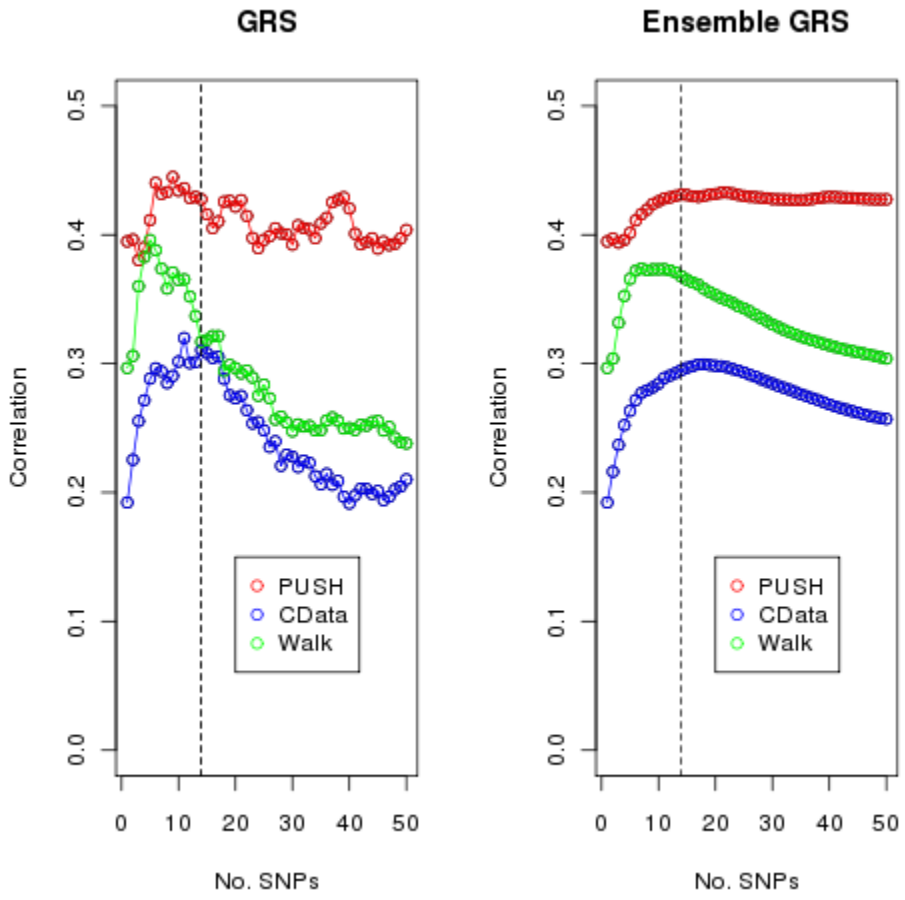
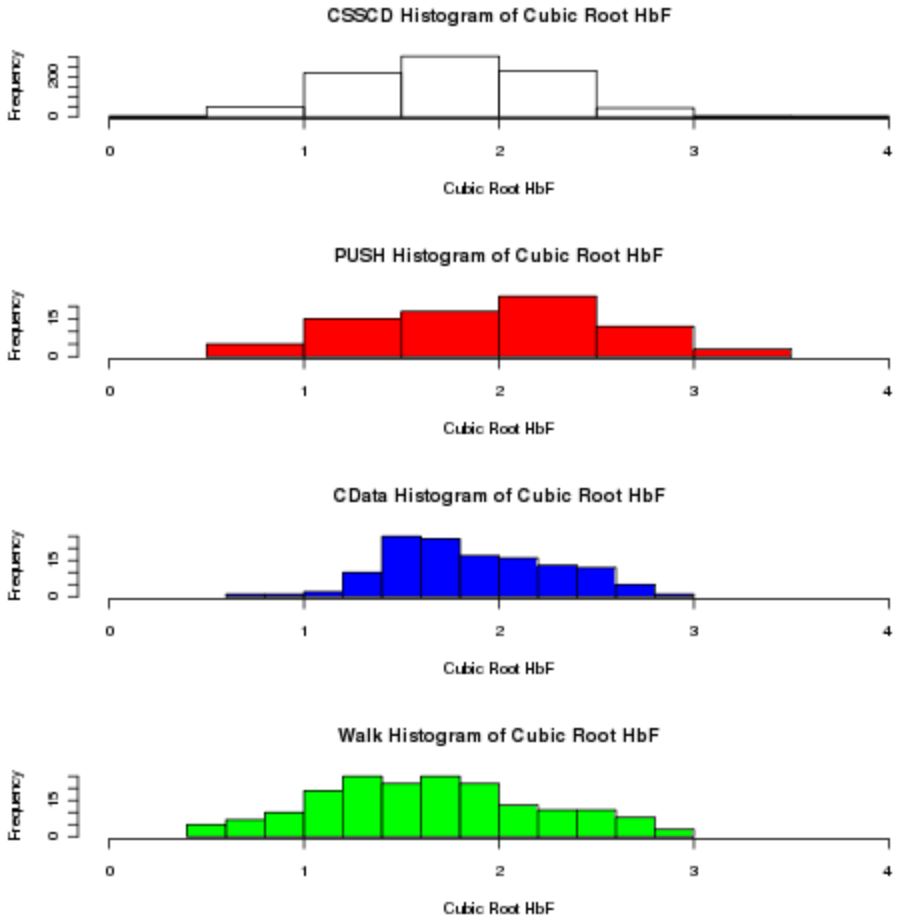
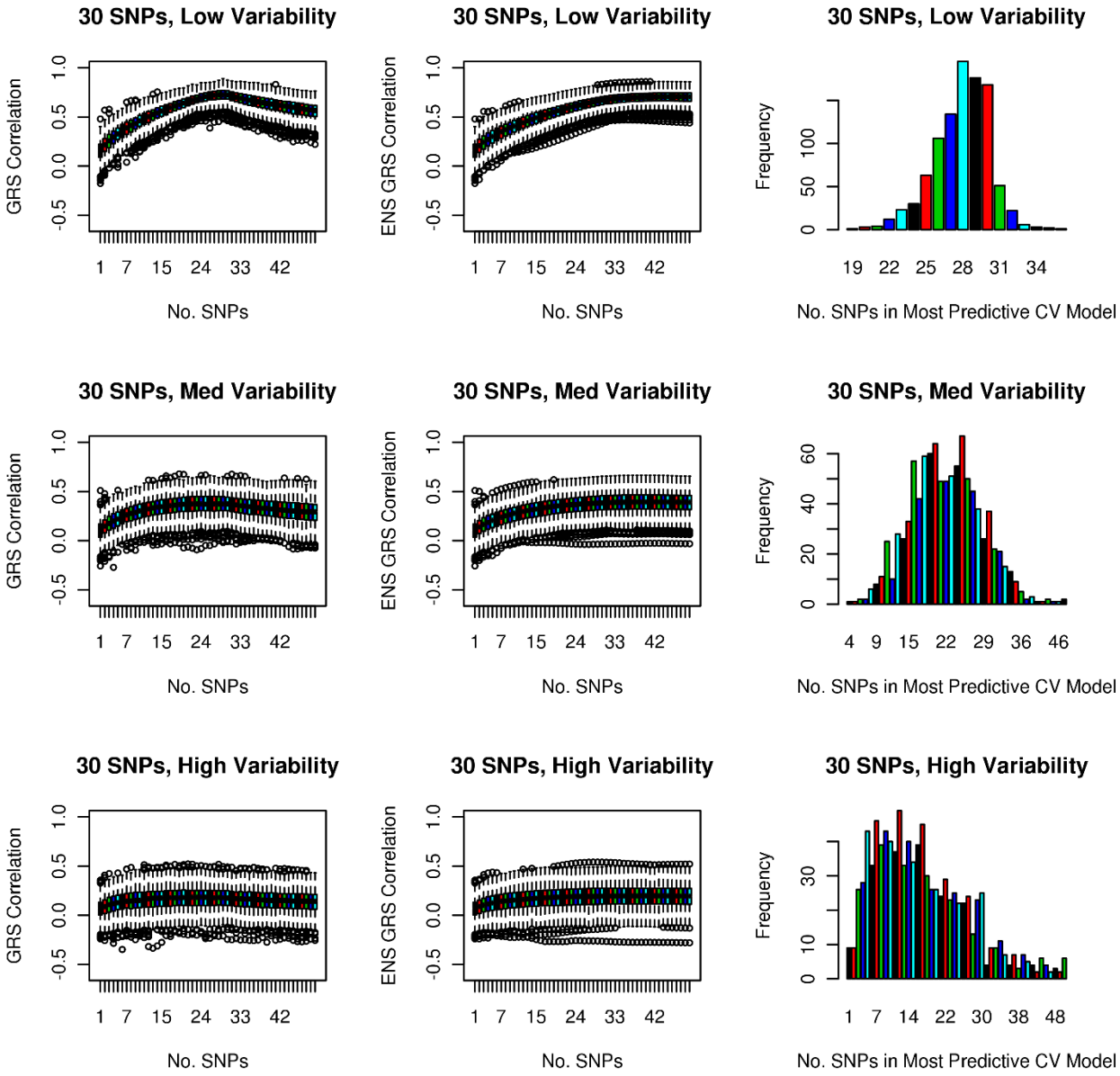
SUPPLEMENTAL MATERIAL



Supplementary Figure 1. Scatterplot of the correlation versus the Number of SNPs in the weighted GRS model. Plot of the correlation between the observed and predicted HbF in the three independent cohorts versus the number of SNPs in the top 50 weighted GRS and ensemble weighted GRS models.



Supplementary Figure 2. Histograms of the age distributions for the CSSCD, PUSH and Walk-PHaSST cohorts.



Supplement Figure 3. Side by side boxplots of the correlation between the observed and predicted phenotype (y-axis) versus the number of SNPs in the model (x-axis) for the GRS model (left panel) and ensemble of GRS models (middle panel). The right panel shows the distribution of the optimal number of SNPs chosen to maximize prediction using 10-fold cross validation. The results are from a simulation study with 1,000 simulated genotype data sets, each

with 1,000 individuals. In each dataset, a continuous phenotype with a heritability of 0.40 was simulated with 30 causal SNPs contributing equally to the genetic effect. The continuous phenotype was simulated to have a variability equal to 1/4 the mean of the phenotype (low variability, row 1), 1/2 the mean of the phenotype (medium variability, row 2) and 3/4 the mean of the phenotype (high variability, row 3). To evaluate the predictive accuracy, each set of 1,000 individuals was randomly split into a training set (N=900) and test set (N=100), and each training set was used to build the GRS and the ensemble of GRS models that were tested in the test set. When 10-fold cross-validation was used, each dataset of 1,000 individuals was randomly partitioned into 10 equal parts where each part was used as test set of the models trained in the other 9 parts, and the results were averaged over the 10 folds. The simulation study shows that for both the GRS and ensemble GRS methods the correlation peaks at 30 SNPs; however, when more than 30 SNPs are included in the GRS model, the correlation rapidly decreases while the decrease in correlation is more gradual for the ensemble of GRS models. The results suggest that if one were to incorrectly choose the wrong model, the ensemble of GRS models would result in smaller loss of prediction accuracy than a single GRS model. In addition, selection of the optimal number of SNPs using cross-validation may be too conservative and produce sub-optimal models. As expected, the prediction accuracy decreases as the variability of the phenotype increases.