Clinical Validation of an NGS Assay for Heritable Mutations

Stephen E. Lincoln[1], Geoffrey Nilsen[1], Mike Anderson[1], Yuya Kobayashi[1], Curtis Kautzer[1], Heather Trumbower[1], Chuck Argentine[1], Kevin Jacobs[1] , Scott Topper[1], Martin Powers[1]

Collaborators: Leif Ellisen[2], Kristen Shannon[2], Andrew Schultz[2], Andrea Desmond[2], Michele Gabree[2], Allison Kurian[3], Meredith Mills[3], Jim Ford[3]

[1]InVitae, San Francisco, CA[3]Massachusetts General Hospital, Boston, MA, , [2]Stanford University, Stanford, CA,

Abstract

Next Generation Sequencing (NGS) technology can help expand clinical use of genetic testing by allowing large numbers of genes to be rapidly tested in patients at low cost. However, broad clinical adoption of NGS requires demonstration of clinical-grade accuracy, in part to demonstrate sensitivity for "difficult" classes of variants, and in part to address clinical concerns about false positive findings from increasingly large gene panels. To address the later concern, most clinical NGS labs, including ours, confirm certain variants in patients using one or more orthogonal technologies, adding to cost and slowing turn-around times.

We have embarked on a series of studies to measure both the analytic and clinical performance of NGS in our diagnostic laboratory. In one study we apply clinical NGS to individuals which have undergone traditional genetic testing. We also analyze a set of reference material (RM) DNAs from public biobanks. We deliberately enriched these cohorts for individuals with known pathogenic variants, in order to help measure diagnostic sensitivity. Among those we further enriched for certain classes of DNA variation known to be challenging for NGS technology (such as indels and CNVs). At the same time, one study cohort is being selected prospectively to include unbiased representation of patients indicated for genetic testing under current clinical guidelines.

We present preliminary results from these studies here.

Study Design

Cases Received	Sequenced to date	Analyzed+ Unblinded	Description
251	241	187	MGH Retrospective Cohort: BRCA1/2 tested,~35% positive.
108	84	0	MGH Prospective Cohort: BRCA1/2 tested.
199	170	170	Stanford Cohort 1: ~30% BRCA1/2 positive.
251	243	243	Stanford Cohort 2: ~5% BRCA1/2 positive.
116	112	112	Public reference materials (RMs) with known mutations *
7	7	7	Public well-characterized genomes (WCGs)
932	**857**	**719**	**Total**

* In hereditary cancer genes: BRCA1, BRCA2, APC, CDKN2A, MEN1, MLH1, MSH2, PTEN, and RET. Other genes: CFTR, ASPA, FANCC, HEXA, GJB1, IKBKAP, HFE, Factor 2, and Factor 5. Large cytogenetic changes associated with Smith-Magenis, Down, Cri-du-chat, Wolf-Hirschhorn, Angelman, and other syndromes are also present.

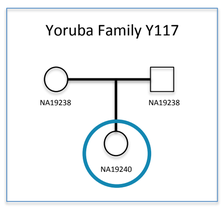
Clinical+ RMs	WCGs	Known Positives
123	2020	Single nucleotide variants
125	50	Small insertions, deletions, and del-ins (≤10 bp)
4		Large insertions and deletions (>10 bp and <1 full exon)
20		Small copy number (del/dup) changes (1 exon to 1 gene)
22		Large (cytogenetic) copy number changes
23		Mutations in/near specific homopolymer regions*

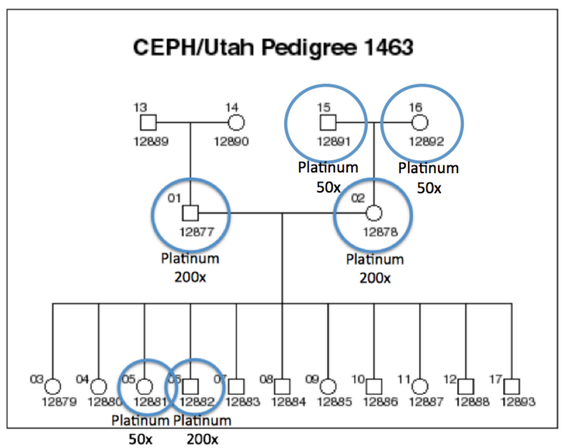
* MSH2 (Intronic Poly-A near splice site) and CFTR (Poly-T/Poly-TG)

MGH and Stanford Cohorts: Candidate Hereditary Breast and Ovarian Cancer (HBOC) patients independently tested for BRCA1 and/or BRCA2 by a well-established diagnostic laboratory using traditional technologies. Both MGH cohorts and Stanford cohort 2 were blinded at the time we processed these data. The completeness of data available to us from the other laboratory varies: In 78% of cases, BRCA1/2 full sequencing was performed, in others only more limited testing (e.g. small mutation panel or a single-site test) was performed. In 60% of cases copy-number data are available, and about 29% of those tests fully assay BRCA1/2 for CNVs, for the remainder only a targeted test is available.

Reference Materials: DNA samples were purchased from Coriell and NIBSC which were independently documented to have mutations in genes of interest, with a particular focus on "difficult" variants (indels and CNVs) .

Well-characterized genomes: These include community standards NA12878 (CEPH/Utah Pedigree 1463) and NA19240 (Yoruban trio child). Independent whole-genome data from Complete Genomics and from the Illumina Platinum set is used for the CEPH samples; and data from Complete Genomics and from the 1000 Genomes Pilot 2 is used for the Yoruba. The public data sets are screened to reduce errors: variant positions with impossible or unlikely inheritance (in the public data only) are considered "no calls", and the two public data sets for each sample are intersected with discordances also considered "no calls". The remaining high confidence variant and invariant positions are then compared to the Invitae data. Unlike the clinical and RM cohorts described above, the majority of the variants compared in these 7 samples will be benign.





Analytic Concordance

Clinical cases and RM samples

Error type	Number observed	Gene x individual combinations with data from 2 labs analyzed
NGS False Positives	**0** **0**	**1193** for sequence variants **439** for small CNVs
NGS False Negatives	**0**	**314** positive variants within test's analytical range *

Result	MGH	Stanford	RMs	Total
Confirmed Positive Variants	**85**	**77**	**155**	**317**
Variants Agree between labs	85	75	149	309
Invitae NGS positive, negative in other data Positive result confirmed by 3rd party lab		1	4	5
"Difficult variant" classes for NGS - Since addressed by new bioinformatics algorithms		1	2	3
Other Results				
Invitae NGS negative, positive in other data Negative result confirmed by 3rd party lab	1		1	2
Novel Invitae CNVs, not tested in other data Pending 3rd party confirmation		1	1	2

*** "Difficult Variants" for NGS and Solutions**

PLP1: del exon 3-4 in a Reference sample initially missed by NGS
- Deletion spanning parts, but not all, of 2 neighboring exons
- Split-read algorithm developed detect this type of event

BRCA2: 9203del126 in a Stanford case initially missed by NGS
- Deletion spanning part of but not all of a full exon
- Split-read algorithm developed detect this type of event

MSH2:c.942+3A>T in a Reference sample initially missed by NGS
- Splice site changing mutation next to an intronic poly-A tract
- PolyMNP algorithm since modified to call this site

Clinical Variant Analytic Concordance: Pathogenic mutations and independently documented VUS in each sample's specific gene(s) of interest are counted in this analysis. A small number of benign variants, when documented independently, are also included, although the vast majority of benign variants are not considered in this analysis.

Conflict Resolution: When conflicts occur between the Invitae NGS results and those reported independently, samples are sent-out to a 3rd lab for confirmation using either Sanger sequencing or aCGH. A negative or positive call is considered confirmed either if it initially matches between both labs, or if it does not but is confirmed by send-out to the 3rd lab.

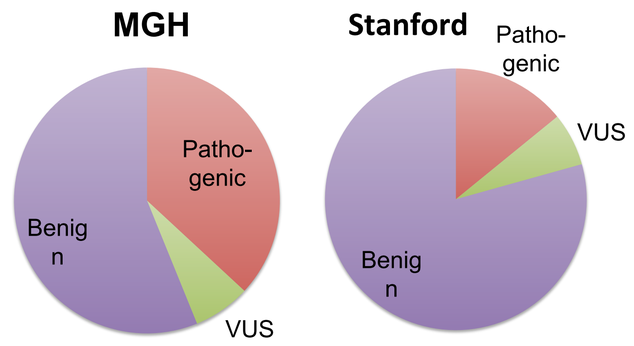
Well-Characterized Genomes

Sample	Relationship	True Positive Sites	Detected by Invitae NGS	True Negative Base Pairs	False Positives
NA12878	Parent	291	100%	621,277	0
NA12877	Parent	300	100%	619,127	0
NA12881	Child	284	100%	620,279	0
NA12882	Child	284	100%	620,373	0
NA12891	Grandparent	302	100%	621,833	0
NA12892	Grandparent	277	100%	621,927	0
NA19240	Child	332	100%	596,798	0

Clinical Concordance

Clinical Sign-Out of BRCA1/2 Cases

	MGH	Stanford
Pathogenic or Likely pathogenic (ACMG Criteria)	69	58
Negative (ACMG Criteria)	105	326
Uncertain (ACMG Criteria)	13	27
Total Cases	187	411



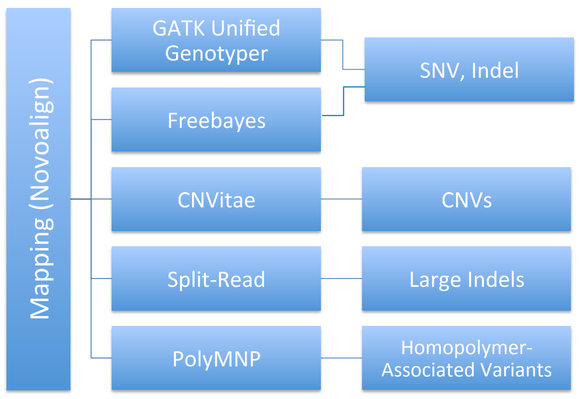
Result	MGH	Stanford	%Total Cases
BRCA1/2 cases signed-out by two labs	**187**	**413**	
Pathogenic Interpretations Agree *	69	59	21.3%
Interpretations with Note	0	1	0.2%
Negative Interpretations Agree	105	322	71.2%
Uncertain (VUS) Interpretations Agree	6	11	2.8%
Invitae VUS not reported by other lab	6	13	3.2%
Other Lab VUS vs. Invitae Likely Benign	0	2	0.3%
Incomplete data from other lab Unable to compare in detail	1	5	1.2%

One Interpretation Difference:

BRCA1: 4986+3G>C in a Stanford case
- Reported in the literature (PMID:12491499) as pathogenic but without evidence meeting ACMG criteria.
- Weak evidence of splicing change using prediction algorithms.
- Reported as as Pathogenic by another diagnostic lab, but without stated evidence.
- We report these facts in full.

Methods

Our panel covers all coding exons, splice-sites and any other regions with known pathogenic variants in 212 genes. These are targeted using hybridization based enrichment with multiple pull-down chemistries to improve coverage of "difficult" targets (for ex. GC-rich regions). Sequencing is performed on the Illumina platform using 150bp paired-ends to an average depth over 400x. We use a custom-built mapping, variant calling and data QC pipeline. Preliminary interpretations are automatically generated for review, and medical specialists then classify variants according to ACMG guidelines. Confirmation on an orthogonal platform and fill-in is used, when needed, to complete a diagnosis.



Conclusions

Our preliminary results from these studies suggest that careful application of targeted NGS can produce data highly concordant with traditional diagnostic methods.

A. We see very **high NGS sensitivity** for the genes and variant classes represented in this study.

B. NGS **false positive rates appear to be very low**. A systematic examination of this in the future may allow us to remove, in some circumstances, the need for orthogonal confirmation in NGS-based genetic diagnostics.

C. By following ACMG guidelines and leveraging widely available tools and information, **clinical interpretations consistent with established laboratories** can be made. As expected, VUS rates are slightly higher but we expect these to come down rapidly as public clinical data accumulate. Note that family members were not available these research cohorts, although they would be tested in our diagnostic process when possible to resolve some of the VUS.

Acknowledgements: We would like to thank and acknowledge Josh Paul , Michele Cargill, Jon Sorenson, Adrienne Ertman, and Emily Hare and for their substantial contributions.