P8 Formula (3) and (4)

P15 Were these the two considered different genetic architectures? Or an example? Is there a place where all considered architectures are explained?

P15 In that sense, PRS-PGx-CT may take a little bit risk to misidentify a few SNPs.

P17 Why was this chosen for comparison here? It seems the other 5 were considered for the majority of the simulation study, but is there a reason to include this one here also? Or alternatively, why were the other methods excluded here?

P20 why were these three methods compared here (vs also including PRS-PGx-GL and PRS-PGx-Unadj)? Is this because the others require individual level data?

P23 Are there any times you would suggest one of the other methods over PRS-PGx-Bayes (e.g., considering other factors than just predictive ability/R2?). Would you recommend always using the PRS-PGx-Bayes approach?

P27 in which condition we say SNP j pops up after Bayesian shrinkage.

P27 but pop up SNPs with large effects.

P28 This seems like a separate topic, i.e., how you applied this method to the data instead of the method definition itself. Could this be moved either to the results section (e.g., where you discuss the simulation/real data analysis) or perhaps a subsection here on application of the method?

From reading further, maybe this could be useful, e.g., in the simulation section below (i.e., within the methods section)?

P28 It seems it may be useful to move this forward in the Methods section as it is the simplest method and can set some kind of backdrop for the other methods? I think it may be helpful to put this, the the CT method, penalization methods, then finally the Bayes method to kind of order in terms of simplicity?

P29 Do you need any detail as to how these quantities are estimated?

P31 Is there any concern that the simulation conditions seem to match with the assumptions in the Bayes method? Have you considered a misspecified model?

P32 what is meant by "1000 or 3000" here? When would one or the other be selected? Did sample size change between 2000 and 4000?

P32 Since this is the methods section, do you want to reiterate where the p-values are from?

P33 The GWAS summary statistics data were generated by running GWAS analysis with the model: $\log \mathbf Y\_1 - \log \mathbf Y\_0 = \beta\_0 + \beta\_{Y\_0}\log \mathbf Y\_0 + \beta\_T \mathbf T +  \mathbf G \upbeta +  \mathbf{(T\times G)} \upalpha +  \mathbf X \upgamma$ where $\mathbf Y\_1$ is the on-treatment LDL-C response, $\mathbf Y\_0$ is the baseline LDL-C response, $\upbeta$ is the prognostic effect, $\upalpha$ is the predictive effect, and the covariate matrix X included age, gender, prior lipid lowering therapy, early Acute Coronary Syndrome (ACS) trial, high risk ACS diagnosis, and the top five principal components estimated from the GWAS data which were used to adjust the population structure. The GWAS summary statistics data including both the prognostics and predictive effects ($\upbeta$ and $\upalpha$) was used for the PRS based drug response analyses.

P33 The effect sizes (after shrinkage) of the selected SNPs were then applied to the IMPROVE-IT data (as the testing set).

P33 what is meant by applying the effect sizes? Do you mean to apply the PRS model based on these effect sizes of the testing set?

P33 he GWAS summary statistics data weregenerated from the training set by running GWAS analysis with the model ??? Conflict with directly using summary statistics data???

P34 Should this be Merck since it is submitted in North America? (Is Nature Genetics headed in Europe?)

Supp P17 what does it mean when the reported trait is missing?