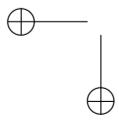
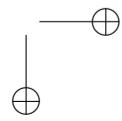
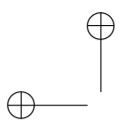
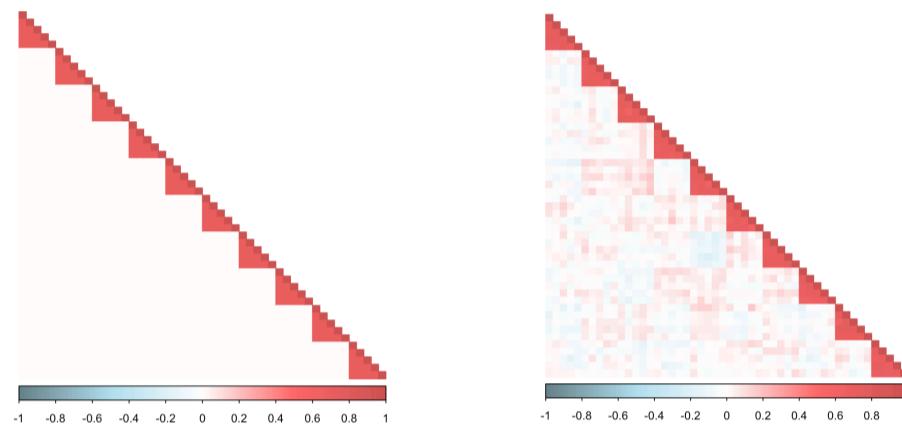
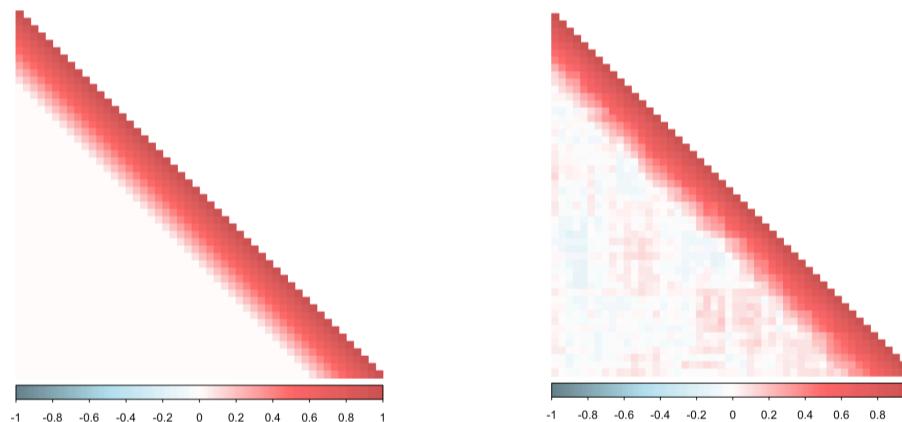
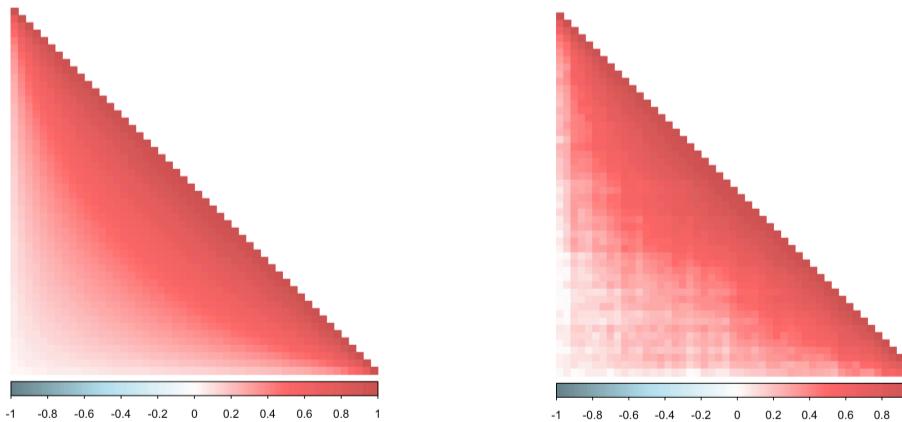


a

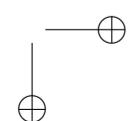
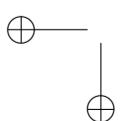
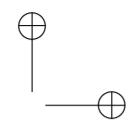
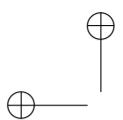
1

**Supplementary Figures**



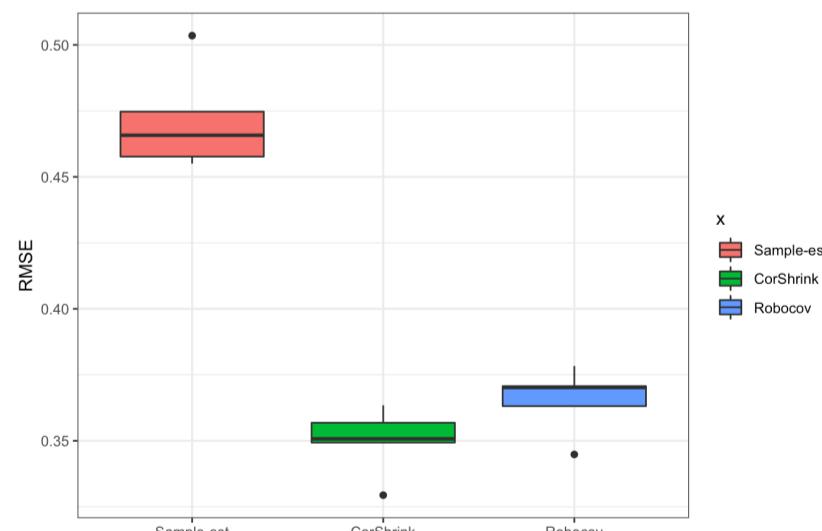
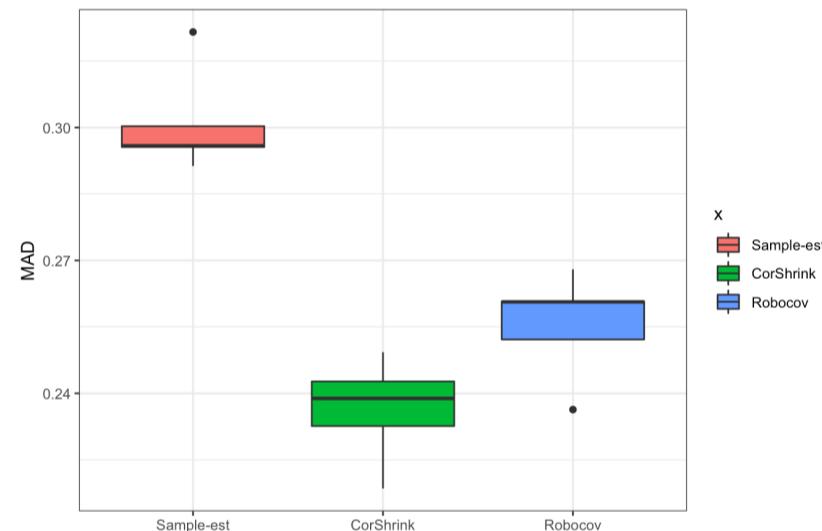
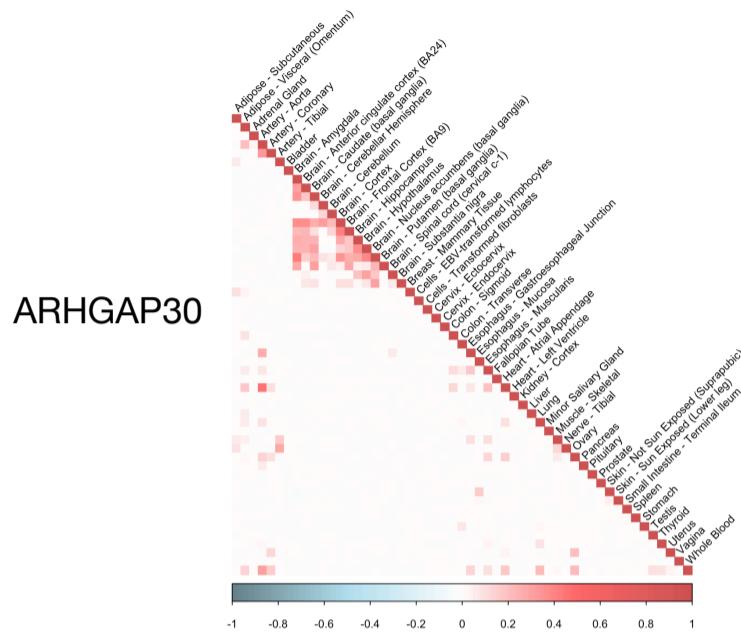
**Population correlation****Standard Pairwise Correlation****Hub****Toeplitz****1-band precision**

**Fig. S1.** We applied standard pairwise correlation estimator on data generated from the simulation models from Figure 1—this comprises of Hub, Toeplitz or 1-band precision matrix-based population models with  $N = 500$  samples,  $P = 50$  features and  $\pi = 50\%$  proportion of missing data.

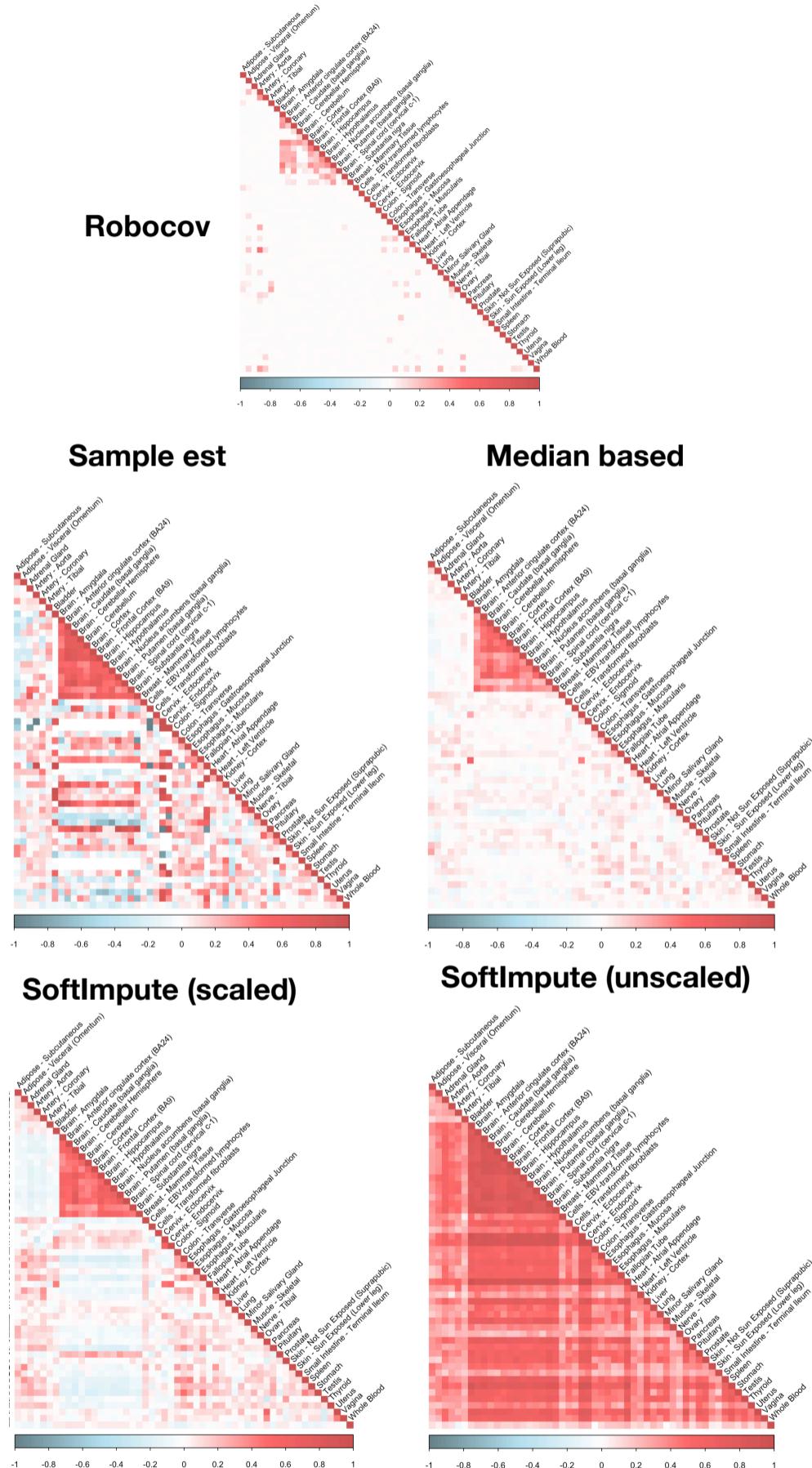


a

3



**Fig. S2.** (Top panel) Robocov correlation estimate of the ARHGAP30 gene. (Middle and bottom panels) We split the data matrix randomly into 2 equal groups. We compare the Robcov, CorShrink and pairwise sample correlation estimators from one half of the data with the pairwise sample correlation matrix on the other half. We use Median Absolute Deviation (MAD) (middle panel) and Root Mean Squared Error (RMSE) (lower panel) metrics. The results are averaged over 50 such random splits. See Table S1 for a numerical summary.

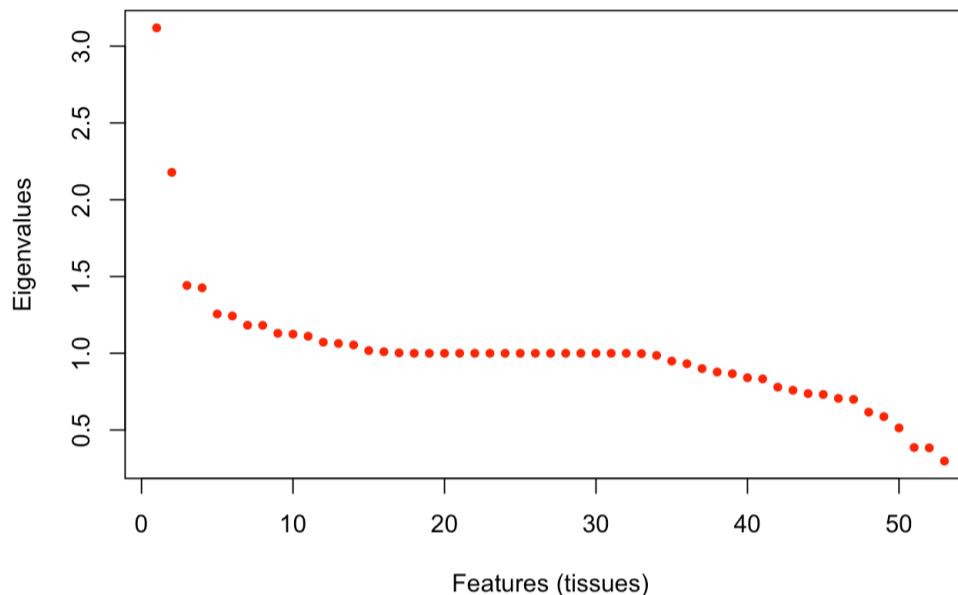


**Fig. S3.** We compare the Robocov correlation estimator for the ARHGAP30 gene with four other estimators. They include the standard pairwise sample correlation estimator, the sample correlation matrix computed over data imputed by either a median-based approach (missing entries of a feature replaced by the median of observed entries), the scaled SoftImpute[7] approach; and an unscaled SoftImpute[7] approach.

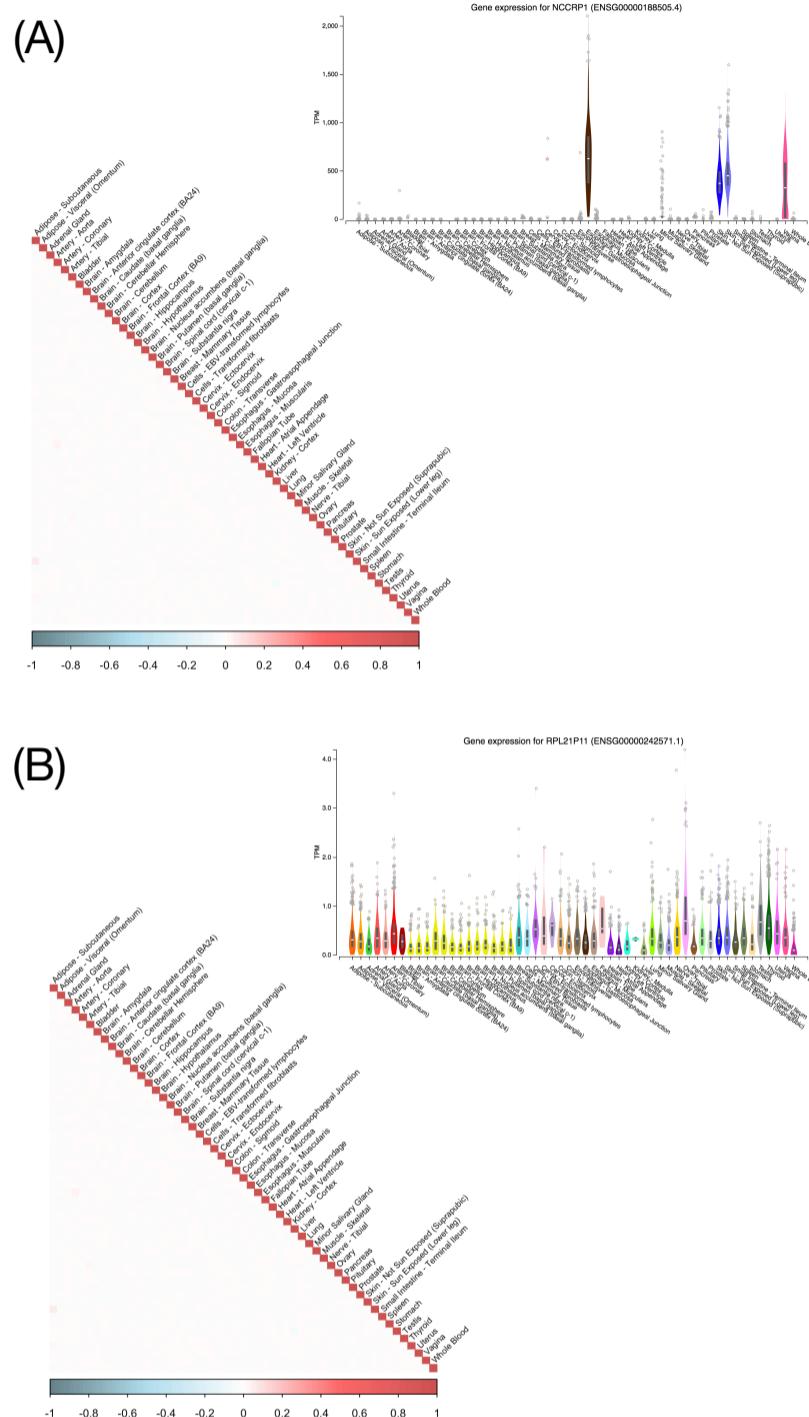
a

5

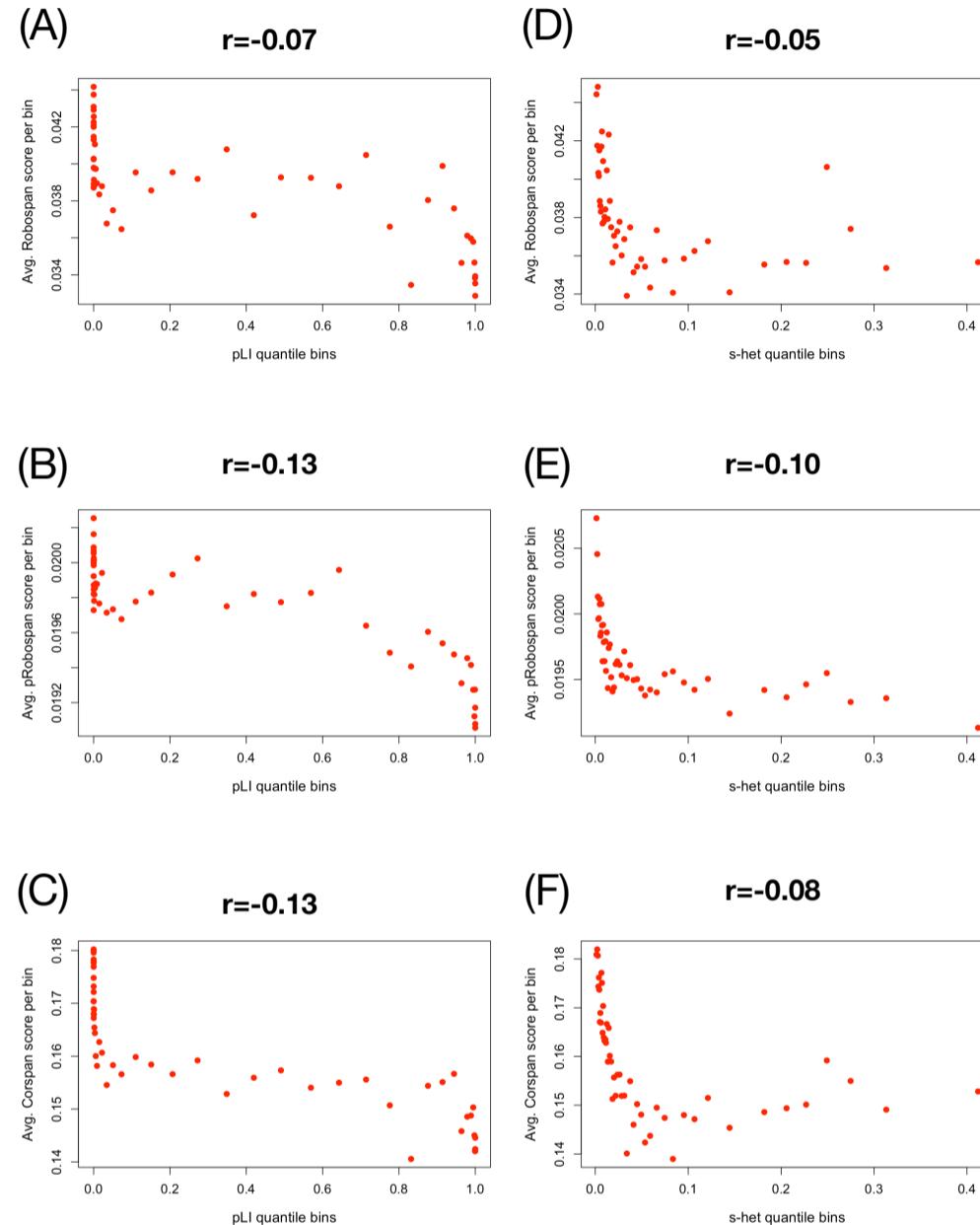
### Eigenvalue trend for ARHGAP30 gene



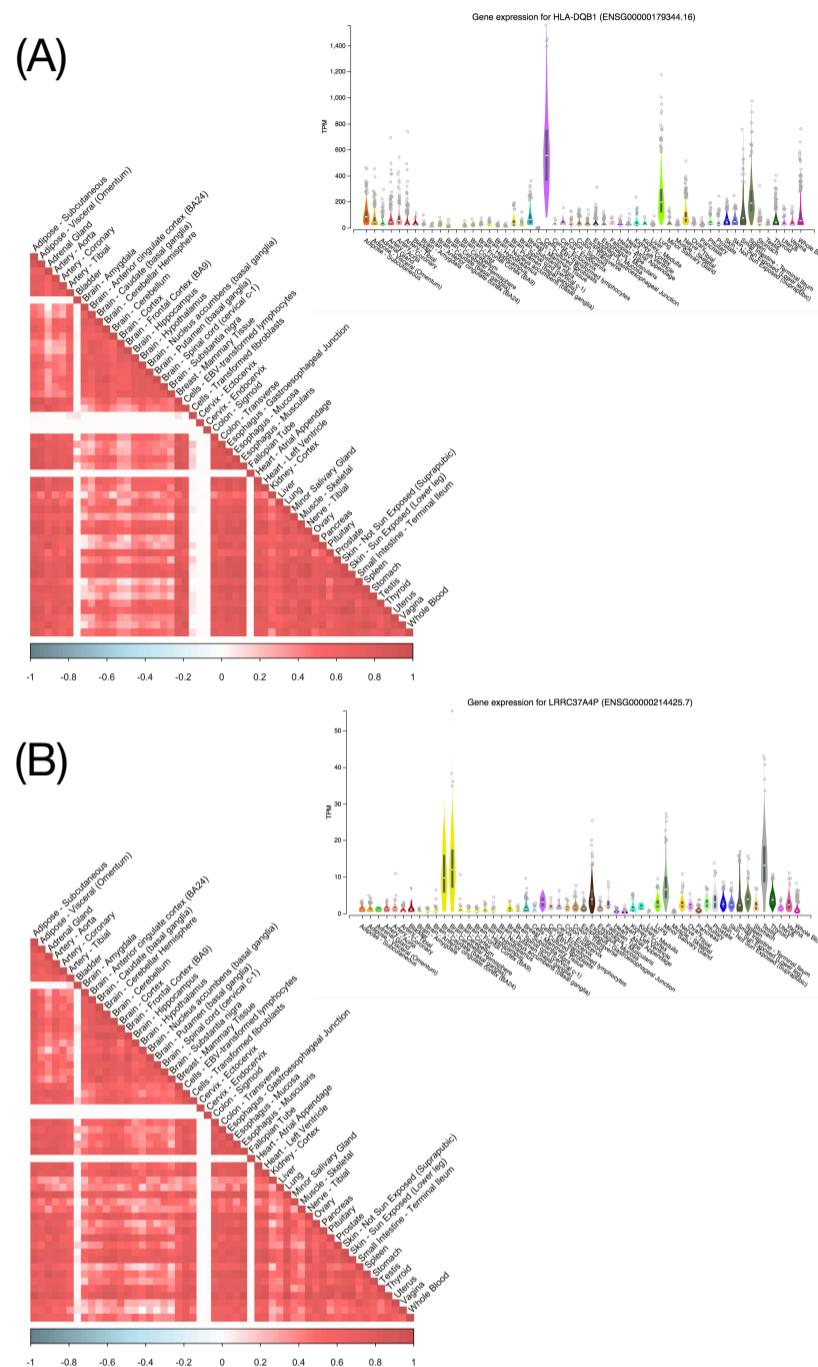
**Fig. S4.** Plot of eigenvalues sorted from highest to lowest in magnitude for tissue-tissue pairwise correlation matrix for a particular gene (ARHGAP30). The eigenvalues do not show any sharp drop close to 0 as one would expect if the matrix allowed a low rank (+noise) structure. This suggests relatively high dimensional structure in the GTEx gene expression data which may explain why a low rank imputation method such as SoftImpute[7] performs poorly in S3.



**Fig. S5.** Examples of two genes, NCCR1 (top) and RPL21P11 (bottom), both of which have close to 0 average correlation in expression across tissue-pairs but having very distinctive expression profiles. NCCR1 has high expression in a few specific tissues including Whole Blood, while RPL21P11 has uniformly low expression across all tissues. The expression profile plots for the genes have been fetched from the GTEx Portal (<https://gtexportal.org/home/>).



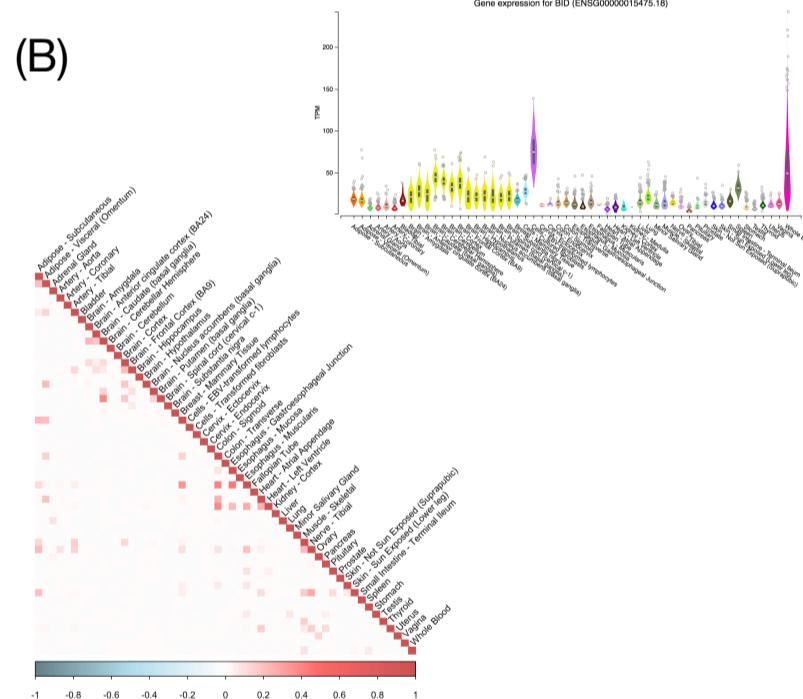
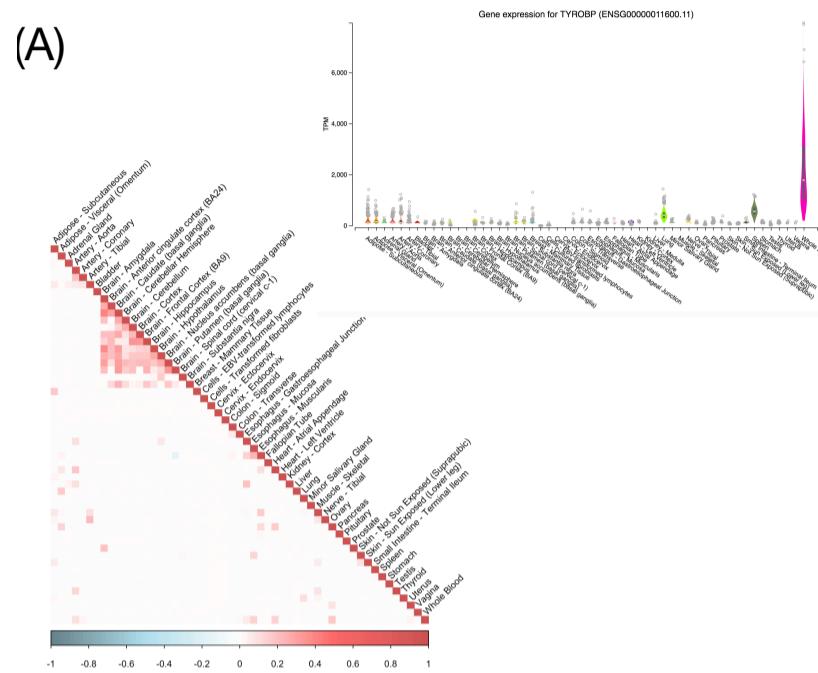
**Fig. S6.** Comparison of pLI gene score with (A) Robospan-score, (B) pRobospan-score and (C) Corspan-score for all genes (See Results section for details). Comparison of s\_het gene score with (A) Robospan-score, (B) pRobospan-score and (C) Corspan-score for all genes (See Results section for details).



**Fig. S7.** Examples of genes high Robospan-score but not specifically expressed in blood or uniformly expressed across tissues. The two genes are HLA-DQB1 (top) and LRRC37A4P (bottom). HLA-DQB1 is specifically expressed in lymphocyte cell line which is related to blood. LRRC37A4P has highest expression in brain cerebellum and testis. The expression profile plots for the genes have been fetched from the GTEx Portal (<https://gtexportal.org/home/>).

a

9



**Fig. S8.** Examples of genes that are specifically expressed in Whole Blood but do not show high Robospan-score. The two genes are TYROBP (top) and BID (bottom). They show tissue specific expression in Whole blood and are in top 10% specifically expressed genes (SEG) as per ref.[29], but do not show consistently high expression correlation across tissue pairs and hence, do not have high Robospan score. The expression profile plots for the genes have been fetched from the GTEx Portal (<https://gtexportal.org/home/>).