Development and Validation of a Residual Risk Score to Predict Breast Cancer Risk in Unaffected Women Negative for Mutations on a Multi-Gene Hereditary Cancer Panel

Elisha Hughes, PhD; Thaddeus Judkins, MS; Susanne Wagner, PhD; Richard Wenstrup, MD; Jerry S. Lanchbury, PhD; Alexander Gutin PhD

Myriad Genetics, Inc., Salt Lake City, Utah

BACKGROUND

- Women who are unaffected with cancer but have a significant family history of breast cancer are frequently referred for hereditary cancer testing with multi-gene panels.

- Despite being at high familial risk for development of breast cancer, fewer than 10% of such patients carry a clinically actionable mutation.

- Large-scale genotyping studies have identified common variants (primarily single-nucleotide polymorphisms, SNPs) that individually confer modest breast cancer risk, but together may explain the genetic susceptibility for breast cancer in many women without monogenic mutations.

- Here, we describe the development and validation of a polygenic residual risk score (RRS) in a large, consecutive cohort of women who tested negative for mutations in known breast cancer susceptibility genes.

METHODS

COHORT

- We assessed a consecutive series of women of European descent who had genetic testing with a multi-gene pan-cancer panel and tested negative for mutations in breast cancer risk genes (*BRCA1*, *BRCA2*, *TP53*, *PTEN*, *STK11*, *CDH1*, *PALB2*, *CHEK2*, *ATM*, *NBN*, *BARD1*).

- Clinical information was obtained from provider-completed test request forms.

- This study was approved with a waiver of informed consent from Quorum Review IRB (31713). Only patients from states which allow retention of DNA were included for study.

- The cohort was divided into a training (July – November 2016, N=11,771) and validation set (November 2016 – March 2017, N=17,205).

NEXT GENERATION SEQUENCING (NGS)

- A screening panel of 94 previously published variants[1-3] and 3 SNPs in linkage disequilibrium with reported associations with breast cancer risk was designed for the RainDance Thunderstorm NGS target enrichment system.

- Variant calls were obtained by NGS of amplicons on an Illumina HiSeq 2500 next generation sequencer to an average sequence depth of >1200X.
  - 2 variants failed due to their location within repetitive elements and were discarded.

- Genotyping calls for the non-failed 95 variants were validated using Sanger sequencing or the IonTorrent Ampliseq platform in 189 DNA samples.
  - There was 100% concordance between the variant calls generated using the RainDance Thunderstorm system and the comparator assays.

STATISTICAL ANALYSIS

- Multivariable logistic regression models were used to evaluate the variants, develop a RRS as a predictor of personal breast cancer history in the training cohort, and assess the performance of the RRS in the validation cohort.

- Independent variables included age, personal/family cancer history, and ancestry (West/Northern European, Central/Eastern European, or Ashkenazi).

- Analysis of the validation cohort was conducted according to a pre-specified Statistical Analysis Plan.

RRS DEVELOPMENT

DEVELOPMENT METHODS

- SNP genotypes were coded as the number of the effect alleles (0, 1, or 2) (Figure 1).

- The corresponding estimated coefficients were combined with log odds ratios (ORs) from the literature[1-3] using weighted averaging with the weights inversely proportional to the squares of the corresponding confidence intervals.

- The resulting coefficients β_i were used to calculate SNP informativeness defined as $2f_i(1-f_i)\beta_i^2$, where f_i is the effect allele frequency for SNP_i.

- To select the optimal number of variants in the final RRS, variants were ordered by informativeness. From 3 groups of linked SNPs in *CCND1*, *ESR1* and *TERT*, we included only the most informative SNP from each group.

- Polygenic scores were constructed for cumulatively increasing numbers of variants and were evaluated with log likelihood ratios from logistic regression models.
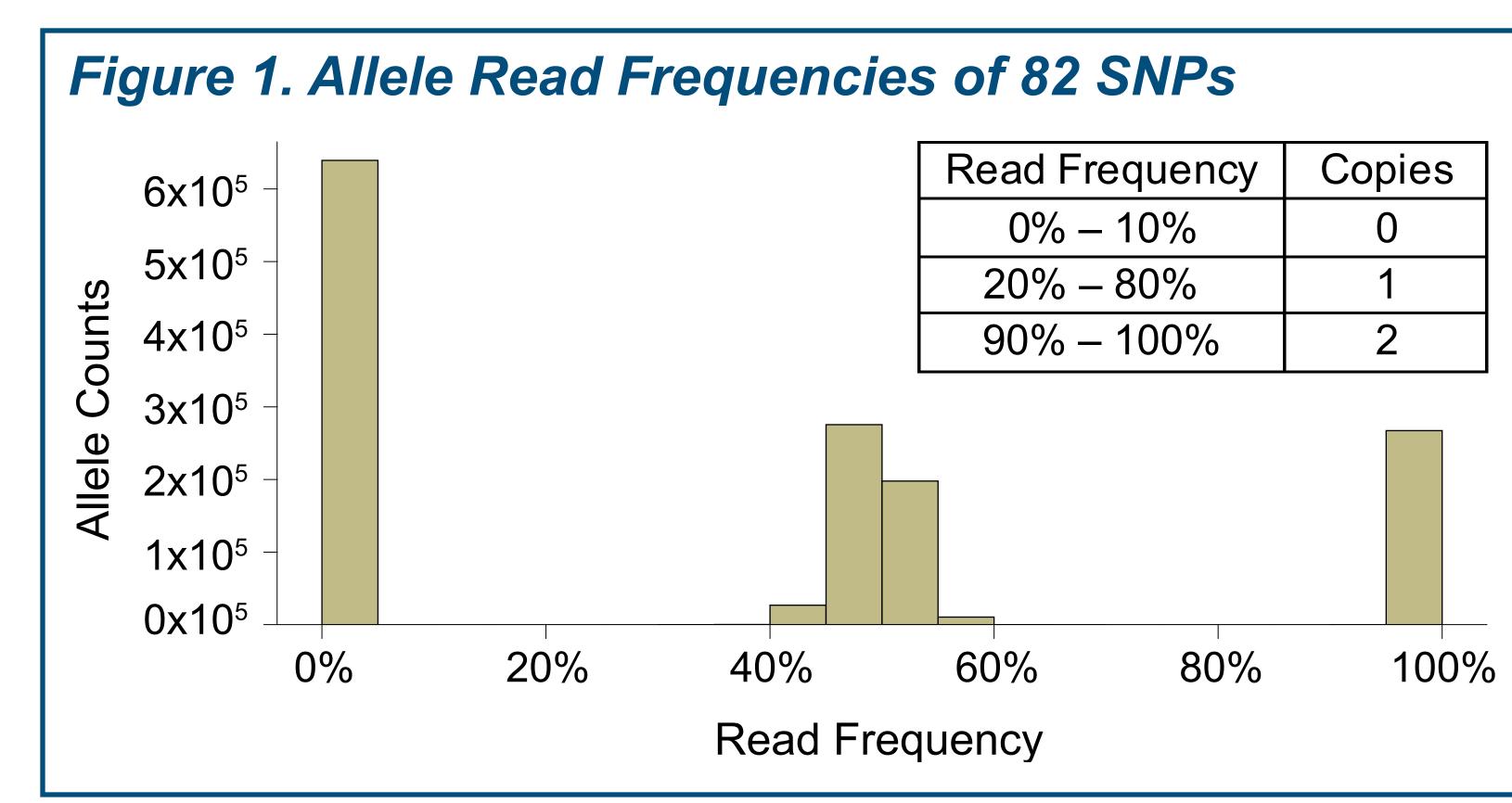
Table 1. Patient Characteristics

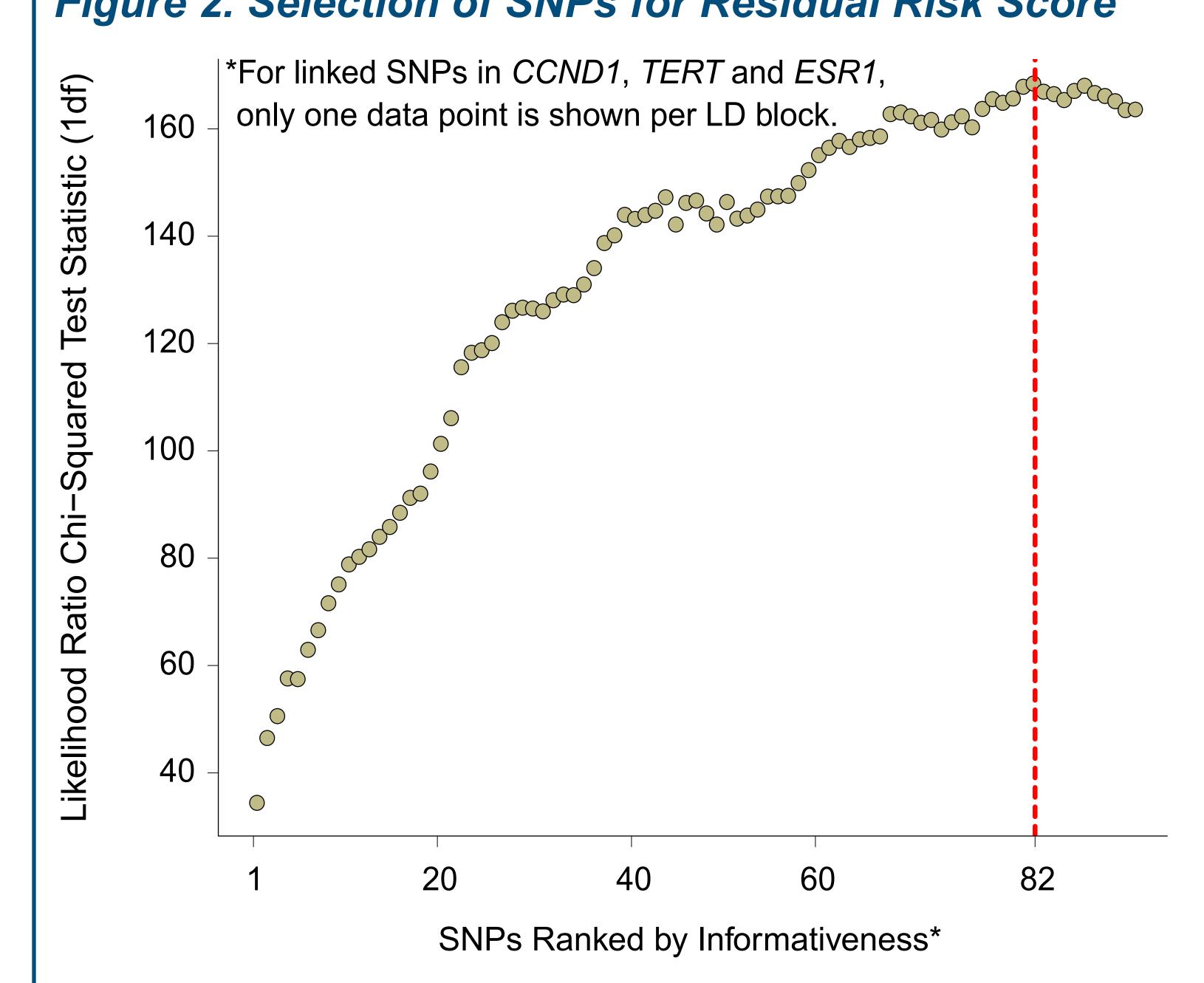|	Training Set		Validation Set	
	All Patients	BC Cases	All Patients	BC Cases
Total Patients	11,771 (100%)	2,089 (18%)	17,205 (100%)	2,917 (17%)
Age at Hereditary Cancer Testing				
Median (years)	18-84	22-84	18-84	22-84
Range (years)	46	54	47	54
Tested ≤ 50	61%	7%	61%	7%
Ancestry				
West/Northern European	8,744 (74%)	1,661 (80%)	12,722 (74%)	2,349 (81%)
Central/Eastern European	2,728 (23%)	385 (18%)	4,008 (23%)	509 (17%)
Ashkenazi	299 (3%)	43 (2%)	475 (3%)	59 (2%)
Cancer History in First Degree Relatives				
No BC or OC	6,367 (54%)	1,349 (65%)	9,507 (55%)	1,919 (66%)
≥ 1 BC	4,262 (36%)	642 (31%)	6,111 (36%)	892 (31%)
≥ 1 OC	1,491 (13%)	132 (6%)	2,042 (12%)	160 (5%)

BC, Invasive ductal breast cancer; OC, Invasive epithelial ovarian cancer

DEVELOPMENT RESULTS

- The maximum of the likelihood ratios was achieved with the 82 most informative SNPs (Figure 2).

- The final RRS score was defined as $RRS = \sum \beta_i n_i$
  - Sum over the 82 most informative SNPs
  - n_i is the number of the effect alleles for SNP_i
  - β_i is the coefficient estimated by combining results from the logistic regression model with 82 SNPs in the training cohort with published estimates as described above for 92 variants
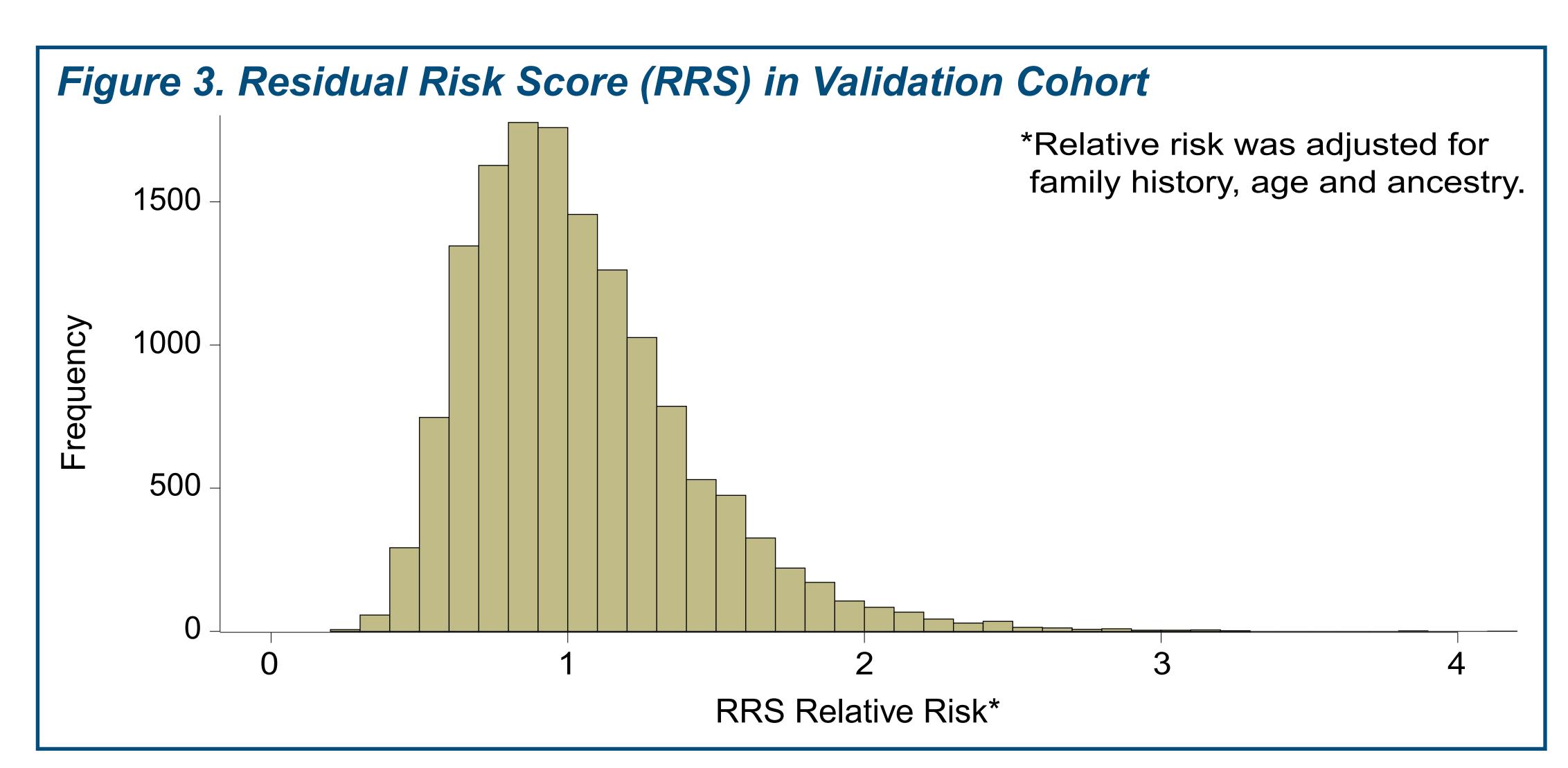

Figure 1. Allele Read Frequencies of 82 SNPs

Read Frequency	Copies
0% – 10%	0
20% – 80%	1
90% – 100%	2

Figure 2. Selection of SNPs for Residual Risk Score

*For linked SNPs in *CCND1*, *TERT* and *ESR1*, only one data point is shown per LD block.

RRS VALIDATION

VALIDATION RESULTS

- The RRS was strongly associated with personal history of breast cancer in the validation cohort (p<10[-50]).
  - OR per unit standard deviation of the RRS: 1.42 (95% CI = 1.36-1.49).

- This OR is lower than that of a published polygenic risk score (PRS) based on 77 SNPs (OR 1.55).[1]
  - This difference can be explained by the fact that OR for RRS was adjusted for family cancer history while the published OR for PRS was not.

- In a model with both scores included, the RRS score was significantly associated with breast cancer (p=3x10[-5]) while the PRS was not (p=0.2).

- Figure 3 shows the distribution of odds ratios associated with RRS in unaffected women from the validation cohort.

Figure 3. Residual Risk Score (RRS) in Validation Cohort

*Relative risk was adjusted for family history, age and ancestry.

CONCLUSIONS

- A residual risk score was developed that is highly predictive of risk of development of future breast cancer in unaffected women with significant family history after testing negative for known high and intermediate risk mutations.

- The residual risk score significantly outperformed the published polygenic risk score in this population of women. This may have been due to inclusion of a larger number of SNPs, inclusion of the most informative SNPs, and refined OR estimates for individual SNPs.

- The clinical testing implementation of a residual risk score in women at risk for hereditary breast cancer may offer significant potential for the management of greater than 90% of high-risk women who test negative for monogenic breast cancer mutations.

REFERENCES
1. Mavaddat et al. *J Natl Cancer Inst.* 2015; 107:djv036.   2. Michailidou et al. *Nat Genet.* 2013;45:353.   3. Michailidou et al. *Nat Genet.* 2015;47:373.