

THE PREDICT Module

Unlocking the power of genomics to identify an individual's genetic risk of disease

In our previous articles on Allelica's PRS pipeline, we ran through how you can use our [DISCOVER module](#) to build your own PRS for a disease or trait of interest, as well as how our [VALIDATE module](#) can be used to test the predictive power of a new or existing PRS on an independent or new population.



In this final article in the series we'll cover our PREDICT module. In many ways, the overarching aim of the entire pipeline is to use genetic information from a new individual or set of individuals to predict their genetic liability for a disease or trait. Our PREDICT module allows just that, and outputs an individual's polygenic risk score in relation to those computed across a reference population.

The PREDICT module is designed to work with a range of input data, so whether you have genotype data generated on a microarray, or low or high coverage whole genome sequencing data, you should have no trouble uploading data into our software. The first step of the PREDICT module is to use our IMPUTE module to increase the number of genetic variants present in your dataset. (You can read more about our [IMPUTE module](#) here.)

With imputed genetic data in hand, users can then apply their own PRS to the new individual or choose from a list of PRSs from the [PGS catalogue](#), which are available in the PREDICT module via API.

In essence, the computation of an individual's PRS is straightforward. A PRS provides an estimate for the contribution of a particular allele to an individual's genetic liability for the trait of interest. These effect sizes are then added up across all variants in a PRS. This gives us a single number which, although has value as an absolute number, it needs to be reported in comparison to a population distribution.

Comparing an individual's score to a distribution of known phenotypes

The most basic approach to PRS comparison involves relating an individual's score with a distribution of scores computed on a large dataset of individuals for whom we have genotype but not phenotype data. An example of this type of approach would be to compare a PRS to a distribution of PRS computed on the 1 Thousand Genomes Project dataset. Whilst this places an individual's PRS in the context of a population distribution so you can see what percentile of risk an individual is in, it provides no real sense of the actual risk of the disease or trait of interest that that individual has.

The proper way is to use a population distribution that includes individuals with known phenotypes. This comparison allows you to estimate a fold increase or decrease in risk to an individual's score, which has important implications for what you can do with the score. Importantly, all of the population distributions used by Allelica's PREDICT module are compared to populations with known phenotypes, and not just a population distribution of genotypes, providing a link between PRS and clinical risk.

We're working with clinicians to incorporate these scores into risk classification models for a number of diseases, including [coronary artery disease](#) and [breast cancer](#). If you'd like to explore ways in which PRS can be used in your clinic or research, we'd love to hear from you.

Please [get in touch](#) for further details about our PREDICT module or to [request a demo](#) of our whole PRS pipeline.