nature research

Corresponding author(s): Daniel J Rader

Last updated by author(s): Aug 31, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Variant annotation was performed using ANNOVAR (version 2018Apr16). Phenotypes were mapped to Phecodes using the R package "PheWAS" via R version 3.3.1 or later. Images of the retinal distribution of PPP1R13L protein were captured using AxioVS40 software version 4.8.1.0. For scRNA-seq of mouse and human aortas, raw FASTQ data alignment was processed using Cell Ranger 3.0. For scRNA-seq of human pancreas, expression data was generated using Cell Ranger 2.1.0.
Data analysis	All statistical analyses of de-identified clinical data were performed using R version 3.3.1 or later. PheWAS was performed using the R package "PheWAS" via R version 3.3.1 or later. Power analyses were conducted using QUANTO version 1.2.4. Public expression datasets from GEO were analyzed using GEO2R and R version 3.6.1. Meta-analysis of differential expression across datasets and its visualization were achieved using the R package "MetaVolcanoR 1.0.1". Public expression data from the Glaucoma Discovery Platform were analyzed using Datgan. For scRNA-seq of human pancreas, Seurat 3.0.2 was used for filtering, UMAP generation, and initial clustering. DoubletFinder 2.0 was used to demarcate and remove potential doublets in the data. Garnett was used for initial cell classification. TooManyCells 2.0.0.0 was then used to cluster and visualize the single cells. Differential genes were found using edgeR 3.24.3. For scRNA-seq of mouse aortas, Seurat 3.1.4 was used for filtering, normalization, and clustering. For scRNA-seq of human aortas, Seurat 3.0 was used for filtering, normalization, and clustering.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All summary statistics for significant gene-phenotype associations from the discovery phase in PMBB as well as significant replications from each replication cohort are fully detailed in the Supplementary Information (Table S1-S16). Data for the individual rare pLOF and missense variants in significant genes that were used for gene burden analyses in the PMBB discovery cohort are also included in the Supplementary Information (Tables S23-S24). In addition, a list of all of the single variants that were used for replication analyses across all the cohorts are provided in the Supplementary Information (Table S25). Each variant in Tables S23-25 is annotated with information regarding genomic location, variant effect, amino acid change, REVEL score (for missense), and minor allele frequency in gnomAD as well as in the PMBB discovery cohort. Additionally, up-to-date summary data for genetic variants captured via whole-exome sequencing in PMBB can be accessed via the Penn Medicine Biobank Genome Browser (https://pmbb.med.upenn.edu/biobank/allele-frequency/). Individual-level data are not made publicly available due to research participant privacy concerns; however, requests from accredited researchers for access to individual-level data relevant to this manuscript can be made by contacting the corresponding author. Additionally, public expression datasets were obtained from the Ocular Tissue Database (https://genome.uiowa.edu/otdb/), Glaucoma Discovery Platform (http://glaucomadb.jax.org/glaucoma), and the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/). From NCBI GEO, we interrogated 11 different microarray and RNA-seq datasets of human fibroblasts from various tissues treated with TGF-ß (GSE1724, GSE65069, GSE64192, GSE39394, GSE79621, GSE68164, GSE97833, GSE97823, GSE135065, GSE125519, GSE40266) as well as microarray data from muscle biopsies in tibial muscular dystrophy patients (GSE42806).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Our discovery experiment in Penn Medicine Biobank (PMBB) included a subset of 10,900 individuals in the PMBB who have undergone whole-exome sequencing following quality control measures. We chose to interrogate gene burdens with at least 25 carriers, which is computationally equivalent to a single variant with minor allele frequency ~ 0.1%. Given that a gene burden of rare variants is expected to have significantly higher effect sizes than common variants, we show through power analyses (Extended Data Figure 2) that this minimum is sufficient for discovery of significant associations across variable odds ratios. For replication studies in other biobanks, including BioMe (N=23,989), DiscovEHR (N=85,450), UK Biobank (N=32,268), and BioVU (N=66,400), the sample size was opportunistic and was determined by the availability of samples in each biobank at the time of genetic sequencing. For qRT-PCR of iPSC-RGCs, we did not use statistical methods to predetermine sample size but are confident our sample numbers give enough confidence to interpret significance when fold change is compared against untreated control sample given the clear differences seen in Extended Figure S5C as well as consistency seen from replication.. For scRNA-seq of human pancreas, sample size calculation was not performed since all high-quality single cells from all donor samples that were collected by HPAP at the time of the preparation of this manuscript with high quality scRNA-seq data were used in this analysis. For scRNA-seq of mouse aorta, sample size was not predetermined. Instead, given that 3 to 5 samples need to be pooled to establish a cDNA library for scRNA-seq studies in mouse aortas, the present study pooled 5 ascending aortas from mice. Similarly, for scRNA-seq of human aorta, sample size was not predetermined. Instead, ascending aorta samples from 11 individuals were pooled for this study.
Data exclusions	For the discovery experiment in PMBB, on the genotypic side, from a total of 11,451 individuals with whole-exome sequencing in PMBB, we removed samples with low exome sequencing coverage (i.e. less than 75% of targeted bases achieving 20x coverage), high missingness (i.e. greater than 5% of targeted bases), high heterozygosity, dissimilar reported and genetically determined sex, genetic evidence of sample duplication, and cryptic relatedness (i.e. closer than 3rd degree relatives), leading to a total of 10,900 individuals following pre-established protocols for quality control of exome sequencing data. On the phenotypic side, patients were determined to have a certain disease phenotype if they had the corresponding ICD diagnosis on two or more dates, while phenotypic controls consisted of individuals who never had the ICD code. Individuals with an ICD diagnosis on only one date as well as individuals under control exclusion criteria based on PheWAS phenotype mapping protocols were not considered in statistical analyses. These pre-established phenotypic exclusion criteria were implemented to add a level of stringency to increase sensitivity for defining cases for disease phenotypes. For replication analyses in all other biobanks, similar exclusion criteria were implemented with regard to both genotypic and phenotypic aspects.
Replication	Evaluation of the robustness of the associations we found during the discovery phase with exome-by-phenome-wide significance (p<E-06) is a central theme of this paper. Firstly, in the same cohort, we interrogated other non-overlapping predicted deleterious variants in the same genes to test whether their association with the originally associated phenotype on discovery withheld. Then, we repeated the same experiments in multiple replication cohorts: 1) a new (non-overlapping) set of 6,432 exomes in African-Americans in the Penn Medicine Biobank (PMBB2); 2) 23,989 exomes (6,470 African, 8,735 European, 8,784 Hispanic) from Mount Sinai's BioMe; 3) 85,450 exomes (European)

from the Geisinger Health System's DiscovEHR cohort; 4) 32,268 European exomes from the UK Biobank; and 5) 66,400 genotypes (10,456 African, 55,944 European) from Vanderbilt's BioVU. Successful replication attempts are detailed in Table S17.

For immunohistochemical localization of PPP1R13L in the retina, the experiment was performed twice independently with consistent results. For qRT-PCR of iPSC-RGCs, we performed the qRT-PCR using RNA derived from a triplicate experiment for each concentration of hydrogen peroxide.

For scRNA-seq of human pancreas, all high-quality single cells from all donor samples that were collected by HPAP at the time of the preparation of this manuscript with high quality scRNA-seq data were used in this analysis. For each donor, a single scRNA-seq experiment was conducted using all high-quality single cells, and thus was carried out without replication.

For scRNA-seq of mouse and human aortas, all high quality single cells with high quality scRNA-seq data were used in this study. A single scRNA-seq experiment was conducted using all high-quality single cells for scRNA-seq of mouse aorta, and thus was carried out without replication. Similarly, a single scRNA-seq experiment was conducted using all high-quality single cells for scRNA-seq of human aorta for each sample, and thus was carried out without replication

Randomization

Individuals in PMBB, BioMe, DiscovEHR, UK Biobank, and BioVU were allocated into experimental groups based on their genotypic and phenotypic statuses, which were variables determined prior to this study. Each disease phenotype was tested for association with each gene burden or single variant adjusted for age, age^2, sex and the first ten principal components of genetic ancestry as covariates.

For transcriptional expression studies in iPSC-RGCs, samples were allocated into experimental groups based on their exposure to a particular concentration of hydrogen peroxide treatment.

For scRNA-seq of human pancreas, participants were allocated into the groups type 1 diabetes (T1D) vs. control based on their medical charts and C-peptide measurements in accordance with the American Diabetes Association guidelines.

For scRNA-seq of mouse aorta, there was no allocation of samples into experimental groups given that the mouse scRNA-seq study was performed using one group of normal aortas from wild type mice. For scRNA-seq of human aorta, control ascending aortic tissue samples were obtained from recipients of heart transplants or lung donors, and diseased aortic tissue samples were obtained from patients with sporadic ATAA excluding those who had ascending aortic dissection, an heritable form of aortopathy (e.g., Marfan syndrome, Loeys-Dietz syndrome, a first-degree relative with ATAA, bicuspid aortic valve), or ATAA related to infection, aortitis, trauma, or isolated pseudoaneurysm. However, gene expression analyses for this study were conducted on the entire cohort regardless of disease state.

Blinding

The investigators were blinded to group allocation during data collection and analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

anti-PPP1R13L (Cat# 51141-1-AP, Proteintech, IL, USA), chicken anti-rabbit IgG conjugated with Alexa Fluor 594 (Cat# A21442, Life Technologies, Carlsbad, CA)

Validation

Specificity of the antibody staining was determined when compared with immunostaining patterns observed with no antibody and a secondary antibody controls.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

The de-identified patient-derived iPSCs from Caucasian individuals were procured from the Human iPSC core facility, University of Pennsylvania.

Authentication

The hiPSCs were generated from keratinocytes or blood cells via polycistronic lentiviral transduction (Human STEMCCA Cre-Excisable constitutive polycistronic [OKS/L-Myc] Lentivirus Reprogramming Kit, Millipore) and characterized with a hES/iPS cell pluripotency RT-PCR kit. Detailed methods about iPSC charecterization and generation are described in PMID:26281015.

	The iPSCs were differentiated into pure iPSC-RGCs cells with structural and functional features characteristic of native RGC cells based on a novel methodology we developed (PMID: 32678240)
Mycoplasma contamination	All the iPSC lines used for the study are negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Details on the mice used for single cell RNA sequencing in this study, including housing conditions, have been included in the Methods section of this manuscript. Mef2c-Cre strain was provided by Dr. Alan Daugherty, University of Kentucky; Rosa26-mTmG mice are available from the Jackson Laboratory (JAX) as strain B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J(stock number 007676). Eight weeks old double-heterozygous Mef2c-Cre;Rosa26-mTmG male mice were used. All of the mice were on a C57BL/6 background.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	Mice were maintained in the Center for Comparative Medicine at Baylor College of Medicine, and procedures were performed according to a protocol AN-4195 approved by the Institutional Animal Care and Use Committee at Baylor College of Medicine. All animal experiments complied with the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	A subset of 10,900 individuals in PMBB were analyzed in the discovery phase of this study, as described in Table 1. They have a median age of 67, and have a variety of disease phenotypes as described in Table 1. The two most prevalent ancestries, Europeans (75.2%) and Africans (19.9%), were analyzed for exome-by-phenome-wide association analyses. For replication analyses in BioMe, the cohort consisted of 23,989 individuals with 59% being female and having a median age of 61. 6,470 individuals were of African ancestry, 8,735 were of European ancestry, and 8,784 were of Hispanic ancestry. They have a variety of disease phenotypes as described in Table S18. For replication analyses in DiscovEHR, the cohort consisted of 85,450 individuals of European ancestry with a median age of 59. They have a variety of disease phenotypes as described in Table S18. For replication analyses in UK Biobank, the cohort consisted of 32,268 individuals of European ancestry with a median age of 59. They have a variety of disease phenotypes as described in Table S18. For replication analyses in BioVU, the cohort consisted of 66,400 genotype individuals with a median age of 56 and 56.4% female. 10,456 individuals are of African ancestry and 55,944 individuals are of European ancestry. The retinal eye sections in this study were obtained from a normal cadaver eye globe from a 68 year-old, female, Caucasian donor. For scRNA-seq of human pancreas cell types in type 1 diabetes (T1D) versus control, 5 individuals with T1D and 6 control individuals were recruited for this study. T1D individuals were 50% female, and had a median age of 29.5 and median BMI of 21.25. Control individuals were 60% female, and had a median age of 13 and median BMI of 17.3. All individuals were of Caucasian race. For scRNA-seq of human aorta, 11 individuals were recruited for this study. Ascending aortic samples were acquired from 3 controls (2 female and 1 male, heart transplant recipient or lung transplant donor) and 8 individuals with ascending thoracic aortic aneurysm (4 female and 4 male). 54.5% were female and the median age was 63. 81.8% were of White race, 9.1% Black, and 9.1% Latino.
Recruitment	All individuals who were recruited for the Penn Medicine Biobank (PMBB) are patients of clinical practice sites of the University of Pennsylvania Health System. Appropriate consent was obtained from each participant regarding storage of biological specimens, genetic sequencing, access to all available electronic health record (EHR) data, and permission to recontact for future studies. All genotypic and phenotypic data were de-identified prior to analyses. For replication analyses in BioMe, samples were ascertained based on the patient population of Mount Sinai who enrolled at various Mount Sinai clinical sites in and around the New York City region. For replication analyses in DiscovEHR, all samples were drawn from MyCode participants, whom provided informed consent that allows their clinical and genomic data to be used for health research (PMID: 26866580 and PMID: 28008009).

For replication analyses in UK Biobank, access to the UK Biobank for this project was from Application 32133.

For replication analyses in BioVU, all samples were drawn from participating Vanderbilt clinic patients who provided consent.

For scRNA-seq of human pancreas, pancreatic islets were procured from the HPAP consortium under Human Islet Research Network (https://hirnetwork.org/). To avoid self-selection bias in donor selection, all high-quality single cells from all donor samples that were collected by HPAP at the time of the preparation of this manuscript with high quality scRNA-seq data were used in this analysis. To avoid self-selection bias in cell clustering, two independent algorithms were employed, Seurat and TooManyCells. To avoid self-selection biases in defining cell types, we used Garnett, a regression-based classifier to assist in the automation of cell type classification

For scRNA-seq of human aorta, ascending aortic samples were acquired from 3 controls (2 female and 1 male, heart transplant recipient or lung transplant donor) and 8 individuals with ascending thoracic aortic aneurysm (4 female and 4 male). Furthermore, two control samples came from heart transplant recipients. Although those patients did not have aortic aneurysms, they may have exhibited molecular or cellular changes in the ascending aorta related to their cardiac disease. However, as this dataset was used to identify gene expression in particular aortic cell types regardless of disease to mirror the scRNA-seq studies of mouse aorta, these differences are unlikely to change the conclusions of the results.

Ethics oversight

The study was approved by the Institutional Review Board of the University of Pennsylvania and complied with the principles set out in the Declaration of Helsinki.

Replication analyses in BioMe were approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai.

For replication analyses in DiscovEHR, the Geisinger IRB and MyCode Governing Board both reviewed and approved the use of MyCode data for this study.

All replication analyses in BioVU were approved under Vanderbilt IRB #200350.

For replication analyses in UK Biobank, access to the UK Biobank for this project was from Application 32133.

The donor eye tissue used for retinal eye sections in this study are exempt from IRB approvals (Exception 4).

For scRNA-seq of human pancreas, pancreatic islets were procured from the HPAP consortium under Human Islet Research Network (https://hirnetwork.org/) with approval from the University of Florida Institutional Review Board (IRB # 201600029) and the United Network for Organ Sharing (UNOS). A legal representative for each donor provided informed consent prior to organ retrieval.

For scRNA-seq of human aorta, the protocol for collecting human aortic tissue samples was approved by the Institutional Review Board at Baylor College of Medicine. Written informed consent was provided by all participants before enrollment. All experiments conducted with human tissue samples were performed in accordance with the relevant guidelines and regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.