Cell Genomics Transparent Peer Review Record

The landscape of expression and alternative splicing variation across human traits

Author list

Raquel Garcia-Perez, Jose Miguel Ramirez, Aida Ripoll-Cladellas, Ruben Chazarra-Gil, Winona Oliveros, Oleksandra Soldatkina, Mattia Bosio, Paul Joris Rognon, Salvador Capella, Miquel Calvo, Ferran Reverter, Roderic Guigó, François Aguet, Pedro G. Ferreira, Kristin G. Ardlie & Marta Melé

Summary

Initial submission:	Received : July 11th 2022
	Scientific editor: Judith Nicholson
First round of review:	Number of reviewers: 3 Revision invited : October 18th 2022 Revision received : November 9th 2022
Second round of review:	Number of reviewers: N/A Accepted : 7th December 2022
Data freely available:	Yes
Code freely available:	Yes

Referees' reports, first round of review

Reviewer 1

Garcia-Perez et. al. have implemented a statistical framework to assess the contribution of multiple population traits and clinical traits to the gene expression and alternative splicing variation based on the GTEx data. Extensive tests are done, and the results cover many aspects, including the extent of DEG sharing, the contribution of cis-eQTL and cis-sQTL, the additive contribution and trait interaction, and the association between traits and cell type composition.

I believe this work is carried out carefully. I just have one major concern in terms of phrasing. The impact or effect of traits on transcriptome is often mentioned, for example, "we took advantage of the GTEx multi-tissue data to assess how type 1 and 2 diabetes affect the transcriptome of multiple tissues", while the tests are for association. It might be better to rephrasing it to avoid possibly giving the reader impression that it is about causation.

One minor point is, from the method part, it seems that two PEER factors are used. It might be better to add some explanation on how the number of PEER factors is decided.

Another minor point is, on the top of page 8, in the sentence:
"DEGs between populations and between sexes are the most tissue- shared (Fig. 3e and Supplementary Fig. 3f),"
Fig. 3e doesn't seem to illustrate the point of this sentence. Supplementary Fig. 3e is the correct one?

Reviewer 2

In this paper, the authors have done an excellent job in systematically characterising how changes in gene expression and splicing correlate with major demographic features (age, sex, ancestry, BMI) as well as number of other traits and diseases. The authors replicate a number of previous findings (mostly from

studies conducted with whole blood samples) and expand their analysis to the most comprehensive collection of cell types and tissues available from the GTEx Consortium. Importantly, they also identify a novel and interesting association between diabetes and gene expression changes in the nerve tissues that correspond to diabetic neuropathy. The paper is well written and all of the analyses have been performed extremely thoroughly and carefully. I don't have any major concerns.

My only minor concern is that inferring percentage-spliced-in (PSI) values from transcript expression estimates (RSEM TPMs) carries a small risk of detecting false positive splicing events. This is caused by the fact that existing transcript annotations often couple together different splicing events that are likely to be regulated independently. For more detailed examples, see Supplemental Figure 1 in the Whippet preprint (https://doi.org/10.1101/158519) and Figure 1 in Alasoo et al, 2019 (https://doi.org/10.7554/eLife.41673). To be clear, I believe that this issues will only add some noise to your quantification results and will not actually invalidate any of your major conclusions which are based on aggregating signal across many independent splicing events. Furthermore, it is not obvious that using any of the alternative approaches will actually give you more accurate PSI estimates for the splicing events of interest. Thus, I think that is perfectly ok for your to keep using the SUPPA2 PSI estimates. However, I think it would be helpful to validate some of the the splicing associations that you see (e.g. the ribosomal genes RPLP2 and RPL10) using complementary approaches such as Leafcutter junction read counts (should be available from GTEx) or read coverage plots (e.g. from ggcoverage, Sashimi Plot, wiggleplotr or others).

Kaur Alasoo

Reviewer 3

Mele et al implemented a statistical framework to assess the contribution of multiple demographic and clinical traits on transcriptome and alternative splicing in human tissues using Genotype Tissue Expression (GTEx) data. Mele et al suggest that type 1 and type 2 diabetes contribute to multiple tissue transcriptome variation with the strongest signal for tibial nerve.
The authors should be commended for their approach to study a large array of human tissues. The study design is well-thought, and the analyses are a tour de force. The disease/diabetes results and discussion sections seem less strong and

do not fully support the conclusions (see below).

Major comments:

The readability of the figures is unfortunately limited by the abbreviations used for the different tissues (e.g. MSCLSK, ESPMCS, etc, see Fig. 1a and throughout) and ancestry (see Fig. 2b). The paper would be much easier to read if the tissues were readily identifiable. Why was kidney not included (see Suppl Fig 1a)?

Type 1 and 2 diabetes are characterized by dysregulation of blood glucose levels. Pancreatic islets and in particular pancreatic beta cells play a key role in the development and progression of diabetes. Pancreatic islets make up few (around 2) percent of pancreas weight. Not surprisingly then, transcriptomic and related association studies have shown that pancreas is not a good proxy for islets and fails to capture islet-specific disease signatures (Alonso et al., 2021; Vinuela et al., 2020). The lack of a tissue that is key to diabetes pathogenesis (i.e. pancreatic islets) in GTEx and the fact that pancreas is not a good proxy thereof are limitations that need to be clearly spelled out. Mele et al could consider adding transcriptome studies in pancreatic islets, which would strengthen their diabetes-related claims. Hundreds of islet transcriptomes are available, see (Fadista et al., 2014; Marselli et al., 2020; van de Bunt et al., 2015; Vinuela et al., 2020) among others.

Taking the above into account, the approach and analyses taken by Mele et al are rather poised to detect consequences of type 1 and type 2 diabetes, i.e. chronic micro- and macrovascular complications caused by long-term exposure to hyperglycemia. Again, this does not clearly appear in the manuscript and should be mentioned. The finding that a strong transcriptomic overlap for type 1 and type 2 diabetes is found in tibial nerve is of interest, and it begs the question why other chronic micro- and macrovascular complications are, apparently, not shared in the tissue transcriptomes (heart, arteries, kidney, etc). This deserves further data analysis and should be discussed. The GTEx data for tibial nerve have been used in previous work on the impact of gender, diabetes etc on transcriptomes (Ray et al., 2019). The existing literature should be acknowledged, the present data compared against it and novelty claims toned down accordingly.

As the number of type 1 diabetes samples is a lot smaller than that of type 2

diabetes, the statistical power for the former is much lower. A threshold free method like Rank-Rank Hypergeometric Overlap (RRHO) is probably more suitable to detect similarities. The authors should consider using RRHO tools for these comparisons, or at least mention the limitation related to sample numbers.

In the star methods, a "Differential gene expression analysis with demographic traits" is included but no section about clinical/phenotypic traits. Did the authors consider confounding ancestry and demographic factors (BMI, age, …) when analyzing type 1 and type 2 diabetes vs controls? This is important as ancestry contributes significantly to pancreas transcriptome variation (Fig 1a).

Additional comments

p.11: "found 79 and 309 DEGs in two or more tissues with either type 1 or type 2 diabetes". Is this correct? We counted 387 lines and 378 distinct genes in Suppl Table 6f.

The finding that INS, IAPP and MAFA are markedly differentially expressed in pancreas in type 1 diabetes is simply the consequence of the (near complete) loss of pancreatic beta cells in this disease. Fig 6e hence does not seem to depict a major discovery.

No data is provided on the expression levels of genes (e.g. mean expression (TPM) in diabetes and controls), while this would be very useful to know for differentially expressed genes. This information should be added to Suppl Table 6f and 6i.

"Functional enrichment analyses revealed that upregulated genes in the tibial nerve are enriched in immune receptor activity, whereas downregulated genes are enriched in ion channel activity (Fig. 6f; Supplementary Table 6h; STAR methods)". Fig 6f does not look like functional enrichment.

Some spelling mistakes need to be corrected, e.g. Abstract: ancestr.

Alonso, L., Piron, A., Morán, I., Guindo-Martínez, M., Bonàs-Guarch, S., Atla, G., Miguel-Escalada, I., Royo, R., Puiggròs, M., Garcia-Hurtado, X., et al. (2021). TIGER: The gene expression regulatory variation landscape of human pancreatic islets.

Cell Rep 37, 109807.
Fadista, J., Vikman, P., Laakso, E.O., Mollet, I.G., Esguerra, J.L., Taneera, J., Storm, P., Osmark, P., Ladenvall, C., Prasad, R.B., et al. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A 111, 13924-13929.
Marselli, L., Piron, A., Suleiman, M., Colli, M.L., Yi, X., Khamis, A., Carrat, G.R., Rutter, G.A., Bugliani, M., Giusti, L., et al. (2020). Persistent or transient human β cell dysfunction induced by metabolic stress: specific signatures and shared gene expression with type 2 diabetes. Cell Rep 33, 108466.
Ray, P.R., Khan, J., Wangzhou, A., Tavares-Ferreira, D., Akopian, A.N., Dussor, G., and Price, T.J. (2019). Transcriptome Analysis of the Human Tibial Nerve Identifies Sexually Dimorphic Expression of Genes Involved in Pain, Inflammation, and Neuro-Immunity. Front Mol Neurosci 12, 37.
van de Bunt, M., Manning Fox, J.E., Dai, X., Barrett, A., Grey, C., Li, L., Bennett, A.J., Johnson, P.R., Rajotte, R.V., Gaulton, K.J., et al. (2015). Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. PLoS Genet 11, e1005694.
Vinuela, A., Varshney, A., van de Bunt, M., Prasad, R.B., Asplund, O., Bennett, A., Boehnke, M., Brown, A.A., Erdos, M.R., Fadista, J., et al. (2020). Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. Nat Commun 11, 4912.

Authors' response to the first round of review

We would like to thank all the reviewers for their time invested and positive feedback, which we found to be very helpful. We have now revised the paper and made the required changes according to the reviewer's comments, as we detail below.

Reviewer 1: Garcia-Perez et. al. have implemented a statistical framework to assess the contribution of multiple population traits and clinical traits to the gene expression and alternative splicing variation based on the GTEx data. Extensive tests are done, and the results cover many aspects, including the extent of DEG sharing, the contribution of cis-eQTL and cis-sQTL, the additive contribution and trait interaction, and the association between traits and cell type composition.

I believe this work is carried out carefully.

We would like to thank the reviewer for their positive evaluation of our work.

I just have one major concern in terms of phrasing. The impact or effect of traits on transcriptome is often mentioned, for example, "we took advantage of the GTEx multi-tissue data to assess how type 1 and 2 diabetes affect the transcriptome of multiple tissues", while the tests are for association. It might be better to rephrasing it to avoid possibly giving the reader impression that it is about causation.

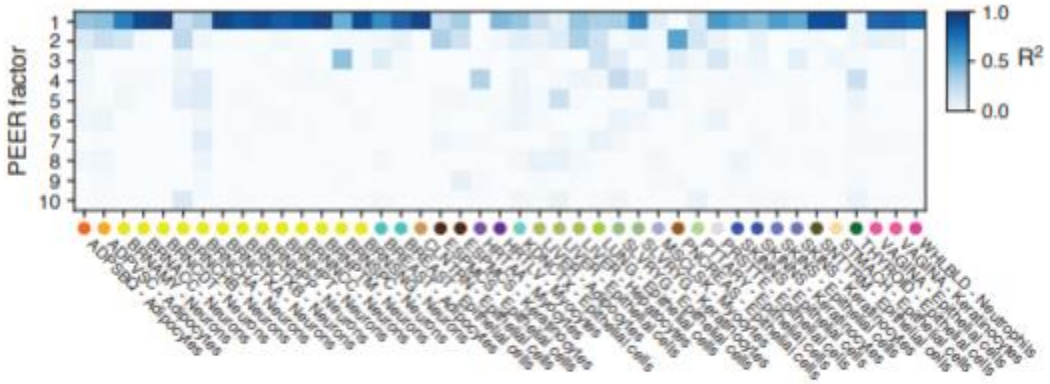
We have revised the text and changed the phrasing according to the reviewer's suggestion. Specifically, we have changed statements that used effect/affect/impact to refer to significant associations between demographic and/or clinical traits and transcriptome variation. We have highlighted those changes in dark blue in the new manuscript version.

One minor point is, from the method part, it seems that two PEER factors are used. It might be better to add some explanation on how the number of PEER factors is decided.

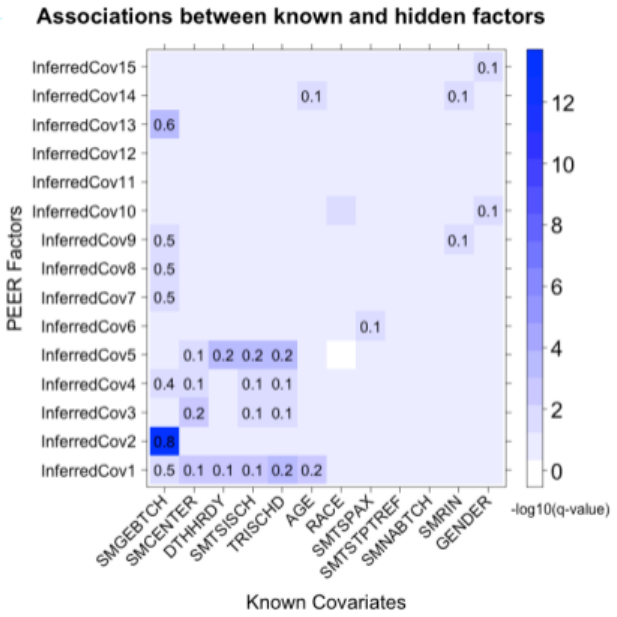
We have included in the methods section a brief explanation on the rationale used to select two PEER factors. Now, it reads as follows:

"To control for unknown sources of variation we explored the expression variance captured by the PEER factors (GTEx Consortium, 2020) and explained by known sample and donor covariates, as well as by the xCell enrichment scores (Kim-Hellmuth et al., 2020). We also investigated the ef ect of including progressively increasing numbers of PEER factors in our model in the identification of DEGs. As previously noted, we found that the first PEER factor was mostly correlated with cell type heterogeneity (see Supplementary Fig. 4a from (Kim-Hellmuth et al., 2020)), and the second PEER factor was mostly correlated with the sequencing batch (see Supplementary Fig. 8a from (GTEx Consortium, 2015). We also noted that, conversely to eQTL discovery, the ef ect of including additional PEER factors on the DEG discovery was variable across tissues and led to reduced power to detect expression dif erences. Thus, to control for unknown sources of variation mainly related to dif erences in tissue composition and sequencing batch, we included the first two PEER factors."

Below we include the figures mentioned above, Supplementary Fig. 4a from (Kim-Hellmuth et al., 2020) and Supplementary Fig. 8a from (GTEx Consortium, 2015), that show the association between Response to Reviewers PEER factors 1 and 2 and cell type heterogeneity and batch effects, respectively.



Supplementary Fig. 4a from (Kim-Hellmuth et al., 2020). PEER factor 1 is mostly associated with tissue heterogeneity. Correlation between xCell enrichment scores for the tissue-cell type combination indicated on the x-axis and the first ten PEER factors computed for each tissue, showing that the top PEER factors capture cell type heterogeneity.



Supplementary Fig. 8a from (GTEx Consortium, 2015). PEER factor 2 is mostly associated with sequencing batch effects. An assessment of correlations between inferred PEER factors and known covariates in adipose tissue as a representative example. The color denotes significance of the association. For significant associations (q value ≤ 0.05), the r 2 value is reported in the cell.

Another minor point is, on the top of page 8, in the sentence:

"DEGs between populations and between sexes are the most tissue- shared (Fig. 3e and Supplementary Fig. 3f),"

Fig. 3e doesn't seem to illustrate the point of this sentence. Supplementary Fig. 3e is the correct one?

The reviewer is right. We have now changed the reference to the correct Supplementary Figures, namely, Supplementary Fig. 3e, that shows the tissue-sharing of genes with additive contributions for each pairwise combination of traits, and Supplementary Fig. 3f, that shows the expression patterns of DEGs between populations and between sexes in each of the tissues where they are DE.

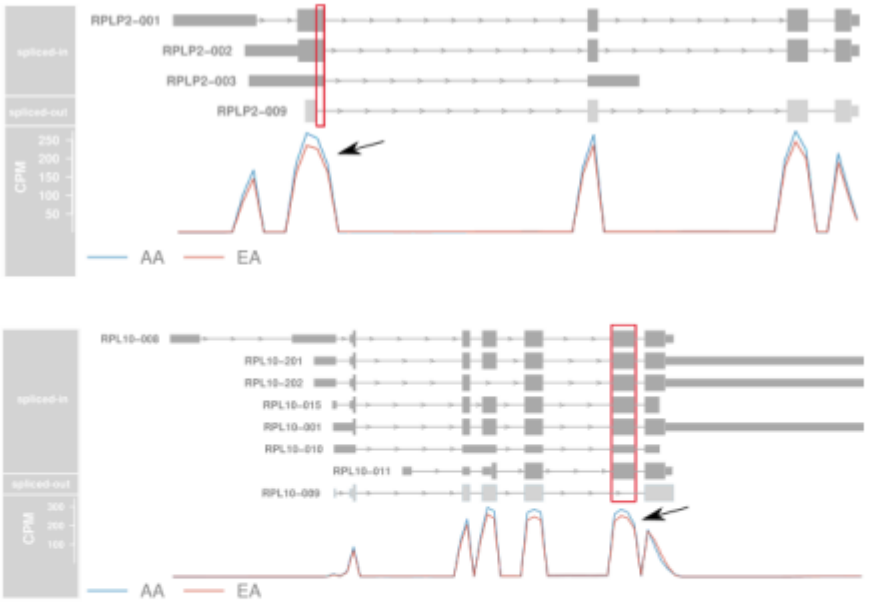
Reviewer 2: In this paper, the authors have done an excellent job in systematically characterising how changes in gene expression and splicing correlate with major demographic features (age, sex, ancestry, BMI) as well as number of other traits and diseases. The authors replicate a number of previous findings (mostly from studies conducted with whole blood samples) and expand their analysis to the most comprehensive collection of cell types and tissues available from the GTEx Consortium. Importantly, they also identify a novel and interesting association between diabetes and gene expression changes in the nerve tissues that correspond to diabetic neuropathy. The paper is well written and all of the analyses have been performed extremely thoroughly and carefully. I don't have any major concerns.

We truly appreciate the reviewer's comments on our manuscript.

My only minor concern is that inferring percentage-spliced-in (PSI) values from transcript expression estimates (RSEM TPMs) carries a small risk of detecting false positive splicing events. This is caused by the fact that existing transcript annotations often couple together different splicing events that are likely to be regulated independently. For more detailed examples, see Supplemental Figure 1 in the Whippet preprint (https://doi.org/10.1101/158519) and Figure 1 in Alasoo et al, 2019 (https://doi.org/10.7554/eLife.41673). To be clear, I believe that this issues will only add some noise to your quantification results and will not actually invalidate any of your major conclusions which are based on aggregating signal across many independent splicing events. Furthermore, it is not obvious that using any of the alternative approaches will actually give you more accurate PSI estimates for the splicing events of interest. Thus, I think that is perfectly ok for your to keep using the SUPPA2 PSI estimates. However, I think it would be helpful

to validate some of the the splicing associations that you see (e.g. the ribosomal genes RPLP2 and RPL10) using complementary approaches such as Leafcutter junction read counts (should be available from GTEx) or read coverage plots (e.g. from ggcoverage, Sashimi Plot, wiggleplotr or others).

We agree with the reviewer and to further validate our findings we have now included coverage plots for the highly tissue-shared splicing events in the ribosomal proteins RPL10 and RPLP2 as suggested (see below and in Supplementary Fig. 6d). In the figure we can see that African Americans have higher coverage than European Americans in the events differentially spliced between populations.



Supplementary Fig. 6d. Read coverage plots spanning the highly tissue-shared events in ribosomal proteins RPLP2 (top) and RPL10 (bottom). Isoform models for the isoforms that include the exonic region (in dark gray) or exclude the exonic region (in light gray) are shown. The exonic regions that are either spliced-in or spliced-out are highlighted with a red box. The average normalized read coverage per population (AA: African American; EA: European American) in SKINS (top) and in SKINNS (bottom) is shown. CPM (per bin) = number of reads per bin / number of mapped reads (in millions). Please note that the isoforms in the plots can also participate in other splicing events involving adjacent exonic and intronic regions.

We have also added a new section in the methods describing the data and approach used to create these plots (see below).

"We downloaded the available RNAseq bam files for the samples from the SkinSunExposedLowerleg and SkinNotSunExposedSuprapubic tissues, which are

part of the GTEx protected data stored in dbGap (accession number phs000424.v8.p2). We used deeptools (Ramírez et al., 2016) to generate normalized coverage tracks (counts per million (CPM)) in 50 base-pairs windows considering uniquely mapped reads. We used the R package Gviz (Hahne and Ivanek, 2016) to plot the average read coverage per population (Supplementary Fig. 6d).”

Reviewer #3: Mele et al implemented a statistical framework to assess the contribution of multiple demographic and clinical traits on transcriptome and alternative splicing in human tissues using Genotype Tissue Expression (GTEx) data. Mele et al suggest that type 1 and type 2 diabetes contribute to multiple tissue transcriptome variation with the strongest signal for tibial nerve. The authors should be commended for their approach to study a large array of human tissues. The study design is well-thought, and the analyses are a tour de force. The disease/diabetes results and discussion sections seem less strong and do not fully support the conclusions (see below).

We thank the review for their thorough review, positive comments and suggestions that we believe have contributed to improving the manuscript.

Major comments: The readability of the figures is unfortunately limited by the abbreviations used for the different tissues (e.g. MSCLSK, ESPMCS, etc, see Fig. 1a and throughout) and ancestry (see Fig. 2b). The paper would be much easier to read if the tissues were readily identifiable.

We understand the reviewer's concern. However, we decided to use tissue abbreviations for two main reasons. First, we wanted to better use the space in the figures, since all of them include multiple panels. Second, this is the consensus approach used in all recent GTEx consortium papers (see for example (GTEx Consortium, 2020; Kim-Hellmuth et al., 2020; Oliva et al., 2020) and the main GTEx leaders that are also co-authors in this publication (Kristin Ardlie and François Aguet) suggested that we keep these abbreviations. Supplementary Fig. 1a shows tissue abbreviations mapped to their corresponding tissue names.

The reviewer is right in pointing to the missing legend for ancestry in Fig. 2d. We have now corrected this figure so it clearly states what AA and EA stand for (AA: African American; EA: European American).

Why was kidney not included (see Suppl Fig 1a)?

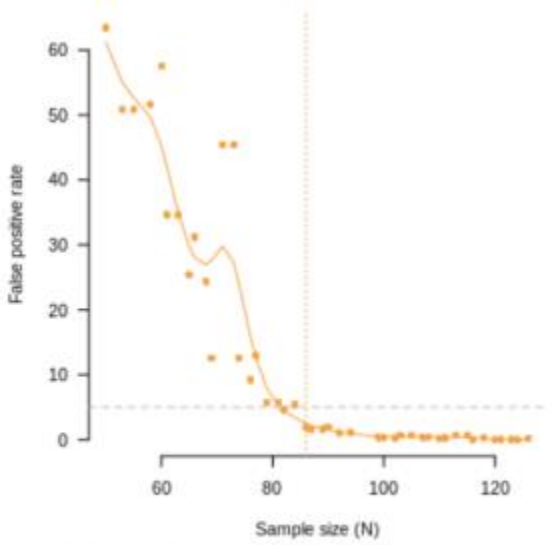
The number of kidney tissue samples available was 73. This number is lower than the conservative 100 RNA-seq samples necessary to obtain reliable splicing results. Because we wanted to perform expression and splicing analysis on the

same samples to be able to compare them, we excluded kidney tissue from all analysis.

In the Methods' section "GTEx data" we state the criteria used to select the tissues and samples used in this study.

"Here, we analyzed data from the 46 tissue sources with at least 100 RNA-seq samples. We only included samples (n=13,684) from donors (n=781) with available metadata for the covariates included in our dif erential expression and splicing analysis, as well as demographic trait information for the donors' genetic inferred ancestry (we only included European American and African American donors), sex, age, and body mass index (BMI)."

Below we explain one of the preliminary analyses we did to determine the minimum number of samples needed to accurately perform differential splicing analysis. We used tissue subsets with different numbers of samples from SkinNotSunExposedSuprapubic and estimated a false positive rate (FPR = False positives (FP) / Positives (P)) for each sample size, where FP are the number of differentially spliced events discovered at a 5% FDR using permuted labels (e.g. randomly shuffling samples' ancestry) and P are the number of differentially spliced events discovered at a 5% FDR using the correct labels.We set the threshold at the number of RNA-seq samples (100 samples) for which the estimated false positive rate was lower than 5%.



Number of samples required to obtain reliable differential splicing results with fractional logit regression models when investigating the effect of a demographic trait (ancestry).

Type 1 and 2 diabetes are characterized by dysregulation of blood glucose levels. Pancreatic islets and in particular pancreatic beta cells play a key role in the development and progression of diabetes. Pancreatic islets make up few (around 2) percent of pancreas weight. Not surprisingly then, transcriptomic and related association studies have shown that pancreas is not a good proxy for islets and fails to capture islet-specific disease signatures (Alonso et al., 2021; Vinuela et al., 2020). The lack of a tissue that is key to diabetes pathogenesis (i.e. pancreatic islets) in GTEx and the fact that pancreas is not a good proxy thereof are limitations that need to be clearly spelled out. Mele et al could consider adding transcriptome studies in pancreatic islets, which would strengthen their diabetes-related claims. Hundreds of islet transcriptomes are available, see (Fadista et al., 2014; Marselli et al., 2020; van de Bunt et al., 2015; Vinuela et al., 2020) among others.

Taking the above into account, the approach and analyses taken by Mele et al are rather poised to detect consequences of type 1 and type 2 diabetes, i.e. chronic micro- and macrovascular complications caused by long-term exposure to hyperglycemia. Again, this does not clearly appear in the manuscript and should be mentioned.

Following the reviewer's suggestion, we have now clearly stated the limitation of studying diabetes in pancreas as well as the fact that the diabetes associated signals we observe in other tissues are likely due to continuous high glucose blood levels.

"When comparing the DE signal across tissues, we found that the tibial nerve is the most af ected tissue in both types of diabetes. Pancreas had fewer DEGs than nerve, likely due to the fact that pancreatic islets, central to the etiology of both types of diabetes (Eizirik et al., 2020; Krentz and Gloyn, 2020), represent only ~3% of the tissue, and thus, the whole pancreas is not representative of pancreatic islets (Alonso et al., 2021; Viñuela et al., 2020). The observation that there are many genes associated with diabetes in other tissues may reflect the consequences of long-term exposure to hyperglycemia across tissues."

We have also acknowledged these limitations in the discussion section:

"We also have reduced statistical power to detect transcriptomic changes associated with certain clinical traits due to both reduced sample sizes and analysis of bulk tissue transcriptomes rather than specific cell types such as pancreatic islet cells for type 1 diabetes (Alonso et al., 2021)."

The finding that a strong transcriptomic overlap for type 1 and type 2 diabetes is found in tibial nerve is of interest, and it begs the question why other chronic micro- and macrovascular complications are, apparently, not shared in the tissue transcriptomes (heart, arteries, kidney, etc). This deserves further data analysis and should be discussed.

We agree with the reviewer that this is an intriguing finding. Although we found DEGs with type 1 and type 2 diabetes in the heart atrial appendage and heart left ventricle (HRTAA and HRTLV, respectively) as well as in the aorta and tibial arteries (ARTAORT and ARTTBL, respectively) (Supplementary Fig. 7e), the number of DEGs is much lower than those in the tibial nerve. Also, contrary to what we observed in the tibial nerve, the overlaps between DEGs with type 1 and type 2 diabetes in these tissues were not statistically significant. The number of individuals with type 1 and 2 diabetes is lower in these tissues compared to the sample size in the tibial nerve, and thus, we might have reduced power to detect statistically significant expression changes -and hence overlaps- in these tissues. Whereas diabetic treatment is known to reduce the risk of diabetic complications (Hicks and Selvin, 2019; Nathan and DCCT/EDIC Research Group, 2014), different treatments vary in their capacity to prevent specific diabetic complications (Call et al., 2022). The treatments between diabetes types also differ, resulting in different reductions of cardiovascular complications (Rawshani et al., 2017). One could also hypothesize that specific diabetic treatment reverses the cardiovascular complications associated with hyperglycemia to non-diabetic levels in hearts and arteries but maybe not so much in the tibial nerve. This could explain why we detect a stronger signal in the tibial nerve compared to heart and arteries. Unfortunately, there is no available information about the drugs prescribed to GTEx donors. Following the reviewer's suggestion we have discussed these limitations in the discussion.

“We also have reduced statistical power to detect transcriptomic changes associated with certain clinical traits due to both reduced sample sizes and analysis of bulk tissue transcriptomes, rather than specific cell types, such as pancreatic islet cells for type 1 diabetes (Alonso et al., 2020). These limitations might explain why we only find the same genes associated with type 1 and type 2 diabetes in the tibial nerve but not in other tissues af ected by long term hyperglycemia, such as the heart or arteries (Nathan and DCCT/EDIC Research Group, 2014; Zheng et al., 2018), Alternatively, disease treatment is known to mitigate diabetic complications (Call et al., 2022; Nathan and DCCT/EDIC Research Group, 2014; Rawshani et al., 2017) and could thus have an ef ect on the number

of genes we observe DE in specific tissues. Hence, collecting information about donor prescribed drugs would be desirable in future studies.

The GTEx data for tibial nerve have been used in previous work on the impact of gender, diabetes etc on transcriptomes (Ray et al., 2019). The existing literature should be acknowledged, the present data compared against it and novelty claims toned down accordingly.

We thank the reviewer for this suggestion. We have toned down the novelty claims and explicitly cited previous papers that reported transcriptomic changes with diabetes in the tibial nerve. The final paragraph of the section "Type 1 and type 2 diabetes are associated with transcriptome variation in multiple tissues, particularly the tibial nerve" now reads as follows:

"Taken together, our findings expand previous work (Gu et al., 2018; Ray et al., 2019; Singh et al., 2017) and suggest that despite their dif erent etiologies, type 1 and 2 diabetes are more strongly associated with transcriptome changes in tibial nerve than in other tissues. These changes are consistent with the high prevalence of diabetic neuropathy in patients (Feldman et al., 2019) due to hyperglycemia and provide novel gene candidates associated with diabetic neuropathy."

As suggested by the reviewer, we also tried to compare our results with those reported by Ray et al. We directly contacted the authors since the list of DEGs with diabetes was not available in the paper. Unfortunately, they replied that since the focus of the paper was not the identification of DEGs with diabetes but rather of DEGs with sex, that gene list was not available. However, we did replicate our findings in a different dataset that used diabetic mice to identify diabetes associated changes in the nerves. This section is in the results section of the paper and it reads as follows:

"*The genes DE in the tibial nerve significantly overlap with those reported as dysregulated in the sciatic nerve of diabetic mice (two-tailed Fisher's exact test, p-value = 1.195e-06)(Gu et al., 2018).*"

As the number of type 1 diabetes samples is a lot smaller than that of type 2 diabetes, the statistical power for the former is much lower. A threshold free method like Rank-Rank Hypergeometric Overlap (RRHO) is probably more suitable to detect similarities. The authors should consider using RRHO tools for these comparisons, or at least mention the limitation related to sample numbers.

As suggested by the reviewer, we have now acknowledged our limitation to identify DEGs with type 1 diabetes compared to DEGs with type 2 diabetes in the text as well as in the discussion.

"The dif erence in the number of DEGs is likely due to decreased statistical power related to the lower number of individuals with type 1 diabetes."

"We also have reduced statistical power to detect transcriptomic changes associated with certain clinical traits due to both reduced sample sizes and analysis of bulk tissue transcriptomes rather than specific cell types such as pancreatic islet cells for type 1 diabetes (Alonso et al.)."

In the star methods, a "Differential gene expression analysis with demographic traits" is included but no section about clinical/phenotypic traits. Did the authors consider confounding ancestry and demographic factors (BMI, age, …) when analyzing type 1 and type 2 diabetes vs controls? This is important as ancestry contributes significantly to pancreas transcriptome variation (Fig 1a).

The reviewer asks a very important question, but we did include a methods section named "Differential expression, differential splicing, and hierarchical partition analysis with clinical traits", in which we stated that "To identify DEGs with clinical traits we used the same approach described in the section Dif erential gene expression analysis, but further including the clinical traits as covariates in the linear models". Thus, we did not only include in the models the technical covariates and demographics traits used in the first section of the paper, but in each tissue, we also included all the clinical traits with evidence of affecting the tissue transcriptome variation to ensure our results were not confounded by the different demographic traits nor by the different diseases. For example, in Lung, the model includes 6 different clinical traits on top of the technical covariates and demographic traits: type 1 and 2 diabetes, atelectasis, emphysema, fibrosis and pneumonia (see Figure 6a).

expression (log2cpm) ~ HardyScale + IschemicTime + RIN + Cohort + NucAcIsoBatch + ExonicRate + PEER1 + PEER2 + Ancestry + Sex + Age + BMI + type 1 diabetes + type 2 diabetes + atelectasis + emphysema + fibrosis + pneumonia

We have included the model in the methods to make it more clear to the reader:

expression (log2cpm)/splicing (PSI) ~ HardyScale + IschemicTime + RIN + Cohort + NucAcIsoBatch + ExonicRate + PEER1 + PEER2 + Ancestry + Sex + Age + BMI + clinical trait(s)

Additional comments p.11:

"found 79 and 309 DEGs in two or more tissues with either type 1 or type 2 diabetes". Is this correct? We counted 387 lines and 378 distinct genes in Suppl Table 6f.

Supplementary Table 6f includes the list of genes that are DE with either type 1

diabetes or type 2 diabetes in two or more tissues. There are 78 DEGs with type 1 diabetes in two or more tissues (we have corrected this number in the manuscript) and 309 DEGs with type 2 diabetes in two or more tissues. Nine of these genes are DE with type 1 and type 2 diabetes in two or more tissues but not necessarily in the same tissues (DMTN, DNAJC12, HADH, MT-ND4, NAT8L, SLC50A1, SPTB, TMEM39A and ZFYVE1). That is why there are 387 lines and 378 genes in Supplementary Table 6f, because nine genes (387-378=9) are DE in two or more tissues with both diabetes.

The finding that INS, IAPP and MAFA are markedly differentially expressed in pancreas in type 1 diabetes is simply the consequence of the (near complete) loss of pancreatic beta cells in this disease. Fig 6e hence does not seem to depict a major discovery.

The reviewer is right. This is not a novel discovery, which we stated in Fig. 6e caption, “As expected, individuals with type 1 diabetes show no expression, and individuals with type 2 diabetes show decreased expression compared to healthy individuals”, but rather a validation showing our analyses identify previously known associated genes. Thus, we have removed this figure.

No data is provided on the expression levels of genes (e.g. mean expression (TPM) in diabetes and controls), while this would be very useful to know for differentially expressed genes. This information should be added to Suppl Table 6f and 6i.

As suggested by the reviewer, we have added the mean gene expression levels (TPM) in individuals with type 1 and type 2 diabetes in Supplementary Tables 6f and 6i.

"Functional enrichment analyses revealed that upregulated genes in the tibial nerve are enriched in immune receptor activity, whereas downregulated genes are enriched in ion channel activity (Fig. 6f; Supplementary Table 6h; STAR methods)". Fig 6f does not look like functional enrichment.

The reviewer is right. Fig. 6f (now Fig. 6e) showed that all the genes DE with type 1 and type 2 diabetes are either upregulated or downregulated, and the upregulated and downregulated genes show different enrichments. Supplementary Table 6f shows the enrichment results. To avoid any confusion we have separated the two messages: we have removed the enriched terms from the figure and we now reference it before and rephrased the sentence that refers to Supplementary table 6f as follows.

“Functional enrichment analysis (Supplementary Table 6h; STAR methods)

revealed that upregulated genes in the tibial nerve are enriched in immune receptor activity, whereas downregulated genes are enriched in ion channel activity."

Some spelling mistakes need to be corrected, e.g. Abstract: ancestr.

We have corrected this typo in the abstract.

Bibliography

Alonso, L., Piron, A., Morán, I., Guindo-Martínez, M., Bonàs-Guarch, S., Atla, G., Miguel-Escalada, I., Royo, R., Puiggròs, M., Garcia-Hurtado, X., et al. (2021). TIGER: The gene expression regulatory variation landscape of human pancreatic islets. Cell Rep. 37, 109807.

Call, J.T., Cortés, P., and Harris, D.M. (2022). A practical review of diabetes mellitus type 2 treatment in primary care. Rom. J. Intern. Med. 60, 14–23. Cole, J.B., and Florez, J.C. (2020). Genetics of diabetes mellitus and diabetes complications. Nat. Rev. Nephrol. 16, 377–390.

Eizirik, D.L., Pasquali, L., and Cnop, M. (2020). Pancreatic β-cells in type 1 and type 2 diabetes mellitus: different pathways to failure. Nat. Rev. Endocrinol. 16, 349–362.

Feldman, E.L., Callaghan, B.C., Pop-Busui, R., Zochodne, D.W., Wright, D.E., Bennett, D.L., Bril, V., Russell, J.W., and Viswanathan, V. (2019). Diabetic neuropathy. Nature Reviews Disease Primers 5, 1–18. GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330.

Gu, Y., Qiu, Z.-L., Liu, D.-Z., Sun, G.-L., Guan, Y.-C., Hei, Z.-Q., and Li, X. (2018). Differential gene expression profiling of the sciatic nerve in type 1 and type 2 diabetic mice. Biomed Rep 9, 291–304.

Hahne, F., and Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. Methods Mol. Biol. 1418, 335–351. Hicks, C.W., and Selvin, E. (2019). Epidemiology of Peripheral Neuropathy and Lower Extremity Disease in Diabetes. Curr. Diab. Rep. 19, 86.

Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al. (2020). Cell type-specific genetic regulation of gene expression across human tissues. Science 369. https://doi.org/10.1126/science.aaz8528.

Krentz, N.A.J., and Gloyn, A.L. (2020). Insights into pancreatic islet cell dysfunction from type 2 diabetes mellitus genetics. Nat. Rev. Endocrinol. 16, 202–212.

Leon, B.M., and Maddox, T.M. (2015). Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research. World J. Diabetes 6, 1246–1258.

Nathan, D.M., and DCCT/EDIC Research Group (2014). The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. Diabetes Care 37, 9–16.

Nathan, D.M., Cleary, P.A., Backlund, J.-Y.C., Genuth, S.M., Lachin, J.M., Orchard, T.J., Raskin, P., Zinman, B., and Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group (2005). Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. N. Engl. J. Med. 353, 2643–2653.

Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., et al. (2020). The impact of sex on gene expression across human tissues. Science 369. https://doi.org/10.1126/science.aba3066.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165.

Rawshani, A., Rawshani, A., Franzén, S., Eliasson, B., Svensson, A.-M., Miftaraj, M., McGuire, D.K., Sattar, N., Rosengren, A., and Gudbjörnsdottir, S. (2017). Mortality and cardiovascular disease in type 1 and type 2 diabetes. N. Engl. J. Med. 376, 1407–1418.

Ray, P.R., Khan, J., Wangzhou, A., Tavares-Ferreira, D., Akopian, A.N., Dussor, G., and Price, T.J. (2019). Transcriptome Analysis of the Human Tibial Nerve Identifies Sexually Dimorphic Expression of Genes Involved in Pain, Inflammation, and Neuro-Immunity. Front. Mol. Neurosci. 12, 37.

Singh, K., Gupta, K., and Kaur, S. (2017). High resolution ultrasonography of the tibial nerve in diabetic peripheral neuropathy. J Ultrason 17, 246–252.

Viñuela, A., Varshney, A., van de Bunt, M., Prasad, R.B., Asplund, O., Bennett, A., Boehnke, M., Brown, A.A., Erdos, M.R., Fadista, J., et al. (2020). Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. Nat. Commun. 11, 4912.

Zheng, Y., Ley, S.H., and Hu, F.B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat. Rev. Endocrinol. 14, 88–98.