The Spanish Polygenic Risk Score Reference Distribution: A Resource for Personalized Medicine

Daniel López-López[1,2,3], Gema Roldán[1], María Peña-Chilet[1,2,3], and Joaquín Dopazo[1,2,3,4]

[1]Computational Medicine Platform, Andalusian Public Foundation Progress and Health (FPS), [2]Computational Systems Medicine, Institute of Biomedicine of Seville (IBiS), [3]Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), [4]FPS-ELIXIR-ES, Andalusian Public Foundation Progress and Health (FPS)

What are Polygenic Risk Scores?

- **Definition**: PRS is a measure of genetic risk for certain diseases or conditions, calculated using many genetic variants.
- **Calculation**: PRSs are derived from data from genome-wide association studies (GWAS).
- **Function**: PRS have demonstrated superior predictive power compared to clinical models in accurately determining the risk of diseases such as breast cancer, prostate cancer, and type 1 diabetes. They can identify individuals at equivalent risk levels as rare mutation carriers.
- **Population specific**: PRS distributions vary significantly according to population structure.
- **Limitations**: Despite the promise of PRS, they cannot yet be widely adopted due to lack of reference distributions for most populations. Other factors not included in the PRS, such as environment and lifestyle, also significantly impact disease development, and their predictive power can vary between populations and diseases.

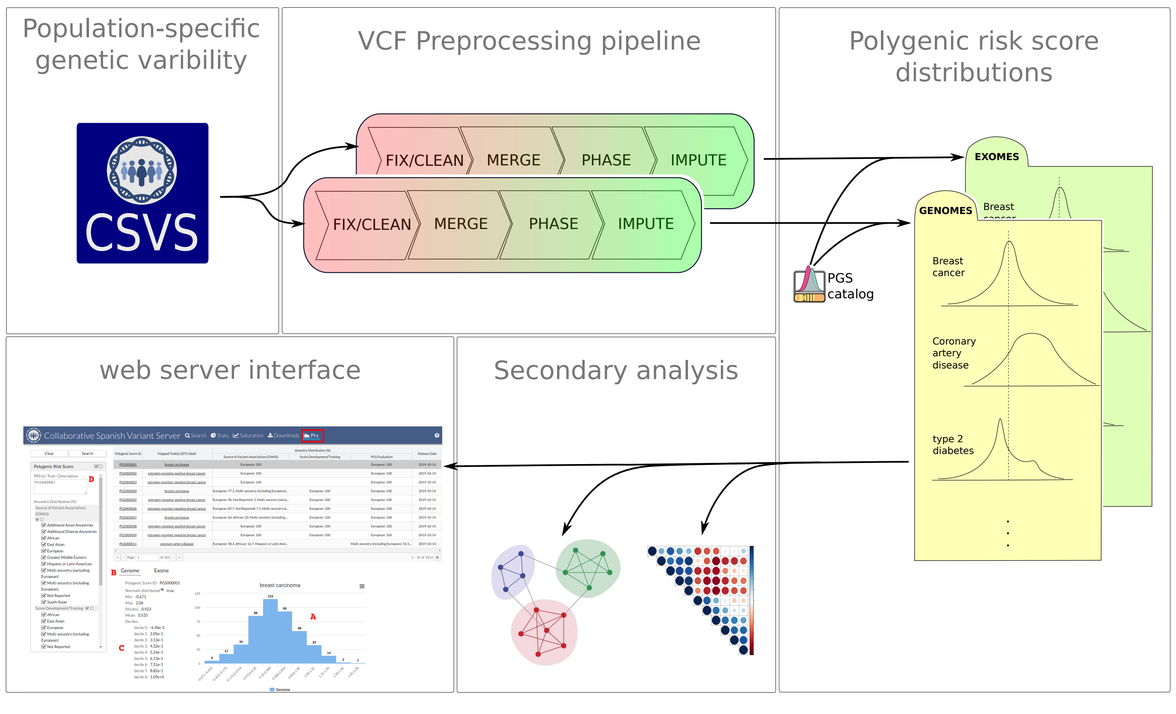
Why do PRSs matters?

- **Personalized Medicine**: PRSs allow for more personalized prevention strategies and treatment plans. They can provide insight into an individual's genetic predisposition to certain conditions, which can inform lifestyle changes, preventive measures, or early interventions to manage risk.
- **Understanding Complex Diseases**: PRSs are particularly useful for understanding complex diseases that are influenced by multiple genes and their interactions. This makes it possible to estimate disease risk more accurately than when looking at single gene mutations.
- **Research Applications**: PRSs aid in participant stratification in research and can assist in drug discovery.
- **Population Health**: PRSs offer insights into health and disease across different populations, informing public health initiatives and interventions.
- **Predictive Tool**: PRSs provide a genetic risk measure, aiding disease prediction when combined with other risk factors.

Results

- We present the Spanish PRS reference distribution, which includes **3124 PRS distributions** for common diseases and quantitative traits.
- The reference includes PRS for various types of cancer, disorders associated with the digestive, cardiovascular, neuronal, and immune systems, as well as quantitative traits.
- A **standardized pipeline** for preprocessing, phasing, and imputing samples has been developed, enabling the computation of PRS for external genomes and exomes.
- The **web interface** allows for the exploration and analysis of the PRS reference distributions, including frequency histograms and distribution parameters.
- The study suggests the potential for establishing population-specific PRS distributions for other populations to facilitate the adoption of PRS in **healthcare systems worldwide**.

Methods



Pairwise correlations

Some diseases and traits commonly used to forecast the onset and progression of the disease are **highly correlated**. An in-depth examination of the pairwise correlations table may uncover intriguing and previously unknown relationships that could be explored in further research.

Clustering

We found **groups of highly similar PRS** when clustered based on their total correlation, even when they don't share most of the variants used to train the model, as measured by the jaccard distance. To ensure effective clinical implementation of PRS, it is crucial to choose **appropriate** ones from the vast public catalog. Therefore, an initiative to select the most suitable PRS for each disease and trait based on their performance and sample origin would be a valuable resource.

Availability

- **web**: http://csvs.clinbioinfosspa.es/?tab=prs
- **pipeline**: https://doi.org/10.5281/zenodo.7919159
- **correlations**: https://doi.org/10.5281/zenodo.7896668
- **clustering**: https://doi.org/10.5281/zenodo.7896698

Junta de Andalucía | Consejería de Salud y Consumo | Fundación Progreso y Salud
IBiS INSTITUTO DE BIOMEDICINA DE SEVILLA
ciberer Centro de Investigación Biomédica en Red Enfermedades Raras
Gobierno de España | Ministerio de Ciencia e Innovación
Instituto de Salud Carlos III
European Union | European Regional Development Fund