

Preview

Clinical traits impacting human tissue transcriptomes

Kaur Alasoo^{1,*}¹Institute of Computer Science, University of Tartu, Tartu, Estonia*Correspondence: kaur.alasoo@ut.ee<https://doi.org/10.1016/j.xgen.2022.100245>

In this issue of *Cell Genomics*, Garcia-Perez et al.¹ report a comprehensive and careful association analysis between gene expression and splicing measured by the GTEx Consortium² in 46 human tissues and 21 demographic and clinical traits.

In this issue of *Cell Genomics*, the study by Garcia-Perez et al.¹ expands on previous work in two major ways. While previous studies have typically considered a single tissue^{3–5} (e.g., blood or brain) or a small number of tissues,⁶ the work by Garcia-Perez et al. investigated 46 different tissues, allowing them to systematically characterize which genes are affected by common demographic features (age, sex, ancestry, body mass index [BMI]) and to determine whether these effects are broadly shared or tissue specific. They found that ancestry, age, sex, and BMI had mostly additive and tissue-specific effects on gene expression, while interactions were rare. Overall, the expression level of most genes was associated with at least one demographic feature in at least one tissue (Figure 1A). Although ancestry had similar effects on splicing to gene expression (both largely explained by *cis*-acting genetic variants), the effects of sex and BMI on splicing were much more limited (Figure 1B), suggesting that largely independent regulatory mechanisms control gene expression levels and splicing. Secondly, the authors expanded their analysis to 17 additional clinical traits extracted from either medical history (diabetes) or donor-matched histopathology images (Figure 1C). The most interesting association was observed between type 1 and type 2 diabetes and gene expression changes in the nerve tissues that correspond to diabetic neuropathy. This association was further validated using histopathology imaging. Together with a previous study focusing on the effect of BMI on gene expression patterns in the adipose tissue,³ these studies highlight

the value of having matched histopathology images together with tissue gene expression samples.

An important limitation of this study, and many of the previous studies, is that while associations with splicing patterns are easy to detect, they are much more difficult to interpret. This is because reference transcriptome annotations are often incomplete and alternative transcripts or splicing events are strongly correlated with each other. As a result, a high expression estimate for a given transcript does not imply that all exons assigned to that transcript are actually expressed in the sample.⁷ Consequently, the observation that ribosomal genes are often differentially spliced between individuals of European and African American ancestries and that these differences are driven by *cis*-acting genetic variants is an interesting one, but it is unclear whether these changes have any discernable functional impact. Future studies based on long-read RNA sequencing combined with better computational models of protein structure and function will hopefully help to address these issues.⁸

This study neatly highlights the value of combining large transcriptomic datasets with rich phenotyping. However, this is very challenging to achieve at scale, because large biobanks such as the UK Biobank, FinnGen, and other resources focus on recruiting healthy volunteers. This means that they can assemble large collections of phenotypes via questionnaires and linking to registries but are limited by the type of samples that they can obtain from those participants, which is typically frozen whole blood. In contrast, studies like GTEx² have focused

on postmortem samples to maximize the number of tissues that can be accessed, but this typically comes at the cost of having access to a much smaller set of organismal phenotypes. Thus, achieving good coverage of both disease phenotypes and tissue samples is likely to remain an ongoing logistical challenge. Most large-scale transcriptomic datasets profiling tissues other than whole blood lack high-resolution phenotyping. A comparable study is perhaps the TwinsUK that has extremely thorough phenotypic information as well as transcriptomic data from four tissues (blood, skin, adipose, and lymphoblastoid cell lines).^{3,6}

The observed association between diabetes and diabetic neuropathy further reaffirms how associations between gene expression levels and disease phenotypes often reflect consequences of disease processes rather than their causes. In a related study, Porcu et al. used a novel design relying on shared genetic associations to link gene expression levels measured in the eQTLGen Consortium⁹ to complex traits measured in completely unrelated cohorts.¹⁰ This Mendelian randomization approach demonstrated that, in whole blood, most observed correlations between gene expression levels and complex traits are consequences of disease processes rather than their cause. Nevertheless, these associations can yield early biomarkers of the disease, provide important insights about disease progression, and reveal whether and how different tissues are affected by the disease.

Because of its comprehensive design and careful execution, the study by Garcia-Perez et al. is likely to remain an



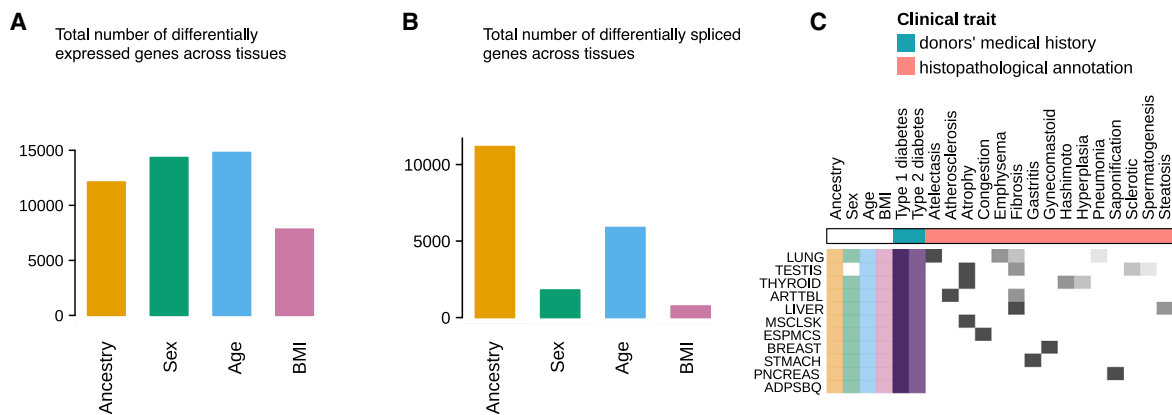


Figure 1. Effects of demographic features and clinical traits on gene expression and splicing levels

(A) Number of unique differentially expressed genes associated with each of the four demographic features in at least one tissue.

(B) Number of unique differentially spliced genes associated with each demographic feature in at least one tissue.

(C) Overview of clinical traits studied and affected tissues.

Adapted from Garcia-Perez et al.¹

important reference for future disease-transcriptome association studies for many years to come. I expect future single-cell and spatial transcriptomic studies, combined with histopathology imaging, to further clarify the molecular and cellular mechanisms of disease-associated changes.³ However, these studies are, at least initially, still likely to focus on easily accessible tissues such as blood. Matching the tissue coverage of GTEx with the phenotypic complexity of large biobanks is likely to remain a major logistical challenge. It is always worth pausing and reflecting on whether the approaches and methodologies that have taken us here are also those that will take us forward. Perhaps there are clever alternatives, like the Mendelian randomization approach proposed by Porcu et al.,¹⁰ that can achieve similar goals with a smaller logistical footprint. The beauty of science is that we do not have these answers yet.

DECLARATION OF INTERESTS

The author declares no competing interests.

REFERENCES

1. Garcia-Perez, R., Ramirez, J.M., Ripoll-Cladellas, A., Chazarra-Gil, R., Oliveros, W., Soldat-

kina, O., Bosio, M., Rognon, P.J., Capella, S., Calvo, M., et al. (2023). The landscape of expression and alternative splicing variation across human traits. *Cell Genomics* 3, 100244. <https://doi.org/10.1016/j.xgen.2022.100244>.

2. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. <https://doi.org/10.1126/science.aaz1776>.

3. Glastonbury, C.A., Couto Alves, A., El-Sayed Moustafa, J.S., and Small, K.S. (2019). Cell-type heterogeneity in adipose tissue is associated with complex traits and reveals disease-relevant cell-specific eQTLs. *Am. J. Hum. Genet.* 104, 1013–1024. <https://doi.org/10.1016/j.ajhg.2019.03.025>.

4. Trabzuni, D., Ramasamy, A., Imran, S., Walker, R., Smith, C., Weale, M.E., Hardy, J., and Ryten, M.; North American Brain Expression Consortium (2013). Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* 4, 2771. <https://doi.org/10.1038/ncomms3771>.

5. Piasecka, B., Duffy, D., Urrutia, A., Quach, H., Patin, E., Posseme, C., Bergstedt, J., Charbit, B., Rouilly, V., MacPherson, C.R., et al.; Milieu Intérieur Consortium (2018). Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. *Proc. Natl. Acad. Sci. USA.* 115, E488–E497. <https://doi.org/10.1073/pnas.1714765115>.

6. Glass, D., Viñuela, A., Davies, M.N., Ramasamy, A., Parts, L., Knowles, D., Brown, A.A.,

Hedman, A.K., Small, K.S., Buil, A., et al.; UK Brain Expression consortium; MuTHER consortium (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* 14, R75. <https://doi.org/10.1186/gb-2013-14-7-r75>.

7. Alasoo, K., Rodrigues, J., Danesh, J., Freitag, D.F., Paul, D.S., and Gaffney, D.J. (2019). Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *Elife* 8, e41673. <https://doi.org/10.7554/elife.41673>.

8. Glinos, D.A., Garborcauskas, G., Hoffman, P., Ehsan, N., Jiang, L., Gokden, A., Dai, X., Aguet, F., Brown, K.L., Garimella, K., et al. (2022). Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* 608, 353–359. <https://doi.org/10.1038/s41586-022-05035-y>.

9. Vösa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al.; BIOS Consortium; i2QTL Consortium (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* 53, 1300–1310. <https://doi.org/10.1038/s41588-021-00913-z>.

10. Porcu, E., Sadler, M.C., Lepik, K., Auwerx, C., Wood, A.R., Weihs, A., Sleiman, M.S.B., Ribeiro, D.M., Bandinelli, S., Tanaka, T., et al. (2021). Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat. Commun.* 12, 5647–5649. <https://doi.org/10.1038/s41467-021-25805-y>.