

A potential of polygenic risk scores as a tool to inform population screening for breast cancer

Arina V. Nostaeva

Institute of Cytology and Genetics, Novosibirsk, Russia; PolyKnomics, 's-Hertogenbosch, The Netherlands avnostaeva@gmail.com

Tatiana I. Shashkova

Institute of Cytology and Genetics, Novosibirsk, Russia

Lennart C. Karssen

PolyKnomics, 's-Hertogenbosch, The Netherlands

Yurii S. Aulchenko

Institute of Cytology and Genetics, Novosibirsk, Russia; PolyKnomics, 's-Hertogenbosch, The Netherlands

A polygenic risk score, or PRS, aggregates the effects of many genetic variants into a single number which predicts genetic predisposition for a phenotype. In the standard approach, a PRS is a linear combination of linear regression effect estimates and genotypes. Many studies have shown that a PRS can predict disease status in research-based case-control studies [1-3]. Here, we develop and validate genome-wide polygenic risk scores for breast cancer (BC) and attempt to quantify their potential to inform population screening programs.

We developed a set of draft PRS models using the SBayesR [4] method that utilised summary statistics from a previously published large genome-wide association study (GWAS) of BC [5]. The models were validated and tested in independent datasets, not overlapping with the samples used for GWAS. The optimal PRS model was chosen based on the maximum area under the receiver-operator curve (AUC) in the UK Biobank [6] validation dataset, which included white British women born in England (513 cases and 9,130 controls). Subsequently, we tested the obtained PRS model in datasets comprised of the subsamples of UK Biobank participants including white British women born in different regions of UK (5 groups including 2,329 cases and 43,319 controls in total).

In the validation dataset, the best PRS model that included effects of 952,794 single nucleotide polymorphisms, demonstrated an AUC of 64%. The model performed equally well in the testing

datasets, with AUCs ranging from 63% to 66%. Next, we estimated lifetime risks of BC for all women from the testing datasets in percentiles of the PRS. The results demonstrate that the estimated lifetime risk by age 75 ranged from 2% for women in the lowest 5% of the PRS distribution to 17% in the highest 5%. At 50 years of age — the recommended age at which to start BC screening — the average risk of having BC was 1.7%, whereas the risk for those in top 5% of the PRS distribution was 4.4%, and for those in the lowest 5% the risk was 0.6%. Notably, the women in the highest 5% of PRS achieved the risk of 1.7% by the age 38, while those in the lowest 5% achieved equivalent risk only by the age of 71.

Thus, while the AUC may appear modest, the predicted risk differences in the tails of the distribution are large. We conclude that PRS has the potential to improve BC population screening programs through prediction of individual risk trajectory.

Acknowledgements: This research has been conducted using the UK Biobank Resource project 41601. This work was funded by PolyKnomics BV.

1. Khera AV et al. (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations., *Nat Genet.*, **50**:1219–24.
2. Mavaddat N et al. (2019) Polygenic risk scores for prediction of breast cancer and breast cancer subtypes., *Am J Hum Genet.*, **104**:21–34.
3. Musliner KL et al. (2019) Association of polygenic liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the Danish population., *JAMA Psychiatry*, **76**:516–25.
4. Lloyd-Jones LR et al. (2019) Improved polygenic prediction by Bayesian multiple regression on summary statistics., *Nat Commun.*, **10**:5086.
5. Zhang H., et al. (2020) Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses., *Nat Genet.*, **52(6)**:572-581.
6. C. Bycroft et al, (2018) The UK Biobank resource with deep phenotyping and genomic data., *Nature*, **562**:203.