A Comprehensive Low-Cost Clinical Diagnostic Test For Hundreds of Inherited Conditions

Jon Sorenson, Michelle Sommargren, Jody Westbrook, Emily Hare, Yuya Kobayashi, Michael Anderson, John Major, Reece Hart, Kevin Jacobs, Jill Hagenkord, Stephen Lincoln, Michele Cargill, Randy Scott, and, far more importantly, the rest of the InVitae team

InVitae, San Francisco, CA, 94107 www.invitae.com

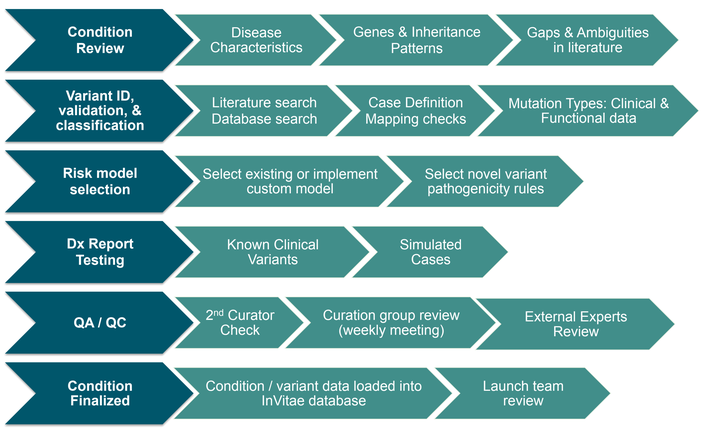
SUMMARY

Historically, diagnostic tests using DNA sequencing have only been offered for a limited set of genes to patients with specific clinical indications. The high cost of *de facto* standard clinical assays (Sanger sequencing, MLPA, etc.), and far more importantly the high cost and challenges in clinical data interpretation have been cited among the reasons for this. Thus, many clinically important genetic conditions often go undiagnosed.

We have developed an in-house infrastructure for NGS-based diagnostic assay development, validation, and operation in a CLIA environment. To date, we have conducted a thorough scientific review of the literature for over 500 genetic conditions, storing validated gene sequences, transcripts, risk models, and over 30,000 clinically characterized variants in a database used both to optimize assay design and to interpret results. We have a hybrid variant calling pipeline employing GATK, Freebayes, and custom algorithms for different variant classes including CNVs. Clinical reports for known and novel variants are automatically generated for review and sign-out by medical specialists. Variants are reported according to ACMG guidelines, but adding an additional category for variants observed in other affected patients but which have otherwise uncertain pathogenicity. In collaboration with other labs and patient advocates we have launched an effort to expand the publicly available set of unpublished clinical variants that we believe will be critical in diagnostic settings. Most importantly, we believe these processes are highly scalable, allowing the assay to grow to report on the vast majority of known genetic conditions with high accuracy.

Validation was performed on a panel of reference samples with >11,000 known variant sites and ~2.1 million non-variant sites. For coding sequence substitutions, we demonstrated 99.7% sensitivity and 99.998% specificity. For insertions/deletions in coding sequence, we demonstrated 98.3% sensitivity and 99.994% specificity. In direct comparisons against established diagnostic laboratories using traditional Sanger sequencing, we saw 100% concordance with those results, both in terms of analytical concordance and clinical interpretation. (Note: these data are updated since the abstract for this meeting).

CURATION PROCESS



GENE SELECTION

Example: Hereditary Cancer

Specialty	Women's Health	Gastroenterology (incl. pancreatic)		Endocrine	Genitourinary (incl. prostate)	Dermatology	Bone and Soft Tissue	Neuro-oncology
Annual US Incidence	300,000	250,000		2,500	300,000	81,000	14,000	23,000
% Hereditary	~10%	~5% (GI)	~10% (panc)	~30%	5% (renal)	unk	unk	~5%
	ATM	APC	APC	APC	BHD/FCLN	APC	APC	CDKN2A
	BRCA1	ATM	ATM	CASR	BRCA1	ATM	ATM	EPCAM
	BRCA2	BMPR1A	BRCA1	HRPT2	BRCA2	BRCA2	HRPT2	MEN1
	BRIP1	CDH1	BRCA2	MAX	EPCAM	EPCAM	MEN1	MLH1
	CHEK2	EPCAM	CDKN2A	MEN1	FH	BHD/FLCN	NF1	MSH2
	EPCAM	MLH1	EPCAM	NF1	HRPT2	BHD/FCLN	PRKAR1A	MSH6
	MLH1	MSH2	MEN1	PRKAR1A	MET	MLH1	RB1	NF1
	MSH2	MSH6	MLH1	PSCA	MLH1	MSH2	TP53	NF2
	MSH6	MUTYH	MSH2	PTEN	MSH2	MSH6	TSC1	PMS2
	PMS2	PMS2	MSH6	RET	MSH6	PMS2	TSC2	PTEN
	PTEN	PSCA	PALB2	SDHA	PMS2	PTCH1		TP53
	RAD51C	PTCH1	PALLD	SDHAF2	PTEN	PTEN		TSC1
	STK11	PTEN	PMS2	SDHB	SDHB	RB1		TSC2
	TP53	SMAD4	PRSS1	SDHC	STK11	XPA		VHL
	BARD1	STK11	PRSS2	TMEM127	TSC 2			RB1
	RAD50	TSC1	STK11	TP53	TSC 1			
	CHEK2	TSC2			VHL			
	RAD51C	TP53						

Note: Not all genes are available in all countries.

ASSAY DESIGN AND VALIDATION

CONDITION-SPECIFIC RULES FOR REPORTING NOVEL VARIANTS

Rule	Variants processed by disease risk model into Likely Pathogenic or Likely Benign	Variant of Unknown Significance ("Pure" VUS)	Example
Typical (Loss of Function)	Initiator codon variant Nonsense Frameshift Missing exon(s) Consensus splice site insertion Consensus splice site deletion	All others	Cystic fibrosis Ataxia with oculomotor apraxia, type 2
No VUS Reported	None	None	CFTR modifier conditions
No risk model	Meets typical rules but inheritance pattern unknown; custom risk output	All others	GDAP1-related peripheral neuropathy
Unknown severity	Meets typical rules but has unknown disease severity; custom risk output	All others	Tay-Sachs disease and variants
Only VUS Reported	None	All novel variants	Amyotrophic lateral sclerosis, type 4
Ranged	Meets 2 requirements: specified mutation type & location in specified range of gene	All others	BRCA1-related hereditary breast and ovarian cancer

FUTURE PUBLIC DATA RELEASES

- The largest source of medically relevant information on human genetic variation lies in diagnostic labs. However most of these data are never published or released.
- InVitae is committed to ongoing public release of deidentified clinical variants, as much as is allowed by law and proper ethical best practices.

DIAGNOSTIC REPORT EXAMPLE

Name Jane Doe	**DOB** 01.01.1980	**Gender** Female	**Ancestry** Caucasian	**MRN** 123456789
Physician Alice Eve Roberts, MD	**Report Date** 01.10.2013	**Sample Type** Blood	**Accession#** RQ00000	**Date Received** 01.02.2013
Test Indication Family and personal history of breast cancer		**Previous Results** BRCA1 and BRCA2 negative		

Primary Findings for Jane Doe

A pathogenic variant was identified in the MLH1 gene associated with Lynch Syndrome, see comment.

We encourage you to visit invitae.com to obtain an electronic copy of this report, view the quality metrics for this sample, view additional information about reported variants, review the clinical validation data for this assay, or to learn more about genetics.

Comment:

- A single base pair insertion was identified in exon 13 of the MLH1 gene, c.1489_1490insC, that has been previously reported in patients with Lynch Syndrome (see PMID: 15849733).
- This insertion leads to a frameshift which results in a truncated protein 11 amino acids downstream (p.Arg497GlyfsX11) and is predicted to result in a non-functional protein product.
- MLH1 encodes a mismatch repair protein that is involved in Lynch Syndrome.
- This patient is at risk for additional cancers, including colorectal, endometrial, and ovarian cancers.
- It is unclear at this time if breast cancer is a component of Lynch Syndrome (see PMID: 22034109).
- See NCCN guidelines for the management of patients with Lynch Syndrome.

Secondary Findings:

- Not applicable.
- Learn more at url/preventative-medicine

Results:

- The genes listed in the table below were evaluated for sequence variants and copy number changes. Results are provided in the table.
- Variants of uncertain significance that are present in >X% of the healthy population are very unlikely to cause rare diseases. Unless otherwise specified, they are not included in this table, but are available at url/technical-report
- Detailed information about the listed variants can be viewed in InVitae's Variant Assessment Dashboard at url/variant-dashboard
- The anatomy of each gene and detailed information regarding the mutation spectrum and variants targeted are available at url/gene-anatomy.

Gene	Coverage	Condition	Variant Name	Protein Effect	Zygosity	Variant Interpretation
MLH1	100%	MLH1-Associated Lynch Syndrome	NM000249.3: c.1489_1490insC	Truncation	Heterozygous	Pathogenic
MSH2	100%	MSH2-Associated Lynch Syndrome	NM_000251.1: c.51C>T	Silent	Heterozygous	VUS, Likely Benign
MSH6	100%	MSH6-Associated Lynch Syndrome	None Identified	-	-	-
MUTYH	99.99%	MYH-Associated Polyposis Syndrome	None Identified	-	-	-
BRIP1	98.79%	BRIP1-Associated Hereditary Breast Cancer	None Identified	-	-	-
APC	100%	Familial Adenomatous Polyposis	NM_000251.1: c.51C>T	Silent	Heterozygous	VUS, Likely Benign
CDKN2A	100%	CDKN2A-related cutaneous melanoma	None Identified	-	-	-
CDK4	99.88%	CDKN2A-related cutaneous melanoma	None Identified	-	-	-

Table continues for additional genes ordered. Methods section of report not shown.
Note: This is not a real patient (simulated data). Pre-release version of clinical report shown.