

**Supplemental information**

**Population screening shows risk of inherited cancer  
and familial hypercholesterolemia in Oregon**

**Timothy D. O'Brien, Amiee B. Potter, Catherine C. Driscoll, Gregory Goh, John H. Letaw, Sarah McCabe, Jane Thanner, Arpita Kulkarni, Rossana Wong, Samuel Medica, Tiana Week, Jacob Buitrago, Aaron Larson, Katie Johnson Camacho, Kim Brown, Rachel Crist, Casey Conrad, Sara Evans-Dutson, Ryan Lutz, Asia Mitchell, Pavana Anur, Vanessa Serrato, Autumn Shafer, Lisa K. Marriott, K.J. Hamman, Amelia Mulford, Wojciech Wiszniewski, Jone E. Sampson, Andrew Adey, Brian J. O'Roak, Christina A. Harrington, Jackilen Shannon, Paul T. Spellman, and C. Sue Richards**

## **Supplemental Note 1: Specific information regarding return of testing results to participants during enrollment and consent.**

1. **Negative:** You will receive these within your secure Healthy Oregon Project account under the “Results” tab. You will also receive an email notification alerting you to check your HOP app as a new result has been uploaded.
2. **Positive:** the majority of participants will not have a positive result (99% will have a negative result). If you have a positive result, you will be contacted by an OHSU genetic counselor by phone. The Genetic Counselors will guide you through the results on the phone and give you the recommended medical guidelines for this result that recommends certain preventive measures that you can choose to take if you wish to decrease your risk of developing cancer in the future. These are publicly established guidelines that are followed by the medical community. The genetic counselor may assist you in helping to find resources for getting the recommended preventive measures if you do not have the financial ability to receive with your insurance or if you do not have insurance. If they cannot reach you by phone or email to coordinate a consultation appointment, you will not receive your results. You may call the genetic counselor at any time to discuss the results if you choose. Positive test results will be uploaded into an OHSU medical record that will be created for you if you do not already have one, this is a requirement for medical testing (that a medical record be created for you). Negative results will not be uploaded into a medical record. If you have a positive result, the clinical laboratory, genetic counselors and the study staff may review your OHSU medical records to manage your result for this study, no information outside of this study will be collected. If you have a significant family history, HOP still recommends consulting with your provider for additional genetic screening that is not covered by HOP.

You will receive results from our current genetic screening panel that is reviewed with you when you activate your saliva sample in the HOP app. See the “Resources” tab in the app to find our current screening panel. The genetic abnormalities that we screen for may change over time. In some instances, we may be able to re-screen your sample if we add new genetic abnormalities to the panel. You are given the choice within the app when you activate your sample if you would like us to send you updated results with any new additions if we are able to re-screen your sample.

Medical guidelines may also change over time as our understanding of genetics evolves and affects what we report back to you, as we only return a result that has medical guidelines that could impact your clinical care. The result you are given will be based on the current medical guidelines at the time a report is issued as they may have changed from the date you initially submitted your sample. Participants that receive a positive result are responsible for contacting their healthcare providers periodically to obtain to the most up-to-date recommendations for their healthcare.



## Supplemental Note 2: Negative HOP report



**Test Result: Negative. No clinically significant variant detected.**

### Interpretation Summary:

This result does not significantly change your personal risk of developing cancer or high cholesterol and heart disease. We did not detect any inherited genetic changes known to increase the risk for cancer or high cholesterol and heart disease in any of the genes covered by this test. Please note, this test does not screen for all genetic changes that cause cancer or high cholesterol and heart disease, only those listed below (in the “Detail for you and your physician” section). Keep in mind that this test does not account for other factors that may increase a person’s risk to develop cancer or high cholesterol and heart disease, such as smoking, poor diet, or obesity. This negative result does not mean you will never develop cancer or heart disease, but it is unlikely that it would be caused by the inherited genetic changes covered by this test.

### Recommendations:

Family history, previous medical history, or lifestyle choices have strong influences on a person’s risk to develop cancer or heart disease. If you have additional concerns or questions about your risk for cancer or high cholesterol and heart disease, we recommend that you talk to your primary care provider about recommended cancer and cardiac screening based on your personal medical history and/or family history.

### Detail for you and your physician:

This screening test did not detect any clinically significant (pathogenic) gene variants.

Test Background: A saliva sample was sent to our laboratory for screening analysis by Next-Generation Sequencing (NGS). This NGS panel test is designed to detect mutations in the coding regions of 32 genes associated with Inherited Cancer Syndromes and Familial Hypercholesterolemia.



Method: DNA was extracted by the OHSU Integrated Genomics Lab (located at 3181 SW Sam Jackson Park Road RJH 5330 Portland, OR 97239) using standard nucleic acid extraction methods. The coding regions and splice sites of the following 32 genes were sequenced using massively parallel sequencing (next-generation sequencing) on the Illumina NextSeq 500 or 550: *APC, ATM, BAP1, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A, CHEK2, LDLR, MEN1, MITF, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, POLD1, PTEN, RAD51C, RAD51D, RB1, RET, SMAD4, STK11, TP53, TSC1, TSC2*. Variants were classified using the 2015 ACMG Standards and Guidelines for Interpretation of Sequence Variants (Richards et al. 2015). Only pathogenic and likely pathogenic variants on autosomal dominant disorders are reported. Heterozygous carriers for *MUTYH* disorders are not reported. Heterozygous carriers for *NBN* disorders are not reported. Mosaic variants are not reported.

Gene Coverage Information: This test was validated for gene coverage sequence analysis of 32 genes associated with the HOP2 Inherited Cancer and Cardio Risk Panel with average coverage of 99.5% across the 32 genes at  $\geq 20X$  with the exception of the pseudogene regions. Please note that the presence of pathogenic variation in genes not analyzed or with incomplete coverage cannot be fully excluded.

Limitations: Unique individual patient history and demographics are not available to the laboratory to use in interpretation. Sequencing does not detect large deletions/duplications and polynucleotide repeats. Sequencing reads with low quality scores may affect variant calling. This test has limited ability to identify small insertions and deletions or mosaicism. This test does not provide complete coverage of all exons; noncoding regions may have limited information and limited ability to interpret. The assay does not detect variants located in regions of insufficient coverage, including introns and promoter regions; pseudogenes; where the reference genome is inaccurate or contains gaps and insertions; and of high GC content. Novel variants in introns that are greater than 10 base pairs from the intron-exon junction are not analyzed. In addition, genes not associated clinically with an inherited cancer or cardiovascular condition at the time this test was performed were not analyzed. Classification of variants in this report utilizes evidence available at the time of analysis and is subject to change over time with additional information.

#### References:

1. Richards et al., 2015; Genet Med. 17(5):405-23.
2. Exome Sequencing Project: <http://evs.gs.washington.edu/EVS/>
3. ExAC: <http://exac.broadinstitute.org/>
4. 1000 Genomes: <http://www.1000genomes.org/>
5. Eldering et al. (2003) Nucl. Acid Res. 31(23):e153:1-9.

NOTE: This database/product contains information obtained from the Online Mendelian Inheritance in Man® (OMIM®) database, which has been obtained through a license from the Johns Hopkins University, which owns the copyright thereto.

#### Disclaimer:

This test was developed and its performance characteristics determined by the OHSU Knight Diagnostic Laboratories. It has not been cleared or approved by the Food and Drug Administration. FDA approval is not required for the clinical use of the test, and therefore validation was done as required under the requirements of the Clinical Laboratory Improvement Act of 1988 (CLIA). The OHSU Knight Diagnostics Laboratories are fully licensed by the state of Oregon under CLIA and are accredited by the College of American Pathologists (CAP). Laboratory Director: Christopher Corless, M.D., Ph.D.

### Supplemental Note 3: Example positive report

#### Participant Results: Heterozygous Pathogenic Variant Detected.

**Interpretation:** This individual carries a pathogenic frameshift variant, c.5946delT (also known as 6174delT), p.Ser1982Argfs\*22 (NM\_000059.4), in *BRCA2* in the heterozygous state. These results should be interpreted in the context of clinical and laboratory findings.

**Recommendation:** Genetic counseling is recommended. Establishment of clinical referrals for appropriate surveillance and risk assessment are highly recommended for this individual. Testing is available for other family members.

#### Evidence for Variant Interpretation:

c.5946delT, p.Ser1982Argfs\*22 in exon 11 of the *BRCA2* gene (NM\_000059.4, hg19, chr13:32914438) is interpreted as pathogenic. This single base deletion c.5946delT, p.Ser1982Argfs\*22 (also reported as c.6174delT) is predicted to cause a frameshift in exon 11 of 27, resulting in the introduction of a premature stop codon in the 22<sup>nd</sup> position after the deletion in a gene and leading to an absent or truncated protein product where loss-of-function is a known mechanism of disease. This variant is a well-known pathogenic variant that is high frequency in Ashkenazi Jewish individuals, and is suggested to be a founder mutation in this population (Finkelman *et al.* 2012, Oddoux *et al.* 1996, Berman *et al.* 1996). Although this variant is a known pathogenic variant in individuals of Ashkenazi Jewish background, it has also been detected in non-Ashkenazi Jewish individuals (Berman *et al.* 1996). This variant was detected 78 times (61 of them in Ashkenazi Jewish) in a population database of presumed healthy adults (gnomAD). Additionally, more than 40 outside clinical and research testing laboratories have classified this variant as pathogenic, including an expert panel (ClinVar: 9325). In summary, this variant meets our criteria for pathogenic. This variant was confirmed by Sanger sequencing.

**Limitations:** This individual may carry *BRCA2* mutations that are not detected by this methodology. No other regions of this gene were analyzed by Sanger sequencing.

#### Disorder and Test Description:

This individual (Primary HOP ID) underwent inherited cancer predisposition and cardio risk screening as part of the Healthy Oregon Project (HOP) under sample ID (Secondary HOP ID). That testing detected this variant in the heterozygous state using a custom-built next-generation sequencing (NGS) assay in our laboratory (KDL). The coding regions and splice sites of the following 32 genes were sequenced using massively parallel sequencing (next-generation sequencing) on the Illumina NextSeq 500 or 550: *APC*, *ATM*, *BAP1*, *BMPR1A*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CDK4*, *CDKN2A*, *CHEK2*, *LDLR*, *MEN1*, *MITF*, *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *NBN*, *PALB2*, *PMS2*, *POLD1*, *PTEN*, *RAD51C*, *RAD51D*, *RB1*, *RET*, *SMAD4*, *STK11*, *TP53*, *TSC1*, *TSC2*. Variants were classified using the 2015 ACMG Standards and Guidelines for Interpretation of Sequence Variants (Richards *et al.* 2015). Only pathogenic and likely pathogenic variants on autosomal dominant disorders were reported. Heterozygous carriers for

*MUTYH* and *NBN* disorders were not reported. Mosaic variants were not reported. No other pathogenic variants were detected.

**Gene Coverage Information for the HOP2 Inherited Cancer and Cardio Panel:** This test was validated for gene coverage sequence analysis of 32 genes associated with the HOP2 Inherited Cancer and Cardio Panel with average coverage of 99.5% across the 32 genes at  $\geq 20X$  with the exception of the pseudogene regions. Please note that the presence of pathogenic variation in genes not analyzed or with incomplete coverage cannot be fully excluded.

**Limitations of the HOP2 Inherited Cancer and Cardio Panel:** Unique individual patient history and demographics are not available to the laboratory to use in interpretation. Sequencing does not detect large deletions/duplications and polynucleotide repeats. Sequencing reads with low quality scores may affect variant calling. This test has limited ability to identify small insertions and deletions or mosaicism. This test does not provide complete coverage of all exons; noncoding regions may have limited information and limited ability to interpret. The assay does not detect variants located in regions of insufficient coverage, including introns and promoter regions; pseudogenes; where the reference genome is inaccurate or contains gaps and insertions; and of high GC content. Novel variants in introns that are greater than 10 base pairs from the intron-exon junction are not analyzed. In addition, genes not associated clinically with an inherited cancer at the time this test was performed were not analyzed. Classification of variants in this report utilizes evidence available at the time of analysis and is subject to change over time with additional information.

### Confirmation of Positive Results:

Targeted mutation analysis was performed on a portion of exon 11 that is specific for this variant (c.5946delT (also known as 6174delT), p.Ser1982Argfs\*22) in the *BRCA2* gene. Pathogenic variants within *BRCA2* (OMIM: 600185) are responsible for Fanconi anemia, complementation group D1 (MIM: 605724); Wilms tumor (MIM: 194070); Breast cancer, male, susceptibility to (MIM: 114480); Breast-ovarian cancer, familial, 2 (MIM: 612555); Glioblastoma 3 (MIM: 613029); Medulloblastoma (MIM: 155255); Pancreatic cancer 2 (MIM: 613347) and Prostate cancer (MIM: 176807). DNA from this individual was used to amplify a portion of exon 11 in the *BRCA2* gene targeting 1 known variant. The PCR products from 1 exon were sequenced in both the forward and reverse directions.

This sequence analysis did identify the previously detected pathogenic variant c.5946delT (also known as 6174delT) (p.Ser1982Argfs\*22), in exon 11 of the *BRCA2* gene (based on the reference sequence GenBank accession number NM\_000059.3 with A of the first ATG being +1).

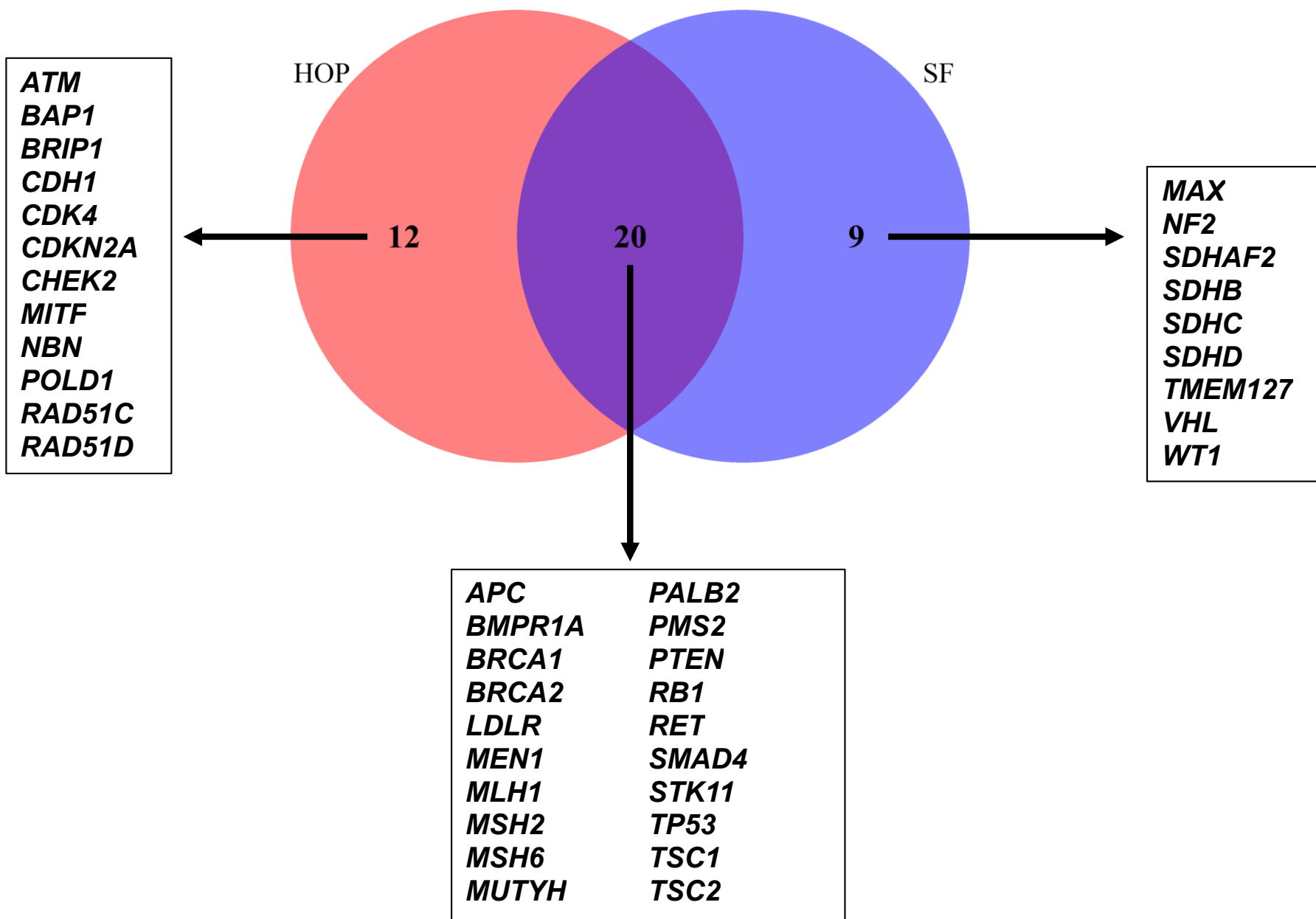
**Sanger Sequencing Test Methodology:** Genomic DNA from this individual was used for PCR amplification of a portion of *BRCA2* exon 11 that is specific for the variant, c.5946delT (also known as 6174delT), (p.Ser1982Argfs\*22). Both PCR products were sequenced in both the forward and reverse directions. The sensitivity is estimated at approximately 99% for the detection of nucleotide base changes, small deletions, and small insertions in the regions analyzed.

### References

1. PMID: 22430266. Finkelman *et al.* 2012. Comparative Study. J Clin Oncol. 2012 Apr 20;30(12):1321-8
2. PMID: 8841192. Oddoux *et al.* 1996. Nat Genet. 1996 Oct;14(2):188-90
3. PMID: 8758903. Berman *et al.* 1996. Cancer Res. 1996 Aug 1;56(15):3409-14
4. gnomAD: <https://gnomad.broadinstitute.org/>
5. ClinVar: 9325: <https://www.ncbi.nlm.nih.gov/clinvar/variation/9325/>
6. PMID: 25741868. Richards *et al.* 2015; Genet Med. 17(5):405-23.

**NOTE:** This database/product contains information obtained from the Online Mendelian Inheritance in Man® (OMIM®) database, which has been obtained through a license from the Johns Hopkins University, which owns the copyright thereto.

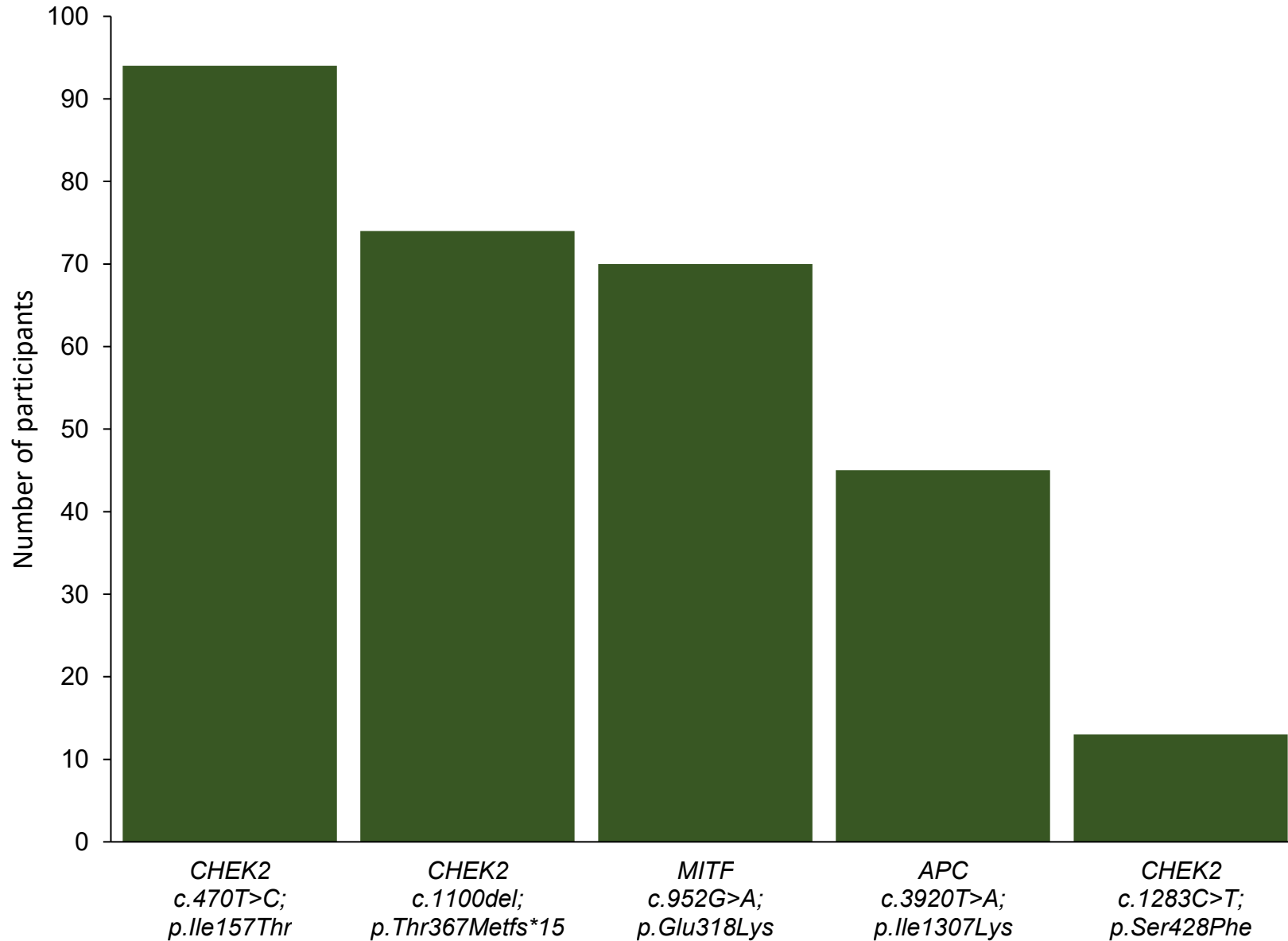
Evaluation of this individual was performed using mutation analysis. Possible diagnostic errors include sample mix-ups, erroneous paternity identification, and genotyping errors. Genotyping errors can result from trace contamination of PCR, from maternal contamination of fetal samples, from rare genetic variants which interfere with analysis, and from other sources. Families being studied should understand that rare diagnostic errors will occur for other reasons. Risk analysis based on one's DNA data could change if molecular data from other family members were available. If this report contains information on family members in addition to the proband, we ask that you maintain the confidentiality of that data. If prenatal diagnostic studies have been performed, we request that follow-up information about the pregnancy outcome be sent to us. Clinical interpretation has been performed by the Clinical Geneticist.



**Figure S1.** Venn Diagram showing overlap between Healthy Oregon Project (HOP) genes and genes on the ACMG Secondary Findings (SF) list. Boxes with genes show unique and shared genes.



### Recurrent P/LP variants identified in > 10 HOP participants



**Figure S2. Recurrent P/LP variants identified in HOP.** Barplot shows the specific recurrent pathogenic variant on the x-axis and the number of times we detected that variant on the y-axis. Table S3 contains the transcript for each variant. P:Pathogenic, LP:Likely Pathogenic

Clinical lab receives DNA plate from core lab

Janus Robot creates 'HOP-Frag' plate

Fragmentation Master mix

Incubation on thermal cycler

Janus Robot creates 'HOP-Ligation' plate

Ligation Master mix

Index Plate broadcasted to 96-well format

Ligation Incubation

Janus Robot performs 2 bead cleanups

Janus Robot creates 'TE PCR' plate

TE PCR Master mix

Incubation on thermal cycler

Pre-PCR

Janus Robot performs bead cleanup

Janus Robot creates 'UPCR' plate

UPCR Master mix

Incubation on thermal cycler

Janus Robot performs bead cleanup

Final Library Plate

QIAxcel library quantitation

Library Quantitation prep

Janus Robot performs Library Normalizations

Next-Generation Sequencing on Illumina NextSeq 500/550

Post-PCR

Figure S3. Overview of the library preparation steps performed by robotics. Illustrated are the steps a HOP sample goes through from a DNA extracted sample received to sequencing. The top panel illustrates the Pre-PCR steps, and the bottom panel shows Post-PCR steps. 4 plates are combined for the final sequencing step. UPCR: Universal PCR, TE PCR: Target Enrichment PCR

- Green – Manual steps
- Blue – Robotic steps
- Grey – Incubation steps
- Pink – Sequencing step

Table S1. Differences between our standard clinical diagnostic test and the HOP test

<b>Diagnostic Test</b>	<b>HOP Test</b>
Patient information	No patient information
Larger set of genes	Select set of genes chosen by clinical side
Report all P/LP/VUS variants	No reporting of VUSs
Report CNVs	No reporting of CNVs
High quality DNA source	Lower quality mouthwash DNA source
All results sent to medical provider	Negative results straight to participant and positive to medical record
Analyze all variants	Heavy filtering of variants before analysis
Consent by genetic counselor	Consent through smartphone app
All results in medical record	Only positive results in medical record
Return all results	Sometimes hold positive results
Deep dive analysis	Rely heavily on ClinVar

P: Pathogenic, LP: Likely Pathogenic, VUS: Variant of Uncertain Significance, CNV: Copy Number Variant

Table S2. Demographics of HOP Participants

	<b>HOP participants</b>
<b>Age</b>	
< 20	0.8%
20-29	8.9%
30-39	28.8%
40-49	29.0%
50-59	18.2%
60-69	8.8%
70-79	4.9%
≥ 80	0.7%
<b>Self-reported sex at birth</b>	
Female	74.2%
Male	23.4%
Undefined <sup>a</sup>	2.4%
<b>Race/Ethnicity</b>	
American Indian or Alaska Native	1.9%
Asian	4.8%
Black or African American	1.1%
Hispanic or Latino/a/x	1.3%
Middle Eastern/North African	0.1%
Native Hawaiian or Other Pacific Islander	0.5%
White	84.4%
Other	1.7%
Undefined <sup>a</sup>	3.8%
Don't Know	0.4%

<sup>a</sup>Blank or declined to answer

\*Percentages may not equal 100 due to rounding

Table S4. Results of RNA studies

Gene	Transcript	Variant	Result	Provisional Classification	Final Classification
<i>APC</i>	NM_000038.6	934-2A>G, p.?	Detected but complex	VUS	VUS
<i>ATM</i>	NM_000051.4	c.2464_2466+2delTTAGT, p.Leu822Ilefs*8	Detected but complex	Likely Pathogenic	Pathogenic
<i>ATM</i>	NM_000051.4	c.7088A>C, p.Lys2363Thr	Splicing not altered	VUS	Benign
<i>ATM</i>	NM_000051.4	c.5919-2A>G, p.?	Splicing altered	N/A	Likely Pathogenic
<i>BRCA2</i>	NM_000059.4	c.425G>A, p.Ser142Asn	Splicing altered	VUS/LP	Likely Pathogenic
<i>BRIP1</i>	NM_032043.3	c.205+1G>T, p.?	Detected but complex	Likely Pathogenic	Likely Pathogenic
<i>TSC2</i>	NM_000548.5	c.641_648+1del, p.?	Splicing altered	N/A	Pathogenic

VUS: variant of uncertain significance, LP: likely pathogenic, N/A: not applicable since RNA sample was requested prior to secondary DNA sample collection.

## Supplemental Methods

### Filtering scheme used for HOP variants

- 1.1. If variant is in input DENYLIST, remove.
- 1.2. If variant is not in our list of reportable genes (HGNC), remove.
- 1.3. If variant has a poor allele balance, remove.
  - 1.3.1. When average allele balance from a set of samples plus 3 standard deviations is below 35%, remove the variant.
  - 1.3.2. If raw allele balance is between 5% and 25%, remove.
  - 1.3.3. At least two samples must call the variant for it to be assessed.
- 1.4. If SnpEff term is “missense\_variant”, and it is not pathogenic in ClinVar, and not in HGMD\*, remove.
- 1.5. If the entry is a “splice variant”, but is not a canonical splice, and is not pathogenic in ClinVar, and is not in HGMD, remove.
- 1.6. If SnpEff term is “synonymous\_variant”, and it is not pathogenic in ClinVar, remove.
- 1.7. If entry is pathogenic, retain.
  - 1.7.1. ClinVar CLNSIG is “Pathogenic” or “Likely\_pathogenic”.
  - 1.7.2. ClinVar CLNSIGCONF contains the term “P(p)athogenic”.
  - 1.7.3. HGMD CLASS is “DM”.
- 1.8. If ClinVar CLNSIG is “Benign”, “Likely\_benign”, or “Benign/Likely\_benign”, remove.
- 1.9. If ClinVar CLNSIG is “Conflicting\_interpretations\_of\_pathogenicity”, remove.
- 1.10. If ClinVar CLNSIG is “Uncertain\_significance”, and we don’t care about given snpEff terms, remove.
  - 1.10.1. Terms we don’t care about: ("3\_prime\_UTR\_variant", "5\_prime\_UTR\_variant", "downstream\_gene\_variant", "intron\_variant", "intergenic\_region", "synonymous\_variant", "upstream\_gene\_variant", "non\_coding\_transcript\_exon\_variant")
- 1.11. If we don’t care about the SnpEff terms (see 1.10.1), and there is no ClinVar record, or CLNSIG is designated “not\_provided”, remove.
- 1.12. If there is a gnomAD AF entry, and it is 2% or greater, remove.
- 1.13. If there is no info (gnomAD, HGMD, ClinVar), and the QUAL score is below 100, remove.
- 1.14. If the local sample set AF is 2% or greater, and ClinVar CLNSIG is “not\_provided”, remove.
- 1.15. If the local sample set AF is less than 2%, and ClinVar CLNSIG is “not\_provided”, retain.
- 1.16. If the local sample set AF is less than 5%, retain.

\* HGMD access was terminated in October 2022, so after this date our filtering scheme no longer uses these data as input. However, all of the results reported here were obtained while we had a valid license.

## **Internal Sanger confirmation**

For samples that contained a pathogenic/likely pathogenic (P/LP) variant(s) but failed variant-specific QC metrics (low depth, imbalanced calls, etc.), an internal Sanger sequencing confirmation was performed on the original DNA extracted from the initial mouthwash sample. We went back to the original plate of DNA that had been normalized to 10ng/μL and extracted 1.5μL from that sample. This DNA was used to perform Sanger sequencing using standard approaches<sup>1</sup>. P/LP variants detected in a gene with known pseudogenes (e.g. *CHEK2*, *PMS2*) were first amplified by long-range PCR specific to the gene, followed by a second targeted PCR and Sanger sequencing if the variant within the pseudogene region could not be specifically targeted by a single set of primers. Upon confirmation by Sanger sequencing, a secondary sample was requested in the same manner for P/LP variants that had initially passed internal QC metrics.

## **Implementation of robotics**

For validating the robots, three HOP plates (288 samples) which had been previously run by the manual setup method, were re-run using a semi-automated protocol on the Janus G3 Workstation CJL8M01/E (Perkin Elmer). Each plate contained 92 mouthwash saliva samples and 4 No Template Controls (NTCs) (unused mouthwash). The mouthwash saliva collections were extracted at the IGL (the core laboratory) on the QIASymphony SP robot, quantitated and normalized to a 10 ng/μL DNA concentration (acceptable range at 5-45 ng/μL), and plated out in 96-well PCR plates at IGL (the core laboratory). All protocols on the Janus G3 Workstation were programmed in the WinPrep Software Ver. 5.4.404 and were run using the Janus Application Assistant Software Ver. 5.4.404, both from Perkin Elmer. Both preparation methods used the QIAseq Targeted DNA Panel (CDHS-15902Z-3798) library prep (Qiagen). The libraries were quantitated using the Agilent TapeStation system, pooled in equimolar proportions, and sequenced using massively parallel sequencing (next-generation sequencing (NGS)) on the Illumina NextSeq 500 or 550 MID output 300 cycle kit in a Paired End run (151+151+8+8, batch capacity 96 samples per run). The coding regions and splice sites of the genes and sample level metrics were assessed for each run. QIAxcel was later validated as an alternate approach for library quantitation.

## **Detailed specifications of the use of robotics in library preparation**

Library preparation for the NGS of HOP samples was a semi-automated process using QIAseq Targeted DNA Custom Panel chemistry programmed on the Janus G3 Workstation Robots (Perkin Elmer). One Pre-PCR robot and one Post-PCR robot were programmed to perform all of the liquid transfer steps with the exception of the preparation of the master mix solutions or index plates, sealing and unsealing the plates, centrifugation, and moving them to a thermal cycler. Thermal cycling steps were performed on external thermocyclers, not on the Janus robots. The Janus robots were specifically equipped with a 96-well pipette head that could transfer liquids from one plate to another in a single step, thus reducing preparation time.

Once the library preparation was completed, the NGS libraries were quantitated on the QIAxcel (Qiagen), and these values were used to program the Post-PCR Janus robot. The libraries from four 96-well plates were then normalized by the robot and pooled into one tube of a 4 nanomolar library pool.

The library preparation process was staggered (a new plate started while the previous plate was running in the Post-PCR space) such that 3-4 plates were processed in a week. 384 samples (4 full plates of libraries) were pooled together and loaded into one NGS run of an Illumina NextSeq 500/550 using a High Output Kit (300 cycles). NTCs were added for every run and assessed for contamination. A typical two-week schedule of the overall major steps that can be followed to sequence > 1,000 samples per month using this approach is outlined below. Specific details of the sub-steps in this process are presented in Figure S3.

### **Example laboratory schedule focused on robotics**

#### **•Day 1**

–Plate 1 Pre PCR

#### **•Day 2**

–Plate 1 Post PCR

–Plate 2 Pre PCR

#### **•Day 3**

–Plate 1 Quantitation QC

–Plate 2 Post PCR

–Plate 3 Pre PCR

#### **•Day 4**

–Plate 2 Quantitation QC

–Plate 3 Post PCR

–Plate 4 Pre PCR

#### **•Day 5**

–Plate 3 Quantitation QC

–Plate 4 Post PCR

#### **•Day 6**

–Plate 4 Quantitation QC



-Plates 1 to 4 Normalization on the Robot

-Plate 5 Pre PCR

**•Day 7**

-Load High Output NGS Plates 1-4 (384 samples, 4 plates together)

-Plate 5 Post PCR

-Plate 6 Pre PCR

**•Day 8**

-Plate 5 Quantitation QC

-Plate 6 Post PCR

-Plate 7 Pre PCR

**•Day 9**

-Plate 6 Quantitation QC

-Plate 7 Post PCR

-Plate 8 Pre PCR

**•Day 10**

-Plate 7 Quantitation QC

-Plate 8 Post PCR

**•Days 11 and 12**

-Plate 8 Quantitation QC

-Plates 5 to 8 Normalization on the Robot

-Load High Output NGS Plates 5-8 (384 samples, 4 plates together)

## References

1. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74, 5463-5467.