysis to aid disease-relevant interpretation and avoid the risk of discovering incidental findings[7].

However, a limitation to prioritising variants based on a virtual gene panel-based approach is that causative variants may lie outside of the applied panel. Another factor for consideration is the range of a patient's phenotypes that may not be covered by a single gene panel. Several studies have evaluated different approaches to the creation of gene panels and their effectiveness in variant prioritisation, for example the evaluation of HPO-driven panels versus the use of large gene panels or whole databases of gene-disease annotations[8]. Stark et al, (2017) demonstrated the added value of clinical input in guiding gene lists versus purely computational methods for variant prioritisation, increasing the efficacy of singleton whole exome sequence analysis and reducing the curation burden for lab scientists receiving results[9]. The approach we have developed for gene panels incorporates clinical expertise in gene panel creation, is adaptable to a large range of diseases and scalable for analysis of a large volume of whole genome or exome sequences.

In addition, we have introduced a number of strategies within our genome interpretation to address the limitations of a gene-panel based approach and mitigate against missed diagnoses. As mentioned in the main manuscript, this includes continual updates to virtual gene panels, shown to be key to identifying new diagnoses in other large-scale sequencing projects[10], and the introduction of PanelAssigner to apply additional panels based on a patient's HPO terms. In parallel to the tiering pipeline described in the main manuscript, Exomiser is applied as a panel-less approach to prioritise variants based on phenotypic characteristics[11]. Both tiering and Exomiser results are presented to the GMCs to allow them to efficiently prioritise the most likely candidates.

Challenges and lessons learnt

PanelApp gene review has stimulated discussion amongst groups and within clinics as to whether genes on diagnostic panels should be re-examined. For example, where only one case with a variant in the gene has been published, or where new data refutes or contradicts the historical evidence. An example is the *FANCM* gene present on many diagnostic laboratory panels, where the original association with fanconi anaemia has been refuted. We are fortunate to have encountered great enthusiasm from reviewers for being able to contribute to the knowledgebase. Experts are extremely busy individuals, and in acknowledgment of this the Genomics England curators are at hand to help reviewers make their contribution. Incorporating user feedback, the user interface and tools have been developed and optimised for the efficient identification of a suitable panel or gene, and to allow reviews to be easily added. PanelApp is being actively developed and improvements in the curation and review tools continue to be made. Large gene panels can be challenging to review, and so to address this we can provide reviewers with a formatted template file that can then be uploaded to capture their reviews in bulk, allowing reviewers to work off-line.

One challenge faced by curators when creating a panel is that the nomenclature used for diseases and genes can differ between sources. Although curators use a range of resources when creating and evaluating a panel, this may still mean that relevant genes are not identified during this initial step. In our experience expert review is therefore key in identifying missing or irrelevant genes, and clinical input is vital for deciding whether particular phenotypes are relevant for inclusion on the panel and whether the recruited patient(s) would present with the associated phenotype. This is

particularly true for broader panels which incorporate multiple phenotypes and disorders, such as ‘Rare multisystem ciliopathy disorders’.

Time constraints are always going to be a consideration for a manually-curated, high-quality database. In addition to the reviewers’ time, the evaluation of reviews by curators is time consuming, with manual curation and clinical input needed to revise and finalise gene panels. The Genomics England curation team prioritise their time by first investigating genes where reviews contradict the current gene rating. It is important for the curation team to keep up to date with new publications, discoveries, and any reclassification of disorders, with external reviews playing a vital role in this process. Curators monitor when an external reviewer leaves a new rating, search for new relevant articles in a subset of key journals and attend relevant conferences, to stay informed about the latest research and clinical opinion. If a version 1 panel undergoes substantial changes based on new data, the panel is promoted to the next major version (Figure 3), an example is the intellectual disability panel which is now version 2 (Extended Data 4).

Much time, effort, and resources are spent contacting potential reviewers, following up and helping individuals interested in contributing and volunteering their time. Resolution of reviews, further curation, and follow-up discussions that contribute to each panel before use in analysis is also an important part of the versioning process. Preliminary examination of genome analysis results for the 100,000 Genomes Project indicates that this process is valuable for establishing likely candidates for patient diagnoses.

Interoperability

To enable interoperability with other databases and allow for data integrity checks, PanelApp utilises HUGO Gene Nomenclature Committee (HGNC)-approved gene symbols and names, Ensembl gene IDs and genomic coordinates (Figure 1)[12, 13]. In the phenotype field, OMIM[6] disease names and identifiers are captured where possible. The mode of inheritance terms in PanelApp were established to remove ambiguity in the Genomics England bioinformatics pipeline and map to commonly used terms such as 'recessive' and 'dominant' which are used by other databases and in the literature. PanelApp data has been integrated into several other key databases, including the Open Targets platform (https://www.opentargets.org/) through mapping PanelApp collected phenotypes to the Experimental Factor Ontology (EFO)[14], Varsome[15], and links from DECIPHER gene pages[16]. We are currently working with members of the European Bioinformatics Institute (EBI) to enable users to plug-in PanelApp panels into the Variant Effect Predictor tool[17]. Links to OMIM, Gene2Phenotype and ClinVar are provided on PanelApp gene pages (Supplementary Table 3).

PanelApp is a member of the Gene Curation Coalition (GenCC), established by the Transforming Genetic Medicine Initiative (TGMI), together with other gene–disease curation efforts ClinGen, Gene2Phenotype, Orphanet, OMIM, HGNC and Genetic Home Reference (Supplementary Table 3). These resources all have independent requirements and audiences, and as a result differences exist between the diseases examined, how and what information is curated, and how data is accessed or displayed. The coalition has been established to produce a consensus for discrepancies between gene evidence ratings and develop standardisation to allow

mapping between resources. A major challenge in the curation of gene-disease evidence is keeping up-to-date with the latest published literature, and sharing curation efforts more effectively can help to address this. The criteria required for a gene to be Green in PanelApp were established based upon the ClinGen[1] and Deciphering Developmental Disorders (DDD)[2,3] project gene guidelines, and so are directly comparable between these resources.

Curation approaches for gene-disease evidence assessment

Curation approaches for gene-disease evidence assessment vary between the GenCC initiatives due to differences in their use-case, objectives and stakeholders. Manual review of the published literature is a foundation of the curation processes by the majority of the initiatives. This approach as the basis of curation was avoided for PanelApp due to time constraints and the resources required to comprehensively assess the literature for more than 200 rare disease categories in a short period of time to enable analysis of participant genomes for the 100,000 Genomes Project. However, information from these resources is incorporated into our curation process, as described above. ClinGen's gene-disease curations are based on a thorough review and collection of extensive information from the published scientific literature to enable semi-quantitative scoring of evidence via a standardised framework[1]. This is then expert reviewed, which can involve iterative rounds, to achieve a classification of clinical validity[1]. This is a comprehensive process for defining gene-disease validity level and provides an extremely valuable resource of gene-disease annotations for the community. However, a downside to this process is it very time and resource consuming. As of May 2019, there are 700 ClinGen gene-disease validity curations, covering 555 unique genes. With some recruitment categories in

the 100,000 Genomes Project having more than 2,000 genes associated with the disease (an example being intellectual disability), a thorough manual literature search and evidence framework curation method was not a suitable option for establishing our gene panels. Since launching in August 2015 up to May 2019, PanelApp has 29,282 gene-disease curations on panels for rare diseases in the 100,000 Genomes Project, covering 4526 unique genes. A drawback to the PanelApp curation process is that we may be missing published evidence that could prove or disprove a gene's role in disease that was not captured through the initial diagnostic test sources, expert review, and curation of information from other database resources. As described in the main manuscript and in Figure 3, additional curation processes are in place for dynamically updating panels in PanelApp, and for prioritising potentially diagnostic variants outside of a gene panel.

Although PanelApp is able to capture all types of evidence, the curation process for PanelApp does not mandate the extensive collection of information used in the ClinGen framework, for example, detail of the full extent of functional evidence underlying the gene association is not necessary for a gene to be considered Green if more than 3 unrelated cases/families have been reported. It is therefore a quicker process as curators capture the essential information for each gene required for the analysis pipeline that can be queried bioinformatically; panel name – gene rating – mode of inheritance, backed by evidence and based on our guidelines. Additional information may be collected and displayed in PanelApp, such as mode of pathogenicity, and though not essential for the genome interpretation pipeline, is valuable for curators reassessing evidence rating and for additional users of PanelApp. Not covered by the ClinGen gene-disease validity curation process, the

PanelApp curation process involves assessing a gene's clinical relevance on a panel to apply for genome analysis, not just evaluating the gene-disease evidence level. This allows variants to be prioritised based on a patient's primary indication as well as reducing the risk of identifying variants related to incidental findings. Decisions and actions made within the curation process are transparent and visible to PanelApp users.

Supplementary Table 1: Percentage coverage of OMIM by phenotype terms in PanelApp taken from different groups of genes

Ratings of PanelApp genes used in OMIM phenotype analysis	Percentage MIM identifiers matched to a MIM identifier in PanelApp
Green	44.7
Green and Amber	48.3
Green, Amber and Red	54.6

Method: Phenotype terms annotated to genes in PanelApp were taken from 189 public, version 1+ panels that are used in the 100,000 Genomes Project (for rare disease and cancer programmes) on 2018-11-08. OMIM[6] (https://www.ncbi.nlm.nih.gov/omim) MIM identifiers were extracted from these phenotype terms. These were then matched to the list of all OMIM identifiers that have the MIM Entry Type of 'phenotype' downloaded from OMIM (https://omim.org/downloads/, file mim2gene.txt, downloaded on 2018-11-08). The OMIM 'phenotype' category corresponds to entries with a # or % symbol before the entry number (# is a descriptive entry, usually of a phenotype, % is an entry describing a a confirmed mendelian phenotype or phenotypic locus for which the underlying molecular basis is not known).

Supplementary Table 2: Biotypes of genes in PanelApp

Biotype	% unique Green genes in public version 1+ panels	% unique genes in public panels (all ratings)
protein_coding	98.60	97.34
mt_tRNA	0.75	0.43
antisense RNA	0	0.04
snRNA	0.04	0.02
snoRNA	0	0.02
mt_rRNA	0.04	0.04
miRNA	0.04	0.12
vaultRNA	0	0.02
IG_C_gene	0.04	0.06
lincRNA	0.04	0.10
processed pseudogene	0	0.02
unprocessed pseudogene	0	0.06
transcribed unitary pseudogene	0	0.02
transcribed unprocessed pseudogene	0	0.04
processed transcript	0	0.06

Method: The biotype for each gene in PanelApp is sourced from Ensembl (version 90)[12]. All genes from public panels were downloaded from PanelApp on 2018-12-05, and biotype numbers calculated for unique gene entries.

Supplementary Table 3: PanelApp sources

Source	Reference
The 4 sources searched to create an initial gene list:	
Radboud University Medical Center Exome sequencing gene panels	https://www.radboudumc.nl/en/patientenzorg/onderzoeken/exome-sequencing-diagnostics
Illumina TruGenome Predisposition Screen	https://www.illumina.com/clinical/illumina_clinical_laboratory/trugenome-clinical-sequencing-services.html#tps
Emory Genetic Laboratory	https://www.egl-eurofins.com
UKGTN	https://ukgtn.nhs.uk/
Other sources frequently utilised during curation:	
OMIM	https://www.ncbi.nlm.nih.gov/omim
Gene2Phenotype	https://www.ebi.ac.uk/gene2phenotype
Orphanet	https://www.orpha.net
Genetics Home Reference	https://ghr.nlm.nih.gov/
HGNC	https://www.genenames.org/
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
Ensembl	https://www.ensembl.org
Genomics England rare disease eligibility criteria statement and data model documents	https://www.genomicsengland.co.uk/library-and-resources/
ClinGen	https://www.clinicalgenome.org/

Supplementary references

1. Strande NT *et al*. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* 100(6):895-906. (2017).
2. Wright CF *et al*. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*. 385(9975):1305-14. (2015).
3. The Development Disorder Genotype - Phenotype Database: https://decipher.sanger.ac.uk/ddd#ddgenes.
4. Landrum M.J. *et al*. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44(D1):D862-8. (2016).
5. Griffith M *et al*. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet*. 31;49(2):170-174. (2017).
6. Amberger, J.S. *et al*. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research.* Jan;43:D789-D798 (2015).
7. Matthijs, G. *et al.* Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet*. Oct;24(10):1515. (2016).
8. Kernohan KD *et al*. Evaluation of exome filtering techniques for the analysis of clinically relevant genes. *Hum Mutat*. Feb;39(2):197-201 (2018).
9. Stark, Z. et al. A clinically driven variant prioritization framework outperforms purely computational approaches for the diagnostic analysis of singleton WES data. *Eur J Hum Genet.* 25(11):1268-1272 (2017).

10. Wright C.F. *et al*. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet Med*. Oct;20(10):1216-1223 (2018).
11. Smedley D. *et al*. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. Dec;10(12):2004-15. (2015).
12. Zerbino DR *et al*. Ensembl 2018. *Nucleic Acids Res.* 46(D1):D754-D761. (2018).
13. Braschi B. et al. Genenames.org: the HGNC and VGNC resources in 2019. Nucleic Acids Res. 47(D1):D786-D792. (2019).
14. Malone J. *et al*. Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics.* 26(8):1112-8. (2010).
15. Kopanos *et al*. VarSome: The human genomic variant search engine. *Bioinformatics.* 1;35(11):1978-1980. (2019)
16. Firth, H.V. *et al*. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources. *Am. J. Hum. Genet*. 84, 524-533. (2009).
17. Thormann *et al*. Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat Commun*. May 30;10(1):2373. (2019).