



Published in final edited form as:

Nat Biotechnol. 2015 July ; 33(7): 689–693. doi:10.1038/nbt.3237.

Good laboratory practice for clinical next-generation sequencing informatics pipelines

Amy S Gargis^{1,29}, Lisa Kalman¹, David P Bick², Cristina da Silva³, David P Dimmock², Birgit H Funke^{4,5}, Sivakumar Gowrisankar^{4,5,29}, Madhuri R Hegde³, Shashikant Kulkarni^{6,7,8}, Christopher E Mason⁹, Rakesh Nagarajan⁸, Karl V Voelkerding^{10,11}, Elizabeth A Worthey², Nazneen Aziz^{12,29}, John Barnes¹³, Sarah F Bennett¹⁴, Himani Bisht¹⁵, Deanna M Church^{16,29}, Zoya Dimitrova¹⁷, Shaw R Gargis¹⁸, Nabil Hafez^{19,29}, Tina Hambuch²⁰, Fiona C L Hyland²¹, Ruth Ann Luna²², Duncan MacCannell²³, Tobias Mann^{24,29}, Megan R McCluskey²⁵, Timothy K McDaniel²⁶, Lilia M Ganova-Raeva¹⁷, Heidi L Rehm^{4,5}, Jeffrey Reid^{27,29}, David S Campo¹⁷, Richard B Resnick¹⁹, Perry G Ridge^{10,29}, Marc L Salit²⁸, Pavel Skums¹⁷, Lee-Jun C Wong²⁷, Barbara A Zehnbaauer¹, Justin M Zook²⁸, and Ira M Lubin¹

¹Division of Laboratory Systems, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

²Department of Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

³Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA

⁴Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, Massachusetts, USA

⁵Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA

⁶Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA

⁷Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri, USA

⁸Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA

⁹Department of Physiology and Biophysics, Cornell University, New York, New York, USA

¹⁰Department of Pathology, University of Utah, Salt Lake City, Utah, USA

¹¹Institute for Clinical and Experimental Pathology, Associated Regional and University Pathologists (ARUP) Laboratories, Salt Lake City, Utah, USA

¹²College of American Pathologists, Northfield, Illinois, USA

²⁹Present addresses: Division of Preparedness and Emerging Infections, Center for Disease Control and Prevention, Atlanta, Georgia, USA (A.S.G.); Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, USA (S.G.); Phoenix Children's Hospital, Phoenix, Arizona, USA (N.A.); Personalis, Menlo Park, California, USA (D.M.C.); Quest Diagnostics, Marlborough, Massachusetts, USA (N.H.); Progenity, Ann Arbor, Michigan, USA (T.M.); Regeneron Pharmaceuticals, Tarrytown, New York, USA (J.R.); and Brigham Young University, Provo, Utah, USA (P.G.R.).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.3237).

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper (doi:10.1038/nbt.3237).

¹³National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

¹⁴Division of Laboratory Services, Centers for Medicare and Medicaid Services, Baltimore, Maryland, USA

¹⁵Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, Maryland, USA

¹⁶National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland, USA

¹⁷Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

¹⁸Division of Select Agents and Toxins, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

¹⁹Genome Quest, Westborough, Massachusetts, USA

²⁰Clinical Services, Illumina, San Diego, California, USA

²¹Thermo Fisher Scientific, South San Francisco, California, USA

²²Texas Children's Microbiome Center, Texas Children's Hospital and Department of Pathology & Immunology, Baylor College of Medicine, Houston, Texas, USA

²³National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia USA

²⁴Illumina, San Diego, California, USA

²⁵Soft Genetics, State College, Pennsylvania, USA

²⁶Oncology, Illumina, San Diego, California, USA

²⁷Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA

²⁸Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland, USA

To the Editor

We report principles and guidelines (Supplementary Note) that were developed by the Next-Generation Sequencing: Standardization of Clinical Testing II (Nex-StoCT II) informatics workgroup, which was first convened on October 11–12, 2012, in Atlanta, Georgia, by the US Centers for Disease Control and Prevention (CDC; Atlanta, GA). We present here recommendations for the design, optimization and implementation of an informatics pipeline for clinical next-generation sequencing (NGS) to detect germline sequence variants in compliance with existing regulatory and professional quality standards¹. The workgroup, which included informatics experts, clinical and research laboratory professionals, physicians with experience in interpreting NGS results, NGS test platform and software developers and participants from US government agencies and professional organizations,

also discussed the use of NGS in testing for cancer and infectious disease. A typical NGS analytical process and selected workgroup recommendations are summarized in Table 1, and detailed in the guidelines presented in the Supplementary Note.

Currently, most clinical NGS tests are offered as laboratory-developed tests (LDTs), which are tests designed, manufactured and used within a single laboratory. These tests use commercially available sequencing platforms to generate raw sequence data that are subsequently analyzed using software algorithms (informatics pipeline). In the United States, LDTs are subject to the Clinical Laboratory Improvement Amendments (CLIA) regulations, which require that laboratories introducing a test system that has not been cleared or approved by the US Food and Drug Administration (FDA; Rockville, MD) establish analytical performance specifications for the accuracy, precision, analytic sensitivity and specificity of the assay, as well as other measures as relevant¹⁻⁴. In 2013, the FDA cleared the Illumina (San Diego, CA) MiSeqDX as a Class II Exempt device, along with its associated reagent kit, and in 2014 two additional sequencing platforms (Life Technologies' (Carlsbad, CA) Ion PGM Dx sequencer and Vela Diagnostics' (Fairfield, NJ) Sentosa SQ301) were registered and listed and can now be marketed under the same regulation (<http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm375742.htm>, <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRL/rl.cfm?lid=427645&lpcd=PPF> and <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRL/rl.cfm?lid=430009&lpcd=PPF>). However, laboratories using these instruments must still establish an informatics pipeline for the intended clinical application(s). The clinical test is therefore an LDT, and its performance specifications must be established and validated under CLIA, even though the FDA has cleared the sequencing platform. The FDA also cleared two tests for screening and diagnosis of cystic fibrosis using NGS. In these instances, no component of either test is laboratory developed, and thus clinical laboratories need not validate these tests, although they do need to verify that the tests can achieve the performance specifications established by the manufacturer.

NGS-based tests use several sequencing technologies to test gene panels, exomes and genomes. These generate large amounts of data that require substantial computational infrastructure for storage, analysis and interpretation. An informatics pipeline identifies a set of sequence variants that are subsequently prioritized to identify those that are relevant to diagnosis. The complexity of informatics pipelines led the workgroup to recommend that laboratories planning to implement NGS budget for, or collaborate with, one or more informaticians with relevant expertise in the design, validation and implementation of a clinical NGS test.

NGS testing can be divided into three phases: primary, secondary and tertiary analysis¹ (Supplementary Fig. 1). Primary analysis includes the production of sequence reads and assignment of base quality scores. This phase was not covered by the Nex-StoCT II workgroup and has been addressed previously^{1,2}. Secondary analysis includes de-multiplexing (computational association of reads with a patient when multiple samples have been multiplexed (pooled) before sequencing), alignment of reads to a reference assembly or sequence(s) and variant calling. Tertiary analysis involves the identification and interpretation of clinically relevant variants. Sometimes laboratories may report incidental

findings that are not related to the patient's indication for testing but are relevant to the patient's health.

Many NGS approaches support multiplexing of multiple patient samples in a single run. DNA fragments generated from each patient are tagged or 'indexed' with a unique oligonucleotide barcode before multiplexing. This unequivocally links each fragment to the correct patient sample during subsequent analysis³. The Nex-StoCT II workgroup provided several recommendations related to ensuring the fidelity of multiplexing and correct assignment of each read to its corresponding patient sample. The workgroup advocated that vendor-supplied indexes, when available, should be used to simplify assay development. For indexes designed in-house, the workgroup provided recommendations for design parameters (for example, index length and composition) to minimize the likelihood of misassignment of reads to incorrect patient samples.

During secondary analysis, software tools are used to map and align each read to the reference human genome assembly, which is available from the Genome Reference Consortium (the latest version (GRCh38) can be found at <http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/>). This website is regularly updated. The workgroup recommended that clinical laboratories document the accession date and data version used for each alignment to ensure that variant positions can be traced back to a reference.

The workgroup also recommended that reads from panel, exome or genome sequencing be aligned against the full human reference assembly, rather than to targeted sequences, to improve specificity. This practice reduces (but does not always eliminate) the potential for read mismapping due to the presence of homologous regions (for example, pseudogenes or paralogs and duplications). In some instances, sequences are missing from the reference assembly, and this can result in reads that do not align or that are misaligned. Sequences may be absent from the assembly for several reasons, such as technical issues related to sample preparation and sequencing and the absence of alternate alleles in the chromosome assembly. Regions with high allelic diversity (for example, the major histocompatibility complex) have multiple alleles. These are represented in the reference as alternative assemblies and are useful for modeling human diversity. Although GRCh37 had three regions with at least one alternate haplotype available, GRCh38 contains >170 alternate alleles. The inclusion of alternate alleles allows better representation of multiple haplotypes; however, these alleles present a challenge for modern analysis pipelines, which cannot distinguish alternate haplotypes from biological events such as segmental duplication. This creates the need for new or updated software tools.

The alignment tools used by clinical laboratories include multiple algorithms that sometimes trade speed for accuracy and can incorporate different levels of sensitivity and specificity. To improve variant calling, the workgroup recommended that initial alignments include additional processing for local realignment, removal of PCR duplicates for genome and exome sequencing (if necessary) and recalibration of base-call quality scores (Supplemental Fig. 2 on p. 32 of guidelines in the Supplementary Note). The specific steps will vary and may depend on the library-preparation and capture methods used. For example, duplicate removal is generally not performed with amplification-based enrichment protocols.

A variety of software tools and strategies are also available for variant calling, the next step of the analysis. Because no single software tool or setting is currently able to identify all variant classes with equal accuracy, the workgroup recommended that several variant callers and/or parameter settings be evaluated to optimize the detection of different variant types during assay development.

Reference materials such as well-characterized DNA from immortalized human cell lines facilitate the design, optimization and validation of informatics pipelines. The workgroup discussed several sources of such materials. The CDC's Genetic Testing Reference Material Coordination Program (GeT-RM; <http://www.cdc.gov/clia/Resources/GetRM/>), together with the US National Center for Biotechnology Information (NCBI; Bethesda, MD; <http://www.ncbi.nlm.nih.gov/>) and the US National Institute of Standards and Technology (NIST; Bethesda, MD) Genome in a Bottle Consortium⁵ (<http://www.genomeinabottle.org>), has developed highly characterized genomic DNA reference materials derived from commonly used HapMap samples (NA12878 and NA19240). Sequence data from these materials displayed on an interactive and multifunctional Web browser (<http://www.ncbi.nlm.nih.gov/variation/tools/get-rm/>) can be used to support assay development and validation. In 2015, NIST released a characterized reference material (RM 8398) developed with reference values that can be used for assessment of performance of variant calling from genome sequencing (see <http://www.nist.gov/srm/>). Several commercial vendors currently produce synthetic human formalin-fixed paraffin-embedded (FFPE) reference materials. NIST is also considering developing human FFPE reference materials.

The purpose of tertiary analysis is to determine which sequence variants are relevant to a patient's clinical presentation. The workgroup recommends that during the design of the tertiary analysis, laboratories select methods for the annotation or labeling of variants with properties that can be used to remove or 'filter' variants that are not relevant to the patient's clinical indication. Annotations include information such as minor allele frequency in the population, predictions of variant effects on protein function or splicing and cross-references to disease-variant databases (for example, the Human Gene Mutation Database (HGMD), <http://www.hgmd.cf.ac.uk/ac/index.php>; ClinVar, <http://www.clinvar.com>; and Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/omim>). Other annotations regarding the function of a gene may be derived from publicly available data sets, such as those developed by NCBI (<http://www.ncbi.nlm.nih.gov>) or the European Molecular Biology Laboratory-European Bioinformatics Institute/Sanger Centre (Hinxton, UK) (<http://useast.ensembl.org/index.html>). These annotations complement other annotations produced during secondary analyses that describe the position of the variant, its associated gene, the type of variant, zygosity and selected metrics such as depth of coverage and confidence score. Recognizing that individual laboratories have differing approaches to tertiary analyses and that these can vary with the diagnostic target application, the workgroup elected not to define a minimum annotation set for tertiary analysis. Even so, the workgroup did suggest that annotations should include the position and type of variant, allele frequency, predicted structural and/or functional consequences and known disease associations or pathogenicity assertions.

The final step in tertiary analysis is a clinical assessment to determine which variants should be reported and how they should be described in the laboratory report. For many analyses, particularly exome and genome testing, a large number of variants must be evaluated for relevance to a patient's clinical presentation. During the design of this assessment step, laboratories select algorithms and optimize settings to predict deleterious variants located in clinically relevant genes. Other algorithms can be used to filter out selected variants, such as those with a high population allele frequency. The workgroup developed a general workflow for integrating data collected during the annotation, assessment and classification process with the patient's clinical presentation to provide a clinically relevant interpretation of the findings (Fig. 1 and Supplementary Fig. 3 on p. 49 of guidelines in Supplementary Note). The workgroup recommended that laboratories consider three essential questions during the design and optimization of annotation and prioritization methods:

1. Does the variant disrupt or alter the normal function of the gene in a manner consistent with the understanding of the disease mechanism?
2. Does this disruption lead to, or predispose a patient to, a disease or other outcome relevant to human health?
3. Does this health outcome have relevance to the patient's clinical presentation and indication for NGS testing?

Laboratories often categorize variants according to predicted functional impact ('often benign', 'likely benign', 'uncertain significance', 'likely pathogenic' or 'pathogenic', although there is some variation among laboratories)⁶. The College of American Pathologists (Northfield, IL), the American College of Medical Genetics and Genomics (Bethesda, MD) and the Association for Molecular Pathology (Bethesda, MD) developed an updated guidance for variant classification (available at https://www.acmg.net/docs/Standards_Guidelines_for_the_Interpretation_of_Sequence_Variants.pdf).

The workgroup did not consider the preparation of the laboratory test result report; however, it recognized that the uses and limitations of the informatics pipeline should be included in the test description communicated in the test result report.

The workgroup also acknowledged that NGS platforms (including firmware associated with the sequencing instrumentation), reference sequences, analysis and interpretation software (for example, for read mapping or variant calling and annotation) and associated reference databases (for example, mutation databases and allele-frequency repositories) are frequently updated, and that these updates can affect the performance specifications of a test. Such changes often require that part or all of a test be re-validated. Many Web-based tools that are accessed during the course of a test are also updated frequently. The workgroup recognized this as a challenge to clinical laboratories attempting to maintain traceability of data analyses. Their recommendation was that if these tools are not archived and versioned online, clinical laboratories should bring the software tools in-house so that modifications can be versioned, documented and referenced for each patient test.

The absence of a variant database that has been curated to provide high-quality data suitable for medical applications is problematic. Existing databases such as the HGMD are valuable

tools for assessing the relevance of variants, but the workgroup emphasized that caution must be applied because many entries are not sufficiently curated and may represent either a false positive variant-disease association or a population-specific risk allele attributable to only a small number of individuals⁴. Such errors may produce incorrect filtering or misassignment of a variant's clinical association with a patient's clinical presentation. Major efforts, such as NCBI's ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>), the Clinical Genome Resource (ClinGen; <http://www.nih.gov/news/health/sep2013/nhgri-25.htm>) and the Human Variome Project (<http://www.humanvariomeproject.org/>), are being made to collect and curate human genetic variants and alleles that affect health. These efforts, and others, are expected to advance the use of NGS in medical practice.

Exome and genome sequencing may reveal clinically reportable, incidental or secondary findings. The topic of incidental findings was initially addressed in the recent American College of Medical Genetics guidelines⁷ and is the subject of considerable discussion⁸⁻¹⁰. The workgroup recommended that each laboratory develop a policy to address this issue. Disclosure to the ordering physician before a test is ordered is consistent with good laboratory practices to communicate which regions of the genome are being targeted for sequencing, the limitations of the test and what is provided in the test result report.

It is challenging for laboratories to share and have a common understanding of variant data sets generated from clinical NGS because there is no standard format and no consistent representation of the variant data and haplotypes. This sharing or interoperability among laboratories and other entities is essential for interlaboratory comparisons of data to determine the concordance of variant calls so that data can be integrated into medical databases, and ultimately into patient medical records. Current variant-call file specifications (for example, VCF, <https://github.com/samtools/hts-specs>; gVCF, <https://sites.google.com/site/gvcftools/home/about-gvcf>; and GVF¹¹, <http://www.sequenceontology.org/>) were designed to be flexible and serve a variety of research needs.

Variants may be encoded in different ways, complicating comparisons of results among laboratories¹¹. To address this issue, the workgroup recommended that a new effort be initiated to establish a 'clinical-grade' variant file format to facilitate data sharing. The workgroup advocated that the standard for data sharing should be compatible with the evolving health information technology framework, which currently includes messaging using HL7 protocols. The CDC, in collaboration with other federal partners, established the Clinical-Grade Variant File Workgroup, which includes stakeholders from the public and private sector, to develop, pilot and design a pathway to implementation for such a file format.

The guidelines developed by the Nex-StoCT II workgroup are, to our knowledge, the first to establish a consensus for the design and optimization of a clinical NGS informatics pipeline. We anticipate that this guidance will aid laboratories in the development and validation of NGS tests for clinical applications. The guidelines will be updated as technology and practices advance. Information about the continued activities of the workgroup is available at http://www.cdc.gov/ophss/csels/dlpss/Genetic_Testing_Quality_Practices/ngsqp.html. We encourage collaborations and ongoing discussions among laboratory researchers, clinicians,

manufacturers, informaticians, software developers, professional organizations and government agencies to ensure the quality of clinical NGS tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by an appointment to A.S.G. to the Research Participation Program at the Centers for Disease Control and Prevention, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the CDC. H.L.R. was supported in part by National Institutes of Health grants U01HG006500 and U41HG006834. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the CDC, the Agency for Toxic Substances and Disease Registry or the FDA. Certain commercial equipment, instruments or materials are identified in this document. Such identification does not imply recommendation or endorsement by the CDC, the Agency for Toxic Substances and Disease Registry, the FDA or NIST, nor does it imply that the products identified are necessarily the best available for the purpose.

References

1. Gargis AS, et al. *Nat. Biotechnol.* 2012; 30:1033–1036. [PubMed: 23138292]
2. Rehm HL, et al. *Genet. Med.* 2013; 15:733–747. [PubMed: 23887774]
3. Craig DW, et al. *Nat. Methods.* 2008; 5:887–893. [PubMed: 18794863]
4. Bell CJ, et al. *Sci. Transl. Med.* 2011; 3:65ra64.
5. Zook JM, et al. *Nat. Biotechnol.* 2014; 32:246–251. [PubMed: 24531798]
6. Richards CS, et al. *Genet. Med.* 2008; 10:294–300. [PubMed: 18414213]
7. Green RC, et al. *Genet. Med.* 2013; 15:565–574. [PubMed: 23788249]
8. Holtzman NA. *Genet. Med.* 2013; 15:750–751. [PubMed: 24008255]
9. McGuire AL, et al. *Science.* 2013; 340:1047–1048. [PubMed: 23686340]
10. Allyse M, Michie M. *Trends Biotechnol.* 2013; 31:439–441. [PubMed: 23664778]
11. Reese MG, et al. *Genome Biol.* 2010; 11:R88. [PubMed: 20796305]

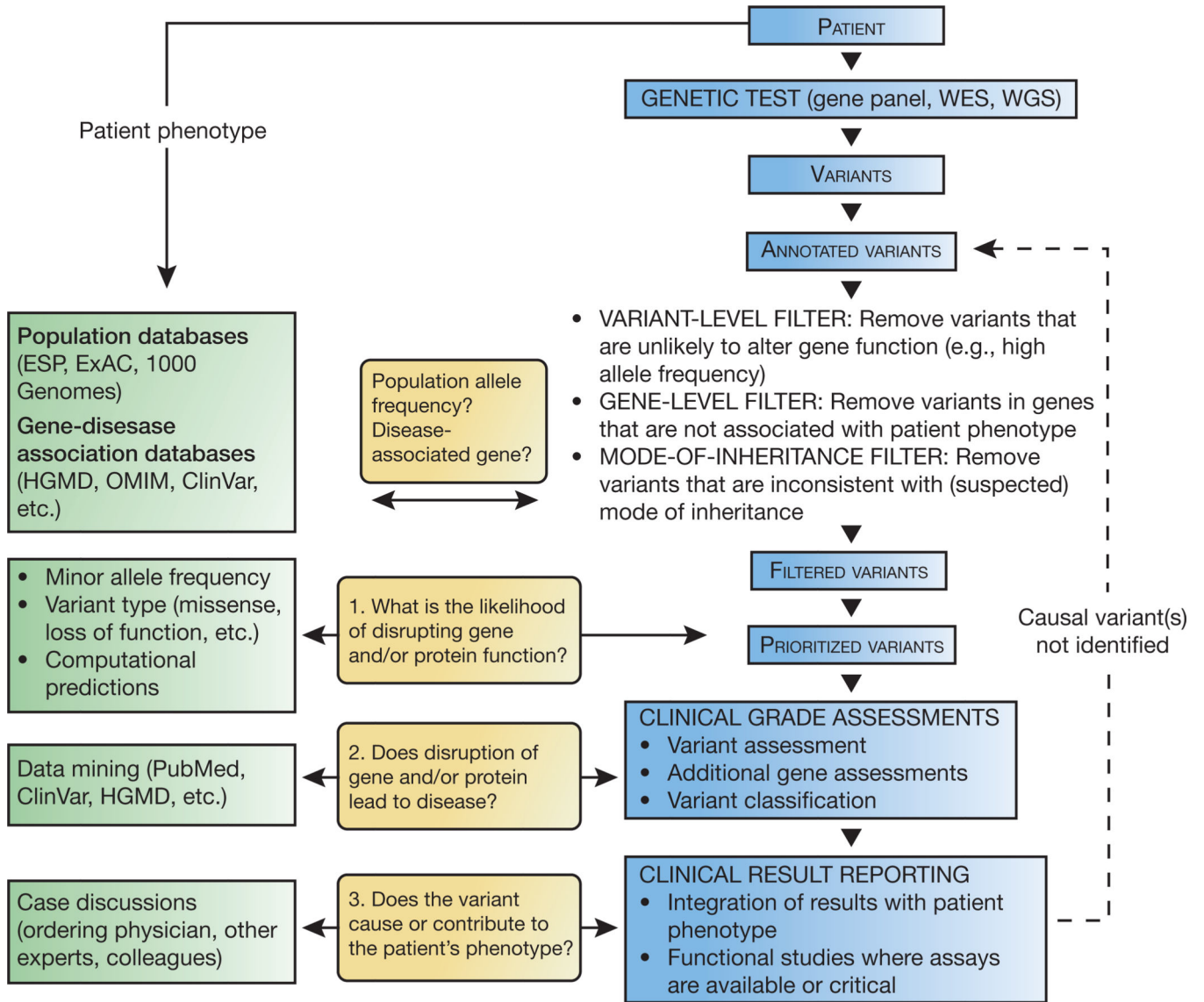


Figure 1. Model for tertiary analysis of sequence variants. The purpose of tertiary analysis is to identify variants to be reported to the physician for medical decision making. The variants identified during secondary analysis are filtered and prioritized, with consideration of their gene associations. ESP, Exome Sequencing Project; WES, whole-exome sequencing; WGS, whole-genome sequencing.

Table 1

Nex-StoCT II workgroup recommendations

Type of recommendation	Description
General	Informaticians should be involved in all steps of informatics pipeline design, optimization and validation.
Primary analysis	Nex-StoCT I workgroup recommendations for test validation and quality-control procedures in primary analysis are published ¹ .
Secondary analysis	Laboratories should use the commercially available indexes and protocols recommended by the platform manufacturers if they can be optimized and validated for the intended clinical application (p. 19 of Supplementary Note). Laboratories should use indexes that differ by more than a single base in the same reaction or lane (p. 20 of Supplementary Note). Laboratories should discard reads with mismatched indexes (p. 21 of Supplementary Note). Samples should be indexed as soon as possible and before targeted capture (p. 21 of Supplementary Note). The fidelity of de-multiplexing should be assessed and validated to ensure the correct assignment of sequence reads to patient samples (p. 21 of Supplementary Note). The assembly accession and version number should be documented for each alignment to enable traceability in both informatics workflow and result reporting (p. 24 of Supplementary Note). Laboratories should evaluate a combination of aligners or the same aligner with different settings to effectively identify the types of variants targeted (p. 27 of Supplementary Note). Laboratories should use the software's default settings initially and modify them (with validation) only when appropriate for the clinical application(s) (p. 28 of Supplementary Note). Changes to default settings and subsequent optimization should be performed in consultation with an informatician (p. 28 of Supplementary Note). Reads should be aligned to the reference assembly to minimize off-target or forced alignments, unless methods to ensure the optimal alignment to targeted regions of the genome have been developed (p. 30 of Supplementary Note). More than one variant caller should be evaluated to identify the combination of settings and/or software able to detect the spectrum of variants targeted by the intended clinical application, and the suitability of a variant caller for the platform should be considered (pp. 35–36 of Supplementary Note). Real and simulated data should be used for optimization of variant and genotype calling (p. 36 of Supplementary Note). Simulated data should not be used in place of data derived from patient samples for optimization and validation of the informatics pipeline (p. 36 of Supplementary Note). Well-characterized reference materials and tools such as those developed by the GeT-RM, the Genome in a Bottle Consortium and similar efforts should be used for test development, optimization and validation (pp. 36–37 of Supplementary Note).
Tertiary analysis and other considerations	The choice of annotation tools should be based on the types of sequence variants that need to be detected by the clinical test and on the strengths and limitations of the software used to detect particular variant types. Annotation and filtration tools should be integrated into the informatics pipeline to enable seamless automation and must be validated (pp. 40–41 of Supplementary Note). Until reliable, medically curated databases are available, data used to annotate variants for clinical assessments should be carefully evaluated to ensure that they are supported by sufficient evidence (p. 42 of Supplementary Note). Laboratory professionals should recognize differences in approaches to variant filtration and consider these in the design of the informatics pipeline (p. 43 of Supplementary Note). Where possible, the laboratory should consider using more than one prediction program, with each taking a different approach to predicting pathogenicity (p. 45 of Supplementary Note). Filtration algorithms should utilize databases containing reported pathogenic variants (for example, HGMD) to minimize the possibility of disease-associated variants being inappropriately filtered (p. 46 of Supplementary Note). The methods selected for annotating a sequence must be evaluated to demonstrate that variant attributes are properly assigned (p. 46 of Supplementary Note).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Type of recommendation	Description
	<p>A revision to a database or analysis algorithm used by a laboratory may affect the annotation process; consequently, the data analysis pipeline must be re-validated before the adoption of any updated data sources or software (p. 47 of Supplementary Note).</p> <p>If Web-based tools are unable to provide version control, the software or data sets should be brought in-house so version changes can be documented (p. 47 of Supplementary Note).</p> <p>Clinical assessment for disease association should be done by personnel with relevant clinical expertise. In some instances, this may be a collaborative activity undertaken by the laboratory professional and others (for instance, physicians, genetic counselors and/or informaticians) (p. 50 of Supplementary Note).</p> <p>Laboratories should consider three essential questions when assessing variants that are identified during the annotation and prioritization process: first, whether the variant(s) disrupts the normal function of the gene; second, whether the variant(s) is associated with or predisposes a patient to a disease or has other health-related implications (for example, carrier status); and third, whether the variant(s) has relevance to the patient’s clinical presentation and indication for testing (p. 51 of Supplementary Note).</p> <p>Pathogenic variants and variants of uncertain significance should be reported for heritable conditions. Reporting of benign variants is not recommended (p. 52 of Supplementary Note).</p> <p>The laboratory should establish strategies to reclassify or to monitor the reclassification of variants as new data become available to inform the analysis of findings (p. 52 of Supplementary Note).</p> <p>The informatics pipeline must be re-validated before the adoption of any new, updated or re-optimized software or databases (p. 54 of Supplementary Note).</p> <p>A new effort should be initiated to establish a ‘clinical-grade’ VCF or equivalent file-format specification to facilitate the interoperability of clinical laboratory and health information technology systems. This will facilitate data sharing among laboratories, with proficiency-testing programs for quality assurance, with databases that are used to support variant interpretation and for other purposes (the CDC established such a workgroup in 2013) (p. 57 of Supplementary Note).</p>