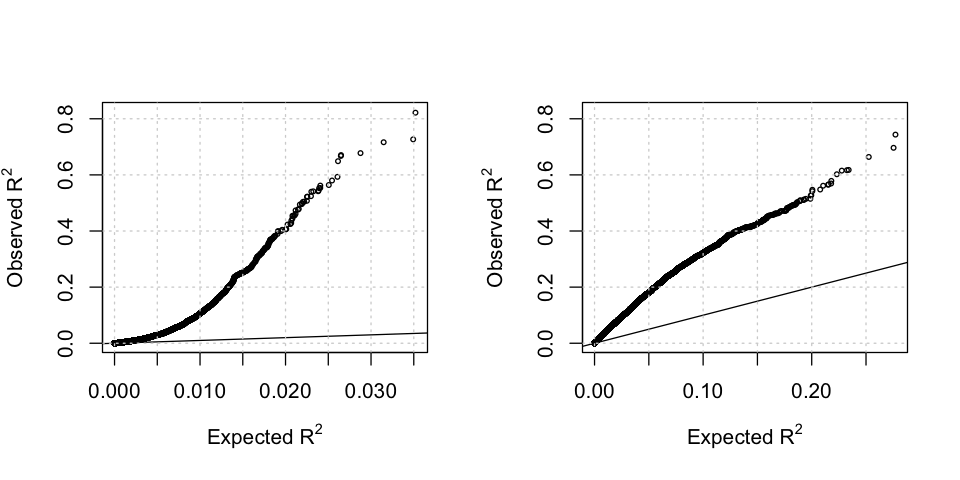
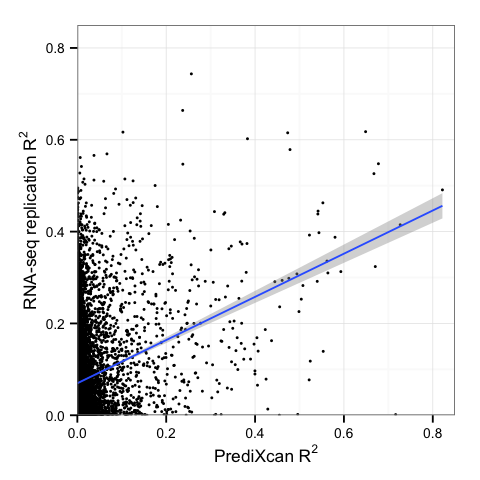
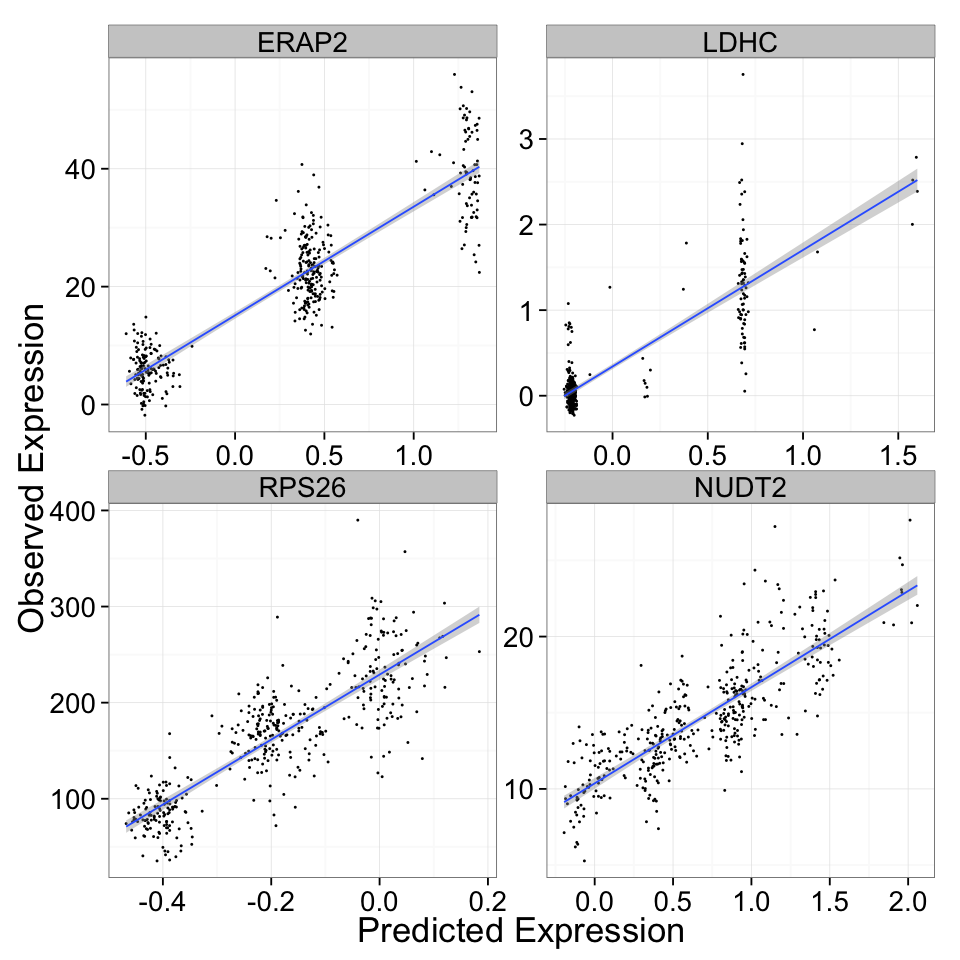


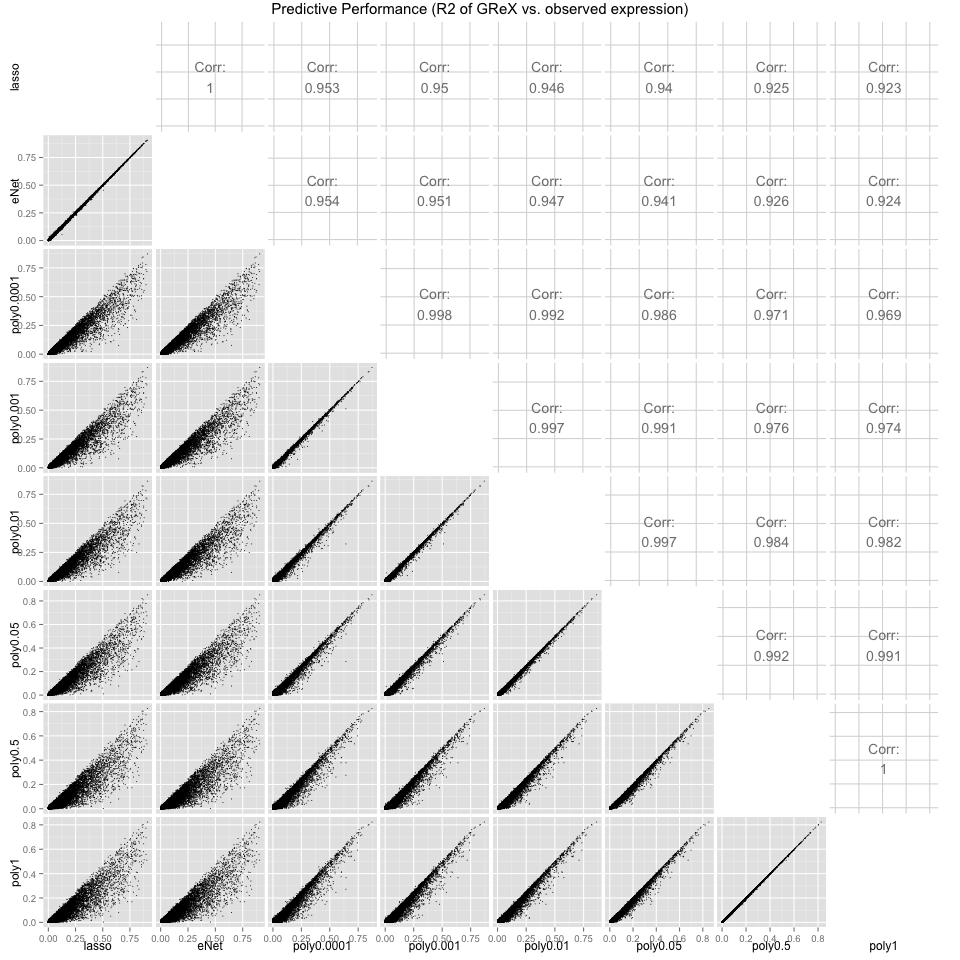
Figure y. PrediXcan performance (R2 of GReX vs. observed expression in red) compared to gene expression heritability estimates (black with 95% confidence interval in gray). Performance was assessed using 10-fold cross-validation in the DGN whole blood cohort (n=922) with LASSO and Polygenic score (p<1e-04).

**Figure 3. Prediction performance of elastic net and replication correlation.** Using whole blood prediction models trained in DGN, we compared predicted levels of expression with observed levels on lymphoblastoid cell lines from the 1000 Genomes project. RNA-sequenced data (n=421) on these cell lines have been made publicly available by the GEUVADIS consortium. Left panel shows the squared correlation, , between predicted and observed levels plotted against the null distribution of . On a subset of the cell lines (n=47) RNA-sequencing data from two different experiments were available (Lappalainen 2013, Pickrell 2010). Right panel shows of the observed gene expression levels between the two replicated experiments, which is a benchmark for the maximum performance expected from gene expression prediction models.



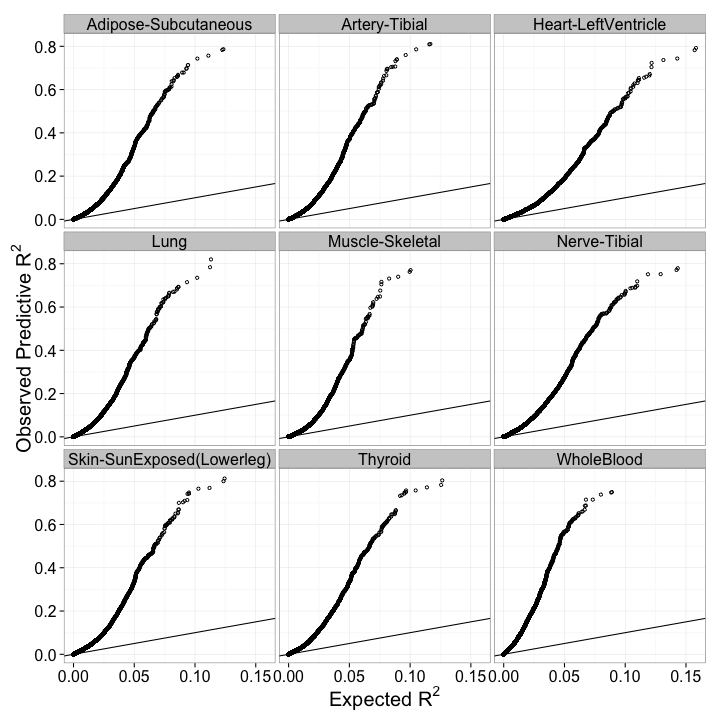


**Figure 4. Examples of well-predicted genes.** These plots show observed vs. predicted levels of 4 genes. Predicted levels were computed using whole blood elastic net prediction models trained in DGN data. Observed levels were RNA-seq data in lymphoblastoid cell lines generated by the GEUVADIS consortium.



Supplemental Figure X. Comparison of 10-fold cross-validated predictive performance

between all tested methods (lasso, elastic net with alpha=0.5, polygenic score at several p-value thresholds) in the DGN whole blood cohort. Predictive performance was measured by the R2 between predicted (GReX) and observed expression.



**Supplemental Figure X. Prediction performance of elastic net in GTEx tissues.** Using whole blood prediction models trained in DGN, we compared predicted levels of expression with observed levels from nine tissues of the GTEx pilot project. The observed squared correlation between predicted and observed gene expression levels, , is plotted against the null distribution of .