

## Introduction

Recently, polygenic risk score (PRS) methods have built on traditional genome-wide association studies to improve disease risk prediction, but possess a few limitations. Adaptively Boosting PRS (AB-PRS) is a method that aims to enhance pre-trained PRSs by detecting missed or biased genetic signals.

In the INSPIRE internship, I:

- 1) integrated AB-PRS functions into an R package for user-friendly access,
- 2) developed a model evaluation function that accepts multiple PRS models and generates an interactive HTML report with various performance plots, and
- 3) applied the AB-PRS method to a simulated dataset.

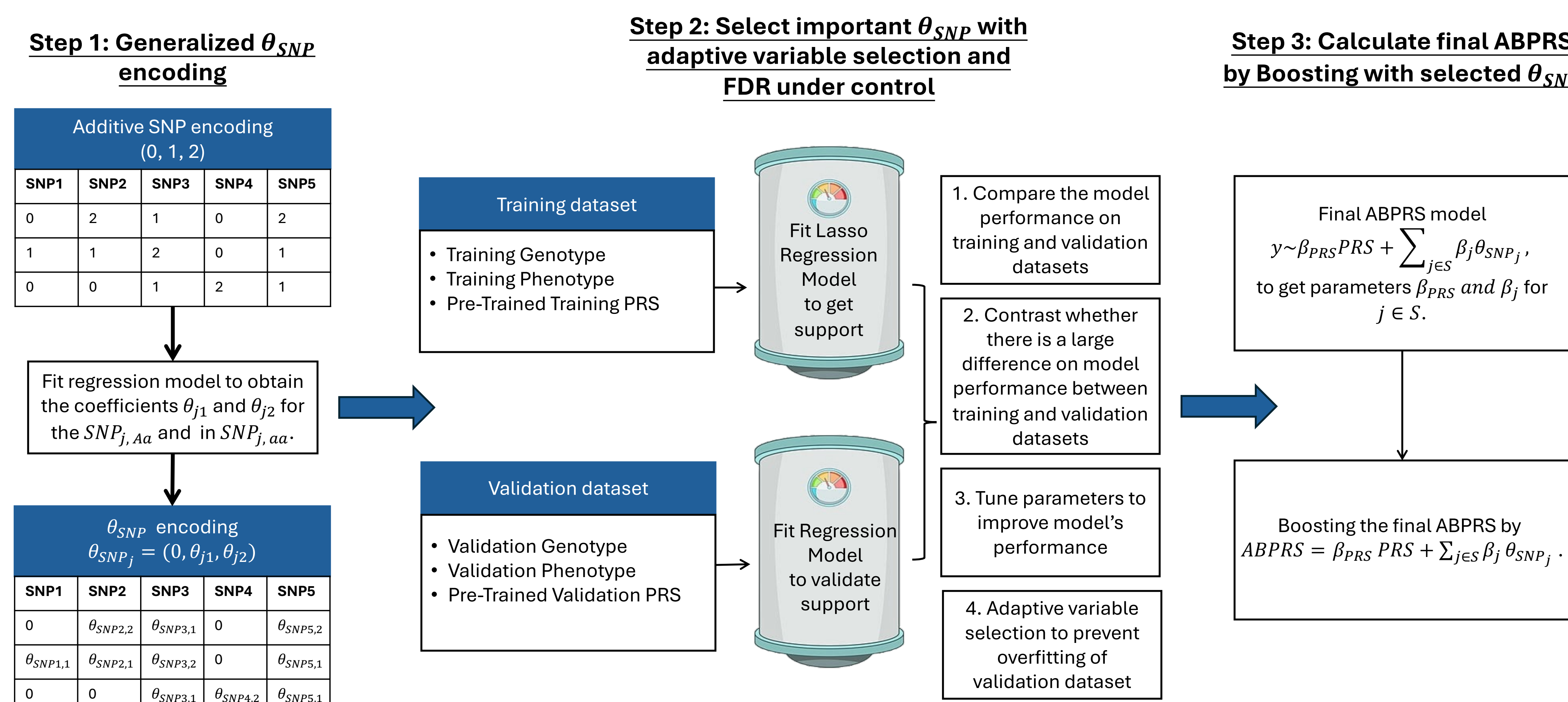


Figure 1: AB-PRS Method Flowchart

## Result 1: AB-PRS R Package

### Objectives

1. Ensure a seamless, customizable user experience with both comprehensive and step-by-step functions.
2. Allow users to compare the performance of different methods and to customize results with downloadable tables and corresponding codes.
3. Understand the AB-PRS method from start to finish.

### Methods

The ABPRS method comprises of 3 parts:

1. Generate  $\theta_{SNP}$  coefficients and convert Single Nucleotide Polymorphisms (SNPs) with additive encoding to  $\theta_{SNP}$  encoding to capture signals missed or biased by the pre-trained PRS.
2. Select important  $\theta_{SNPs}$  through adaptive variable selection while controlling false discovery rates.
3. Calculate the final AB-PRSs by enhancing pre-trained PRSs with the selected  $\theta_{SNPs}$  and their corresponding weights.

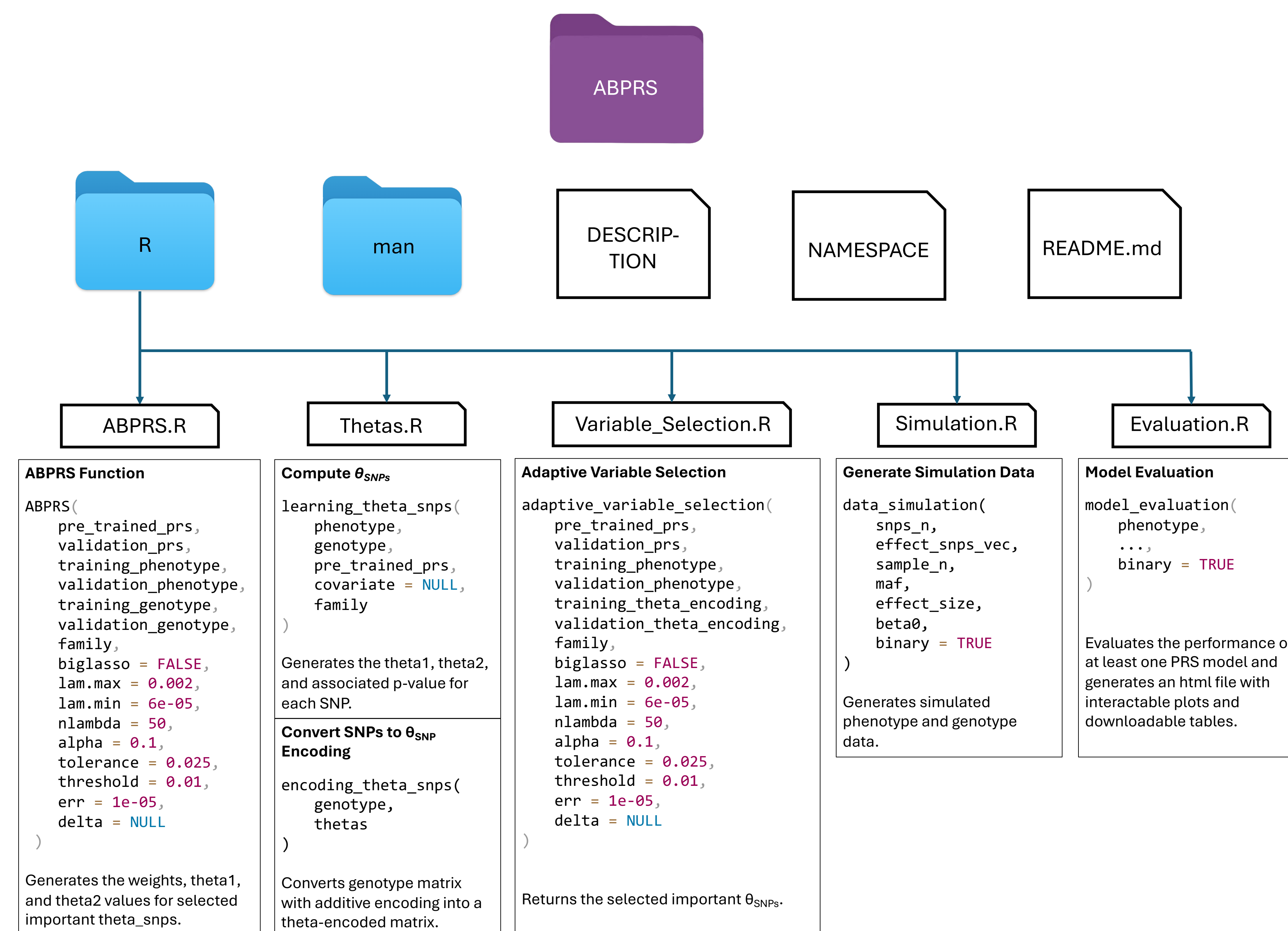


Figure 2: AB-PRS R Package Structure Diagram. The R package consists of 2 main folders and 6 main functions. The "R" folder stores the functions while the "man" folder stores the user manuals for each function. The package consists of a comprehensive AB-PRS function, three step-by-step functions, a simulation function, and a model evaluation function.

## Result 2: Application of AB-PRS on Simulated Dataset

- 1) Generate simulated datasets of binary phenotype, and split the data into 50% training, 25% validation, and 25% testing datasets.
- 2) Generate pre-trained polygenic risk scores.
- 3) Retrieve weights of important  $\theta_{SNPs}$  from the ABPRS function.
- 4) Calculate AB-PRSs through matrix multiplication.
- 5) Evaluate and compare the pre-trained PRS and AB-PRS model performances.

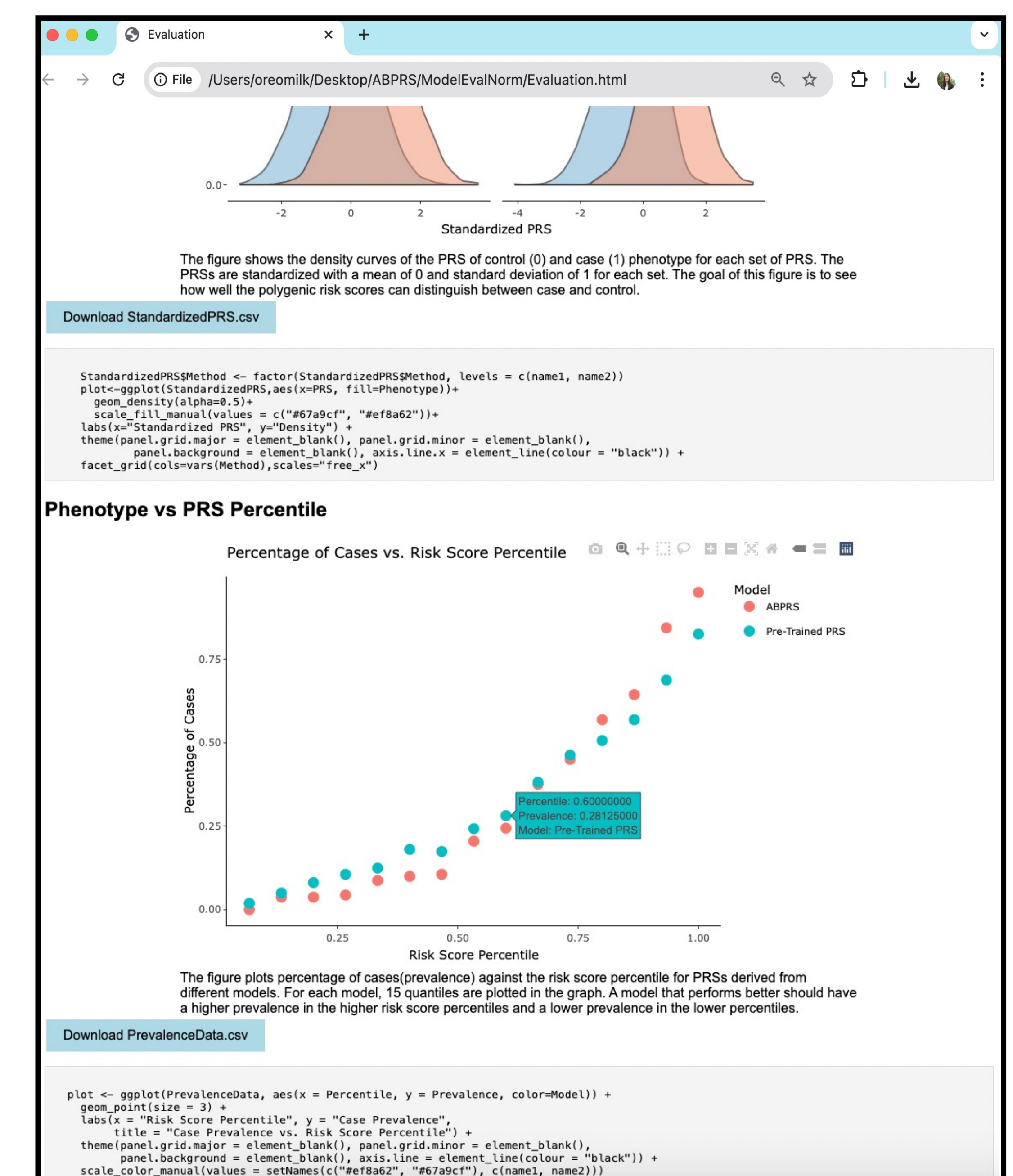


Figure 3: Snapshot of the resulting interactive evaluation html file

## Future Direction

Run the AB-PRS method to boost the pre-trained PRSs of type 2 diabetes and fasting glucose phenotypes from the Microbiome and Insulin Longitudinal Evaluation Study (MILES) by Dr. Mark Goodarzi, MD.