Irish Association of Clinical Scientists (IACS) draft guidance on clinical Next Generation Sequencing (NGS) testing

O'Brien CP,[1] and Mullaney, BG[2]

[1]Cancer Molecular Diagnostics, St James's Hospital, Dublin
[2]Haemostasis Molecular Diagnostics, St James's Hospital, Dublin

October 30, 2019

Summary

- This (draft) guideline is not intended to serve as a replacement for the substantial volume of guidelines in the public domain. The intent of this guidelines is to provide clarification on issues that are not clearly dealt with by one or multiple guidelines that may result in difficulties in interpretation for clinical laboratory accreditation.
- Recommended best practice guidelines are noted in the text and should be considered to be appropriate for the guidance of local service implementation and audit.
- This guideline does not specifically cover research studies or genotyping-only services covered by ISO17025. However, it is recommended that any research study covering inherited disease variants comply with best practice guidelines, preferably those endorsed by this document, and national legal requirements regarding genetic testing.
- This guideline does not deal specifically with microbial genome sequencing as this should not result in the generation of human genotyping data provided any human DNA in a mixed sample can be reliably excluded from analysis prior to variant calling.
- These guidelines are currently in draft form and feedback from clinical service users and laboratories is encouraged.

1 Clinical Service Considerations

1.1 The gene panel must be appropriate to the clinical service need

The clinical utility of a diagnostic panel, and each gene included on the panel should be discussed at the beginning of the NGS panel design. The criteria for including a gene into a panel have to be defined when developing a diagnostic test. Although NGS has led to improved diagnostic yield in comparison with traditional methods, an NGS assay may include newly emerging genes which have little related clinical knowledge or evidence. For diagnostic purposes, only established disease genes should be included in the analysis. There are a number of established guidelines for assessing the validity and utility of genetic tests (Haddow JE, 2003, Teutsch et al., 2009). More recently the Clinical Genome Resource (ClinGen) has developed a framework to define and evaluate the clinical validity of gene-disease pairs across a variety of Mendelian disorders (Strande et al., 2017)

1.2 The targeted regions-of-interest (ROIs) should be defined and documented

The reference sequence used for each gene should be documented and justified. If an Locus Reference Genomic (LRG) is available it should be used, alternatively a NCBI refseq should be chosen. The chosen reference sequence will then act as a basis for the design of the panel. For gene screening panels all coding regions of a gene should be included. Additionally a portion of the intron surrounding each coding exon should be included. At a minimum this should include the invariant splice donor/acceptor sites but it is good practice to extend 10 base pairs into each intron as this should detect the majority of intronic pathogenic variants (Association for Clinical Genetic Science, 2015). Care should be taken to include any well characterised pathogenic variants that may lie outside this minimally defined region (e.g.CFTR c.3718-2477C>T (legacy name: c. 3849+10 kb C>T)).

Part of a region of interest may not be adequately covered during validation (this may happen due to high GC content, pseudogenes, repetitive regions). If this occurs a redesign of the may boost the performance in that region (additional/shorter amplicons in PCR methods, additional tiling in hybridisation methods). If, after optimisation, a region continues to not be adequately covered the region should be targeted by an alternative method (e.g. Sanger) or else explicitly excluded from the assay. Regions that contain reported pathogenic variation should be targeted by alternative methods.

1.3 Reportable range should be appropriate to the clinical need and may include additional secondary analyses, if appropriate.

Regardless of the size of the sequencing panel, be it targeted, clinical exome, whole exome or indeed whole genome, it is highly likely that a virtual panel will be created to focus only on those genes with a high prior probability of disease causation. The composition of the panel analysed should be appropriate to the clinical query. Reports should clearly detail the absence of conclusive results for any gene on the panel, or the incidental findings list, that fails to meet the appropriate quality cutoffs.

Secondary analyses, unrelated to the original clinical query, e.g. inclusion of additional genes from the ACMG incidental findings lists (Green et al., 2017, Green et al., 2013, Kalia et al., 2017) should only be performed if indicated by a clinical geneticist.

1.4 Samples should not be accepted for testing outside of pre-agreed referral services

Laboratories accepting samples for the investigation of inherited disease must ensure that the referral is only accepted from clinicians who can ensure that proper care and genetic counselling is available to patients. A service level agreement or other documentation to support the appropriateness of referring services should be stored locally.

Reporting of genetic results for research purposes may carry the same implications as for clinical cases. Laboratories providing genetic analyses for research purposes which will be used to guide patient's treatment or diagnosis (e.g. for clinical trials) should follow the recommendations included in this document. This does not apply to anonymised research studies, which do not fall under the remit of ISO15189 accreditation.

1.5 Somatic mutation analysis should be designed to minimise anticipatable incidental findings

Clinically, somatic mutation profiling is used to identify acquired pathogenic mutations that are commonly found in cancer. A highly focused analysis of mutation hotspots in disease associated oncogenes is unlikely to yield a variant with implications for inherited disease. Laboratories applying NGS to tumour samples must also be aware of the potential for anticipatable incidental findings and should ensure that sufficient clinical supports are available for such cases.

Subtractive analysis describes the process of simultaneously testing the mutation status of both tumour samples and a matched blood sample. Mutations present in the blood are subtracted from the total of the mutations found in the tumour to leave a filtered list containing only acquired variants. This requires the identification of germline mutations which can 'subtracted' from the

complete mutation list to generate the list of somatic variants. Laboratories performing subtractive analysis should not perform direct germline analysis unless resourced to do so and should follow the guidelines suggested in this document if performing germline analysis.

1.6 Bacterial genome profiling may not require precautions employed for clinical genetics

Provided that human genetic material is not used for variant calling, and human NGS reads are not stored by the analysis platform it can be assumed that any data generated using NGS for bacterial or viral sequencing does not need to follow the same precautions as those identified for human germline sequencing.

2 Analytical platforms

Clinical laboratories may use NGS testing for a number of applications including but not limited to:

1. Inherited disease gene testing
2. Somatic mutation testing
3. Mitochondrial genome testing
4. Non-invasive prenatal testing (NIPT)
5. Pathogen detection and characterisation
6. Circulating tumour DNA testing

Within these applications whole genome, exome, targeted gene panels or known hotspot targeting methodologies may be used. A detailed review of the specific applications is outside the scope of these guidelines; however some key considerations are listed in the following text.

Target enrichment

The most common target enrichment methods are PCR-based and hybridisation-based methods. Both methods have limitations and these should be considered depending on the application. PCR-based methods may be preferable when the amount of input material is low (e.g. when DNA is extracted form FFPE material) or when a region of interest is small (e.g. hotspot panel, single gene test). Hybridisation-based methods may be preferable for larger gene panels and also when dosage analysis may be required.

PCR-based target enrichment is generally based on multiplexed PCR primers to generate PCR products with the required adapters for a given sequencing platform. In a similar manner to traditional PCR based methods, PCR based NGS methods are susceptible to allele dropout as a result of variations in primer

binding sites. PCR-based methods are also susceptible to amplification bias. Amplification bias can be overcome by the use of molecular barcoding which must be factored into the analysis stage to identify and discount PCR duplicates.

Hybridisation-based target enrichment uses complementary target-specific DNA or RNA oligonucleotide 'baits' to hybridise and capture genomic DNA (fragmented enzymatically or via physical shearing). Random shearing during genomic DNA preparation ensures that allele dropout due variation at specific sites is not an issue. Amplification bias can also be an issue in hybridisation based methods but removal of PCR duplicates during the analysis stage ensures bias is removed from the results.

Sequencing technology

A review of the available sequencing platforms is outside the scope of these guidelines. Currently short-read NGS technologies are most frequently employed in diagnostic laboratories. Each sequencing technology has specific issues that must be considered during test design and implementation. Particular attention should be paid to the spectrum of mutations within a panel and the sequence context of those mutations. Laboratories should have an in-depth understanding of the limitations of the selected NGS technology, and any corresponding validation should challenge the known technology-specific issues relevant to the test. Non end-terminating NGS (e.g. Ion Torrent) of homopolymer regions can lead to problems in the detection of the correct number of repeated nucleotides at homopolymer stretches which can limit the utility of such platforms for mutation calling in homopolymer tracts. Any laboratory using non end-terminating NGS should analyse the region of interest of all panels to identify any homopolymer stretches. The validation cohort should include samples with variants present within representative homopolymer stretches.

2.1 Bioinformatics pipelines and software should be appropriately validated and version controlled

The range of bioinformatics tools that can be employed for the analysis of NGS data is beyond the scope of this article, however, we would recommend that the validation and version control of bioinformatics software follow the AMP/CAP guidelines (Roy et al., 2018). We have also made recommendations on staff training, data retention and cloud computing later in the document.

2.2 The laboratory should be able to provide a comprehensive service even if this requires using supplemental techniques

Clinical panels should not be offered where the service is insufficient to provide full and complete coverage of all high-risk genes for the disease under investigation. Taking a regularly investigated disease such as Lynch syndrome as an

example, it is necessary to bear in mind that some genes of interest may not be easily analysed using NGS. PMS2 is a gene that presents such a problem (Mandelker et al., 2016). A laboratory offering an NGS panel for Lynch syndrome diagnosis would therefore need to ensure that it had an alternative approach in place to complement its NGS based panel and provide complete PMS2 sequencing or enrichment. Similarly, where large genomic rearrangements are known to be associated with disease pathogenesis (e.g. BRCA1) it would be necessary to supplement an NGS-based panel with a companion technique such as MLPA. Clearly defining the scope of testing may also identify additional analytical requirements e.g. RNA sequencing for genes with a high proportion of variants in non-coding regions e.g. NF1

If a laboratory is unable to provide the required testing in-house it would be appropriate to source this from an accredited tertiary service. The laboratory that is initially referred the case should provide an integrated report with all relevant results (both in-house and referred) and an overall conclusion to the requesting clinician. Rewriting of clinical reports is discouraged; if integrated into a final report, the original report should be referenced and a copy provided.

2.3 Confirmation testing, or an appropriate alternative, should be available if required

Confirmation testing (using Sanger sequencing for example) is often required by many clinical genetics services before a pathogenic variant will be reported. Given the long term implications of some findings this is reasonable and provides a mechanism to confirm the lineage of the primary sample in addition to validating the finding using an orthogonal technique. This step would not be required where an alternative but equally rigorous approach is put in place.

3 Laboratory Staffing and Training

3.1 Variant interpretation should be performed in line with best practice guidelines

All variants detected in a targeted gene set should be assessed by a clinical laboratory team with the appropriate training and resource to correctly classify the variant in line with best practice guidelines. The ACMG variant classification guidelines (Richards et al., 2015) are recommended by the IACS for variant interpretation. Current updates and refinements of the guidelines are published by the Sequence Variant Interpretation Working Group (SVI WG). The ACGS publish annual updates on the implementation of the ACMG guidelines with guidance for using the ACMG guidelines and the current version should be used.

3.2 Staff should be appropriately supported to ensure that analysis and variant interpretation skills are optimal for clinical service delivery

In spite of massive developments in NGS and bioinformatics, variant classification remains demonstrably labour intensive. Similarly, the rapid growth of NGS based diagnostics has meant that training in emerging technologies can be lacking. Staff expertise needs to be developed and training needs to be appropriate for service continuity. While some of the planning needs to take place at a local level, there is a distinct lack of a structured post-graduate training programme in genetics. Laboratories cannot assume that national coordination will be forthcoming and should be able to demonstrate engagement at a local and national level to meet the needs of the genetics service. Successful participation in appropriate variant interpretation schemes should be demonstrable for key staff.

3.3 Staff support and training should be sufficient to ensure that bioinformatics does not become a service 'blind-spot'

As the role of bioinformatics continues to grow it is necessary to consider the biological assumptions implicitly coded into the bioinformatic pipelines for NGS data analysis. It is necessary to ensure that any biological interpretations of the data between base-calling and the variant calling are explicitly understood and documented by the laboratory staff. A dedicated bioinformatics scientist can be a key addition to a team and can automate a number of repetitive steps making the overall process quicker, more reproducible and ultimately safer. The biological assumptions made when defining a bioinformatic pipeline should be well documented and the process should be documented and validated prior to offering the service. There should be evidence in place that the bioinformatic pipeline has been discussed with and approved by the service leads.

4 Data retention from NGS analyses

4.1 Data should be retained only as long as is necessary to ensure availability for anticipatable primary and secondary analyses and integrity of authorized results

Data retention times will vary from site-to-site in line with clinical needs. Ultimately, data should be retained only as long as those data are likely to be of benefit to the patient – unless consent has been obtained for long term retention for clinical or research studies. This means that data should be deleted once the result has been reported and the window for any clinically anticipatable secondary analyses has passed. It may also be necessary to retain data longer beyond this point to ensure a buffer for re-analysis is maintained should there

be any primary data analysis issues requiring data re-analysis. Retention of the data for a longer time period may also be clinically warranted in cases where the primary sample is either scant or cannot be re-acquired (e.g. cancer samples, skin biopsies, hair root, or a deceased patient). In all cases a data retention and discard policy should be put in place to detail the local data retention policy.

The following examples can be used for guidance purposes but should be interpreted based on local needs.

Raw data: Binary files etc.

Temporary storage only. Can be deleted after result authorization or held anonymized and secured for future validations.

Sequence data: FASTQ, BAM etc

Sequence data should only need to be held for as long as the clinical investigation of the primary sample is likely to require. In other words should a likely disease causing variant be detected and confirmed the laboratory would have no further use for the data and can (and should) delete it. However, if no variant is identified using the virtual panel and the referring clinical service has expressed an interest in performing full genome or exome analysis, it would be reasonable to hold the data for a time frame that would allow such a discussion to take place at a clinical level.

For all other investigations it is relatively simple to repeat sequencing and analysis so the retention of data beyond the primary and confirmatory investigations is not necessary and represents a level of risk to the service if the data is personally identifiable.

VCF files

VCF files (or their equivalent) should only be generated for the genes of interest on a virtual panel/exome/genome when that is the entity under investigation i.e. a whole genome VCF should not be generated and then filtered into a virtual panel. The VCF file should be stored in line with the RCPath guidance on the storage of laboratory worksheets as these files and documents are deemed to be equivalent in function. The current recommendation states for worksheets 'Keep for same length of time as related permanent or semi-permanent specimens or preparations.'(Royal College of Pathologists, 2015). This would mandate the retention of the VCF files as long as the extracted material is held on file.

4.2 Cloud computing can be used for data analysis if appropriately secured

The use of cloud computing is growing and is a logical approach where local data processing infrastructure is insufficient to meet the demands for analysis or turn-around-time. When using cloud storage it is necessary to ensure that

patient confidentiality is protected. Patient data should be stored locally where possible and not submitted to cloud services.

Where patient data is submitted to cloud services it is necessary to ensure that the cloud service has put in place a standard of data security (e.g. ISO 27107) sufficient for the protection of cloud data (International Organisation for Standardisation, 2015). Separately, the cloud service provider should be ISO 27108:2019 within 2 years of the date of this guideline if processing identifiable patient data (International Organisation for Standardisation, 2019) – it should be noted that a whole genome or exome sequence could be considered to be identifiable patient data. The data security element of the cloud service should be accredited by an entity recognized by the laboratory's accreditation body (INAB) for this purpose.

5 Data processing, consent and disclosure

5.1 Consent for genetic testing should be documented and this documentation should be available within the clinical laboratory

It is a legal requirement that informed consent is obtained prior to genetic testing. Documentation detailing that patient consent for genetic testing has been obtained must be provided for each request and must be retained by the laboratory. The Royal College of Pathologists has recently published updated guidelines on results disclosure (Royal College of Pathologists, 2019).

5.2 The availability of genetic information should be restricted to clinical services with a medical need for the data

Results should only be communicated directly to the requesting clinical team who can discuss the results and their implications with the patient. The introduction of the shared MedLIS results record will represent a breakthrough for continuity of care across hospitals, allowing for an online database of pathology results for each patient regardless of where the patient's sample has been processed. However, there is currently no mechanism to secure patient results within the database. Until such a mechanism is put in place it is not recommended that genetic results (whether generated using NGS or an alternative method) are added to MedLIS.

5.3 Off-target testing should be minimised or eliminated

Given the legal and ethical implications of holding on-file a clinically relevant result that is not released to the clinical service it is imperative that off-scope testing be eliminated from the assay design stage. This minimization can only be made possible by close collaboration with the clinical genetics team.

References

ASSOCIATION FOR CLINICAL GENETIC SCIENCE, T. 2015. Practice guidelines for Targeted Next Generation Sequencing Analysis and Interpretation.

GREEN, R. C., BERG, J. S., GRODY, W. W., KALIA, S. S., KORF, B. R., MARTIN, C. L., MCGUIRE, A. L., NUSSBAUM, R. L., O'DANIEL, J. M., ORMOND, K. E., REHM, H. L., WATSON, M. S., WILLIAMS, M. S. & BIESECKER, L. G. 2017. CORRIGENDUM: ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med, 19, 606.

GREEN, R. C., BERG, J. S., GRODY, W. W., KALIA, S. S., KORF, B. R., MARTIN, C. L., MCGUIRE, A. L., NUSSBAUM, R. L., O'DANIEL, J. M., ORMOND, K. E., REHM, H. L., WATSON, M. S., WILLIAMS, M. S., BIESECKER, L. G., AMERICAN COLLEGE OF MEDICAL, G. & GENOMICS 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med, 15, 565-74.

HADDOW JE, P. G. 2003. ACCE: A Model Process for Evaluating Data on Emerging Genetic Tests. In: KHOURY M, L. J., BURKE W (ed.) Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease.

INTERNATIONAL ORGANISATION FOR STANDARDISATION 2015. ISO 20170:2015 Information technology – Security techniques – Code of practice for information security controls based on ISO/IEC 27002 for cloud services.

INTERNATIONAL ORGANISATION FOR STANDARDISATION 2019. ISO/IEC 27018:2019 Information technology – Security techniques – Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors.

KALIA, S. S., ADELMAN, K., BALE, S. J., CHUNG, W. K., ENG, C., EVANS, J. P., HERMAN, G. E., HUFNAGEL, S. B., KLEIN, T. E., KORF, B. R., MCKELVEY, K. D., ORMOND, K. E., RICHARDS, C. S., VLANGOS, C. N., WATSON, M., MARTIN, C. L. & MILLER, D. T. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med, 19, 249-255.

MANDELKER, D., SCHMIDT, R. J., ANKALA, A., MCDONALD GIBSON, K., BOWSER, M., SHARMA, H., DUFFY, E., HEGDE, M., SANTANI, A., LEBO, M. & FUNKE, B. 2016. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. Genet Med, 18, 1282-1289.

RICHARDS, S., AZIZ, N., BALE, S., BICK, D., DAS, S., GASTIER-FOSTER, J., GRODY, W. W., HEGDE, M., LYON, E., SPECTOR, E., VOELKERDING, K., REHM, H. L. & COMMITTEE, A. L. Q. A. 2015. 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, 17, 405-24.

ROY, S., COLDREN, C., KARUNAMURTHY, A., KIP, N. S., KLEE, E. W., LINCOLN, S. E., LEON, A., PULLAMBHATLA, M., TEMPLE-SMOLKIN, R. L., VOELKERDING, K. V., WANG, C. & CARTER, A. B. 2018. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. J Mol Diagn, 20, 4-27.

ROYAL COLLEGE OF PATHOLOGISTS. 2015. The retention and storage of pathological records and specimens (5th edition).

ROYAL COLLEGE OF PATHOLOGISTS. 2019. Consent and confidentiality in genomic medicine.

STRANDE, N. T., RIGGS, E. R., BUCHANAN, A. H., CEYHAN-BIRSOY, O., DISTEFANO, M., DWIGHT, S. S., GOLDSTEIN, J., GHOSH, R., SEIFERT, B. A., SNEDDON, T. P., WRIGHT, M. W., MILKO, L. V., CHERRY, J. M., GIOVANNI, M. A., MURRAY, M. F., O'DANIEL, J. M., RAMOS, E. M., SANTANI, A. B., SCOTT, A. F., PLON, S. E., REHM, H. L., MARTIN, C. L. & BERG, J. S. 2017. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. Am J Hum Genet, 100, 895-906.

TEUTSCH, S. M., BRADLEY, L. A., PALOMAKI, G. E., HADDOW, J. E., PIPER, M., CALONGE, N., DOTSON, W. D., DOUGLAS, M. P., BERG, A. O. & GROUP, E. W. 2009. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med, 11, 3-14.