

THE VALIDATE MODULE

Essential testing of your PRS on an independent population

In our previous articles outlining Allelica's PRS pipeline, we described how users can [impute genotype or low coverage sequencing data](#) and [build their own PRS](#) using our cloud-based computing platform. A key aspect of the DISCOVER module is that it allows you to trial multiple different methodologies for computing a PRS so that the one that has the best predictive power can be chosen.

The next critical step in developing a PRS is validating it on an independent dataset. In this article, we'll cover how Allelica's VALIDATE module can be used to help users understand the applicability of their PRS to a new population. Importantly, the same infrastructure can be used to test the transferability of PRS to a population that is different from the one on which the PRS was developed.

HELPING BRIDGE THE PRS DIVERSITY GAP

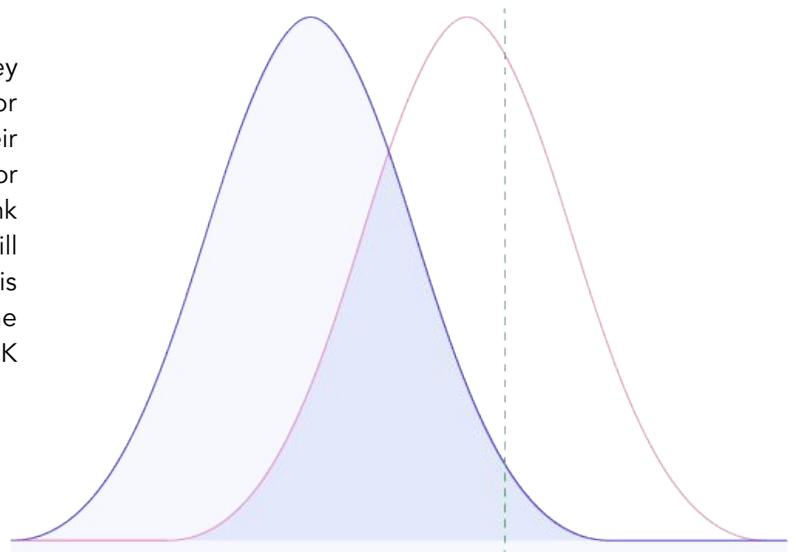
Most PRS published so far have concentrated on using genomic datasets from populations of predominantly western European genetic ancestry. This is due to there being much more data available from these populations. Several researchers have highlighted this lack of diversity and we support all efforts to increase datasets to produce a more equitable approach to precision medicine. In the meantime, our VALIDATE module allows users to quantify how well a PRS developed on a different population transfers to a new population of interest.

TESTING A POLYGENIC RISK SCORE ON AN INDEPENDENT POPULATION

To validate a PRS, users need a new, independent set of genomic data on which the phenotype of interest has been measured.

Users who have a new genomic dataset with matched phenotype data on which to test a previously generated PRS will want to jump into the pipeline here. It's quick and easy to use the VALIDATE module to run any of the over 1700 PRS that are available on the online [PGS Catalog](#).

Alternatively, users might want to validate a PRS they have themselves built using [the DISCOVER module](#). For users with large datasets, we suggest that they split their dataset prior to starting with the DISCOVER module. For those with smaller datasets, we provide the UK Biobank dataset as a means of validating their PRS, which will suffice as long as the phenotype under investigation is present in the UK Biobank, and the PRS built in the DISCOVER module is not itself based on the UK Biobank.



QUANTIFYING THE VALIDITY OF A PRS

As with our approach to quantifying the predictive power of a PRS in the DISCOVER module, the VALIDATION module provides three standard metrics of model fit. These are the Area Under the Curve of the Receiver Operator Curve, which measures how well the model classification works; the Odds Ratio per standard deviation, which measures how the model captures the gradient of risk of the disease in question; and the percentiles of the dataset that are at 3 fold or greater increased risk relative to the remainder of the population. Armed with these statistics, users can quantify how well a PRS works in an independent population, providing a measurement of the likely predictive power of the PRS across different populations.

FROM VALIDATION TO INDIVIDUAL PREDICTION

The final step of Allelica's PRS pipeline is the PREDICT module which lets users predict PRS for a given disease in a set of new individuals, placing them into the context of their population genetic risk. We'll write more about this in the final article of this series on our SaaS pipeline.

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