The landscape of gene expression variation in human populations

Intro

Gene expression variation drives phenotypic difference.

Here we use public gene expression data sets to evaluate how the differences in gene expression variation is structures across several independent studies. We collected and compared the gene expression variation across several studies, and used the similarities across these studies to create a gene expression variation ranking, which orders genes from least variable to most variable. We then explore the expected drivers of this gene expression ranking, showing that both cis and trans regulation are involved with the determination of gene expression variance.

Methods

Data sources

We selected 60 studies with large sample sizes from public gene expression repositories recount3 (Wilks et al. 2021) and Expression Atlas (Papatheodorou et al. 2020).

Data processing

Filtering by min cpm and mean cpm. Fixed effects correction. Outlier removal using (Chen et al. 2020).

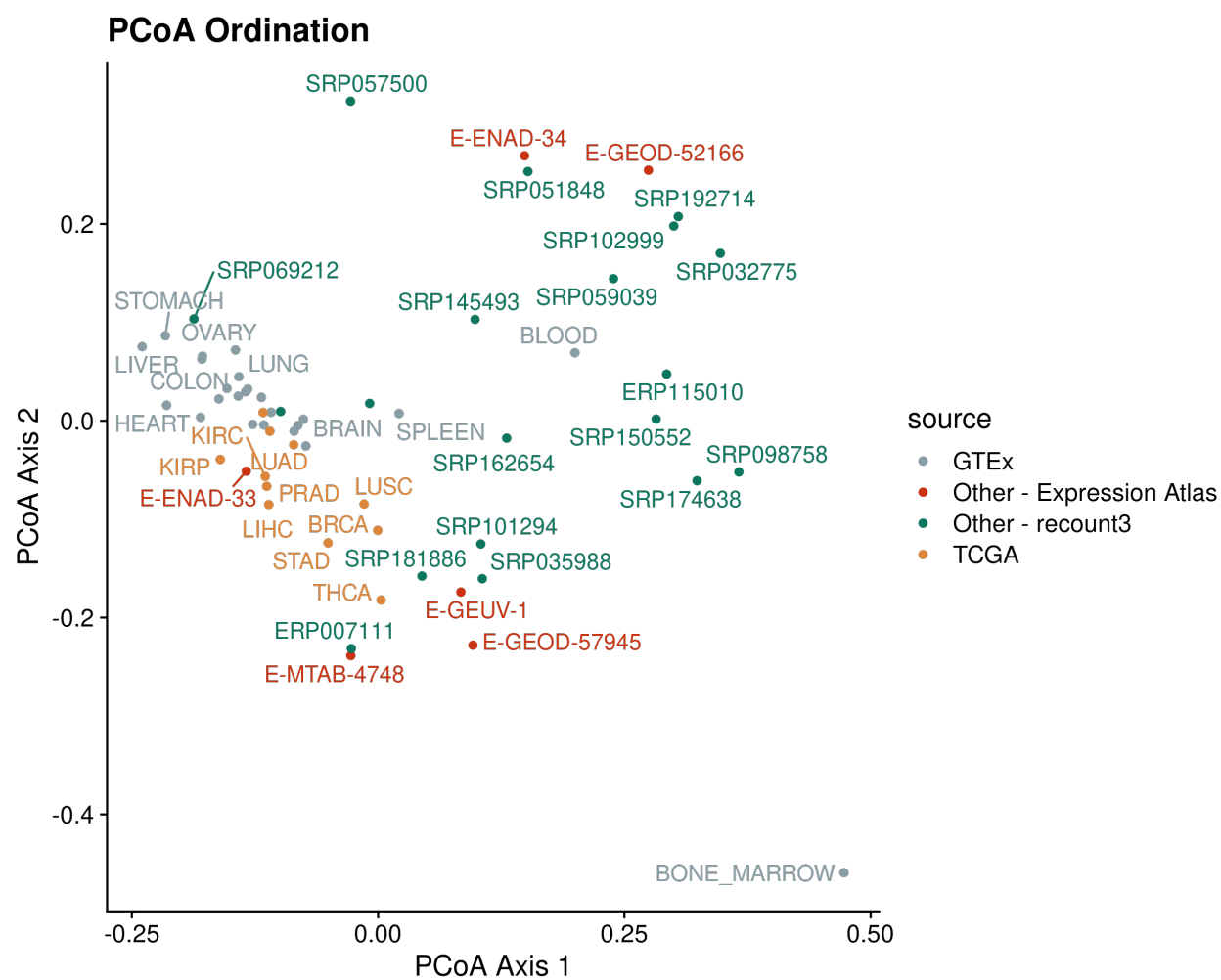Variance stabilizing transformation from DESeq2 (Love et al. 2014)

Results

Figure 1: Standard deviation correlation PCoA

Discussion

Gene expression variance is reasonably conserved across studies. Gene expression variance is predictive of biological function. Gene expression variance can be partially explained by genetic variation and genetic associations between gene expression.

References

Chen, X., B. Zhang, T. Wang, A. Bonni, and G. Zhao. 2020. Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. BMC Bioinformatics 21:269.

Love, M., S. Anders, and W. Huber. 2014. Differential analysis of count data–the DESeq2 package. Genome Biol. 15:10–1186.

Papatheodorou, I., P. Moreno, J. Manning, A. M.-P. Fuentes, N. George, S. Fexova, N. A. Fonseca, A. Füllgrabe, M. Green, N. Huang, L. Huerta, H. Iqbal, M. Jianu, S. Mohammed, L. Zhao, A. F. Jarnuczak, S. Jupp, J. Marioni, K. Meyer, R. Petryszak, C. A. Prada Medina, C. Talavera-López, S. Teichmann, J. A. Vizcaino, and A. Brazma. 2020. Expression atlas update: From tissues to single cells. Nucleic Acids Res. 48:D77–D83.

Wilks, C., S. C. Zheng, F. Y. Chen, R. Charles, B. Solomon, J. P. Ling, E. L. Imada, D. Zhang, L. Joseph, J. T. Leek, A. E. Jaffe, A. Nellore, L. Collado-Torres, K. D. Hansen, and B. Langmead. 2021. recount3: Summaries and queries for large-scale RNA-seq expression and splicing.