DATA SHEET

Polygenic risk score software: Predict module

A powerful data analysis solution for PRS research

- Streamlined genotype-to-risk assessment solution with broad disorder coverage
- Easy-to-use interface simplifies data analysis and genetic risk score reporting
- Ancestry-informed genetic risk scores provide highly accurate disease risk prediction

illumina®

Introduction

Polygenic diseases, such as type 2 diabetes, breast and prostate cancers, hypertension, coronary artery disease, and Alzheimer's disease, are caused by the combined effects of hundreds to thousands of susceptibility genes. Each of these common disease-associated variants contributes to the overall disease risk, which can also be influenced by environmental factors, making the prediction of disease susceptibility challenging. Polygenic risk scores (PRS), or genetic risk scores (GRS), represent the total number of genetic variants that increase the individual risk of developing a specific disease. This promising precision medicine tool can be used to estimate an individual's likelihood of developing a disease, stratify populations according to disease risk, and identify individuals who would benefit from additional monitoring or preventive interventions.[1,2] However, PRS is an emerging field, requiring more research before it becomes mature.

PRS studies have demonstrated the ability to identify individuals with a high genetically determined risk of developing complex polygenic diseases (Figure 1).[3-7]

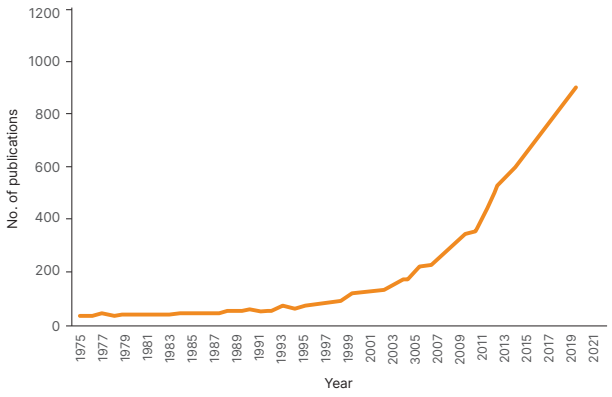

Figure 1: Number of publications with polygenic risk scores—A keyword search for 'PRS' or 'GRS' in publications in the PubMed database shows a marked increase in studies using this tool to determine disease risk. Current as of February 2022.

Researchers can leverage data from genome-wide association studies (GWAS) to compute PRS, defined as the weighted sum of allele counts across disease-associated variants (Figure 2).[8] However, due to the vast number of variants identified in large-scale population studies, per-individual PRS construction is computationally intensive. Deep bioinformatics expertise is required to run the PRS pipeline, from imputing genomic data to quality control and reporting, limiting the widespread use of PRS in clinical research.

The Illumina Predict module, developed in partnership with Allelica, simplifies the process of PRS calculation, analysis, and reporting with an easy-to-use interface to generate individualized PRS at scale. The Predict module is available as part of a comprehensive toolkit that includes the Infinium™ Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip, offering a streamlined genotype-to-risk prediction solution to help expand research in the emerging field of PRS assessment.

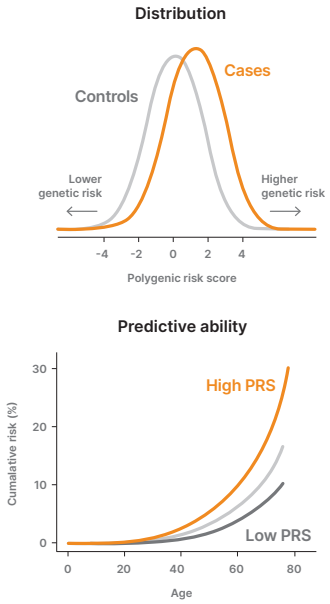

Figure 2: Representative results for disease risk prediction using PRS—The top panel shows distributions of PRS for cases (orange line) and controls (gray line). The lower panel shows the predictive ability of PRS for an individual sample demonstrating an increased risk of developing disease earlier than reference samples with a low or average PRS percentile.

Accurate risk assessment with phenotype-based PRS distribution

The Predict module computes individual PRS scores and reports disease risk in relation to a reference population. Though comparing a numerical PRS to a population distribution with genotype data provides a percentile risk score relative to the reference cohort, it does not provide information on the absolute disease risk. To link PRS to actual disease risk, all population distributions used in the Predict module contain phenotype data in addition to genotype. The Predict module includes scores from the Polygenic Score (PGS) Catalog and Allelica-developed risk scores. Quality control measures specific to each PRS are built into the Predict module to maintain high predictive performance.

Ancestry-informed disease risk calculation

Heritability of polygenic diseases is often variable across diverse populations, which makes ancestry-informed risk assessment crucial for accurate PRS determination.[9] The Predict module, which is intended for research use only, includes all PRS included in the PGS Catalog with ~1K precompiled risk models comprising more than 20 traits and over 80 PRS for each of the five continental ancestries. As a result, the software provides comprehensive coverage of potentially pathogenic variants that contribute to disease risk. This tool provides fine-scale ancestry calibration by leveraging a harmonized principal component analysis (PCA) space of ~800K individual samples. The Predict module also enables researchers to generate automated PRS reports with ancestry-specific risk prediction and offers broad disorder coverage including oncology, cardiovascular disease, and general wellness.

Simplified data analysis workflow

The Predict module is compatible with input data generated on microarrays or whole-genome sequencing systems. Data files from Illumina sequencing systems are sent to the Illumina cloud platform and scores of interest are selected, either from over 2600 scores included in the Predict module or custom laboratory-generated PRS, to compute an individualized PRS for each sample (Figure 3). After a score is computed, the Predict module generates personalized reports with comprehensive risk assessments that include other variables, such as sex and ethnicity, for accurate risk prediction.

For low coverage sequencing data, the Predict module offers an imputation option to increase the number of genetic variants present in the data set. With the imputation engine, ~10K samples can be imputed in less than three hours with precision exceeding 99% and high concordance with the Infinium Global Screening Array.

Figure 3: Simplified workflow for PRS determination using the Predict module—The Predict module offers a comprehensive risk prediction solution for fast and accurate disease risk assessments without the need for extensive bioinformatics resources.

Easy-to-use interface

Raw genotype data from Illumina sequencing systems integrate with the intuitive BaseSpace™ Sequence Hub interface, enabling easy point-and-click analysis and PRS computation. Up to 24 PRS can be calculated in 15 minutes using the Predict module. A standardized PRS report is generated for every sample. This comprehensive report includes ancestry-informed numerical polygenic scores and relative lifetime risk of disease, providing critical context for PRS interpretation (Figure 4).

Secure, cloud-based data analysis platform

The Predict module is embedded in the Illumina Connected Analytics (ICA) cloud-based infrastructure, which is optimized for quality and performance. In addition to providing a scalable and secure platform, ICA enables parallel computing of up to 1152 samples to be assessed in a single analysis, for fast, accurate, ancestry-informed PRS determination. Data generated using the Predict module can be integrated with other tools also hosted on ICA, including pharmacogenomics and genotyping analysis pipelines.

Comprehensive genotype-to-risk prediction solution

The Predict module is available as part of a high-performance PRS toolkit that includes the Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip for highly accurate risk prediction and standardized reporting. The BeadChip features updated content with ~160K new PRS markers, developed in collaboration with Allelica, added to the expansive Infinium Global Diversity Array genome-wide backbone. The comprehensive BeadChip content was carefully selected to provide broad coverage of key polygenic disorders, including type 1 and type 2 diabetes; breast, colon, and prostate cancers; coronary artery disease; and Alzheimer's disease. The iScan™ System, Infinium LCG Assay technology, and integrated Predict module work together to create a comprehensive, high-performance genotype-to-risk prediction solution to support PRS research (Figure 5).

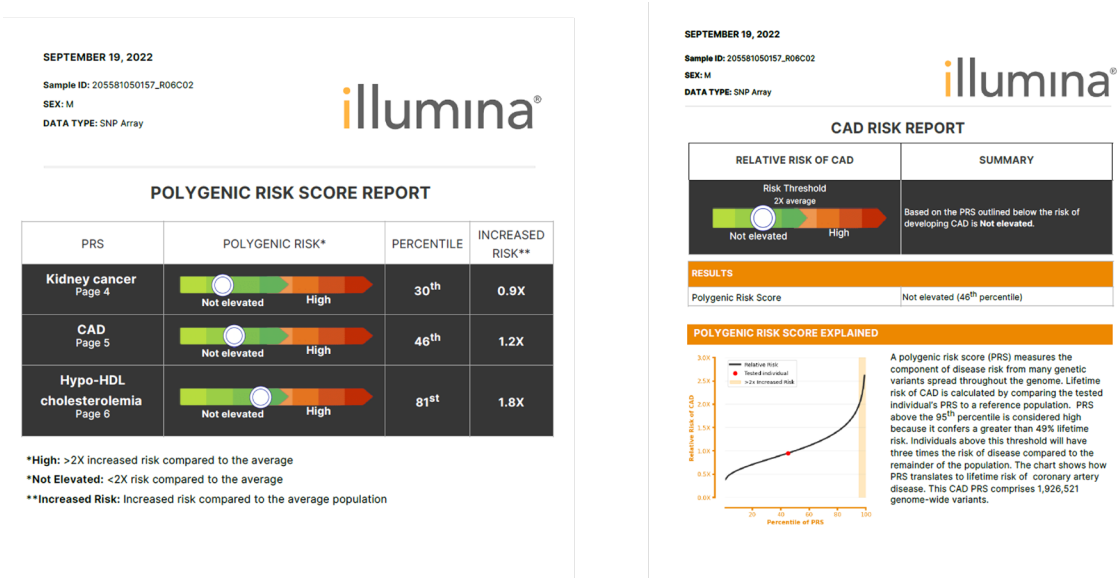

Figure 4: Example risk prediction report—The Predict module provides individualized reports that convey the relative risk percentile for each sample in an easy-to-visualize format with information about the PRS that was used for risk calculation.

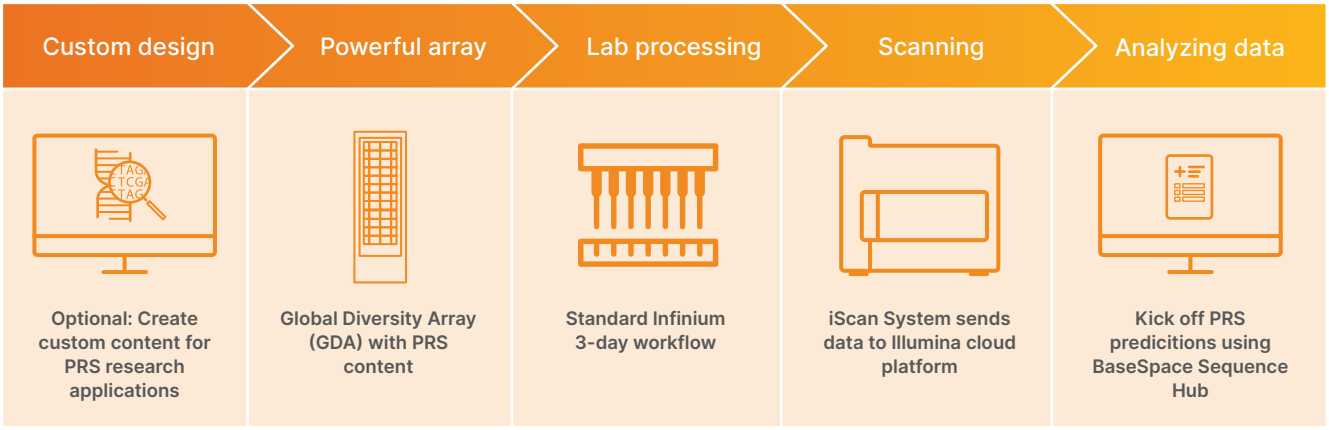

Figure 5: Complete toolkit for PRS research—The Predict module combined with the Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip offers a streamlined genotype-to-risk prediction solution for highly accurate PRS assessment and standardized reporting.

Summary

The Predict module uses powerful analysis algorithms and bioinformatics pipelines developed in partnership with Allelica to simplify PRS assessment and reporting. The easy-to-use BaseSpace Sequence Hub user interface offers fast, accurate, ancestry-informed PRS determination with broad polygenic disorder coverage. The Predict module leverages data from the expansive PGS Catalog to compute relative risk scores. This software module is available as part of a comprehensive toolkit that includes the Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip, offering a streamlined genotype-to-risk prediction solution for precision medicine research.

Learn more

Polygenic risk scores, illumina.com/areas-of-interest/complex-disease-genomics/polygenic-risk-scores

Illumina Connected Analytics, illumina.com/products/by-type/informatics-products/connected-analytics

References

1. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9(3):e1003348. doi:10.1371/journal.pgen.1003348
2. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12(1):44. doi:10.1186/s13073-020-00742-5
3. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z
4. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*. 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
5. Huynh-Le MP, Fan CC, Karunamuni R, et al. Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. *Nat Commun*. 2021;12(1):1236. doi:10.1038/s41467-021-21287-0
6. Elliott J, Bodinier B, Bond TA, et al. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *JAMA*. 2020;323(7):636-645. doi:10.1001/jama.2019.22241
7. Ge T, Irvin MR, Patki A, et al. Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Med*. 2022;14(1):70. doi:10.1186/s13073-022-01074-2

8. Collister JA, Liu X, Clifton L. Calculating polygenic risk scores (PRS) in UK Biobank: A practical guide for epidemiologists. *Front Genet*. 2022;13:818574. doi:10.3389/fgene.2022.818574.
9. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*. 2019;10(1):3328. doi:10.1038/s41467-019-11112-0

Ordering information

Product	Catalog no.
Predict module	20086666
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip (16 samples)	20090683
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip (48 samples)	20090684
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip (96 samples)	20090685
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 BeadChip (384 samples)	20090686

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-01376 v1.0