A peer-reviewed version of this preprint was published in PeerJ on 28 June 2016.

View the peer-reviewed version (peerj.com/articles/2162), which is the preferred citable publication unless you specifically need to cite this preprint.

Design and validation of a next generation sequencing assay for hereditary breast and ovarian cancer


Hereditary breast and ovarian cancer syndrome, caused by a germline deleterious variant in the BRCA1 or BRCA2 genes, is characterized by an increased risk for breast, ovarian, pancreatic and other cancers. Identification of those who have a BRCA1/2 mutation is important so that they can take advantage of genetic counseling, screening, and potentially life-saving prevention strategies. We describe the design and analytic validation of the Counsyl Inherited Cancer Screen, a next-generation-sequencing-based test to detect pathogenic variation in the BRCA1 and BRCA2 genes. We demonstrate that the test is capable of detecting single-nucleotide variants (SNVs), short insertions and deletions (indels), and copy-number variants (CNVs, also known as large rearrangements) with zero errors over a 96-sample validation set consisting of samples from cell lines and deidentified patient samples, including the well-characterized NA12878 sample from HapMap/1000 Genomes.
Design and validation of a next generation sequencing assay for hereditary breast and ovarian cancer


Counsyl Inc., South San Francisco, CA, USA

* Correspondence should be addressed to Hyunseok P. Kang (peter@counsyl.com)

ABSTRACT

Hereditary breast and ovarian cancer syndrome, caused by a germline deleterious variant in the BRCA1 or BRCA2 genes, is characterized by an increased risk for breast, ovarian, pancreatic and other cancers. Identification of those who have a BRCA1/2 mutation is important so that they can take advantage of genetic counseling, screening, and potentially life-saving prevention strategies. We describe the design and analytic validation of the Counsyl Inherited Cancer Screen, a next-generation-sequencing-based test to detect pathogenic variation in the BRCA1 and BRCA2 genes. We demonstrate that the test is capable of detecting single-nucleotide variants (SNVs), short insertions and deletions (indels), and copy-number variants (CNVs, also known as large rearrangements) with zero errors over a 96-sample validation set consisting of samples from cell lines and deidentified patient samples, including the well-characterized NA12878 sample from HapMap/1000 Genomes.

INTRODUCTION

Clinical Scenario and Public Health Importance

Hereditary breast and ovarian cancer syndrome (HBOC) is associated with mutations in tumor suppressor genes BRCA1 and BRCA2. Genetic analysis for individuals who are at risk for HBOC has become widely accepted. Several professional organizations and expert panels, including the National Comprehensive Cancer Network (NCCN) (National Comprehensive Cancer Network, 2014), the American Society of Clinical Oncology (ASCO) (Robson et al., 2010), the American Society of Human Genetics (ASHG) (Statement of the American Society of Human Genetics on genetic testing for breast and ovarian cancer predisposition, 1994), the American College of Medical Genetics and Genomics (ACMG) (Hampel et al., 2015), the National Society of Genetic Counselors (NSGC) (Hampel et al., 2015), the U.S. Preventive Services Task Force (USPSTF) (Nelson et al., 2014), the Society of Gynecologic Oncologists (SGO) (Lancaster et al., 2007), and the European Society for Medical Oncology (ESMO) (Balmaña et al., 2011) have developed clinical criteria and practice guidelines for identifying individuals who may benefit from BRCA1 or BRCA2 mutation testing. A selection of these is summarized below.

According to the NCCN guidelines, personalized risk assessment, genetic counseling, and often BRCA1/2 testing and management are recommended for individuals with a significant personal
and/or family history of breast, ovarian, pancreatic and/or prostate cancer (National Comprehensive Cancer Network, 2014).

ASCO recommends genetic testing when there is personal or family history suggestive of genetic cancer susceptibility, the test can be adequately interpreted, and the results will aid in diagnosis or medical management of the patient or family member at hereditary risk for cancer. It also recommends genetic testing only when pre-test and post-test counseling are included (Robson et al., 2010).

The USPSTF guidelines recommend that primary care providers prescreen women with a family history of breast or other cancers to identify individuals at an increased risk for germline mutations in the \textit{BRCA1} and \textit{BRCA2} genes. Women with positive screening results should receive genetic counseling and, if indicated after counseling, \textit{BRCA} testing (Grade B) (Nelson et al., 2014). The USPSTF recommends against routine genetic counseling or \textit{BRCA} testing for women whose family history is not associated with an increased risk for mutations in the \textit{BRCA1} or \textit{BRCA2} genes (Grade D) (Nelson et al., 2014).

SGO recommends genetic risk assessment for individuals with a personal risk of more than approximately 20% to 25% for an inherited predisposition to cancer and states that it may be helpful for patients with more than approximately 5% to 10% risk. Genetic testing for cancer predisposition requires informed consent that should encompass pre-test education and counseling about the risks, benefits, and limitations of testing, including the implications of both positive and negative genetic test results (Lancaster et al., 2007).

The ESMO clinical practice guidelines indicate that \textit{BRCA} testing criteria may differ between countries based on mutation prevalence (Balmaña et al., 2011). Widely accepted clinical criteria for referral include: three or more breast and/or ovarian cancer cases, at least one <50 years; two breast cancer cases <40 years; male breast cancer and ovarian cancer or early onset female breast cancer; Ashkenazi Jew with breast cancer of <60 years; young onset bilateral breast cancer; and breast and ovarian cancer in the same patient. In some countries, the criterion for testing is based on an \textit{a priori} 10–20% probability of finding a mutation based on predictive models such as \textit{BRCAPRO}, \textit{BOADICEA} or Manchester Score (Fischer et al., 2013; Kast et al., 2014). The performance of the models can vary in specific ethnic groups. For instance, the \textit{BRCAPRO} model appeared to best fit a series of French Canadian families (Oros et al., 2006).

As suggested by various guidelines, individuals identified with \textit{BRCA1} or \textit{BRCA2} mutation are at significantly increased risk for breast, ovarian, prostate, pancreatic and possibly other cancers: a 12% general population risk for breast cancer rises to 50-80% for \textit{BRCA1} mutation carriers or 40-70% for \textit{BRCA2} mutation carriers (Petrucelli, Daly & Feldman, 2015). Recommended risk-reducing options include increased screening, chemoprevention and/or prophylactic surgery (Balmaña et al., 2011; Hampel et al., 2015; Lancaster et al., 2007; National Comprehensive Cancer Network, 2014; Nelson et al., 2014; Robson et al., 2010; Statement of the American Society of Human Genetics on genetic testing for breast and ovarian cancer predisposition, 1994). Table 1 summarizes these options and their effect on cancer risks.

Genetic testing for \textit{BRCA} mutation status has the potential to offer multiple benefits, including: identification of high-risk individuals who will benefit from the initiation of cancer risk
management; identification of noncarriers in families with a known mutation, who do not need to have rigorous cancer screening; and perhaps relief of anxiety through increasing the understanding of medical options. However, 20-73% of mutation carriers may not be identified with current guidelines (Alsop et al., 2012; Brozek et al., 2012; Frank et al., 2002; Kang et al., 2014; Norquist et al., 2013) or only meet current guidelines once they are diagnosed with ovarian cancer or early onset breast cancer, resulting in some researchers to call for more inclusive guidelines or even population screening (Finch et al., 2014; Gabai-Kapara et al., 2014; Metcalfe et al., 2013). It is also important to consider limitations and pitfalls of BRCA mutation testing, including the possibility of uncertain or uninformative results, potential for psychological distress, and effect on family members.

MATERIALS AND METHODS

Ethics Statement
The study was approved by Western Institutional Review Board (IRB number 1145639) and complied with the Health Insurance Portability and Accountability Act (HIPAA). The information associated with patient samples was de-identified in accordance with the HIPAA Privacy Rule. A waiver of informed consent was requested and approved by the IRB.

Test Description
The reportable range of the test is all coding exons of BRCA1 and BRCA2, 20 bp into the introns from intron/exon junctions, and selected intronic regions where pathogenic variants have been reported in the literature. DNA from a patient’s blood or saliva sample is isolated and then fragmented by sonication. The fragmentet DNA is converted to an adapter-ligated sequencing library; samples are multiplexed and identified by molecular barcodes. Hybrid capture-based enrichment for BRCA1/2 targeted regions is performed on these multiplexed samples, after which next generation sequencing of the selected targets is performed with sequencing-by-synthesis on the Illumina HiSeq 2500 instrument. All SNPs, insertions/deletions, and large deletions/duplication within the reportable range are analyzed and classified by the method described in the section “Variant Classification”.

All target nucleotides are required to be covered with a minimum depth of 50 reads. Sequence reads are aligned to the hg19 human reference genome using the BWA-MEM algorithm (Li, 2013), also which trims sequencing adapters. Automated statistical analysis is used to identify and genotype single-nucleotide variants (SNVs) and short insertions and deletions (indels) following methods in GATK and FreeBayes (Garrison & Marth, 2012; McKenna et al., 2010).

The calling algorithm for copy number variants (insertions or deletions longer than 100bp) is described below. Ancillary quality-control metrics, including fraction of sample contamination, library complexity, and bias, are computed on the final output and used to exclude and re-run failed samples. All reportable calls are reviewed by licensed clinical laboratory personnel.

CNV Calling Algorithm
Reads are extracted from the Illumina instrument output, and aligned to the human reference genome using BWA.

Analysis is performed on a per-lane basis. A matrix of counts of reads for each putative CNV in each sample is created. Reads for all probes targeting the same CNV region are added together.
Let $d_{i,j}$ be a matrix representing the number of reads observed from the $i$th sample for the $j$th variant.

This matrix must be normalized. To protect against normalization issues due to individual samples with very large CNVs (such as a whole-gene deletion), we generate a normalization matrix $n_{i,j}$ by removing the highest variance probes from the total data set $D$ via the invariant set method described in (Li & Hung Wong, 2001).

The data matrix $d$ is then normalized in two steps:

$$d_{i,j}' = \frac{d_{i,j}}{\text{mean}(n_{i,j} \text{ for all } j)}$$

$$d_{i,j}'' = d_{i,j}' / \text{mean}(n_{i,j} \text{ for all } i)$$

For each putative CNV $j$ in sample $i$, a hypothetical copy number and corresponding $Z$-score is computed:

$$c_{i,j} = 2 * d_{i,j}''$$

$$z_{i,j} = \frac{(d_{i,j}'' - \text{mean}(d_{i,j}'' \text{ for all } i))}{\text{stdev}(d_{i,j}'' \text{ for all } i)}$$

A CNV call is considered confidently non-reference if $\text{abs}(z) \geq 4$ and the estimate $c$ is $<1.2$ or $>2.8$.

Variation Classification

We have designed custom curation software to compile information from a wide range of sources. For each variant, information is collected from the following: entries in public databases such as ClinVar (Landrum et al., 2014), the Human Gene Mutation Database (HGMD) (Stenson et al., 2003), and selected locus-specific databases (e.g., BIC (the Breast Cancer Information Core) (Szabo et al, 2000) and UMD-BRCA1/2 (Caputo et al., 2012)); population-specific frequencies in ESP (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), 2013), 1000 Genomes (Abecasis et al., 2012), and internal data; results of computational algorithms based on evolutionary conservation, structural modeling, and splice site predictors. A curation team also reviews articles in the medical literature that mention each variant and collects additional information from them such as numbers and clinical characteristics of cases and controls the variant was seen in and the results of functional assays. All of this information is analyzed and variants are categorized according to ACMG Standards and Guidelines for the Interpretation of Sequence Variants (American College of Medical Genetics and Genomics, 2015) to arrive at a final classification of benign, deleterious, or unknown. All variants that are known or predicted to be deleterious are reported; patients and providers have an option to have variants of uncertain significance reported as well. Final variant classifications are regularly uploaded to ClinVar.

RESULTS

Evidence Overview
Data to calculate the validation metrics were compiled by testing three classes of samples: deidentified blood samples (N=25), deidentified paired blood and saliva samples (3 pairs), genomic DNA reference materials obtained from Coriell (N=56), and deidentified DNA samples provided by external laboratories (N=15) (Table 2).

Sequence data of 41 Coriell samples was compared to reference data obtained from the 1000 Genomes project and sequence data for NA12878 (a Coriell sample) was compared to high-quality reference data published by Illumina, Inc. (http://www.illumina.com/platinumgenomes/). Sequence data of 15 samples from the BIC BRCA1/BRCA2 Mutation Panel, available from Coriell, was analyzed to confirm the detection of documented variants in BRCA1/BRCA2. Data for copy number calls was compared to calls provided by reference labs, when available, and MLPA assays otherwise, on 56 samples: 15 samples from reference labs; 15 samples from the BIC BRCA1/BRCA2 Mutation Panel available from Coriell; 25 random blood samples; and NA12878. Sequence data and CNV calls for the saliva samples were compared to the respective paired blood sample.

Analytic Validity
The results of the Counsyl Inherited Cancer Screen validation are presented below in Tables 3 and 4. For SNPs and small insertions/deletions, 536 true positive calls, 12,920 true negative calls and no false positive or false negative calls were observed from the analysis of 57 samples. For copy number variants, 60 true positive calls, 2,736 true negative calls and no false positive or false negative calls were observed from the analysis of 40 analyzed samples. The accuracy, sensitivity and specificity are therefore all 1.0 for SNPs, small insertions/deletions, and copy number variants. The results from paired blood and saliva samples (n=3) were 100% concordant.

However, there are some limitations to the study presented here. The validation was limited to mostly blood-derived and Coriell cell line samples and included only three saliva samples.

DISCUSSION

Clinical Validity And Utility
Deleterious mutations in the BRCA genes are known to be associated with increased risk for breast, ovarian and other cancers. For women, the risk of developing breast cancer by age 70 is approximately 60-70% for BRCA1 and 45–55% for BRCA2 mutation carriers. The cumulative ovarian cancer risk by age 70 (including fallopian tube and primary peritoneal carcinomas) is 40% for BRCA1 and 20% for BRCA2 mutation carriers respectively (Antoniou et al., 2003; Chen & Parmigiani, 2007; King, Marks, & Mandell, 2003). Identification of those who have a BRCA1/2 mutation is important so that they can take advantage of genetic counseling, screening, and potentially life-saving prevention strategies.

The optimal cancer risk management approach for BRCA1/2 mutation carriers continues to evolve. For breast cancer risk management, current options include intensive screening, chemoprevention, and risk-reducing surgery (National Comprehensive Cancer Network, 2014; Nelson et al., 2015; Petrucelli, Daly & Feldman, 2015). Prophylactic bilateral mastectomies (PBM) showed an 85%–100% reduction in breast cancer risk in retrospective and prospective studies (Hartmann et al., 1999; Hartmann et al., 2001; Meijers-Heijboer et al., 2001; Rebbeck et
The Prevention and Observation of Surgical Endpoints study is the largest, prospective cohort study performed to estimate the risk reduction benefit of PBM in women with BRCA mutations (Rebbeck et al., 2004). Results of this trial supported a 90% reduction in risk with breast cancer being diagnosed in 2% of BRCA carriers undergoing PBM compared to 49% of carriers who did not. Risk reduction was increased to 95% in women undergoing prior or concurrent prophylactic bilateral oophorectomy (Rebbeck et al., 2004).

Intensive screening for early detection of breast cancer is an alternative approach for a woman who does not desire surgery. Screening guidelines are available from numerous organizations, including the NCCN (National Comprehensive Cancer Network, 2014) and USPSTF (Nelson et al., 2014). The addition of breast magnetic resonance imaging (MRI) to screening mammography has been shown to significantly increase sensitivity and lead to earlier detection of breast cancers (Hagen et al., 2007; Kriege et al., 2004; Kuhl et al., 2005; Lehman et al., 2005; Rijnsburger et al., 2010; Sardanelli et al., 2011; Warner et al., 2004; Warner et al., 2008). However, the impact of any surveillance strategy (including MRI) on breast cancer mortality has not been established.

Chemoprevention, specifically prophylactic use of tamoxifen, is recommended for BRCA1/2 carriers. The randomized, double-blind, Breast Cancer Prevention Trial (BCPT) demonstrated that tamoxifen reduced breast cancer incidence among healthy BRCA2 carriers by 62%. In contrast, tamoxifen use beginning at age 35 years or older did not reduce breast cancer incidence among healthy women with inherited BRCA1 mutations (King et al., 2001). A differential effect of tamoxifen in BRCA2 as compared to BRCA1 mutation carriers may be attributed to estrogen receptor (ER) status of BRCA1- and BRCA2-associated tumors. Tamoxifen might be expected to have an impact only against ER-positive tumors, and BRCA2-associated tumors have a greater likelihood than BRCA1-associated tumors of being ER-positive. However, in other settings, tamoxifen has shown benefit for both BRCA1- and BRCA2-associated tumors, irrespective of ER-status (Foulkes et al., 2002; Gronwald et al., 2006; Narod et al., 2000).

Ovarian cancer risk management options are more limited, with no proven effective early detection method available. Risk reducing salpingo-oophorectomy (RRSO) has been shown to reduce the risk of developing breast cancer by approximately 50%, with higher benefits associated with earlier age at surgery, and that of ovarian/fallopian tube cancer by approximately 80% to 90% (Domchek et al., 2010; Finch et al., 2014; Haber, 2002; Kauff et al., 2002; Rebbeck et al., 2002). In addition, one study showed a 69% reduction in all-cause mortality associated with RRSO among BRCA1/2 mutation carriers (Domchek et al., 2010).

For chemoprevention of ovarian cancer, oral contraceptive use has been associated with a decrease in ovarian cancer risk. A meta-analysis of 18 studies, which were either case-control or retrospective cohort studies, of oral contraceptive use in BRCA1 and BRCA2 mutation carriers and included 2855 breast cancer cases and 1503 ovarian cancer cases, demonstrated a significantly reduced risk of ovarian cancer [summary relative risk (SRR), 0.50, 95% CI 0.33–0.75]. For each additional 10 years of oral contraceptive use, there was a significantly reduced ovarian cancer risk (SRR 0.64, 95% CI 0.53–0.78) (Iodice et al., 2010).
In conclusion, *BRCA1* and *BRCA2* are the most prevalent high-penetrance breast/ovarian cancer susceptibility genes identified to date. It is important to identify individuals who have mutations in these genes so that they can benefit from surveillance and preventative options, primarily for breast and ovarian cancers.

**Competing Interests**
All authors are employees and shareholders of Counsyl, Inc.

**REFERENCES**


Table 1. *BRCA1* and *BRCA2* cancer risk management options and effectiveness

<table>
<thead>
<tr>
<th>Risk Management Options</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prophylactic mastectomy</td>
<td>Up to 90% reduction in breast cancer risk (Hartmann et al., 1999; Meijers-Heijboer et al., 2001)</td>
</tr>
<tr>
<td>Prophylactic oophorectomy</td>
<td>~50% reduction in breast cancer risk when performed premenopausally (more pronounced effect for <em>BRCA2</em> mutation carriers compared to <em>BRCA1</em>) (Kauff et al., 2002; Kauff et al., 2008) Up to 96% reduction in ovarian cancer risk (Olopade &amp; Artioli, 2004; Rebbeck et al., 2002; Rutter et al., 2003)</td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>Up to 62% reduction in breast cancer risk among <em>BRCA2</em> mutation carriers Up to 50% contralateral breast cancer risk reduction in both <em>BRCA1</em> and <em>BRCA2</em> Limited data but appears to be more effective in <em>BRCA2</em> mutation carriers compared to <em>BRCA1</em> (King et al., 2001; Metcalfe et al., 2005; Narod et al., 2000)</td>
</tr>
<tr>
<td>Oral contraceptives</td>
<td>Up to 50% reduction in ovarian cancer risk (Iodice et al., 2010)</td>
</tr>
<tr>
<td>Breast MRI/mammogram</td>
<td>No risk reduction, but earlier detection (Kuhl et al., 2010; Sardanelli et al., 2011; Warner et al., 2011)</td>
</tr>
<tr>
<td>Ovarian cancer screening (transvaginal ultrasound and serum cancer antigen 125 (CA-125))</td>
<td>No risk reduction and no effect on cancer mortality (Buys et al., 2011; Clarke-Pearson, 2009)</td>
</tr>
</tbody>
</table>
Table 2. Source of samples and reference data used in validation.

<table>
<thead>
<tr>
<th>Mutation Type</th>
<th>Test Samples</th>
<th>Reference Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP/Indel</td>
<td>41 Coriell Cell Line Samples</td>
<td>1000 Genomes Project Exomes</td>
</tr>
<tr>
<td></td>
<td>NA12878</td>
<td>Illumina Platinum Genome</td>
</tr>
<tr>
<td></td>
<td>15 BIC samples</td>
<td>BIC reference data</td>
</tr>
<tr>
<td>CNV</td>
<td>15 reference lab samples</td>
<td>Reference lab results</td>
</tr>
<tr>
<td></td>
<td>25 random anonymized samples</td>
<td>Orthogonal confirmation by MLPA</td>
</tr>
</tbody>
</table>
Table 3. Performance of Counsyl Inherited Cancer Screen for SNPs and small insertions/deletions.

<table>
<thead>
<tr>
<th>Term or Formula</th>
<th>Value (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive calls TP</td>
<td>536</td>
</tr>
<tr>
<td>True negative calls TN</td>
<td>12920</td>
</tr>
<tr>
<td>False positive calls FP</td>
<td>0</td>
</tr>
<tr>
<td>False negative calls FN</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>((TP + TN)/(TP + FP + TN + FN))</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>(TP / (TP + FN))</td>
</tr>
<tr>
<td>Specificity</td>
<td>(TN / (TN + FP))</td>
</tr>
</tbody>
</table>
Table 4. Performance of Counsyl Inherited Cancer Screen for Copy Number Variants.

<table>
<thead>
<tr>
<th>Term or Formula</th>
<th>Value (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive calls</td>
<td>TP = 60</td>
</tr>
<tr>
<td>True negative calls</td>
<td>TN = 2736</td>
</tr>
<tr>
<td>False positive calls</td>
<td>FP = 0</td>
</tr>
<tr>
<td>False negative calls</td>
<td>FN = 0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>(TP + TN) / (TP + FP + TN + FN) = 1.0 (0.998628 – 1.0)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>TP / (TP + FN) = 1.0 (0.9398281 – 1.0)</td>
</tr>
<tr>
<td>Specificity</td>
<td>TN / (TN + FP) = 1.0 (0.9985979 – 1.0)</td>
</tr>
</tbody>
</table>