High Accuracy and Expanded Yield from Next-Generation Testing of Multiple Cancer Risk Genes

Stephen Lincoln, Allison Kurian, Leif Ellisen, Kristen Shannon, Martin Powers, James Ford, and Federico Monzon

1. Invitae, San Francisco, CA  2. Stanford University Medical Center, Palo Alto, CA  3. Massachusetts General Hospital, Boston, MA

Abstract

Background: Gene panels assayed using next-generation sequencing technologies (NGS) are moving from research labs into clinical use, with the potential to provide improved diagnostic yield quickly and at low cost. It is important to understand the performance of these tests in detail by comparison with traditional sequential tests, and to articulate the differences between research and clinical uses of NGS.

Methods: 712 germ-line DNA samples were collected: (a) 600 from two clinical cancer centers, all previously tested for BRCA1 and/or BRCA2 using traditional methods; (b) 112 reference samples from public biobanks, chosen both to have broad coverage of genes previously tested and also enriched for variants known to be challenging for NGS (e.g. large indels and small copy-number deletions and duplications). These were blinded and tested on a custom NGS panel including 27 hereditary cancer genes with laboratory protocols and bioinformatics methods specifically adapted for clinical use. The NGS and previous results were compared and, for the very few discrepancies initially observed, samples were sent to a 3rd party laboratory for resolution.

Results: 100% analytical sensitivity and 100% analytical specificity were observed between NGS and all confirmed previous results including sequence and copy-number changes. Clinical interpretations were also highly concordant, although with a slightly higher rate of VUS (Variants of Unknown Significance) in the NGS data (6% vs. 4%). This was expected following ACMG (American College of Medical Genetics) classification guidelines with public, and not proprietary reference databases. Roughly 40% more cases saw a validated positive reportable finding with the NGS panel.

Conclusions: NGS, when carefully applied, can produce high quality clinical results compared to traditional testing. The high specificity observed suggests that it may become possible, in some circumstances, to remove the need for orthogonal confirmation and further improve time to diagnosis and reduce costs using NGS.

Analytic Performance

All previous reported variants in the tested samples were detected, with the only exception being 2 sites which confirmed as negative when tested by a 3rd party lab. The results were concordant for both sequence changes and also all deletion/ duplication events. These results indicate a 100% sensitivity for our test. Counting all gene x individual combinations where data were available from both Invitae and from the previous lab for comparison, 1192 individual gene evaluations were performed. From these, no false positives were detected. Every Invitae finding was either matched in the “gold-standard” or confirmed by the 3rd party lab. These results indicate a 100% specificity for our test.

Certain sequence variants posed specific challenges for NGS methods that led to development of new bioinformatic solutions:

• BRCA2: 9203del126 in Stanford case 8178, initially missed by NGS
  - Challenge: Deletion spanning part of but not all of a full exon
  - Solution: Split-read algorithm implemented to detect this type of event

• PLP1: del exon 3-4 in NA13434, initially missed by NGS
  - Challenge: Deletion spanning parts, but not all, of 2 neighboring exons
  - Solution: Split-read algorithm developed to detect this type of event

• MSH2: 942+3A>T in NBSC-2, initially missed by NGS
  - Challenge: Splice site changing mutation next to a 25b intronic poly-A tract
  - Solution: PolyMNP algorithm modified to call this site

Clinical Performance

Clinical interpretations for variants in the BRCA1/2 genes were 99.8% concordant with prior results. One single variant was discrepant (see box below). The observed rate for variant of unknown significance (VUS) interpretations in our laboratory was 6% which compares well with the 4% VUS rate for the reference laboratory in this patient cohort. The distribution of results by diagnostic categories is shown below:

<table>
<thead>
<tr>
<th>BRCA1 and BRCA2</th>
<th>Stanford</th>
<th>NCCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathogenic or Likely pathogenic</td>
<td>66</td>
<td>105</td>
</tr>
<tr>
<td>Negative</td>
<td>105</td>
<td>315</td>
</tr>
<tr>
<td>Unverifiable Significance</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>Total Cases</td>
<td>187</td>
<td>411</td>
</tr>
</tbody>
</table>

Exactly One Interpretation Difference for a Pathogenic Variant:

BRCA1: 4986+3G>C in Stanford case 6396

• Reported in the literature as pathogenic but without evidence meeting ACMG criteria (PubMed:12491499)
• Weak evidence of splicing change using prediction algorithms
• Diagnostically reported as both pathogenic by another lab, but without stated evidence

Study Design

<table>
<thead>
<tr>
<th>Sample</th>
<th>Analyzed in this Report</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>105</td>
<td>Public Reference Materials*</td>
</tr>
<tr>
<td>243</td>
<td>243</td>
<td>Stanford Reference Genotype 1</td>
</tr>
<tr>
<td>241</td>
<td>187</td>
<td>Stanford Reference Genotype 2</td>
</tr>
<tr>
<td>230</td>
<td>0</td>
<td>MGH Reference Genotype 1</td>
</tr>
<tr>
<td>211</td>
<td>0</td>
<td>MGH Reference Genotype 2</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>Public Well-Characterized Genomes</td>
</tr>
<tr>
<td>Total</td>
<td>712</td>
<td>Total to Date</td>
</tr>
</tbody>
</table>

Clinical Case Study

**Patient:** Female, diagnosed with unilateral breast cancer in her mid-40s. Indicated for BRCA1/2 testing under NCCN guidelines based on family history and age of onset.

**Previous Results**

The patient received a negative BRCA1/2 test report from an independent laboratory.

**New Results**

Approximately 8 years later, we found a pathogenic germ-line mutation in MTHFR associated with Lynch syndrome. The patient was recontacted and offered additional counseling.

**Outcome**

In the time between the BRCA1/2 and NGS tests the patient had been diagnosed with endometrial cancer. Additionally, a baseline colonoscopy had been performed at age 50 with negative results, and she had been told to come back at age 60. Based on her MTHFR status, she underwent an early second colonoscopy and a tubular adenoma was caught and removed 7 years earlier than if no broad genetic test had been performed.

* MGH and Stanford Cohorts: Candidate Hereditary Breast and Ovarian Cancer (HBOC) patients who met NCCN guidelines for BRCA1/2 testing were selected. In the three retrospective cohorts about 23% of the patients are BRCA1/2 positive. Most (>90%) were independently tested for BRCA1 and/or BRCA2 by a well-established laboratory using traditional technologies. The completeness of prior data available to us varies: BRCA1/2 full sequencing was performed for 75% of tested cases, while only limited testing (e.g. small mutation panel or a single site test) was performed for the remainder. In 40% of cases copy-number data are available, and ~30% of those tests fully analyzed BRCA1/2, while 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 resources NA12878 (CEPH/Utah Pedigree 465) and NA12451 (Yoruban trio child). Whole-genome sequences from Complete Genomics and from the Illumina Platinum set is used for comparison in the CEPH samples. Data from Complete Genomics and from the 1000 Genomes Pilot 2 is used for the Yoruba. The public data sets are screened to exclude errors: variant positions with impossible or unlikely inheritance in the public data are considered “no calls”, and the two public DNA sets for each sample are intersected with discrepancies also considered “no calls”. The remaining high confidence variant and invariant positions are then compared to the Invitae data. Unlike the cohorts described above, the majority of the variants compared in these samples are benign.

* Certain sequence variants posed specific challenges for NGS methods that led to development of new bioinformatic solutions:

  - BRCA2: 9203del126 in Stanford case 8178, initially missed by NGS
  - Challenge: Deletion spanning part of but not all of a full exon
  - Solution: Split-read algorithm implemented to detect this type of event

  - PLP1: del exon 3-4 in NA13434, initially missed by NGS
  - Challenge: Deletion spanning parts, but not all, of 2 neighboring exons
  - Solution: Split-read algorithm developed to detect this type of event

  - MSH2: 942+3A>T in NBSC-2, initially missed by NGS
  - Challenge: Splice site changing mutation next to a 25b intronic poly-A tract
  - Solution: PolyMNP algorithm modified to call this site

* Variant interpretation for both sequence variations and deletion duplication events was based on criteria from the 2013 draft “ACMG/CAP/AMP Guidelines for Interpretation of Sequence Variants” with consultation of the scientific/medical literature and public databases. Clinical performance was evaluated by comparing the variant interpretation issued by the Invitae clinical team and the BRCA1/2 clinical testing report provided by an independent lab. Benign variants were not included in these comparisons because they are usually not reported.