**SUPPLEMENTARY MATERIALS AND METHODS**

**Genome and exome sequencing data**

The 2,284 genome sequencing (GS) samples were subjected to PCR-free DNA preparation and sequenced at the Broad Institute (Cambridge, MA, USA) using an Illumina HiSeq X Ten device (150-bp paired-end reads; mean coverage, 30×). The 2,500 exome sequencing (ES) samples were sequenced at Nestlé Institute of Health Sciences (Lausanne, Switzerland) using the Agilent SureSelect Human All Exon V5+UTRs Kit according to the manufacturer’s recommendations, with a mean target coverage of 67×. Sequenced reads were aligned to the GRCh37 human reference genome assembly.

Outlier samples were excluded based on high degree (>5%) of contamination, large proportion (>5%) of chimeric alignment, low call rate (<99%) for single nucleotide variants (SNVs), and mismatches between phenotypic and genotypic sex. Additional outlier samples (mean ± three standard deviations [SDs]) were excluded based on the numbers of variants, non-reference variants, and heterozygous variants, and the heterozygous/homozygous and transition/transversion ratios. In addition, extremely low-coverage samples (below the mean –3 SDs) and extreme values (mean ± 3 SDs) of multiallelic SNVs and inbreeding coefficient *F* were removed in ES quality control. After filtering and exclusion of duplicates, the final set contained 4,594 unique samples (GS, *n* = 2,240; ES, *n* = 2,354). As adjustment for relatedness, first-degree relatives (identity by descent > 0.4) were excluded using PLINK v1.9.1 Samples were excluded by choosing the minimal list of related individuals to break all kinship ties using the RELOUT5 tool from Allele (http://www.toomashaller.com/allele.html).

The Genome Analysis Toolkit’s2 variant quality score recalibration was used to filter variants with truth sensitivities of 99.8% for SNVs and 99.9% for indels. Variants with inbreeding coefficients < –0.3, quality by depth scores < 2 for SNVs and <3 for indels, call rates < 95%, and Hardy-Weinberg equilibrium *P* values < 1×10-6 were removed. Sequenced variants were annotated with Variant Effect Predictor3 version 87 (Gencode v19 on GRCh37.p13) and ANNOVAR.4

**Variant detection and evaluation of variant pathogenicity**

We used a custom pipeline to annotate and prioritize clinically important variants. This pipeline incorporates standard annotations, including variant effect, protein functional predictions, and conservation and allele frequencies, along with publicly available variant data. For estimation of the pathogenicity of likely pathogenic (LP) (and known pathogenic [KP]) variants, the following *in silico* prediction algorithms were considered: SIFT,5 PolyPhen-2, HumDiv, Hvar,6 CADD,7 PhyloP,8 MutationTaster2,9 MutationAssessor,10 PROVEAN,11 MetaSVM, and MetaLR.12 In this manner, we compiled a list of clinically actionable LP and KP *BRCA1*/*2* variants. Potentially important variants were identified, and manually assessed.13 They were classified as pathogenic, likely pathogenic, of uncertain significance, likely benign, and benign.14 Selected candidate variants were confirmed by conventional PCR amplification and Sanger sequencing.

References

1. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell S and, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Am J Hum Genet*. 2015;4(7). doi:10.1186/s13742-015-0047-8

2. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. doi:10.1101/gr.107524.110

3. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):1-14. doi:10.1186/s13059-016-0974-4

4. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015;10(10):1556-1566. doi:10.1038/nprot.2015.105

5. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11:1-9. doi:doi:10.1038/nprot.2015.123

6. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248

7. Kircher M, Witten DM, Jain P, O’roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892

8. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15(7):901-913. doi:10.1101/gr.3577405

9. Schwarz JM, Cooper DN, Schuelke M, Seelow D. Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11(4):361-362. doi:10.1038/nmeth.2890

10. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations : application to cancer genomics. 2011;39(17):37-43. doi:10.1093/nar/gkr407

11. Choi Y, Chan AP. PROVEAN web server : a tool to predict the functional effect of amino acid substitutions and indels. 2015;31(16):2745-2747. doi:10.1093/bioinformatics/btv195

12. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125-2137. doi:10.1093/hmg/ddu733

13. Freeman PJ, Hart RK, Gretton LJ, Brookes AJ, Dalgleish R. VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Hum Mutat*. 2018;39(1):61-68. doi:10.1002/humu.23348

14. Richards S, Aziz N, Bale S, et al. 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. *Genet Med*. 2015;17(5):405-424. doi:10.1038/gim.2015.30