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

Jon Sorensen, 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collaboration with other labs and patient advocates we have launched an effort to expand the publicly available set of unpublished clinical variants that we believe will be critical in diagnostic settings.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, these processes are highly scalable, allowing the assay to grow 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 99.98% specificity. For insertions/deletions in coding sequence, we demonstrated 98.3% sensitivity and 99.94% 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

- Condition Review
- Chinese Genetics
- Genetics & Inheritance Patterns
- Gaps & Antidotes: literature
- Variant ID, validation, Disruption
- Literature search database search
- clone Deletion
- Mapping view
- Mutation Type: Clinical & Functional data
- Risk model selection
- Select coding or implement custom model
- Select novel variant pathogenicity rules
- DA/QC
- 3rd-Party Check
- Published data, Simulated cases, MeSH
- KConGEN
- External Experts review
- Condition
- Condition variant data loaded into securely database
- Launch team review

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: Doe
DOB: 01/01/1980
Gender: Female
Ancestry: Caucasian
MIN: 123456789

Physician: Alice Eve Roberts, MD
Report Date: 01/10/2013
Sample Type: Blood
Accession#: R000000
Date Received: 01/02/2013

Test Indication: Family and personal history of breast cancer
Previous Results: BRCA1 and BRCA2 negative

Primary Findings for Doe Jane

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

Comment:

- A single base pair insertion was identified in exon 13 of the MLH1 gene, c.1498+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.Arg479GlyX11) 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.
- VUS: variants of uncertain significance that are present in ≥1% 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.

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.