Analytical Validation of Clinical Whole Genome Sequencing for Germline Disease Diagnostics: Best Practices and Performance Standards

Christian R. Marshall[1*], Shimul Chowdhury[2], Ryan J. Taft[3], Mathew S. Lebo[4,5], Jillian G. Buchan[6], Steven M. Harrison[4,5], Ross Rowsey[7], Eric W. Klee[7,8], Pengfei Liu[9], Elizabeth A Worthey[10], Vaidehi Jobanputra[11,13], David Dimmock[2], Hutton M. Kearney[7], David Bick[10], Shashikant Kulkarni,[9,12], John W. Belmont[3], Dmitri J. Stavropoulos[1], Niall J. Lennon[5], on behalf of the Medical Genome Initiative.

[1]Genome Diagnostics, The Hospital for Sick Children, Toronto, ON, Canada, [2]Rady Children's Institute for Genomic Medicine, San Diego, CA, USA., [3]Illumina Inc., San Diego, CA, USA, [4]Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA, USA, [5]Broad Institute of MIT and Harvard, Cambridge, MA, USA, [6]Stanford Medicine Clinical Genomics Program, Stanford Health Care, Stanford, CA, USA, [7]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, [8]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA, [9]Baylor Genetics Laboratories, Houston, TX, USA., [10]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA, [11]New York Genome Center, New York, New York, USA, [12]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA, [13]Department of Pathology and Cell Biology, Columbia University Irving Medical Center (CUIMC), New York, New York, USA

THE MEDICAL GENOME INITIATIVE

BACKGROUND

- Whole genome sequencing (WGS) can detect most forms of clinically relevant variants and is emerging as a first-tier diagnostic test for patients with rare genetic disorders.[1]
- Standards addressing the definition and deployment of a clinical WGS test are beginning to emerge but have not been fully defined to date.
- The **Medical Genome Initiative** was formed to address challenges in the implementation of high quality clinical WGS with an aim to expand access.
- The goals of the **Medical Genome Initiative** include:
  1. Developing and publishing laboratory and clinical best practices for implementing clinical WGS for diagnosis of rare germline disorders.
  2. Aligning clinical research community on endpoints for measuring WGS clinical and economic utility.
- The Initiative is comprised of nine equally participating member institutions:

BAYLOR GENETICS · illumina · MAYO CLINIC · HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY · BROAD INSTITUTE · SickKids THE HOSPITAL FOR SICK CHILDREN · Rady Children's Institute Genomic Medicine · NEW YORK GENOME CENTER · Stanford MEDICINE

WORKING GROUPS & PUBLICATION ROADMAP

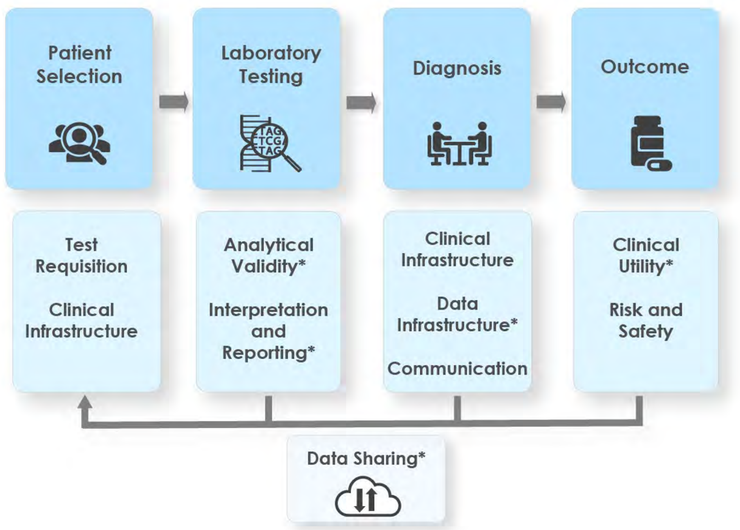


FIGURE 1: Concepts in clinical whole genome sequencing (*in progress)

- Working groups focused on key topic areas spanning the use of clinical WGS have been established.
- Current working groups include Analytical Validity, Clinical Utility, and Clinical Data Infrastructure & Sharing.
- **The Analytical Validity working group is focused on best practices for the analytical validation of WGS based on practical experience and, where possible, consensus of initiative members**

KEY STEPS IN ANALYTICAL VALIDATION

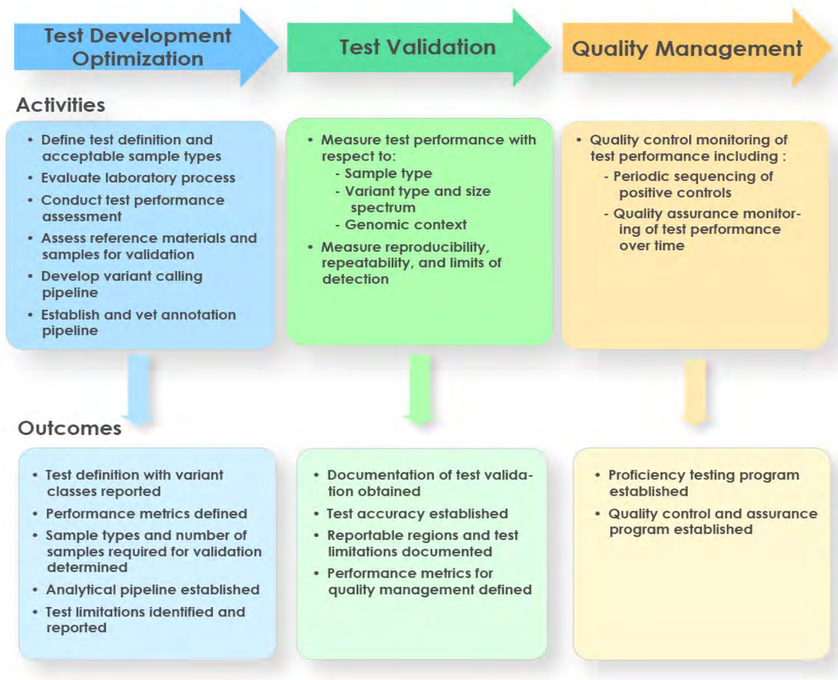


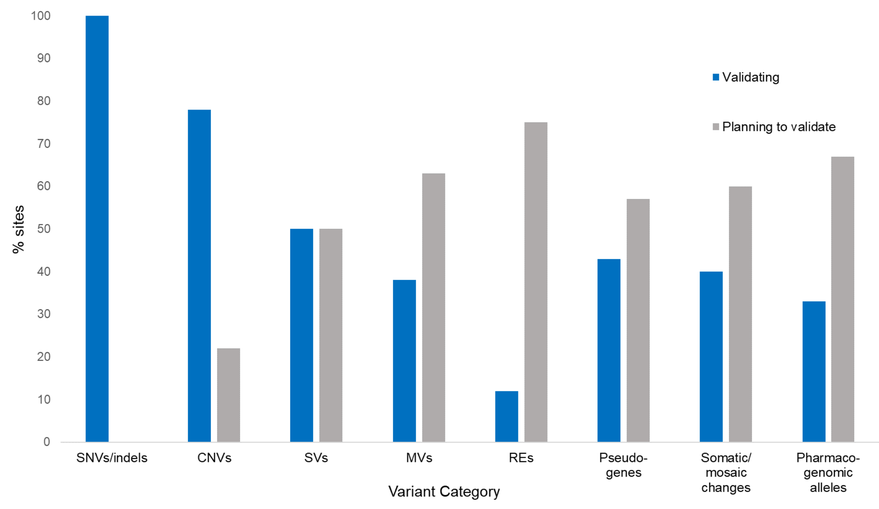
FIGURE 2: Steps in the Analytical Validation of WGS with Activities and Expected Outcomes

REFERENCES

[1]Lionel, A. C. et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet Med, doi:10.1038/gim.2017.119 (2017). [2]Zook, J. M. et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. Sci Data 3, 160025, doi:10.1038/sdata.2016.25 (2016). [3]Eberle, M. A. et al. A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. Genome Res 27, 157-164, doi:10.1101/gr.210500.116 (2017). [4]Trost, B. et al. A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. Am J Hum Genet 102, 142-155, doi:10.1016/j.ajhg.2017.12.007 (2018). [5]Deveson, I. W. et al. Representing genetic variation with synthetic DNA standards. Nat Methods 13, 784-791, doi:10.1038/nmeth.3957 (2016). [6]Escalona, M., Rocha, S. & Posada, D. A comparison of tools for the simulation of genomic next-generation sequencing data. Nat Rev Genet 17, 459-469, doi:10.1038/nrg.2016.57 (2016). [7]Duncavage, E. J. et al. A Model Study of In Silico Proficiency Testing for Clinical Next-Generation Sequencing. Arch Pathol Lab Med 140, 1085-1091, doi:10.5858/arpa.2016-0194-CP (2016).

SURVEY OF WGS PRACTICES ACROSS SITES

FIGURE 3: Types of variants offered as part of WGS test across initiative sites. CNV: copy number variant; SV: structural variants including balanced changes; MV: mitochondrial variants; RE: repeat expansions.



The majority of Initiative groups have validated sequence level variants with most planning on validating more complex variant types in the near future.

FIGURE 4: Reference Standards and Positive Control Resource Table

	Variant Type								
	SNVs & Indels	CNVs (>10Kb)	SVs	MVs	Pseudo-genes	REs	Somatic/ mosaic changes	Literature/Data	Source
Reference Samples: NA12878	100%	40%	0	0	0	0	0	Zook et al[2] FTP Directory	NIST Reference Materials Link
Reference Samples: Other NIST standard (e.g. Asian trios)	71%	40%	50%	0	0	0	0	Zook et al[2] FTP Directory	NIST Reference Materials Link
Reference Samples: Platinum Genomes	29%	0	0	0	0	0	0	Eberle et al[3]	Platinum Genomes Link
Reference Samples: Venter/HuRef	14%	40%	0	0	0	0	0	Trost et al[4]	HuRef Link
Positive Controls: Disease specific positive controls[2]	86%	80%	50%	100%	100%	100%	50%	GeT-RM Link	GeT-RM Link
Positive Controls: Synthetic controls	0	0	0	33%	0	0	50%	Deveson et al[5]	Sequins Standards Link
Positive Controls: In silico data	0	20%	0	0	0	0	0	Escalona et al;[6] Duncavage et al[7]	
Positive Controls: No. positive control samples	10—85	7—40	>10	4—20	4—40	18—175	N/A		

Percentages indicate the proportion of Initiative sites using a reference sample or positive control for select variant types.

ANALYTICAL VALIDATION KEY STATEMENTS

A series of discussions and surveys among working group members yielded the following key statements:

1. A clinical WGS test should aim to analyze and report on all possible variant types covered by its reportable range. This group recommends SNVs, Indels, and large CNVs as a minimum viable set of variants types for reporting.
2. Clinical WGS test performance should aim to meet or exceed that of any tests that it is replacing and clearly note established gaps and limitations on the report (e.g. detection of mosaic SNVs).
3. Metrics that measure genome completeness, including uniformity and depth of coverage, should be used to define the performance of clinical WGS and should be continually monitored
4. For variant types commonly addressed by the field (e.g. SNVs and indels) a low number of controls can be utilized if these include well-accepted reference standards. For variant types where standards are still evolving (e.g. REs),) a larger number of samples should be employed.
5. The analytical validation framework should include metrics that account for genome complexity, with special attention to sequence content and variant type.
6. Ongoing quality control of a clinical WGS test should include identification of a comprehensive set of performance metrics, continual monitoring of these metrics across samples over time, and the use of positive controls on a periodic basis dependent on overall sample volume.

CONCLUSIONS AND FUTURE DIRECTIONS

- Based on member experiences, the Medical Genome Initiative developed consensus recommendations and suggested best practices for the analytical validation of clinical WGS to aid other laboratories in its deployment.
- Arriving at consensus on select topics proved to be challenging due to different but equally valid methods and definitions.
- Future and ongoing Medical Genome Initiative Projects:
  - Clinical Utility of Genetic Testing Measurement Toolkit Publication (in progress)
  - Clinical WGS Data Infrastructure Best Practices Publication (in progress)
  - Interpretation and Reporting Publication (in progress)

CONTACT

Website: https://medgenomeinitiative.org/
Email: info@medgenomeinitiative.org
Twitter: @medical_genome
Corresponding author: Christian Marshall (crm@sickkids.ca)