Interpretation & Reporting of Clinical Diagnostic Next-Generation Sequencing Results

Dr David Hunt
Consultant in Clinical Genetics
Wessex Clinical Genetics Service

28th February 2017

Overview

- Next-generation sequencing (NGS) as a diagnostic tool
- Standardising variant nomenclature/terminology
- Prioritisation strategies for variant interpretation – the 100KGP approach
- The role of the genetics MDT
- Types of evidence for variant interpretation
- ACMG variant interpretation guidelines
- EuroGentest guidelines for diagnostic NGS services
- Key features of a lab report
- UK laboratory accreditation for clinical diagnostics

NGS as a Diagnostic Tool

- NGS tests are now very cheap and accessible
- At ~£600, an exome costs less than many conventional single gene tests
- Whilst NGS has its limitations, it offers significant benefits to the laboratory, clinician and patient alike
- It is, however, driving a paradigm shift in clinical practice – from the old model of deducing genotype from phenotype (with a high prior probability of pathogenicity for a given gene), to a new era of linking genotype back to phenotype

A Typical 'European' Genome

Variant type	Median number of autosomal variants
SNPs	3.53M
Indels	546k
Large deletions	939
CNVs	157
Nonsynon	10.2k
Synon	11.2k
Intron	1.68M
UTR	30.0k
Promoter	82.2k
Filtered LOF	149
HGMD-DM	18

The 1000 Genomes Project Consortium (2015)

Standardisation of Nomenclature/Terminology

- Official gene names, assigned by the HUGO Gene Name Committee (HGNC), should be used:
  - http://www.genenames.org/
- Variants should be described using the nomenclature approved by the Human Genome Variation Society:
  - http://varnomen.hgvs.org/
  - if in doubt, Mutalyzer (https://www.mutalyzer.nl/) is a useful online resource…

Reference Sequences

- Variants should be described in relation to an appropriate reference sequence:
- **NCBI RefSeq** accession number and version (e.g. NM_005859.4): https://www.ncbi.nlm.nih.gov/refseq/
- **Locus Reference Genomic (LRG)** sequence (now available for >1000 genes, and recommended by HGVS): http://www.lrg-sequence.org/
- **Genome build** (e.g. GRCh38/hg38)

Reference Sequences

- In addition to listing the reference sequence, the variant should be prefaced by a letter indicating the type of reference sequence:
  - **"c."** for a **coding DNA** sequence, e.g. c.578C>T
  - **"g."** for a **genomic** sequence, e.g. g.2456932G>A
  - **"m."** for a **mitochondrial** sequence, e.g. m.1555A>G
  - **"p."** for a **protein** sequence, e.g. p.(Leu525Pro)
- When a sequence variant may affect multiple transcripts, it is usually described in relation to the most clinically relevant transcript

Prioritisation

- The current 'bottle neck' in high-throughput sequencing is in variant interpretation and clinical reporting
- Given the potentially huge number of variants encountered, an effective strategy for prioritisation of variants is essential
- For most high-throughput sequencing strategies, phenotype-driven virtual gene panels are applied in the first instance in conjunction with various filters (for MAF, variant effect on the protein, inheritance, etc.)

100,000 Genomes Project (100KGP)

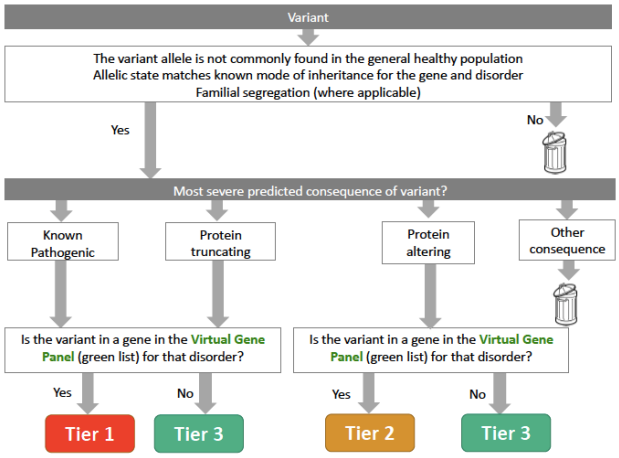
- Virtual gene panels (assembled using PanelApp) are applied to the WGS data by GeL, based on patient phenotype (expressed in Human Phenotype Ontology terms)
- Variants are essentially annotated, filtered and prioritised according to frequency, location (inside or outside of clinically relevant gene panel), genotype (allelic state), mode of inheritance and predicted consequence

100KGP

- Uses Ensembl's Variant Effect Predictor (VEP) to determine the most severe consequence any given variant will have across all transcripts

- Categorised into:
  - **High impact** (mostly protein truncating): nonsense, frameshift, canonical splice site, etc.
  - **Moderate impact** (mostly protein altering): missense, inframe insertions/deletions, splice region, etc.

100KGP 'Tiering' System



http://www.acgs.uk.com/media/1013939/11.45_-_01_-_validation_of_100_000_genomes_results__from_wgs_to_the_genomics_england_result_emma_baple_genomics_england_acgs_2016.pdf

100KGP

- Tier 1 & 2 variants need to be **clinically evaluated** by GMCs
  - ~0-2 expected per trio
  - ~0-20 expected per singleton
- Tier 3 variants do not need to be routinely evaluated (mainly of research interest)
  - ~10-20 expected per trio
  - ~100-400 expected per singleton

Clinical Interpretation Partners

- 100KGP has contracted a small number of commercial partners to aid in variant interpretation
- The Wessex Genome Medicine Centre (GMC) has been partnered with Congenica, and therefore receives 100KGP results through their interface, Sapientia
- Congenica are contracted to review all Tier 1 & 2 variants and check that appropriate panels have been applied in relation to the phenotypic terms
- They are able to apply additional gene panels, if appropriate, and also use Exomiser to rank variants in the dataset according to patient phenotype

Clinical Evaluation

- Within the Wessex GMC, it is intended that the **primary review** will be performed by a clinical geneticist and a clinical scientist (as well as the referring clinician, if necessary)
  - Decision is made whether the variant warrants further consideration in full genetics MDT
  - Limited genetics MDT capacity, so this cannot constitute a forum for primary review of every variant returned to us by 100KGP

The Wessex Genetics MDT

- Cases referred using a standard proforma
- Allocated amongst a pool of appropriately trained clinical scientists on a rolling basis for work-up
- Referral to MDT discussion, typically 2-3 weeks
- MDT held weekly and attended by a quorum of clinical scientists, clinical geneticists and our NHS bioinformatician
- 4-5 cases usually presented and discussed in 1 hour allocated slot
- Clinical scientist preparation time for a single case typically 4-5 hours
- Currently trialing **ACMG guidelines for the interpretation of sequence variants**
- MDT outcome form jointly signed off by a clinical scientist and clinical geneticist, and designed to be incorporated into patient medical records as a formal document

Some Key Questions

- Is the gene linked to a plausible phenotype?
- Does it fit with the expected inheritance?
- Is the type of variant consistent with a known mechanism of pathogenicity (e.g. dominant negative effect)?
- Is the allelic frequency of the variant compatible with the prevalence of the disease?
- Is the variant in a known functional region?
- What is the predicted effect on the protein?
- Is there any suggestion of an effect on splicing?

Some Key Questions

- For missense variants:
  - Does the substitution affect a well conserved amino acid (GERP score)?
  - Does it constitute a radical change in physicochemical properties of the residue at that position (Grantham distance)?
  - Is it within a functional domain of the protein?
  - Is it within a mutation hotspot?
- Is the variant previously reported in the medical/scientific literature or curated in a relevant database (Clinvar, HGMD Pro, LOVD, DECIPHER NHS Consortium, etc)?
- Has the variant been seen in 'control' databases (ExAC, dbSNP, 1000 genomes, EVS)?
- Do *in silico* tools (such as SIFT, PolyPhen, MaxEntScan, etc.) provide any supporting lines of evidence?

The ACMG Guidelines (2015)

ACMG STANDARDS AND GUIDELINES | Genetics inMedicine

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

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

ACMG Guidelines

- **Pathogenic criteria weightings:**
  - Very strong (PVS1)
  - Strong (PS1-4)
  - Moderate (PM1-6)
  - Supporting (PP1-5)
- **Benign criteria weightings:**
  - Stand-alone (BA1)
  - Strong (BS1-4)
  - Supporting (BP1-7)

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

acgs Association for Clinical Genetic Science
Part of the British Society for Genetic Medicine

- The ACGS issued practice guidelines in 2013 for variant interpretation and reporting

Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics.

Yvonne Wallis[1], Stewart Payne[2], Ciaron McAnulty[3], Danielle Bodmer[4], Erik Sistermans[5], Kathryn Robertson[6], David Moore[7], Stephen Abbs[8], Zandra Deans[9] and Andrew Devereau[6]

- However, in Nov 2016, all UK labs agreed to trial the ACMG guidelines

Classification System

- Both the ACMG and ACGS advocate a 5 class system, based on the system originally proposed by Plon et al (2008) for the IARC Unclassified Genetic Variants Working Group:

Class	Description	Probability of being Pathogenic
5	Definitely Pathogenic	>0.99
4	Likely Pathogenic	0.95–0.99
3	Uncertain	0.05–0.949
2	Likely Not Pathogenic or of Little Clinical Significance	0.001–0.049
1	Not Pathogenic or of No Clinical Significance	<0.001

- However, the current system is not truly quantitative

EuroGentest Guidelines

- EuroGentest (www.eurogentest.org):

  "EuroGentest is a project funded by the European Commission to harmonize the process of genetic testing, from sampling to counseling, across Europe. The ultimate goal is to ensure that all aspects of genetic testing are of high quality thereby providing accurate and reliable results for the benefit of the patients."

EuroGentest ESHG HARMONIZING GENETIC TESTING ACROSS EUROPE

EuroGentest Guidelines

European Journal of Human Genetics (2016) 24, 2–5

www.nature.com/ejhg

EJHG Open

POLICY

Guidelines for diagnostic next-generation sequencing

Gert Matthijs[*,1,8], Erika Souche[1,8], Mariëlle Alders[2], Anniek Corveleyn[1], Sebastian Eck[3], Ilse Feenstra[4], Valérie Race[1], Erik Sistermans[5], Marc Sturm[6], Marjan Weiss[5], Helger Yntema[4], Egbert Bakker[7], Hans Scheffer[4] and Peter Bauer[6]

We present, on behalf of EuroGentest and the European Society of Human Genetics, guidelines for the evaluation and validation of next-generation sequencing (NGS) applications for the diagnosis of genetic disorders. The work was performed by a group of laboratory geneticists and bioinformaticians, and discussed with clinical geneticists, industry and patients' representatives, and other stakeholders in the field of human genetics. The statements that were written during the elaboration of the guidelines are presented here. The background document and full guidelines are available as supplementary material. They include many examples to assist the laboratories in the implementation of NGS and accreditation of this service. The work and ideas presented by others in guidelines that have emerged elsewhere in the course of the past few years were also considered and are acknowledged in the full text. Interestingly, a few new insights that have not been cited before have emerged during the preparation of the guidelines. The most important new feature is the presentation of a 'rating system' for NGS-based diagnostic tests. The guidelines and statements have been applauded by the genetic diagnostic community, and thus seem to be valuable for the harmonization and quality assurance of NGS diagnostics in Europe.

European Journal of Human Genetics (2016) **24**, 2–5; doi:10.1038/ejhg.2015.226; published online 28 October 2015

EuroGentest Guidelines

- Devised for the evaluation and validation of various aspects of next-generation sequencing applications for the diagnosis of genetic disorders
- Endorsed by the Board of the European Society of Human Genetics on 1st July 2015, and published in EJHG in January 2016
- 38 key statements, 6 of which relate to 'reporting'

Matthijs et al. (2016)

Statement 26

"The report of a NGS assay should summarize the patient's identification and diagnosis, a brief description of the test, a summary of results, and the major findings on one page"

"It is essential that NGS results are reported in a clear and consistent manner, as laboratory reports may be read by both experts and nonexperts. From a practical standpoint, the clinically significant conclusions and the relevant test and test quality data should feature on the first page."

Matthijs et al. (2016)

Statement 27

"A local policy, in line with international recommendations, for reporting genomic variants should be established and documented by the laboratory prior to providing analysis of this type."

"All pathogenic (class 5) and likely pathogenic (class 4) variants have to be reported. Whether or not Unclassified Variants (UVs – class 3) are reported will depend on local practice. The latter has to be clear for the laboratory scientists, as well as for the referring clinicians."

Matthijs et al. (2016)

Statement 28

"Data on UVs have to be collected, with the aim to eventually classify these variants definitively."

"A community activity is needed to collect and share the available information, with the aim to definitely classify the variants into pathogenic (class 5) or benign (class 1)."

Matthijs et al. (2016)

Statement 29

"Laboratories should have a clearly defined protocol for addressing unsolicited and secondary findings prior to launching the test."

"The policy that has been adopted by the laboratory or institute, with respect to unsolicited and secondary findings, has to be reflected in the laboratory practice and in the report."

Matthijs et al. (2016)

Statement 30

"The laboratory is not expected to re-analyze old data systematically and report novel findings, not even when the core disease gene panel changes."

"A diagnostic request is a contract at a certain point in time. A laboratory will only be able to offer what is known, and validated, at a given point in time."

Matthijs et al. (2016)

Statement 31

"To be able to manage disease variants, the laboratory has to set up a local variant database for the different diseases for which testing is offered on a clinical basis."

"On the other hand, if at a particular moment, it is decided – by the labor by the community of experts in the disease – to change a variant from one class to another, the lab is responsible for reanalyzing the available data, re-issuing a report on the basis of the novel evidence, and also re-contacting referring clinicians for the patients that are possibly affected by the new status of the variant. A system effectively linking patients and variants, and allowing for the retrieval of the affected cases when variants are re-classified is necessary in such a situation."

Matthijs et al. (2016)

The lab report – some key features

- Appropriate patient identifiers
- Main clinical features or indication for the test
- Basic description of the test and when it was performed
- Main findings on front page, using HGNC gene name and HGVS nomenclature in relation to the most appropriate reference sequence
- The disease associated with the gene should be listed, as should the mode of inheritance
- Variant classification should be given using accepted phraseology (e.g. 'likely pathogenic' rather than 'class 4') to avoid confusion
- Summary of evidence supporting classification should be included

The lab report – some key features

- Confirmation that the main NGS finding has been validated by Sanger sequencing
- An assertion regarding the extent to which it might be contributing to the patient's phenotype
- General statements about:
  - any additional samples necessary to determine inheritance, if appropriate
  - offspring risk
  - Referral for genetic counselling
- Technical description of the test and limitations (small print)
- Percentage gene coverage data at minimum accepted read depth
- A list of any 'variants of unknown significance' that were identified may also be included in small print or tabular form
- Reports should be appropriately signed off

UK Laboratory Accreditation

- UK clinical diagnostic laboratories must be accredited by UKAS (ISO 15189 for Medical Laboratories)
- **External Quality Assessment (EQA)** is conducted by UK NEQAS
  - EQA is "A system of objectively assessing the laboratory performance by an outside agency. EQA is a system whereby a set of reagents and techniques are assessed by an external source and the results of the testing laboratory are compared with those of an approved reference laboratory. The main objective of external quality assessment is to establish inter-laboratory compatibility" (WHO 1981)

References

- **Matthijs et al.** (2016). Eur J Hum Genet 24: 2-5
- **Plon et al.** (2008). Hum Mutat 29: 1282-1291
- **Richards et al.** (2015). Genet Med 17: 405-24
- **The 1000 Genomes Project Consortium** (2015). Nature 526, 68-74