Development and validation of a 36-gene sequencing assay for hereditary cancer risk assessment

Valentina S. Vysotskaia [Corresp., 1], **Gregory J. Hogan** [1], **Genevieve M. Gould** [1], **Xin Wang** [1], **Alex D. Robertson** [2], **Kevin R. Haas** [1], **Mark R. Theilmann** [1], **Lindsay Spurka** [1], **Peter V. Grauman** [1], **Henry H. Lai** [1], **Diana Jeon** [1], **Genevieve Haliburton** [1], **Matt Leggett** [1], **Clement S. Chu** [1], **Kevin Lori** [1], **Jared R. Maguire** [1], **Kaylene Ready** [1], **Eric A. Evans** [1], **H. Peter Kang** [Corresp., 1], **Imran S. Haque** [1]

[1] Counsyl Inc, South San Francisco, California, United States

[2] Color Genomics Inc, Burlingame, California, United States

Corresponding Authors: Valentina S. Vysotskaia, H. Peter Kang
Email address: valentina@counsyl.com, peter@counsyl.com

The past two decades have brought many important advances in our understanding of the hereditary susceptibility to cancer. Numerous studies have provided convincing evidence that identification of germline mutations associated with hereditary cancer syndromes can lead to reductions in morbidity and mortality through targeted risk management options. Additionally, advances in gene sequencing technology now permit the development of multigene hereditary cancer testing panels. Here, we describe the 2016 revision of the Counsyl Inherited Cancer Screen for detecting single-nucleotide variants (SNVs), short insertions and deletions (indels), and copy number variants (CNVs) in 36 genes associated with an elevated risk for breast, ovarian, colorectal, gastric, endometrial, pancreatic, thyroid, prostate, melanoma, and neuroendocrine cancers. To determine test accuracy and reproducibility, we performed a rigorous analytical validation across 341 samples, including 118 cell lines and 223 patient samples. The screen achieved 100% test sensitivity across different mutation types, with high specificity and 100% concordance with conventional Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA). We also demonstrated the screen's high intra-run and inter-run reproducibility and robust performance on blood and saliva specimens. Furthermore, we showed that pathogenic Alu element insertions can be accurately detected by our test. Overall, the validation in our clinical laboratory demonstrated the analytical performance required for collecting and reporting genetic information related to risk of developing hereditary cancers.

Development and validation of a 36-gene sequencing assay for hereditary cancer risk assessment

Valentina S. Vysotskaia,[1]* Gregory J. Hogan,[1]* Genevieve M. Gould,[1]* Xin Wang,[1] Alex D. Robertson,[2] Kevin R. Haas,[1] Mark R. Theilmann,[1] Lindsay Spurka,[1] Peter V. Grauman,[1] Henry H. Lai,[1] Diana Jeon,[1] Genevieve Haliburton,[1] Matt Leggett,[1] Clement S. Chu,[1] Kevin Lori,[1] Jared R. Maguire,[1] Kaylene Ready,[1] Eric A. Evans, [1]^ H. Peter Kang,[1]^ Imran S. Haque[1]^

[1] Counsyl Inc., South San Francisco, CA, USA
[2] Color Genomics Inc., Burlingame, CA, USA
* These authors contributed equally to this work.
^ These authors jointly supervised this work.
Correspondence should be addressed to H. Peter Kang (research@counsyl.com).

ABSTRACT

The past two decades have brought many important advances in our understanding of the hereditary susceptibility to cancer. Numerous studies have provided convincing evidence that identification of germline mutations associated with hereditary cancer syndromes can lead to reductions in morbidity and mortality through targeted risk management options. Additionally, advances in gene sequencing technology now permit the development of multigene hereditary cancer testing panels. Here, we describe the 2016 revision of the Counsyl Inherited Cancer Screen for detecting single-nucleotide variants (SNVs), short insertions and deletions (indels), and copy number variants (CNVs) in 36 genes associated with an elevated risk for breast, ovarian, colorectal, gastric, endometrial, pancreatic, thyroid, prostate, melanoma, and neuroendocrine cancers. To determine test accuracy and reproducibility, we performed a rigorous analytical validation across 341 samples, including 118 cell lines and 223 patient samples. The screen achieved 100% test sensitivity across different mutation types, with high specificity and 100% concordance with conventional Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA). We also demonstrated the screen's high intra-run and inter-run reproducibility and robust performance on blood and saliva specimens. Furthermore, we showed that pathogenic Alu element insertions can be accurately detected by our test. Overall, the validation in our clinical laboratory demonstrated the analytical performance required for collecting and reporting genetic information related to risk of developing hereditary cancers.

INTRODUCTION

Tremendous advances in our knowledge of evaluating and treating patients with germline mutations associated with hereditary cancer syndromes have been realized in the past two decades. Multiple studies demonstrate the feasibility and clinical utility of genetic testing (Norton et al., 2007; Domchek et al., 2010; Kurian et al, 2014; Lynce and Isaacs, 2016). Most importantly, studies have provided convincing evidence that identification of hereditary cancer syndromes can lead to reductions in morbidity and mortality through targeted risk management options. For example, for unaffected women who carry a *BRCA1* or *BRCA2* mutation, risk-reducing salpingo-oophorectomy results in a significant reduction in all-cause mortality (3% vs. 10%; hazard ratio [HR] 0.40; 95% CI, 0.26–0.6), breast cancer-specific mortality (2% vs. 6%; HR 0.44; 95% CI, 0.26–0.76) and ovarian cancer–specific mortality (0.4 vs. 3%; HR 0.21; 95% CI, 0.06–0.8) when compared with carriers who chose not to undergo this procedure (Domchek et al., 2010).

Until recently, the traditional approach for germline testing was to test for a mutation in a single gene or a limited panel of genes (syndrome-based testing) using Sanger sequencing (Sanger et al., 1977), quantitative PCR (Barrois et al., 2004), and MLPA (Hogervorst et al., 2003). With advances in next-generation DNA sequencing (NGS) technology and bioinformatics analysis, testing of multiple genes simultaneously (panel-based testing) at a cost comparable to traditional testing is possible. NGS-based, multigene panels of 25 to 79 genes have been developed and are offered by several clinical diagnostic laboratories (Easton et al., 2015; Kurian & Ford, 2015; Lynce & Isaacs, 2016). Panel-based testing has proven to provide improved diagnostic yield (Castéra et al., 2014; Cragun et al., 2014; Kurian et al., 2014; LaDuca et al., 2014; Lincoln et al., 2015; Minion et al., 2015; Rehm, 2013). Among clinic-based studies that collectively assessed more than 10,000 patients who tested negative for *BRCA1/2* mutations, mutation prevalence in non-*BRCA* genes ranged from 4% to 16% (Castéra et al., 2014; LaDuca et al., 2014; Kurian et al., 2014; Maxwell et al., 2015; Tung et al., 2015). Some mutations were clinically unexpected (e.g., a *MSH6* mutation, consistent with Lynch syndrome, was found in a patient with triple-negative breast cancer) (Kurian et al., 2014), prompting calls for a change in screening and prevention recommendations.

Published validation studies demonstrate high analytical concordance between results from NGS and the traditional Sanger method for detection of sequence level variations (single-nucleotide variants, small deletions and insertions) (Bosdet et al., 2013; Chong et al., 2014; Judkins et al., 2015; Lincoln et al., 2015; Strom et al., 2015). However, detection of exon-level copy number variations and larger indels might be relatively challenging for NGS (Lincoln et al., 2015). To address this concern, some laboratories complement NGS with microarrays (Chong et al., 2014). Other laboratories achieve high accuracy of NGS-based copy number variation and indel detection using sophisticated bioinformatics pipelines (Lincoln et al., 2015; Kang et al., 2016; Schenkel et al., 2016). Although this is encouraging, it is important to consider the potential limitations of NGS for detection of larger insertions/deletions (indels) and copy number variants (CNVs, also known as deletions and duplications or large rearrangements). Samples with technically challenging classes of mutations should be included in analytical validation.

Here, we describe the development and validation of the 2016 revision of the Counsyl Inherited Cancer Screen, an NGS-based test to identify single nucleotide variants (SNVs), indels, and copy number variants in 36 genes associated with an elevated risk for breast, ovarian, colorectal, gastric, endometrial, pancreatic, thyroid, prostate, melanoma, and neuroendocrine cancers. To evaluate analytical performance of the test and ensure quality of results, we followed

the American College of Medical Genetics and Genomics (ACMG) guidelines for analytical validation of NGS methods (Rehm et al., 2013). The validation study included both well-characterized cell lines (N=118) and de-identified patient samples (N=223) with clinically relevant variants.

MATERIALS AND METHODS

Institutional Review Board Approval

The protocol for this 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.

Multigene Panel Design

Thirty six genes associated with hereditary forms of cancer, including breast, ovarian, colorectal, gastric, endometrial, pancreatic, thyroid, prostate, melanoma, and neuroendocrine, were selected for development of the Counsyl Inherited Cancer Screen panel. The genes are: *APC, ATM, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A, CHEK2, EPCAM, GREM1, MEN1, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, POLD1, POLE, PTEN, RAD50, RAD51C, RAD51D, RET, SDHA, SDHB, SDHC, SMAD4, STK11, TP53,* and *VHL* (Table 1). Twenty eight of the 36 genes were specifically included due to the availability of patient management guidelines by NCCN or other professional societies. Further details regarding the panel are available in Table S1.

The selected genes are tested for SNVs, indels, and CNVs throughout coding exons and 20 bp of flanking intronic sequences. Additionally, known deleterious variants outside the coding regions are sequenced. In *EPCAM*, only large deletions that include exon 9 are reported as these mutations are known to silence the *MSH2* gene (Tutlewska, Lubinski, and Kurzawski, 2013). In *GREM1*, specific pathogenic duplications in the promoter, which are commonly associated with individuals of Ashkenazi Jewish descent, are covered. Specifically, the screen targets the three most common promoter duplications in *GREM1* (coordinates with respect to GRCh37/hg19 reference assembly)*:*

- chr15:32,964,939-33,004,759 (40kb)
- chr15:32,986,220-33,002,449 (16kb)
- chr15:32,975,886-33,033,276 (57kb)

For *PMS2*, exons 11-15 are excluded from the reportable region of interest (ROI) because of high similarity between this portion of *PMS2* and its highly homologous pseudogene *PMS2CL*. In *RET*, exon 1 is not sequenced due to high guanine-cytosine (GC) content.

Next Generation DNA Sequencing

Our application of next-generation DNA sequencing is performed as described previously (Kang et al., 2016). Briefly, DNA from a patient's blood or saliva sample is isolated, quantified by a dye-based fluorescence assay and then fragmented to 200-1000 bp by sonication. The fragmented DNA is converted to a sequencing library by end repair, A-tailing, and adapter ligation. Samples are then amplified by PCR with barcoded primers, multiplexed, and subjected to hybrid capture-based enrichment with 40-mer oligonucleotides (Integrated DNA Technologies, Coral, IL) complementary to targeted regions. Next generation sequencing of the

selected targets is performed with sequencing-by-synthesis on the Illumina HiSeq 2500 instrument to a mean sequencing depth of ~650x. All target nucleotides are required to be covered with a minimum depth of 20 reads.

Bioinformatics Processing

Sequencing reads are aligned to the hg19 human reference genome using the BWA-MEM algorithm (Li, 2013). Single-nucleotide variants and short indels are identified and genotyped using GATK 1.6 and FreeBayes (McKenna et al., 2010; Garrison & Marth, 2012). The calling algorithm for copy number variants is described below. All SNVs, indels, and large deletions/duplications within the reportable range are analyzed and classified by the method described in the section "Variant Classification". All reportable calls are reviewed by licensed clinical laboratory personnel.

CNV Calling Algorithm

Copy number variants for samples are determined by inspecting the number of mapped reads observed at targeted positions in the genome across samples in a flowcell lane. Our method is based upon previous successful approaches applying hidden Markov models (HMMs) to exome sequencing data (Plagnol et. al. 2012) with modifications presented below that have been optimized for accurate resolution of CNVs based on the particulars of the sequencing technology. As sequencing depth is linearly proportional to the number of copies of the genome at that position, we construct a statistical model for the likelihood of observing a given number of mapped reads $d_{i,j}$ at a given genomic position i for sample j with copy number $c_{i,j}$.

The expected number of reads is dependent upon 3 factors: the average depth for that targeted location across samples μ_i, the average depth for that particular sample across targeted positions μ_j, and the local copy number of the sample's genome at that targeted position. These are first determined by finding the median depth at targeted region across all N_s samples in an analyzed flowcell lane

$$\mu_i = \frac{\sum_j \ d_{i,j}}{N_s}$$

then the sample dependent factor μ_j is found by taking the median across all N_p positions in genome after normalizing for the expected number of reads at each position

$$\mu_j = \frac{\sum_i \ d_{i,j}/\mu_i}{N_p}$$

Combining these factors the observed data are modeled by the negative binomial distribution

$$p(d_{i,j}|c_{i,j}) = NegBinom(d_{i,j}|\mu = c_{i,j}\mu_i\mu_j, r = r_i)$$

This characterization has been found to accurately model the observed number of reads from previous targeted sequencing experiments (Anders & Huber, 2010).

In the negative binomial model, the variance parameter r_i accounts for regions of the genome where sequencing depth is observed to follow idealized Poisson statistics in the limit that $r \to \infty$ and regions that are excessively noisy with respect to observed number of reads when $r \to 0$. r_i may be estimated as

$$r_i = \frac{{\mu_i}^2}{Var_j[d_{i,j}] - \mu_i}$$

which is found to closely model the empirical distribution over several orders of magnitude in read depth.

Because duplications and deletions will simultaneously impact the expected depth of all genomic positions encompassing the variant, depth data from spatially adjacent positions are correlated. We leverage the HMM to account for this correlation. The HMM's state transition probabilities between wild-type and copy-number-variant are parameterized by matching the average length of such variations observed in human population (Sudmant et. al. 2015) through setting $p_{CNV \to WT} = 1/6200$ between each subsequent base-pair and a prior on the frequency of such variations

$$\frac{p_{WT \to CNV}}{p_{CNV \to WT}} = p_{CNV}$$

The prior $p_{CNV} = 0.001$ was determined by balancing the thresholds for confident calling and retesting of calls to achieve the desired sensitivity and specificity, and the prior was set independently of this validation.

Detecting CNVs using this probabilistic framework invokes the Viterbi algorithm (Korn et. al., 2008) to determine the most likely number of copies at every targeted region within a sample. Any contiguous regions of duplication or deletion produce a reported variant, and the confidence of that call is determined by aggregating the posterior probability of the call $\sum_{i \epsilon CNV} p\,(c_{i,j} \neq 2)$ not being wildtype over the called region.

All copy-number called variants are inspected for quality of raw data by human review, and observed positive variants are rerun in our production SOP for verification of the call. Samples that emit low confidence called variants are additionally rerun to resolve a confident genotype.

Detection of Alu Insertions

Alu positives were detected by looking for Alu sequences in reads overlapping with Alu insertion positions. All insertions were only tested for at positions where the sequence had been previously confirmed by Sanger sequencing. At the site of an Alu insertion, the Alu sequence is soft-clipped by BWA alignment. These soft-clipped reads were compiled; duplicate reads were discarded; and the remaining reads with sequences matching the known Alu sequence at this site were tallied. Sites with three unique reads matching the Alu sequence were called as Alu positive.

Pre- and Post-sequencing Quality Metrics

To ensure the quality of the results obtained from the assay, 27 different review checkpoints (Table S2) were developed. Ancillary quality-control metrics are computed on the sequencing output and used to exclude and re-run failed samples, and include the fraction of sample contamination (<5%), extent of GC bias, read quality (percent Q30 bases per Illumina specifications), depth of coverage (per base minimum coverage >=20x and mean coverage of >250x), and region of interest (ROI) coverage (100%). Calls that do not meet criteria listed in

Table S2 are set to "no-call". To ensure clinical calling accuracy, all calls and no-calls for potentially deleterious, variants of unknown significance, and uncurated variants are manually reviewed by laboratory personnel and are subject to override if warranted, based on a pre-established protocol.

Variant Classification

Variants are classified using multiple lines of evidence according to the ACMG Standards and Guidelines for the Interpretation of Sequence Variants (American College of Medical Genetics and Genomics, 2015; Richards et al., 2015). Variants that are known or predicted to be pathogenic 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 (Landrum et al., 2014), a peer-reviewed database created with a goal of improving variant interpretation consistency between laboratories.

Statistical Analysis

Variant calls were defined as true positive for variants identified by the Counsyl Inherited Cancer Screen and by independent testing (the 1000 Genomes Project or MLPA/Sanger data), false positive for variants identified by the Counsyl test but not by the independent data, and false negative for variants identified by the independent data but not by the Counsyl test. To estimate true negatives, we counted polymorphic sites (positions at which we observed non-reference bases in any sample) with concordant negative results across all considered samples. No-calls were censored from the analysis. As no-calls have the potential to introduce clinically relevant false negatives, we separately examined the no-calls containing potentially deleterious alleles by treating no-calls as homozygous reference and comparing to the 1000 Genomes calls. We found all no-calls when treated as homozygous reference were concordant with the exception that one comparison was inconclusive due to low allele balance in both our data and the exome data from the 1000 Genomes Project (Table S8).

Validation metrics were defined as: Accuracy = (TP + TN) / (TP + FP + TN + FN); Sensitivity = TP / (TP + FN); Specificity = TN / (TN + FP); FDR = FP / (TP + FP), where TP=true positives, TN=true negatives, FP=false positives, FN=false negatives, and FDR=false discovery rate. The confidence intervals (CIs) were calculated by the method of Clopper and Pearson (Clopper & Pearson, 1934). To estimate reproducibility within and between runs, the ratio of concordant calls to total calls was calculated.

Study Samples

The validation sample set comprised (a) 111 genomic DNA reference materials purchased from the Coriell Cell Repositories (Camden, NJ), (b) MLH1/MSH2 exon copy number reference panel from the National Institute for Biological Standards and Control (NIBSC) (N=7), and (c) 223 deidentified patient samples used for MLPA- and Sanger-based confirmation (Tables 2, S3, and S4).

The validation set included samples with reference data for SNVs and indels (the 1000 Genomes Project), a broad range of indels (both short <=10 bp and long >10 bp) characterized by Sanger sequencing, homopolymer-associated variants, Alu element insertions, and both single- and multi-exon copy-number variants characterized by MLPA (Table 3). Validation material was derived from cell lines, blood, and saliva samples. Collectively, the validation set provides broad

coverage of known relevant types of genomic variation across the reportable region of the test (Table 3). A list of the validation samples from Sanger and MLPA confirmation is provided in Table S4.

RESULTS

Test description

We developed an NGS-based test that interrogates 36 genes associated with hereditary cancer risk (Table 1). The majority of the 36 genes were selected based on the availability of patient management guidelines developed by NCCN or other professional societies. The reportable region of interest (ROI) of the test is 124,245 bp representing coding exons, intron boundaries and non-exonic mutation-containing regions (Table 1). The wet lab protocols and reagents are carefully optimized to ensure 100% coverage of targeted base pairs at an average depth of 650 reads and a minimal depth of 20 reads sufficient for robust detection of multiple classes of genomic alterations: single-nucleotide variations, indels, and copy number variations.

Validation approach

Several regulations, including the Clinical Laboratory Improvement Act of 1988 (CLIA), the ACMG guidelines for analytical validation of NGS methods (Rehm et al., 2013), as well as various quality standards for diagnostic laboratories require rigorous analytical validation of panel tests for clinical use. In contrast to diagnostic assays for a single gene or a limited panel of genes (syndrome-based testing), analytical validation of a NGS-based test assaying 36 genes for multiple types of genomic alterations is a complex task. To address this challenge, we developed a representative validation approach with reference samples selected to cover variant and specimen variability that may affect test accuracy and reproducibility for clinical use.

To measure the accuracy of SNV and indel detection, we tested samples from the 1000 Genomes Projects with reference data for SNVs and indels in all 36 genes. Testing on the 1000 Genomes Project samples allows us to assess the ability to call commonly observed variant types and the ability to test calling in regions that may be difficult for NGS due to considerable sequence homology (e.g. *CHEK2*, *SDHA*, and *PMS2*) or low complexity (homopolymer runs). However, the 1000 Genomes reference samples provide limited validation for technically challenging variants like CNVs, larger indels, and Alu insertions. To build a collection of reference material to test such challenging variants, we identified relevant patient samples tested with a previous version of the Counsyl test (a 24-gene panel) and orthogonally confirmed each of the positive samples by either Sanger or MLPA. Using these cohorts of reference samples (e.g. samples with CNVs), we could then assess call accuracy for each type of technically challenging variant on this newly designed 36-gene panel. Finally, to validate test reproducibility, we examined SNV, indel, and CNV calls in cell line and patient (blood and saliva) samples processed independently in several batches (inter-run reproducibility) or tested repeatedly in the same batch (intra-run reproducibility).

Analytical validation for SNVs and indels

The analytical validation of the Inherited Cancer Screen was performed according to ACMG guidelines (Rehm et al., 2013) and in accordance with the requirements of CLIA for medical laboratories. SNV and indel detection was examined on a 101-sample validation set consisting of reference samples from the 1000 Genomes Project with known SNV and indel sites across the targeted regions (Tables 3 and S5). Counsyl sequence data for 36 genes were compared to

reference data obtained from the 1000 Genomes Projects. Out of 42,925 total calls validated, 18 calls were discordant between Counsyl and the 1000 Genomes Project (Table S6). One of the 18 discordances was a potential false positive variant call, identified as a variant by the Counsyl test, but identified as reference by the 1000 Genomes Project. The remaining 17 calls were potential false negative variants identified by the 1000 Genomes Project, but not by the Counsyl test. Manual review of the 1000 Genomes reference data for each of the discordant sites using the Integrated Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdóttir, Robinson, Mesirov, 2013) found that a large portion of the discordant calls came from hard-to-sequence (e.g., highly homologous *SDHA* gene) or low-coverage regions, which is a reported limitation in the 1000 Genomes Project (1000 Genomes Project Consortium, 2012). With that in mind, each of the discordant sites was subjected to Sanger sequencing as an independent testing method and the data from Sanger sequencing supported all 18 of Counsyl's calls as true positives or true negatives (Table S6).

Analytical validation results of Counsyl's test for SNV and indel detection is presented in Table 4. Counsyl's test identified 5182 true positive calls, 37,743 true negative calls, and no false positive nor false negative calls, resulting in 100% sensitivity (95% CI, 100%-99.93%), 100% specificity (95% CI, 100%-99.99%) and 0% FDR (95% CI, 0-0.0007%) of the test for detecting SNVs and indels.

Validation of challenging variants

CNVs

To assess the accuracy of CNV detection, we measured the concordance between Counsyl's test results on 44 blood and saliva samples with CNV positives confirmed by MLPA (N=43) or Sanger (N=1) (Tables 2 and S4b). For one CNV positive sample (Counsyl_147), Sanger sequencing was used for orthogonal confirmation; MLPA analysis of this sample failed to identify the partial deletion of exon 15 in *APC* because the deletion was relatively small and fell between the MLPA probes (Table S4b). For the patient sample Counsyl_128, two duplications affecting exons 8-9 of *EPCAM* and exons 1-16 of *MSH2* were detected and confirmed by MLPA. Additionally, 5 NIBSC reference samples with known CNVs in the *MLH1* and *MSH2* genes were included in the validation. Among the 49 tested samples (a total of 50 CNVs), 12 had a single-exon deletion or duplication, which can be technically challenging for a NGS-based assay (Table 3).

As shown in Table 5, we detected all 50 CNVs, including 12 single-exon events, demonstrating the high sensitivity of the assay (100%; 95% CI, 100%-93%). Furthermore, no additional CNV calls were made in the 49 sample cohort, resulting in 100% specificity (Table 5).

Challenging indels

To measure accuracy for detecting indels, we built a cohort (N=82) of patient samples with variants of a range of sizes, including both short (≤10bp) and the more technically challenging long (>10bp) deletions or insertions (Tables 3 and S6a). These samples were identified using a previous version of the Counsyl test (a 24-gene panel) and orthogonally confirmed by Sanger. We then tested these samples with the newly developed 36-gene panel and confirmed all of the expected indel calls; no false-positives nor false-negatives were observed in the 36-gene panel results (Table 5).

Alu insertions

Alu elements represent a special class of insertions and are known to be clinically important (Belancio et al., 2010). Alu insertions have been reported in *ATM*, *BRCA1, BRCA2*, and *BRIP1* (Belancio et al., 2010; Kennemer et al., 2016), including known examples of Alu insertion founder mutations (e.g., c.156_157insAlu in *BRCA2* exon 3 in Portuguese populations) (Peixoto et al., 2014). Accurate detection of Alu insertions is challenging, especially for traditional Sanger sequencing where longer Alu-containing alleles are usually out-competed during PCR (De Brakeleer et al., 2013). To test the sensitivity of our assay and bioinformatics pipeline for Alu insertion detection, we included 7 positive cases (Portuguese founder mutation in exon 3 of *BRCA2,* Alu insertion in *BRCA2* exon 25 and intronic Alu insertions in *ATM* and *MSH6)* in our validation study (Table 6). We confirmed that the Alu insertions identified by the Counsyl Inherited Cancer Screen were also detected by Sanger sequencing.

Reproducibility

In addition to establishing the test's analytical sensitivity and specificity, Counsyl's Inherited Cancer Screen was validated for intra- and inter-run call reproducibility. Intra-run reproducibility of SNV and indel calls was established by testing 8 cell lines and 13 blood or saliva samples in 2-3 replicates in the same batch, split across sequencer lanes. Inter-run reproducibility was validated by testing 8 cell lines and 84 patient blood or saliva samples in 2-3 different batches (Table S7a). Concordance between replicates was > 99.99%, with just one discordant call at a known benign homopolymer site in an intron of *ATM* (Table S7a).

For CNVs, intra-run and inter-run reproducibility was established using the Coriell sample NA14626 with a duplication of *BRCA1* exon 12 (Table S7b). Concordance between 8 replicates was 100%, with no differences between inter- and intra-run replicates observed.

DISCUSSION

The evidence base for genetic testing, counseling, risk assessment and management for hereditary cancer syndromes is rapidly evolving. The expansion of knowledge regarding cancer-risk associated genes and advances in gene sequencing technology now permit the development of multigene hereditary cancer testing panels. Recently, we have expanded the Counsyl Inherited Cancer Screen to 36 genes known to impact inherited risks for ten important cancers: breast, ovarian, colorectal, gastric, endometrial, pancreatic, thyroid, prostate, melanoma, and neuroendocrine. Twenty eight of the 36 genes were specifically selected for inclusion due to the availability of patient management guidelines by NCCN or other professional societies.

Accurate detection of clinically relevant genomic alterations in the targeted genes is critical and requires the interrogation of coding exons as well as selected non-coding regions with known pathogenic mutations. Furthermore, robust detection of a broad range of clinically relevant genomic alterations in routine clinical specimens, such as blood and saliva, is also required for a clinical-grade test. To address these challenges, we developed a clinical-grade, targeted NGS test for 36 genes. We carefully optimized and validated the probe design and NGS-based workflow using reference cell lines and clinical samples. We performed a comprehensive validation study and did not identify any false positives or false negatives. High sensitivity, specificity, accuracy and call reproducibility were observed across all call types, including those challenging for NGS, such as single- and multi-exon deletions/duplications (N=50), >10 bp indels (N=19) and Alu insertions (N=7).

Although some NGS validation studies report a higher false positive rate and require orthogonal confirmation of positive calls (Chong et al., 2014; Mu et al., 2016), high sensitivity

and specificity consistent with this report have been achieved in similar studies, both in our laboratory (Kang et al, 2016) and in other laboratories (Bosdet et al., 2013; Judkins et al., 2015; Lincoln et al., 2015; Strom et al., 2015). No false negatives were observed in our study, corroborating previous reports of high analytic accuracy of NGS relative to Sanger sequencing (99.965%) (Beck et al., 2016). However, another recent publication uses data from 20,000 NGS panel tests performed in a clinical setting (Ambry Genetics, Aliso Viejo, CA) to claim the necessity of Sanger confirmation of variants detected by NGS (Mu et al., 2016). This study observed a 99/7845 (1.3%) false positive rate and concluded that Sanger confirmation is needed to maintain high accuracy, particularly in difficult-to-sequence regions. In contrast to other work in the field, Mu et al. state that it was impossible with their pipeline to reach a zero false negative rate when filtering NGS variant calls for a zero false positive rate. For example, the *MSH2*:c.942+3A>T variant, which falls at the end of a stretch of 27 adenines, was missed by Mu et al. in 5 of 6 patients when they tuned their false positive rate to zero.

The results presented here support the high accuracy for NGS calls, including challenging variants in hard-to-sequence regions, and demonstrate that the requirement for secondary confirmation is a property of each particular NGS pipeline, not a generic property of all NGS protocols. The *MSH2*:c.942+3A>T variant, highlighted as difficult in the Mu et al. publication, was included and correctly called in our validation data. Indeed, our cell line and patient validation cohorts included 3,421 pathogenic and nonpathogenic variants (Table S5) in the gene set that exhibited false positives in Mu et al.'s study; for all 3,421 variants, we observed 100% analytical concordance with reference (1000 Genomes) and orthogonal confirmation (Sanger/MLPA) data.

The high accuracy reported here underlines the importance of using metrics beyond simple base and variant call quality to assess NGS variant calls. Table S2 shows the comprehensive set of metrics by which we assess each variant call. As one example, information on read directionality ("strand bias LOD") is incorporated into our pipeline, and would have eliminated many of the false positives encountered by Mu et al (in particular, the *MSH2* homopolymer site) without sacrificing sensitivity. Finally, the call review process described here includes visual inspection of all potentially deleterious calls.

For copy number variants, the low throughput of non-NGS-based CNV analysis methods combined with the low prevalence of CNVs makes it difficult to assess CNV calling sensitivity with precision. While in principle orthogonal testing of all negative CNV calls using MLPA, qPCR, or microarrays may uncover additional samples with copy number variants, this would constitute a large discovery effort with low probability of discovering a false negative. The development of a set of reference samples with a diverse deeply-characterized collection of copy number variants (analogous to the efforts of the Genome in a Bottle project) would be a great benefit to laboratory validation procedures.

In conclusion, we developed a 36-gene sequencing test for hereditary cancer risk assessment. We assessed test performance across a broad range of genomic alteration types and clinical specimen properties to support clinical use. We confirmed high analytical sensitivity and specificity in this validation study consisting of 5315 variants, including many technically challenging classes. The test is now offered by Counsyl's laboratory, which is CLIA certified (05D1102604), CAP accredited (7519776), and NYS permitted (8535).

REFERENCES

1. **1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012.** An integrated map of genetic variation from 1,092 human genomes. Nature 491:56-65 DOI:10.1038/nature11564.
2. **American College of Medical Genetics and Genomics. 2015.** Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Available at https://www.acmg.net/ (accessed 27 February 2015).
3. **Anders S, Huber W. 2010.** Differential expression analysis for sequence count data. Genome Biology 11:R106 DOI: 10.1186/gb-2010-11-10-r106. Article R106.
4. **Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkäs K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomäki K, Moilanen JS, Turnbull C, Seal S, Mannermaa A, Kallioniemi A, Lindeman GJ, Buys SS, Andrulis IL, Radice P, Tondini C, Manoukian S, Toland AE, Miron P, Weitzel JN, Domchek SM, Poppe B, Claes KBM, Yannoukakos D, Concannon P, Bernstein JL, James PA, Easton DF, Goldgar DE, Hopper JL, Rahman N, Peterlongo P, Nevanlinna H, King M-C, Couch FJ, Southey MC, Winqvist R, Foulkes WD, Tischkowitz M. 2014.** Breast-Cancer Risk in Families with Mutations in PALB2. New England Journal of Medicine 371(6):497-506 DOI:10.1056/NEJMoa1400382.
5. **Apostolou P, Fostira F. 2013.** Hereditary Breast Cancer: The Era of New Susceptibility Genes. BioMed Research International 2013:747318 DOI: 10.1155/2013/747318.
6. **Barrois M, Bièche I, Mazoyer S, Champème MH, Bressac-de Paillerets B, Lidereau R. 2004.** Real-time PCR-based gene dosage assay for detecting BRCA1 rearrangements in breast-ovarian cancer families. Clinical Genetics 65(2):131-6.
7. **Bartkova J, Tommiska J, Oplustilova L, Aaltonen K, Tamminen A, Heikkinen T, Mistrik M, Aittomäki K, Blomqvist C, Heikkilä P, Lukas J, Nevanlinna H, Bartek J. 2008.** Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. Molecular Oncology 2(4):296-316 DOI: 10.1016/j.molonc.2008.09.007.
8. **Beck TF, Mullikin JC, NISC Comparative Sequencing Program, Biesecker LG. 2016.** Systematic evaluation of Sanger validation of next-generation sequencing variants. Clinical Chemistry 62(4):647–654 DOI 10.1373/clinchem.2015.249623.
9. **Belancio VP, Roy-Engel AM, Deininger PL. 2010.** All y'all need to know 'bout retroelements in cancer. Semin Cancer Biol. 20(4):200–210 DOI: 10.1016/j.semcancer.2010.06.001.
10. **Bellido F, Pineda M, Aiza G, Valdés-Mas R, Navarro M, Puente DA, Pons T, González S, Iglesias S, Darder E, Piñol V, Soto JL, Valencia A, Blanco I, Urioste M, Brunet J, Lázaro C, Capellá G, Puente XS, Valle L. 2016.** POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genetics in Medicine 18(4):325-32 DOI: 10.1038/gim.2015.75.
11. **Bosdet IE, Docking TR, Butterfield YS, Mungall AJ, Zeng T, Coope RJ, Yorida E, Chow K, Bala M, Young SS, Hirst M, Birol I, Moore RA, Jones SJ, Marra MA, Holt**

R, Karsan A. 2013. A clinically validated diagnostic second-generation sequencing assay for detection of hereditary BRCA1 and BRCA2 mutations. Journal of Molecular Diagnostics 15(6):796-809. DOI: 10.1016/j.jmoldx.2013.07.004.

12. **Castéra L, Krieger S, Rousselin A, Legros A, Baumann J-J, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, Baert-Desurmont S, Tinat J, Bera O, Dugast C, Berthet P, Polycarpe F, Layet V, Hardouin A, Frébourg T, Vaur D. 2014.** Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. European Journal of Human Genetics 22: 1305–1313.
13. **Chong HK, Wang T, Lu H-M, Seidler S, Lu H, Keiles S, Chao EC, Stuenkel AJ, Li X, Elliott AM. 2014.** The Validation and Clinical Implementation of BRCAplus: A Comprehensive High-Risk Breast Cancer Diagnostic Assay. *PLoS ONE* 9(5): e97408 DOI:10.1371/journal.pone.0097408.
14. **Clopper CJ, Pearson ES. 1934.** The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika 26 (4): 404-413 DOI:10.1093/biomet/26.4.404.
15. **Cragun D, Radford C, Dolinsky JS, Caldwell M, Chao E, Pal T. 2014.** Panel-based testing for inherited colorectal cancer: a descriptive study of clinical testing performed by a US laboratory. Clinical Genetics 86: 510–520 DOI: 10.1111/cge.12359.
16. **Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegele C, Young EL, Robinot N, Forey N, Durand G, Vallée MP, Tao K, Roane TC, Williams GJ, Hopper JL, Southey MC, Andrulis IL, John EM, Goldgar DE, Lesueur F, Tavtigian SV. 2014.** Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. Breast Cancer Research 16(3):R58 DOI:10.1186/bcr3669.
17. **De Brakeleer S, De Grève J, Lissens W, Teugels E. 2013.** Systematic detection of pathogenic alu element insertions in NGS-based diagnostic screens: the BRCA1/BRCA2 example. Human Mutation 34(5):785-91 DOI: 10.1002/humu.22297.
18. **De Brakeleer S, De Grève J, Loris R, Janin N, Lissens W, Sermijn E, Teugels E. 2010.** Cancer Predisposing Missense and Protein Truncating BARD1 Mutations in Non-BRCA1 or BRCA2 Breast Cancer Families. Human Mutation 31(3):E1175-85 DOI:10.1002/humu.21200.
19. **Domchek SM, Friebel TM, Singer CF, Evans DG, Lynch HT, Isaacs C, Garber JE, Neuhausen SL, Matloff E, Eeles R, Pichert G, Van t'veer L, Tung N, Weitzel JN, Couch FJ, Rubinstein WS, Ganz PA, Daly MB, Olopade OI, Tomlinson G, Schildkraut J, Blum JL, Rebbeck TR. 2010.** Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. JAMA 304:967-975 DOI: 10.1001/jama.2010.1237.
20. **Eng C. PTEN Hamartoma Tumor Syndrome (PHTS) 2001 Nov 29 [Updated 2014 Jan 23].** In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2014. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1488/.
21. **Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DG, Chenevix-Trench G, Rahman N, Robson M, Domchek SM, Foulkes WD. 2015.** Gene-panel

sequencing and the prediction of breast-cancer risk. New England Journal of Medicine 372(23):2243–57 DOI: 10.1056/NEJMsr1501341.

22. **Frantzen C, Links TP, Giles RH. Von Hippel-Lindau Disease. 2000 May 17 [Updated 2012 Jun 21].** In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2014. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1463/.
23. **Garrison E, Marth G. 2012.** Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN].
24. **Giusti F, Marini F, Brandi ML. Multiple Endocrine Neoplasia Type 1. 2005 Aug 31 [Updated 2012 Sep 6].** In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2014. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1538/.
25. **Helgason H, Rafnar T, Olafsdottir HS, Jonasson JG, Sigurdsson A, Stacey SN, Jonasdottir A, Tryggvadottir L, Alexiusdottir K, Haraldsson A, le Roux L, Gudmundsson J, Johannsdottir H, Oddsson A, Gylfason A, Magnusson OT, Masson G, Jonsson T, Skuladottir H, Gudbjartsson DF, Thorsteinsdottir U, Sulem P, Stefansson K. 2015.** Loss-of-function variants in ATM confer risk of gastric cancer. Nature Genetics 47(8):906-10 DOI: 10.1038/ng.3342.
26. **Hirotsu Y, Nakagomi H, Sakamoto I, Amemiya K, Oyama T, Mochizuki H, Omata M. 2015.** Multigene panel analysis identified germline mutations of DNA repair genes in breast and ovarian cancer. Molecular Genetics and Genomic Medicine 3(5):459-66 DOI:10.1002/mgg3.157.
27. **Hogervorst FB, Nederlof PM, Gille JJ, McElgunn CJ, Grippeling M, Pruntel R, Regnerus R, van Welsem T, van Spaendonk R, Menko FH, Kluijt I, Dommering C, Verhoef S, Schouten JP, van't Veer LJ, Pals G. 2003.** Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. Cancer Research 63: 1449–1453.
28. **Jasperson KW, Burt RW. APC-Associated Polyposis Conditions. 1998 Dec 18 [Updated 2014 Mar 27].** In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2015. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1345/
29. **Judkins T, Leclair B, Bowles K, Gutin N, Trost J, McCulloch J, Bhatnagar S, Murray A, Craft J, Wardell B, Bastian M, Mitchell J, Chen J, Tran T, Williams D, Potter J, Jammulapati S, Perry M, Morris B, Roa B, Timms K. 2015.** Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. BMC Cancer 15:215 DOI:10.1186/s12885-015-1224-y.
30. **Kang HP, Maguire JR, Chu CS, Haque IS, Lai H, Mar-Heyming R, Ready K, Vysotskaia VS, Evans EA. 2016.** Design and validation of a next generation sequencing assay for hereditary BRCA1 and BRCA2 mutation testing. PeerJ 4:e2162 DOI:10.7717/peerj.2162.
31. **Kaurah P, Huntsman DG. Hereditary Diffuse Gastric Cancer. 2002 Nov 4 [Updated 2014 Jul 31].** In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2015. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1139/.

32. **Kennemer M, Pa N, Zeman M, Powers M, Ouyang K. 2016.** Detection of novel Alu insertions by next-generation sequencing of hereditary cancer genes. Poster presented at the 2016 American College of Medical Genetics Annual Clinical Genetics Meeting. Retrieved from https://marketing.invitae.com/acton/attachment/7098/f-03e5/1/-/-/-/-/Invitae_ACMG-2016_Detection-Novel-Alu-Insertions-by-NGS.pdf.
33. **Kirmani S, Young WF. Hereditary Paraganglioma-Pheochromocytoma Syndromes. 2008 May 21 [Updated 2014 Nov 6].** In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2015. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1548/
34. **Klonowska K, Ratajska M, Czubak K, Kuzniacka A, Brozek I, Koczkowska M, Sniadecki M, Debniak J, Wydra D, Balut M, Stukan M, Zmienko A, Nowakowska B, Irminger-Finger I, Limon J, Kozlowski P. 2015.** Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example. Scientific Reports 5:10424 DOI: 10.1038/srep10424.
35. **Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D. 2008.** Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nature Genetics 40(10):1253-60 DOI: 10.1038/ng.237.
36. **Kurian AW, Ford JM. 2015.** Multigene Panel Testing in Oncology Practice: How Should We Respond? JAMA Oncology 1(3):277-278 DOI:10.1001/jamaoncol.2015.28.
37. **Kurian AW, Hare EE, Mills MA, Kingham KE, McPherson L, Whittemore AS, McGuire V, Ladabaum U, Kobayashi Y, Lincoln SE, Cargill M, Ford JM. 2014.** Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. Journal of Clinical Oncology 32: 2001–2009 DOI: 10.1200/JCO.2013.53.6607.
38. **LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, Chen E, Gau CL, Palmaer E, Shoaepour K, Shah D, Speare V, Gandomi S, Chao E. 2014.** Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients Genetics in Medicine 6(11):830-7 DOI: 10.1038/gim.2014.40.
39. **Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014.** ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Research 42 (Database issue): D980-5 DOI: 10.1093/nar/gkt1113.
40. **Larsen Haidle J, Howe JR. Juvenile Polyposis Syndrome. 2003** May 13 [Updated 2014 May 22]. In: Pagon RA, Adam MP, Amemiya A, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2015.Available from: http://www.ncbi.nlm.nih.gov/books/NBK1469/.
41. **Li H. 2013.** Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
42. **Li J, Meeks H, Feng BJ, Healey S, Thorne H, Makunin I, Ellis J; kConFab Investigators, Campbell I, Southey M, Mitchell G, Clouston D, Kirk J, Goldgar D1, Chenevix-Trench G. 2016.** Targeted massively parallel sequencing of a panel of putative breast cancer susceptibility genes in a large cohort of multiple-case breast and ovarian cancer families. Journal of Medical Genetics 53(1):34-42 DOI: 10.1136/jmedgenet-2015-103452.

43. **Lincoln SE, Kobayashi Y, Anderson MJ, Yang S, Desmond AJ, Mills MA, Nilsen GB, Jacobs KB, Monzon FA, Kurian AW, Ford JM, Ellisen LW. 2015.** A Systematic Comparison of Traditional and Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Genes in More Than 1000 Patients. Journal of Molecular Diagnostics 17(5): 533–544 DOI: 10.1016/j.jmoldx.2015.04.009.
44. **Lynce F, Isaacs C. 2016.** How Far Do We Go With Genetic Evaluation? Gene, Panel, and Tumor Testing. American Society of Clinical Oncology Educational Book 35:e72-8 DOI: 10.14694/EDBK_160391.
45. **Maxwell KN, Wubbenhorst B, D'Andrea K. 2015.** Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. Genetics in Medicine 11:176 DOI: 10.1038/gim.2014.176.
46. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010.** The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20(9):1297-303 DOI: 10.1101/gr.107524.110.
47. **Minion LE, Dolinsky JS, Chase DM, Dunlop CL, Chao EC, Monk BJ. 2015.** Hereditary predisposition to ovarian cancer, looking beyond BRCA1/BRCA2 Gynecologic Oncology 137(1):86-92 DOI: 10.1016/j.ygyno.2015.01.537.
48. **Moline J, Eng C. Multiple Endocrine Neoplasia Type 2. 1999 Sep 27 [Updated 2013 Jan 10].** In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2014.Available from: http://www.ncbi.nlm.nih.gov/books/NBK1257/.
49. **Mu W, Lu HM, Chen J, Li S, Elliott AM. 2016.** Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. Journal of Molecular Diagnostics S1525-1578(16)30143-X DOI:10.1016/j.jmoldx.2016.07.006.
50. **National Cancer Institute. 2016.** PDQ Genetics of Skin Cancer. Bethesda, MD: National Cancer Institute. Date last modified 02/16/2016. Available at: http://www.cancer.gov/types/skin/hp/skin-genetics-pdq. Accessed 05/09/2016.
51. **National Comprehensive Cancer Network. 2016.** Gastric Cancer. Version 3.2016. Available at https://www.nccn.org/professionals/physician_gls/pdf/gastric.pdf (accessed August 30, 2016).
52. **National Comprehensive Cancer Network. 2016.** Genetic/familial high risk assessment: breast and ovarian. Version 2.2016. Available at http://www.nccn.org/ professionals/physician_gls/ pdf/genetics_screening.pdf (accessed August 30, 2016).
53. **National Comprehensive Cancer Network. 2016.** Genetic/familial High Risk Assessment: Colorectal. Version 1.2016. Available at https://www.nccn.org/professionals/physician_gls/pdf/genetics_colon.pdf (accessed August 30, 2016).
54. **National Comprehensive Cancer Network. 2016.** Neuroendocrine tumors. Version 2.2016. Available at https://www.nccn.org/professionals/physician_gls/pdf/neuroendocrine.pdf (accessed August 30, 2016).
55. **Norton JA, Ham CM, Van Dam J, Jeffrey RB, Longacre TA, Huntsman DG, Chun N, Kurian AW, Ford JM. 2007.** CDH1 truncating mutations in the E-cadherin gene: an

indication for total gastrectomy to treat hereditary diffuse gastric cancer. Annals of Surgery 245(6):873-9 DOI: 10.1097/01.sla.0000254370.29893.e4.

56. **Ollier M, Radosevic-Robin N, Kwiatkowski F, Ponelle F, Viala S, Privat M, Uhrhammer N, Bernard-Gallon D, Penault-Llorca F, Bignon YJ, Bidet Y. 2015.** DNA repair genes implicated in triple negative familial non-BRCA1/2 breast cancer predisposition. American Journal of Cancer Research 15;5(7):2113-26.
57. **Peixoto A, Santos C, Pinto P, Pinheiro M, Rocha P, Pinto C, Bizarro S, Veiga I, Principe AS, Maia S, Castro F, Couto R, Gouveia A, Teixeira MR. 2014.** The role of targeted BRCA1/BRCA2 mutation analysis in hereditary breast/ovarian cancer families of Portuguese ancestry. Clinical Genetics 88(1):41-8 DOI: 10.1111/cge.12441.
58. **Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, Wood NW, Hambleton S, Burns SO, Thrasher AJ, Kumararatne D, Doffinger R, Nejentsev S. 2012.** A Robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics 28(21):2747-54 DOI: 10.1093/bioinformatics/bts526.
59. **Potrony M, Badenas C, Aguilera P, Puig-Butille JA, Carrera C, Malvehy J, Puig S. 2015.** Update in genetic susceptibility in melanoma. Annals of Translational Medicine 3(15):210 DOI: 10.3978/j.issn.2305-5839.2015.08.11.
60. **Rafnar T, Gudbjartsson DF, Sulem P, Jonasdottir A, Sigurdsson A, Jonasdottir A, Besenbacher S, Lundin P, Stacey SN, Gudmundsson J, Magnusson OT, le Roux L, Orlygsdottir G, Helgadottir HT, Johannsdottir H, Gylfason A, Tryggvadottir L, Jonasson JG, de Juan A, Ortega E, Ramon-Cajal JM, García-Prats MD, Mayordomo C, Panadero A, Rivera F, Aben KK, van Altena AM, Massuger LF, Aavikko M, Kujala PM, Staff S, Aaltonen LA, Olafsdottir K, Bjornsson J, Kong A, Salvarsdottir A, Saemundsson H, Olafsson K, Benediktsdottir KR, Gulcher J, Masson G, Kiemeney LA, Mayordomo JI, Thorsteinsdottir U, Stefansson K. 2011.** Mutations in BRIP1 confer high risk of ovarian cancer. Nature Genetics 43(11):1104-7 DOI: 10.1038/ng.955.
61. **Rehm HL. 2013.** Disease-targeted sequencing: a cornerstone in the clinic. Nature Reviews Genetics 14(4):295-300 DOI: 10.1038/nrg3463.
62. **Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E. 2013.** ACMG clinical laboratory standards for next-generation sequencing. Genetics in Medicine 15, 733–747 DOI: 10.1038/gim.2013.92.
63. **Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015.** Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine 17: 405–423 DOI: 10.1038/gim.2015.30.
64. **Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, Gallinger S, Schwartz AG, Syngal S, Cote ML, Axilbund J, Schulick R, Ali SZ, Eshleman JR, Velculescu VE, Goggins M, Vogelstein B, Papadopoulos N, Hruban RH, Kinzler KW, Klein AP. 2012.** ATM mutations in hereditary pancreatic cancer patients. Cancer Discovery 2(1):41-6 DOI: 10.1158/2159-8290.CD-11-0194.

65. **Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011.** Integrative Genomics Viewer. Nature Biotechnology 29(1):24-6 DOI: 10.1038/nbt.1754.
66. **Sanger F, Nicklen S, Coulson AR. 1977.** DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74: 5463–5467.
67. **Schenkel LC, Kerkhof J, Stuart A, Reilly J, Eng B, Woodside C, Levstik A, Howlett CJ, Rupar AC, Knoll JH, Ainsworth P, Waye JS, Sadikovic B. 2016.** Clinical Next-Generation Sequencing Pipeline Outperforms a Combined Approach Using Sanger Sequencing and Multiplex Ligation-Dependent Probe Amplification in Targeted Gene Panel Analysis. Journal Of Molecular Diagnostics 18(5):657-67 DOI: 10.1016/j.jmoldx.2016.04.002.
68. **Schneider K, Zelley K, Nichols KE, et al. Li-Fraumeni Syndrome. 1999 Jan 19 [Updated 2013 Apr 11].** In: Pagon RA, Adam MP, Amemiya A, et al., editors. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2015. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1311/.
69. **Seemanová E, Jarolim P, Seeman P, Varon R, Digweed M, Swift M, Sperling K. 2007.** Cancer risk of heterozygotes with the NBN founder mutation. Journal of the National Cancer Institute 99(24):1875-80 DOI: 10.1093/jnci/djm251.
70. **Soura E, Eliades PJ, Shannon K, Stratigos AJ, Tsao H. 2016.** Hereditary melanoma: Update on syndromes and management. Genetics of familial atypical multiple mole melanoma syndrome. Journal of the American Academy of Dermatology 74:395-407 DOI: 10.1016/j.jaad.2015.08.038.
71. **Strom CM, Rivera S, Elzinga C, Angeloni T, Rosenthal SH, Goos-Root D, Siaw M, Platt J, Braastadt C, Cheng L, Ross D, Sun W. 2015.** Development and Validation of a Next-Generation Sequencing Assay for BRCA1 and BRCA2 Variants for the Clinical Laboratory. *PLoS One* 10:e0136419 DOI: 10.1371/journal.pone.0136419.
72. **Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, Coe BP, Baker C, Nordenfelt S, Bamshad M, Jorde LB, Posukh OL, Sahakyan H, Watkins WS, Yepiskoposyan L, Abdullah MS, Bravi CM, Capelli C, Hervig T, Wee JT, Tyler-Smith C, van Driem G, Romero IG, Jha AR, Karachanak-Yankova S, Toncheva D, Comas D, Henn B, Kivisild T, Ruiz-Linares A, Sajantila A, Metspalu E, Parik J, Villems R, Starikovskaya EB, Ayodo G, Beall CM, Di Rienzo A, Hammer MF, Khusainova R, Khusnutdinova E, Klitz W, Winkler C, Labuda D, Metspalu M, Tishkoff SA, Dryomov S, Sukernik R, Patterson N, Reich D, Eichler EE. 2015.** Global diversity, population stratification, and selection of human copy-number variation. Science 349(6253):aab3761 DOI: 10.1126/science.aab3761.
73. **Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013.** Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14(2):178-92 DOI: 10.1093/bib/bbs017.
74. **Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, Timms K, Garber JE, Herold C, Ellisen L, Krejdovsky J, DeLeonardis K, Sedgwick K, Soltis K, Roa B, Wenstrup RJ, Hartman AR. 2015.** Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next- generation sequencing with a 25-gene panel. Cancer 121:25-33 DOI: 10.1002/cncr.29010.

75. **Tutlewska K, Lubinski J, Kurzawski G. 2013.** Germline deletions in the EPCAM gene as a cause of Lynch syndrome – literature review. Hereditary Cancer in Clinical Practice 11(1):9 DOI: 10.1186/1897-4287-11-9.
76. **Zhang G, Zeng Y, Liu Z, Wei W. 2013.** Significant association between Nijmegen breakage syndrome 1 657del5 polymorphism and breast cancer risk. Tumor Biology 34(5):2753-7 DOI: 10.1007/s13277-013-0830-z.

760
761 **Table 1:** List of 36 genes included in the Inherited Cancer Screen panel.
762

Gene	Transcript:Exon Sequenced	SNV/Indel Reportable ROI, bp	Variants Reported
APC	NM_000038: 2-16	9433	SNVs, indels, CNVs
ATM	NM_000051: 2-63	11853	SNVs, indels, CNVs
BARD1	NM_000465: 1-11	2776	SNVs, indels, CNVs
BMPR1A	NM_004329: 3-13	2046	SNVs, indels, CNVs
BRCA1	NM_007294: 2-23	7351	SNVs, indels, CNVs
BRCA2	NM_000059: 2-27	11652	SNVs, indels, CNVs
BRIP1	NM_032043: 2-20	4556	SNVs, indels, CNVs
CDH1	NM_004360: 1-16	3350	SNVs, indels, CNVs
CDK4	NM_000075: 2-8	1229	SNVs, indels, CNVs
CDKN2A	NM_000077: 1-3	1343	SNVs, indels, CNVs
CHEK2	NM_007194: 2-15	2199	SNVs, indels, CNVs
EPCAM	NM_002354: 9		CNVs
GREM1	NM_013372: upstream duplications		CNVs
MEN1	NM_000244: 2-10	2306	SNVs, indels, CNVs
MLH1	NM_000249: 1-19	3295	SNVs, indels, CNVs
MRE11A	NM_005591: 2-20	2897	SNVs, indels, CNVs
MSH2	NM_00025: 1-16	3692	SNVs, indels, CNVs
MSH6	NM_000179: 1-10	4566	SNVs, indels, CNVs
MUTYH	NM_001048171: 1-16	2321	SNVs, indels, CNVs
NBN	NM_002485: 1-16	2905	SNVs, indels, CNVs
PALB2	NM_024675: 1-13	4090	SNVs, indels, CNVs
PMS2	NM_000535: 1-10	1649	SNVs, indels, CNVs
POLD1	NM_001256849: 2-27	4435	SNVs, indels, CNVs
POLE	NM_006231: 1-49	8823	SNVs, indels, CNVs
PTEN	NM_000314: 1-9	1866	SNVs, indels, CNVs
RAD50	NM_005732: 1-25	4944	SNVs, indels, CNVs
RAD51C	NM_058216: 1-9	1509	SNVs, indels, CNVs
RAD51D	NM_002878: 1-10	1862	SNVs, indels, CNVs
RET	NM_020975: 2-20	4167	SNVs, indels, CNVs
SDHA	NM_004168: 1-15	2606	SNVs, indels, CNVs
SDHB	NM_003000: 1-8	1188	SNVs, indels, CNVs
SDHC	NM_003001: 1-6	864	SNVs, indels, CNVs
SMAD4	NM_005359: 2-12	2148	SNVs, indels, CNVs
STK11	NM_000455: 1-9	1717	SNVs, indels, CNVs
TP53	NM_000546: 2-11	1818	SNVs, indels, CNVs
VHL	NM_000551: 1-3	789	SNVs, indels, CNVs

763
764

765 **Table 2:** Source of samples and reference data used in validation.
766
767

Measures	Variant Type	Test Samples	Reference Data
Accuracy Sensitivity Specificity	SNV Indel	101 Coriell cell line samples	1000 Genomes project exomes
		2 Coriell cell lines with specific mutations	Coriell data
		2 NIBSC samples	NIBSC reference data
		82 mutation-positive patient samples	Orthogonal confirmation by Sanger
Accuracy Sensitivity Specificity	CNV	5 NIBSC samples	NIBSC reference data
		44 CNV-positive patient samples	Orthogonal confirmation by MLPA
Intra-run reproducibility	SNV Indel CNV	8 Genome-in-a-Bottle (GiaB) cell line samples	
		13 patient samples	
Inter-run reproducibility	SNV Indel CNV	8 GiaB cell line samples	
		84 patient samples	

768
769
770

771 **Table 3:** Variants in validation study.
772
773

Variant Type	Deletion/Insertion Size	Number of Variants	
		Reference Data	Orthogonal Confirmation
SNV		5182	
Indel	Indels ≤ 10 bp		57
	Indels >10 bp		19
Alu insertion			7
CNV	Single-exon deletions or duplications	3	9
	Multiple exon deletions or duplications	2	36

774
775
776
777
778
779
780

Table 4: Performance of Counsyl Inherited Cancer Screen for SNVs and indels.

	Counsyl test	1000 Genomes Project data		Results (95% confidence interval)
		Variant present	Variant not present	
SNV & Indel	Variant detected	5182 true positives	0 false positives	100% accuracy (99.991- 100%) 100.0% sensitivity (99.93-100%) 100% specificity (99.990- 100%) 0% FDR (0-0.0007%)
	Variant not detected	0 false negatives	37743 true negatives	

Validation metrics were defined as: Accuracy = (TP + TN) /(TP + FP + TN + FN); Sensitivity = TP / (TP + FN); Specificity = TN / (TN + FP); FDR = FP / (TP + FP). For true negative calculations, all polymorphic positions (positions at which we observed non-reference bases in any sample) across all samples were considered.

Table 5: Performance of Counsyl Inherited Cancer Screen for indels and CNVs.

	Counsyl test	Sanger or MLPA reference data		Results (95% confidence interval)
		Variant present	Variant not present	
Indel	Variant detected	76 true positives	0 false positives	100% accuracy (99.88-100%) 100% sensitivity (95-100%) 100% specificity (99.88-100%) 0% FDR (0-5%)
	Variant not detected	0 false negatives	3040 true negatives	
CNV	Variant detected	50 true positives	0 false positives	100% accuracy (99.5-100%) 100% sensitivity (93-100%) 100% specificity (99.5-100%) 0% FDR (0-7.1%)
	Variant detected	0 false negatives	685 true negatives	

Validation metrics were defined as: Accuracy = (TP + TN) /(TP + FP + TN + FN); Sensitivity = TP / (TP + FN); Specificity = TN / (TN + FP); FDR = FP / (TP + FP). For indels, true negatives defined as the number of homozygous reference calls made at sites for which an alternative variant was observed in at least one sample in the cohort. For CNVs, true negatives defined as the number of genes assigned the reference copy number in the CNV validation cohort, and the summation included only genes for which a known CNV positive was tested (N=15 genes with a CNV positive).

808
809 **Table 6:** List of Alu insertions confirmed in validation.
810

Sample ID	Gene	Variant Description
Counsyl 24	*ATM*	Intron 54-55, NM_000051.3: c.8010+13_8010+14insAlu
Counsyl 25	*ATM*	Intron 54-55, NM_000051.3: c.8010+13_8010+14insAlu
Counsyl 26	*ATM*	Intron 54-55, NM_000051.3: c.8010+13_8010+14insAlu
Counsyl 27	*BRCA2*	Exon 3, NM_000059.3: c.156_157insAlu
Counsyl 28	*BRCA2*	Exon 3, NM_000059.3: c.156_157insAlu
Counsyl 85	*BRCA2*	Exon 25, NM_000059.3:c.930_931insAlu
Counsyl 84	*MSH6*	Intron 2-3, NM_000179: c.458-19_458-18insAlu

811