

## RESEARCH ARTICLE

## Genetically regulated gene expression and proteins revealed discordant effects

Janne Pott<sup>1,2\*</sup>, Tarcyane Garcia<sup>1</sup>, Stefanie M. Hauck<sup>3</sup>, Agnese Petrera<sup>3</sup>, Kerstin Wirkner<sup>1,2</sup>, Markus Loeffler<sup>1,2</sup>, Holger Kirsten<sup>1,2</sup>, Annette Peters<sup>3,4,5</sup>, Markus Scholz<sup>1,2\*</sup>

**1** Institute for Medical Informatics, Statistics and Epidemiology, Medical Faculty, University of Leipzig, Leipzig, Germany, **2** LIFE Research Center for Civilization Diseases, Medical Faculty, University of Leipzig, Leipzig, Germany, **3** Research Unit Protein Science and Metabolomics and Proteomics Core Facility, Helmholtz Zentrum Munich - German Research Center for Environmental Health, Neuherberg, Germany, **4** German Center for Diabetes Research (DZD), München-Neuherberg, Germany, **5** Chair of Epidemiology, Institute for Medical Information Processing, Biometry and Epidemiology, Medical Faculty, Ludwig-Maximilians-Universität München, Munich, Germany

\* [janne.pott@imise.uni-leipzig.de](mailto:janne.pott@imise.uni-leipzig.de) (JP); [markus.scholz@imise.uni-leipzig.de](mailto:markus.scholz@imise.uni-leipzig.de) (MS)



## OPEN ACCESS

**Citation:** Pott J, Garcia T, Hauck SM, Petrera A, Wirkner K, Loeffler M, et al. (2022) Genetically regulated gene expression and proteins revealed discordant effects. *PLoS ONE* 17(5): e0268815. <https://doi.org/10.1371/journal.pone.0268815>

**Editor:** Jie V Zhao, University of Hong Kong, HONG KONG

**Received:** June 18, 2021

**Accepted:** May 10, 2022

**Published:** May 23, 2022

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0268815>

**Copyright:** © 2022 Pott et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All summary statistics are publicly available from Zenodo (DOI: [10.5281/zenodo.6045694](https://doi.org/10.5281/zenodo.6045694)). Scripts used in the secondary analyses are included at [https://github.com/GenStatLeipzig/LWAS\\_Olink](https://github.com/GenStatLeipzig/LWAS_Olink). Complete data

## Abstract

## Background

Although gene-expression (GE) and protein levels are typically strongly genetically regulated, their correlation is known to be low. Here we investigate this phenomenon by focusing on the genetic background of this correlation in order to understand the similarities and differences in the genetic regulation of these omics layers.

## Methods and results

We performed locus-wide association studies of 92 protein levels measured in whole blood for 2,014 samples of European ancestry and found that 66 are genetically regulated. Three female- and one male-specific effects were detected. We estimated the genetically regulated GE for all significant genes in 49 GTEx v8 tissues. A total of 7 proteins showed negative correlations with their respective GE across multiple tissues. Finally, we tested for causal links of GE on protein expression via Mendelian Randomization, and confirmed a negative causal effect of GE on protein level for five of these genes in a total of 63 gene-tissue pairs: BLMH, CASP3, CXCL16, IL6R, and SFTPD. For IL6R, we replicated the negative causal effect on coronary-artery disease (CAD), while its GE was positively linked to CAD.

## Conclusion

While total GE and protein levels are only weakly correlated, we found high correlations between their genetically regulated components across multiple tissues. Of note, strong negative causal effects of tissue-specific GE on five protein levels were detected. Causal network analyses revealed that GE effects on CAD risks was in general mediated by protein levels.

sets including genetic data cannot be shared publicly due to ethical and legal restrictions, as they are sufficient to identify study participants. This is not covered by the informed consent form of the LIFE-Adult study. Data are available from the LIFE Research Center (contact via Dr. Matthias Nüchter, Head of Managing Office, E-mail: [matthias.nuechter@life.uni-leipzig.de](mailto:matthias.nuechter@life.uni-leipzig.de)) for researchers who meet the criteria for access to confidential data.

**Funding:** LIFE-Adult is funded by the Leipzig Research Center for Civilization Diseases (LIFE). LIFE is an organizational unit affiliated to the Medical Faculty of the University of Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by funds of the Free State of Saxony within the framework of the excellence initiative. Olink measurements were funded by the HI-MAG project "Serum proteome biomarkers as mediators of cardiometabolic disease development" of the Medical Faculty of the University Leipzig and the Helmholtz Zentrum München. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Several large-scale genome-wide association studies (GWASs) identified more than 150 genetic risk loci of coronary artery disease (CAD) [1–6]. However, for the majority of loci, the underlying molecular pathology remains to be elucidated. High-throughput proteomics could contribute to our understanding of molecular patho-mechanisms by providing functional causal links between genetic loci, proteome expressions and cardiovascular disease traits.

While there is typically a strong relationship between genetics and transcriptomics via cis expression quantitative trait loci (eQTLs), some studies have shown only weak correlations between the transcriptomic and the proteomic layer [7–10]. Possible reasons comprise different half-lives of mRNA and respective protein, posttranscriptional modifications and tissue and compartment specificity [11]. Nevertheless, genetic effects on expression and protein (proteome quantitative trait loci—pQTL) levels are partly overlapping suggesting common genetic drivers. For example, the Framingham Heart Study detected 26 pQTLs in cis that overlapped with respective eQTLs in whole blood, liver and heart tissues [12]. He et al. [13] analyzed the liver-specific proteome on genome-wide scale and found for about 40% of all tested genes an overlap with known eQTLs in liver.

In this study, we aimed at identifying cis-pQTLs for a panel of 92 biomarkers of CAD measured by proximity extension assays in blood. To characterize these cis loci in more detail, we compared the effects of cis-eQTLs and pQTLs at these loci in more detail. For this purpose, we analyzed the overlap of eQTLs and pQTLs by co-localization analyses and tested for association of genetically regulated gene expression (GE) across tissues and respective blood protein expression (PE). Finally, the identified genetic associations were used to establish causal chains of genetics, transcriptomics, proteomics and CAD via concatenated Mendelian Randomization analyses.

## Material and methods

### Cohort description

All analyses were performed in participants of the LIFE-Adult study. In LIFE-Adult, 10,000 residents of the city of Leipzig, Germany were randomly recruited in an age- and sex-stratified manner. All participants were deeply examined with respect to civilization diseases such as obesity, diabetes, cardiovascular diseases, cognitive impairment and mental disorders as well as contributing environmental and life-style factors. Details can be found in Loeffler et al. [14]. Blood samples were taken from all participants after an overnight fasting and were stored in the Leipzig Medical Biobank for subsequent analyzes and measurements of genetic, transcriptomic and proteomic data. Overlap of OMICs data is displayed in S1 Fig in [S1 File](#).

LIFE-Adult meets the ethical standards of the Declaration of Helsinki and is approved by the Ethics Committee of the Medical Faculty of the University Leipzig, Germany (Reg. No 263-2009-14122009). Written informed consent including agreement with genetic analyses was obtained from all participants. A basic description of samples used in this study can be found in [Table 1](#).

### Protein biomarker measurement

For proteomic profiling, we selected EDTA plasma probes of 2,024 elderly LIFE-Adult participants. Measurement of 92 CVD-related protein biomarkers was performed with the proximity extension assay (PEA) [15] using the Olink CVD Panel III. Measurements were performed in 23 batches each including 88 samples and two identical controls each. For eight samples, Olink

**Table 1. Basic sample description.**

Variable	Overall	Female	Male	P-value
	(n = 2,014)	(n = 974)	(n = 1,040)	
Age (years)	62.5 (11.5)	62.0 (11.3)	62.9 (11.6)	<b>2.47E-02</b>
BMI (kg/m <sup>2</sup> )	27.7 (4.5)	27.2 (4.8)	28.1 (4.1)	<b>1.02E-06</b>
Current smoker	320 (16.7%)	147 (16.2%)	173 (17.3%)	5.49E-01
Hypertension <sup>a</sup>	1166 (58.7%)	496 (51.9%)	670 (65.1%)	<b>2.91E-09</b>
Type 2 diabetes <sup>a</sup>	479 (23.8%)	214 (22.0%)	265 (25.5%)	7.25E-02
Statin therapy <sup>b</sup>	319 (15.9%)	124 (12.8%)	195 (18.8%)	<b>2.99E-04</b>
TC (mmol/l)	5.70 (1.06)	5.87 (1.06)	5.53 (1.04)	<b>3.35E-13</b>
LDL-C (mmol/l)	3.58 (0.95)	3.59 (0.96)	3.57 (0.95)	5.88E-01
HDL-C (mmol/l)	1.62 (0.46)	1.81 (0.46)	1.44 (0.39)	<b>2.86E-73</b>

For continuous parameters, the unit is given in parenthesis, and arithmetic mean and standard deviation values are shown. For binary variables, total number and percentages are provided. Differences between sexes were tested with a chi-squared test for all binary parameters, and with Mann-Whitney U test for all continuous parameters. Abbreviations: BMI, body mass index; TC, total cholesterol; LDL-C, low-density lipoprotein; HDL-C, high-density lipoprotein.

<sup>a</sup> anamnestic, medication or determined by HbA1c > 6.5%;

<sup>b</sup> ATC-code beginning with C10.

<https://doi.org/10.1371/journal.pone.0268815.t001>

measurement failed, and additional two samples were excluded as outliers (Mahalanobis Distance > 3 IQR), resulting in N = 2,014 samples available for further analyses.

Measurements of biomarkers are available for all of these samples except for three biomarkers. Two assays (BLMH, CTSD) failed at one plate, resulting in N = 1,926 for these traits. A single missing value of metalloproteinase 4 was mean-imputed. Across all 92 assays in CVD III, the mean intra-assay (within run) and inter-assay (between runs) variations expressed as coefficients of variation are reported to be 8.1% and 11.5%, respectively. We used normalized protein expression units as semi-quantitative traits. Genetic data were available for all of the samples. An overview of all biomarkers including their distribution and genetic regions is given in S1 Table in [S2 File](#).

## Gene expression measurement

Isolated mRNA from whole blood of 3,527 samples was hybridized to Illumina HT-12 v4 Expression BeadChips (Illumina, San Diego, CA, USA) and gene expression (GE) was measured on the Illumina HiScan (47,231 raw GE probes). We then processed the data by log<sub>2</sub>-transformation, quantile-normalization [16, 17] and correction for batch effects [18] using R/Bioconductor.

Probes were excluded if they were (1) expressed in less than 5% of the samples, (2) still significantly associated with batch effects, or (3) unable to map to a gene according to ingenuity pathway analyses (IPA, QIAGEN Inc., accessed on 2019-04-04). In summary, 20,972 valid GE probes remained, corresponding to 15,950 genes. We looked for transcripts corresponding to the 92 proteome features of the PEA. There were 91 probes with sufficient QC, matching to 68 unique genes.

Samples were removed if (1) the number of detected GE probes deviated more than 3\*IQR from the median, (2) the Mahalanobis distance of several quality characteristics deviated more than 3\*IQRs from the median [19], or (3) the Euclidean distance of expression values deviated more than 4\*IQRs from the median [16]. Overall, of the assayed 3,527 samples, 110 had to be

removed for quality reasons. Of those remaining, 3,194 samples had also genetic data available in high quality.

## Genotyping & Imputation

A total of 7,838 participants of LIFE-Adult were genotyped on the genome-wide SNP array Axiom CEU1 (Affymetrix). Genotype calling was performed using the software Affymetrix Power Tools (version 1.20.06). We conducted calling and quality control according to Affymetrix's best practice steps [20].

SNPs were excluded if (1) their call rate was less than 97%, (2) there was a significant violation of Hardy-Weinberg equilibrium ( $p < 1 \times 10^{-6}$  for autosomal SNPs,  $p < 1 \times 10^{-4}$  for X-chromosomal SNPs in women only), (3) significant plate association ( $p < 1 \times 10^{-7}$ ), or (4) cluster plot specific parameters according to Affymetrix's recommendation [20].

Samples were removed if (1) their signal contrast on the array was low ( $< 0.82$ ), (2) their call rate was less than 97%, (3) the estimated sex differed from the sex retrieved from the databank, (4) cryptic relatedness was observed ( $> 0.6$  [21]), or (5) the estimated genetic ethnicity was out of range ( $> 6 \times \text{SD}$  in any of the first 10 principal components). There were 33 ethnic outliers, which were removed for all further analyses (see S2 Fig in S1 File). After filtering, LIFE-Adult was genetically homogeneous and we therefore refrained from correcting for population stratification via PCs in the main analyses, but included the first ten PCs in a sensitivity analysis of all lead SNPs per protein.

We imputed our SNP data on the reference 1000 Genome Phase 3 [22] using SHAPEIT [23] v2r900 for prephasing and IMPUTE2 [24] v2.3.2 for genotype estimation. For this study, all SNPs with minor allele frequency (MAF)  $< 1\%$  or imputation info score  $< 0.8$  were excluded, resulting in 9,033,656 SNPs for further analyses.

## Statistical analysis

An overview of our analysis plan is given in S3 Fig in S1 File.

**Genetic association analyses for 92 protein biomarkers.** For each of the 92 biomarkers we performed genetic association analyses at the regions of the gene coding for the biomarker, i.e. we searched for cis-pQTLs, only. The region between gene start -500 kb and gene stop +500 kb was considered (see S1 Table in S2 File for the assumed starts and stops of genes). Primary genetic association analysis was done in all subjects ( $n = 2,014$ ) with PE adjusted for age and sex. In a secondary analyses, we ran sex-stratified analyses ( $n = 974$  female,  $n = 1,040$  male) adjusting PE for age. For the analyses, we used the additive frequentist model with expected genotype counts as implemented in PLINK 2.0 [25]. We lifted our data from hg19 to hg38 using the GWAS Summary Statistics harmonization tool [26].

We pooled all cis-regions of our primary analysis and performed a hierarchical FDR correction as suggested for eQTLs by Peterson et al. [27]. In more detail, we first applied Benjamini & Hochberg (BH) [28] correction of all SNPs associations calculate for a specific PE and identified the SNP with the minimal corrected p-value (Simes p-value). Next, we applied BH on the 92 Simes p-values and tested with  $\alpha_1 = 0.05$  to determine the  $k$  proteins showing significant associations. We then used  $\alpha_2 = 0.05 \times k/92$  as significance threshold on the first level as proposed by Benjamini & Bogomolov (BB) [29]. The SNP with lowest and significant p-value was denoted as lead cis-pQTL of the respective protein. We then merged all significant associations and pruned the variants to a subset of markers that are in approximate linkage equilibrium with each other ( $r^2 < 0.1$ ). Linkage disequilibrium (LD) was calculated using all LIFE-Adult participants. Finally, we annotated these independent variants with (1) other nearby genes (Ensemble, +/- 250 kb of SNP position) [30], (2) known traits associations from the GWAS

Catalog (LD  $r^2 > 0.3$ ) [31], (3) known cis-eQTLs (LD  $r^2 > 0.3$ ,  $\alpha_{cis} = 0.05$ ) [32–35], and (4) CADD scores as measure of deleteriousness [36]. We defined novel loci as regions whose lead SNP was not in LD with a variant reported for blood protein biomarker levels in the GWAS catalog (LD  $r^2 \leq 0.3$ ).

For all lead cis-pQTLs, we checked for sex-specific effects on PE and compared effect sizes between females and males applying t-tests of beta estimates [37]. We also looked for sex-specific significant loci by applying the same hierarchical FDR correction as mentioned above. Finally, we looked up the eQTL summary statistics of all GTEx v8 [38] tissues for all lead SNPs and their associated genes and compared their effect direction with our pQTL findings. The GTEx data used for the analyses described in this manuscript were obtained from the GTEx Portal on 09.06.2020. We reported those for whole blood and the (second) best associated tissue. In addition, per locus we retrieved the best cis-eQTLs (defined by lowest p-value per tissue and gene) and calculated pairwise LD ( $r^2$ ) to the respective lead pQTLs. To validate our findings, we performed whole blood cis-eQTL analysis in our LIFE data (using  $n = 3,194$  samples with gene expression in whole blood and genetic data).

#### **Co-localization and association analyses between gene-expression and protein levels.**

In order to investigate the link between gene expression and protein levels in more detail, we performed three locus-wise analyses: First, we performed a pairwise co-localization test [39] between our pQTLs and eQTLs obtained from GTEx v8. In more detail, this method tested if two trait associations share the same causal variant, regardless of effect direction. Five hypotheses were tested in parallel, of which H4 states that the traits share the causal SNP, while H3 assumes two independent signals. As threshold for co-localization and for independence, we used a posterior probability for H4 and H3 of  $\geq 0.75$ , respectively. The region of co-localization was defined as the position of the lead pQTL +/- 500 kb. We used the R-package “coloc” for this analysis [39].

In a second analysis, the summary statistics of all proteomic features with significant cis-pQTLs were used to search for correlations with respective genetically estimated gene-expression (gGE) using the MetaXcan approach [40]. The expression prediction models were downloaded from the github repository [41] (see also PredictDB [42]; GTEx v8 models using elastic net algorithm). PredictDB contains only models that passed stringent criteria (e.g., number of SNPs used, posterior probability for being an eQTL). Hence, not all gene—tissue combinations were available for this analysis. In total, we tested 2,242 tissue-specific gGE for protein association. To adjust for multiple testing of several tissues per protein, we performed a hierarchical FDR correction as mentioned above. The first level were the tissues per protein, the second level were the analyzed proteins. We report findings in whole blood and the (second) best associated tissue.

Finally, we validated the MetaXcan results obtained for blood tissue using our measured gene-expression profiles. Raw gene-expression data were available for 45 of the 64 genes with pQTLs and paired GE/proteome data were available for 1048 samples. We estimated both Pearson’s correlation and Pearson’s partial correlation controlling for sex, age, percentage of lymphocytes and percentage of monocytes on total white blood cells. We repeated this analysis in the sex-stratified subsets.

**Mendelian randomization analyses.** To investigate whether observed associations between GE and PE were causal, we performed Mendelian Randomization (MR) analyses. As MR requires strong instruments, we used the best-associated cis-eQTLs per tissue with lowest p-value and  $p < 5 \times 10^{-8}$  ( $n = 428$  SNPs for a total of 58 genes). To adjust for multiple testing, we performed a hierarchical FDR correction as mentioned above. The first level were the tissues, the second level were the analyzed genes.



Since the proteins on our array were supposed to be cardio-vascular biomarkers, we also estimated the causal effects of protein levels on CAD. We considered only lead pQTLs reaching  $p < 5 \times 10^{-8}$  as instruments ( $n = 48$  proteins). P-values of MR were adjusted using Bonferroni correction for 48 tests. For proteins with significant causal effect on CAD, we tested for causal chains  $GE \rightarrow PE \rightarrow CAD$  [43].

In all analyses, we used the ratio method and estimated the standard error using the first two terms of the delta method [44]. Summary statistics were obtained from our pQTL-analyses, from GTEx v8 [38] and from van der Harst et al. [6].

## Results

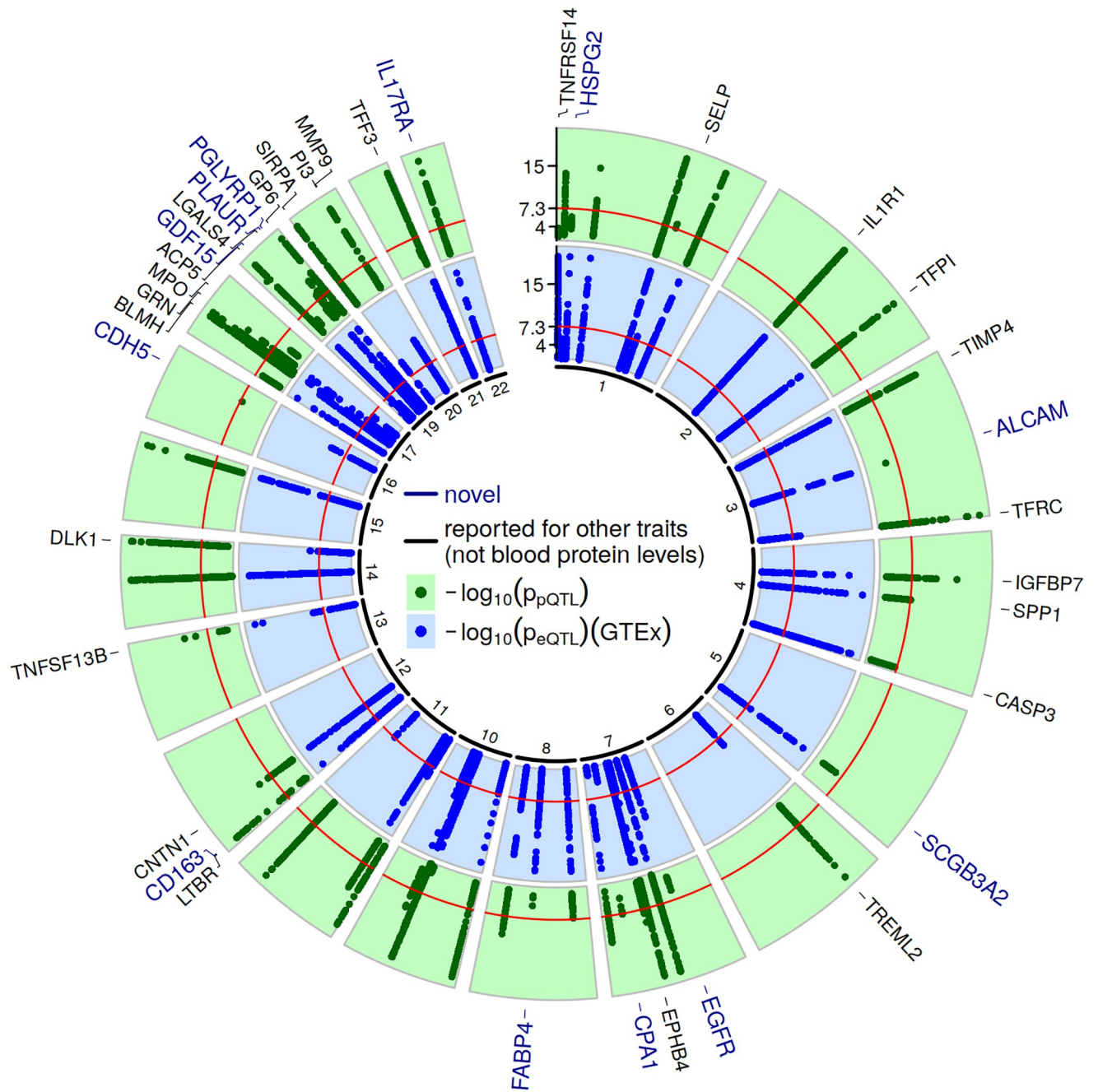
An overview of all 92 analyzed proteins, their abbreviations and full name is given in S1 Table in [S2 File](#). In the following, only the gene name abbreviations are used, with the regular written names referring to PE and italic written names to GE. In addition, all main results are included in this table as TRUE/FALSE vectors, which summarize the genetic associations, the GE correlations and the causal analyzes per protein (for the combined setting).

### SNP level results

After applying hierarchical FDR, we detected for 64 biomarkers significant associations in or nearby the corresponding gene (23,951 unique SNPs, see S2 Table in [S2 File](#) for an overview of all Simes p-values). Priority pruning revealed 758 independent SNPs (see S3 Table in [S2 File](#) for summary statistics and full annotation with nearby genes, GWAS catalog traits and enriched pathways, and S4 Table in [S2 File](#) for lead pQTLs per protein). Several of these loci were already described for association with blood protein levels ( $n = 27$  loci) [45–49]. Of the remaining 37 loci, 25 were previously reported for other traits (e.g. lipids, CAD related traits, or blood fractions), while 12 loci were not reported for any trait associations so far. These 37 loci are considered novel for our protein traits. [Fig 1](#) shows a circular plot of all cis regions and  $-\log_{10}$  transformed association p-values of our association study and those of the best GTEx tissue per gene. Although we did not perform a classic GWAS, 16 of the 37 novel and 22 of the other loci also reached the classic genome-wide significance threshold of  $p < 5 \times 10^{-8}$  (regarding the Simes p-value).

In our sex-stratified approach, we detected 54 proteins significantly associated in men, and 48 in women (see S2 Table in [S2 File](#)). Five proteins were associated in males, but not in females, and had lower Simes p-values compared to the combined setting, suggesting male-specific loci (GDF15, MPO, PAI, OPN, and TFF3). Of note, PAI was only associated in the male setting, not in the combined one. Similarly, NOTCH3 was only associated in the female setting, but not in the combined or male setting. In the following, all 66 proteins with association in at least one setting are analyzed. Regarding the 110 lead SNPs of all settings (both from the combined and sex-stratified analyses, if other lead SNPs were detected here), we observed for 18 of them a significant difference in effect size in men and women, but only five of them survived multiple testing correction (S4 Fig in [S1 File](#)), including NOTCH3 and MPO. We reported the sex-stratified results in S4 Table in [S2 File](#). In our sensitivity analyses additionally adjusting for the first ten principal components, we found no significant bias in our results, as the effect estimates were the same, and their p-value increased only slightly. All associations remained significant according to our FDR threshold (see S4 Table in [S2 File](#)).

Next, we searched the GTEx database for associations of our 110 lead pQTLs with GE of the corresponding genes. We found 7,714 such associations across all 49 GTEx tissues. We checked the direction consistency of the e- and pQTLs, and surprisingly, detected for 41 proteins at least one discordant direction. For 13 of them, this discordant direction was observed



**Fig 1. Circular plot of cis-associations.** Log-transformed p-values for cis-pQTLs and eQTLs are shown in the green respectively blue circle. We obtained the statistics for eQTLs from GTEx and present the results of the tissue with the strongest eQTL per gene (see S3 Table in S2 File). For plotting, the y-axis was restricted to  $-\log(p) = 20$ , i.e. all larger  $-\log(p)$  values were set to 20. The red circles mark the classical genome-wide significance threshold ( $p = 5 \times 10^{-8}$ ). Gene names are added for loci not yet described for blood protein levels, and are colored with respect to the novelty level (blue: not described for any other traits, black: reported for other traits except for blood protein levels).

<https://doi.org/10.1371/journal.pone.0268815.g001>

in most of the associated tissues (more than 75% of tissues in which the eQTL was observed, see Table 2 and Fig 2). Restricted to whole blood, there were ten QTLs with discordant effects. To validate this finding in whole blood, we replicated the eQTL analysis in our LIFE data (GE available for 45 genes). Here, four of the ten SNPs were associated with  $p < 0.05$  and showed

Table 2. Comparison of effect direction of cis-pQTLs from our GWAS and cis-eQTLs from GTEx.

Locus Information		pQTL		eQTL GTEx whole blood and (sec.) best tissue		
Protein (ratio)	pQTL effect allele / EAF	beta	p-value	beta	p-value	Tissue
IL6R (17/18)	rs4129267 T / 0.380	0.421	2.96x10 <sup>-323</sup>	-0.087	5.40x10 <sup>-09</sup>	Artery Tib.
CCL15 (20/21)	rs41436444 CAGGGCAG / 0.080	0.939	4.56x10 <sup>-292</sup>	-	-	Lung
CCL16 (3/4)	rs10445391 G / 0.072	-0.852	8.01x10 <sup>-155</sup>	-	-	Thyroid
<b>SFTPD</b> (42/42)	rs721917 G / 0.406	-0.475	1.37x10 <sup>-95</sup>	0.113	3.39x10 <sup>-03</sup>	Artery Tib.
<b>BLMH</b> (24/26)	rs7214248 A / 0.346	0.193	4.29x10 <sup>-61</sup>	-0.072	3.80x10 <sup>-04</sup>	Artery Tib.
<b>ACP5</b> (35/37)	rs897811 C / 0.116	-0.163	1.70x10 <sup>-20</sup>	0.501	5.78x10 <sup>-19</sup>	Thyroid
TIMP4 (16/17)	rs392394 A / 0.782	0.140	4.54x10 <sup>-17</sup>	-0.024	4.36x10 <sup>-01</sup>	Artery Tib.
TNFRSF11B (2/2)	rs11300005 C / 0.493	0.073	2.70x10 <sup>-11</sup>	-	-	Eso. Mus.
AXL (28/28)	rs3786556 T / 0.184	-0.446	9.39x10 <sup>-16</sup>	-0.002	9.68x10 <sup>-01</sup>	Artery Tib.
CPA1 (7/9)	rs35454128 C / 0.120	0.197	3.79x10 <sup>-08</sup>	0.117	8.18x10 <sup>-02</sup>	Adipose Sub.
CASP3 (14/15)	rs6845294 A / 0.687	0.153	1.09x10 <sup>-07</sup>	0.026	1.54x10 <sup>-01</sup>	Cells fibro.
CDH5 (3/3)	rs16956504 C / 0.110	0.105	1.10x10 <sup>-07</sup>	-0.248	5.35x10 <sup>-17</sup>	Pituitary
<b>CXCL16</b> (25/25)	rs145042193 T / 0.208	-0.064	8.22x10 <sup>-07</sup>	-0.161	2.14x10 <sup>-02</sup>	Cells fibro.
				0.104	2.47x10 <sup>-03</sup>	
				0.362	4.92x10 <sup>-17</sup>	

We show results of the 13 genes for which discordant effect directions between pQTL and most of the respective eQTLs (more than 75% of all significant eQTLs across tissues) were observed. We also report eQTLs of whole blood and the best-associated tissue in GTEx. The effect allele and its frequency is given below of the respective SNP ID. For four genes we could replicate the different effect direction in our LIFE data (marked in bold, see S2 Table in S2 File for more details).

<https://doi.org/10.1371/journal.pone.0268815.t002>

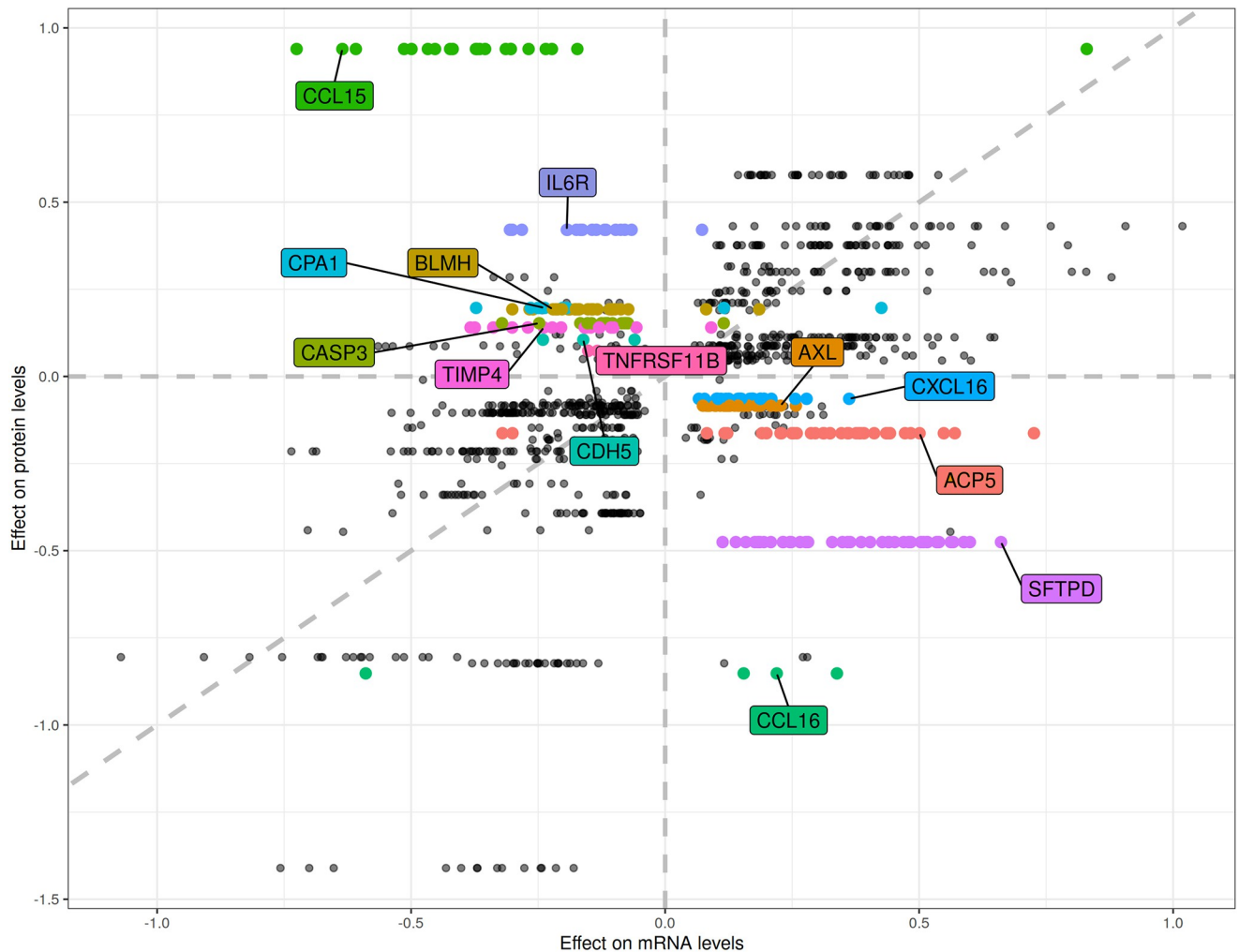
discordant effect directions when compared to the corresponding pQTL (SFTPD, BLMH, ACP5, and CXCL16). The other SNPs had the same effect direction or showed no significant effect in our data.

### Locus level results

Although all lead pQTLs were associated with GE in at least one tissue, only in 4% of the GE-tissue combinations the best eQTL was also the best pQTL, and in 30% the two SNPs were in some LD ( $r^2 > 0.1$ ). To determine whether these signals are inter-related, we performed co-localization analyses and tested for an association of genetically regulated gene expression and protein levels. A summary of these tissue-specific analyses is shown as Venn diagram (S5 Fig in S1 File) and in S6 Table in S2 File.

We observed 50 proteins with at least one shared (PP4 > 75%) GE signal, and 42 with at least one independent signal (PP3 > 75%) across tissues. A total of 34 proteins show both, dependent and independent signals in different tissues. Posterior probabilities for all pairs can





**Fig 2. Scatter plot of effect estimates of the 63 lead pQTL-SNPs on gene expression and protein levels.** As the focus was on the direction only, we did not normalize the effect estimates. Only SNPs with a significant eQTL in at least one tissue are displayed ( $p < 0.05$  in GTEx). Results of the 13 genes showing discordant pQTL and eQTL directions in more than 75% of eQTL tissues are labeled (see also Table 2 and S5 Table in S2 File).

<https://doi.org/10.1371/journal.pone.0268815.g002>

be found in S7 Table in S2 File and are displayed in S6 Fig in S1 File. We compared the gene-tissue combinations of shared and independent signaling with those of high and low LD between best eQTL and pQTL. Regarding the higher LD ( $r^2 > 0.1$ ) combinations, the distribution between shared and independent signals was almost the same ( $n = 212$  with  $PP3 > 0.75$ ,  $n = 295$  with  $PP4 > 0.75$ ). This demonstrates that LD does not guarantee co-localization. In contrast, for low LD pairs, there was a clear trend to independent signals ( $n = 189$  with  $PP3 > 0.75$ ,  $n = 14$  with  $PP4 > 0.75$ ).

We estimated the genetically regulated gene expression in all GTEx v8 tissues using MetaXcan. Here, several SNPs at each gene locus were selected and included into the GE prediction model. The predicted GE was then tested for association with the respective protein. After applying hierarchical FDR, we detected significant associations for 58 of the 64 considered biomarkers in at least one tissue ( $n = 1,474$  significant tests of a total of 2,242). Most genes showed this association in about half of all tissues (median of 27 associated tissues). Counterexamples are LBTR and IL17RA, showing GE-PE association across many tissues (49, respectively 42 tissues).

**Table 3. Proteins showing predominantly negative correlation between tissue-specific GE and blood PE.**

Protein	#tissues (neg/tot)	Effect	P-value	Tissue
IL6R	25/29	-2.177	$7.11 \times 10^{-303}$	Artery Tib.
		-2.965	$8.21 \times 10^{-197}$	WB
SFTPD	33/40	-6.003	$7.69 \times 10^{-77}$	Colon Trans.
BLMH	12/15	-0.985	$1.48 \times 10^{-58}$	Artery Tib.
		-1.052	$8.11 \times 10^{-15}$	WB
TIMP4	29/36	-0.371	$8.00 \times 10^{-17}$	Brain Ant. Cingulate Cortex
CASP3	25/29	0.865	$1.51 \times 10^{-7}$	Brain Putamen
		-109.229	$2.78 \times 10^{-7}$	WB
CXCL16	41/41	-0.242	$6.70 \times 10^{-7}$	Kidney Cortex
		-0.067	$1.54 \times 10^{-2}$	WB
TREML2	13/14	-40.610	$5.85 \times 10^{-6}$	Thyroid
		-0.876	$3.58 \times 10^{-4}$	WB

MetaXcan results are shown for the best-associated tissue with negative effect estimate and whole blood (WB), if a prediction model was available. Neg = Number of significant negative correlations across tissues. Tot = Total number of significant associations.

<https://doi.org/10.1371/journal.pone.0268815.t003>

We compare the MetaXcan results with our results obtained by co-localization analyses. The intersection of significant associations and co-localization comprised  $n = 531$  gene-tissue pairs, of which  $n = 288$  pairs showed co-localization and  $n = 243$  indicated independent signaling. This demonstrates that results of MetaXcan-based gene-expression association analysis are only loosely related to those of co-localization. We summarized all MetaXcan results in S8 Table in [S2 File](#).

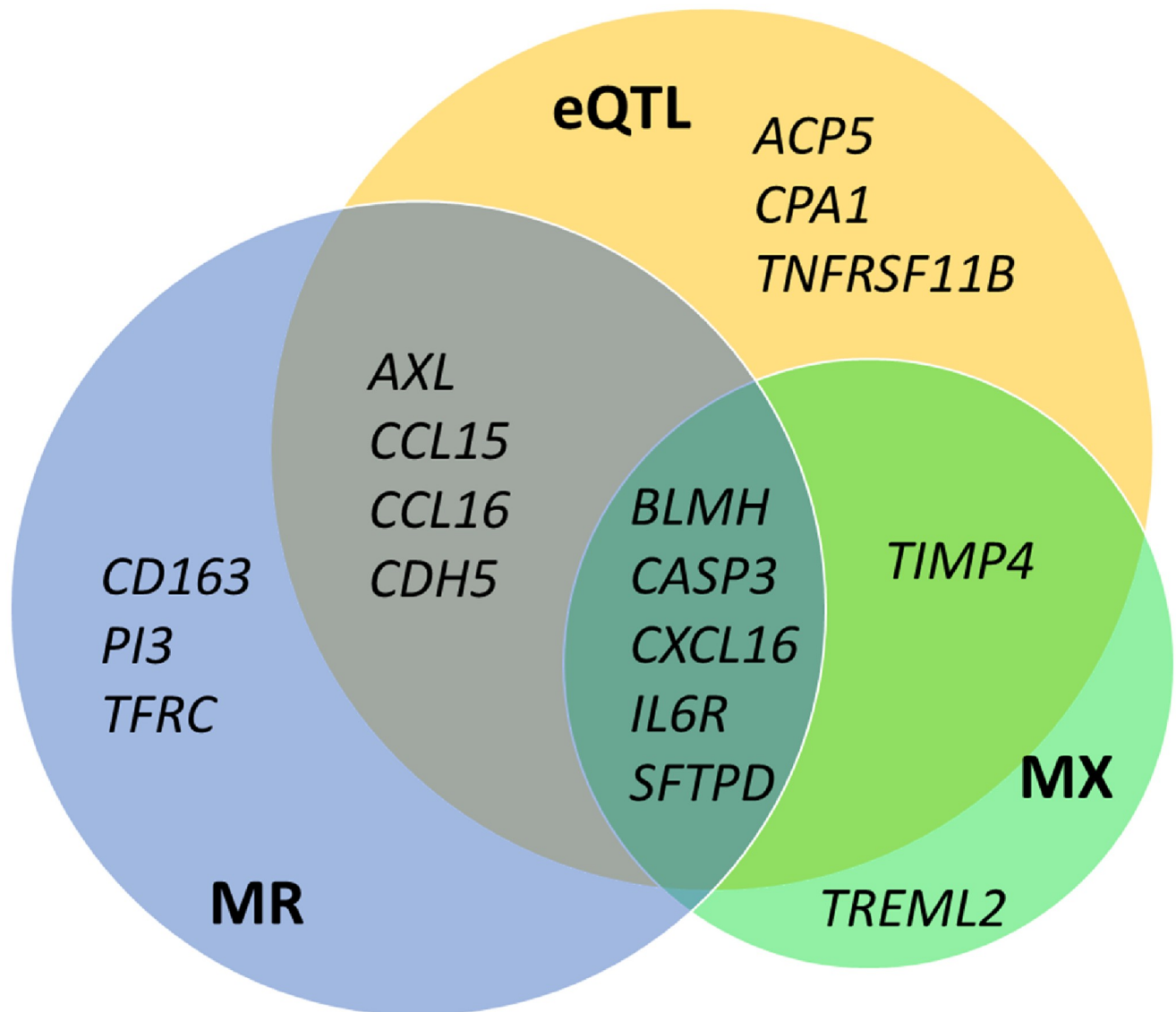
We checked the direction of the correlation of tissue-specific GE and PE. A total of 42 proteins showed opposite direction of effects in at least one tissue. Seven of them showed this negative association in most of the tissues (see [Table 3](#)), including six which were also found based on the LD considerations performed in the previous paragraph (see [Table 2](#)).

Finally, we compared the MetaXcan-derived GE-protein associations with GE-protein associations based on raw GE data of whole blood from our LIFE study. GE data were available for 45 genes. Among those, we detected 21 significant partial correlations controlling for age, sex and white blood cell counts (S9 Table in [S2 File](#)). Comparing these results with the respective MetaXcan results of whole blood, we found eight pairs that are significant in both analyses. For all of them, the same effect direction was observed, with negative correlation of PE, and both, total GE and genetically estimated GE in whole blood.

### Causal network of gene expression, protein levels and CAD

We performed Mendelian Randomization analyses of the causal relationship between GE and PE for all tissues for which a strong eQTL ( $p < 5 \times 10^{-8}$ ) was available. Accordingly, we tested 58 genes in up to 27 tissues ( $n = 670$  tests in the combined setting). We detected causal links for 51 genes (501 genes-tissue pairs). There were predominantly positive effects (364 pairs with positive causal effect of GE on PE). A summary of all instruments, tissues, and causal effects is given in S10 Table in [S2 File](#).

For 419 pairs of GE and protein, we found both, significant MetaXcan association and MR effect. Among those, 398 showed concordant effect directions between GE and PE (92 pairs with negative effect, 306 with positive effect), i.e. effect directions are in large agreement. The 92 pairs with negative effect comprise 20 unique genes. Twelve of them show this relation in >75% of associated tissues, including five genes described in the previous sections (BLMH,



**Fig 3. Overlap of genes with negative correlation to protein levels according to different analysis strategies.** We present the genes with negative protein correlations for more than 75% of the significant tissues according to MetaXcan (MX), negative causal effect estimates according to Mendelian Randomization (MR), and opposite effect direction of eQTLs and pQTLs (see also Tables 2–4).

<https://doi.org/10.1371/journal.pone.0268815.g003>

CASP3, CXCL16, IL6R, and SFTPD). Summaries of analysis results of these genes are shown in Fig 3 and S7 Fig in S1 File, and Table 4.

Next, we tested the causal link between the biomarkers and coronary artery disease. We restricted the analyses to lead pQTLs with  $p < 5 \times 10^{-8}$  and available CAD statistics from van der Haarst et al. [6]. This left us with 47 biomarkers. Results are given in S11 Table in S2 File and a scatter plot for all pairs is shown in S8 Fig in S1 File. Four protein showed a significant effect in the combined setting: IL6R ( $\beta_{IV} = -0.094$ ,  $p = 4.90 \times 10^{-14}$ ), PCSK9 ( $\beta_{IV} = 0.540$ ,  $p = 4.72 \times 10^{-10}$ ), TFPI ( $\beta_{IV} = -0.116$ ,  $p = 7.83 \times 10^{-5}$ ), and AXL ( $\beta_{IV} = 0.487$ ,  $p = 6.09 \times 10^{-5}$ ). IL6R had significant causal estimates in the sex-stratified settings as well. The estimated causal TFPI effect was also significant in females, but in males it reached only nominal significance that did not

Table 4. Proteins with predominantly negative causal links of GE and PE.

Protein	#tissues (neg/tot)	Causal effect estimate	P-value MR	Association effect estimate	P-value MetaXcan	Tissue
IL6R	5/5	-0.713	5.57x10 <sup>-27</sup>	0.126	4.63x10 <sup>-06</sup>	Testis
		-3.008	1.35x10 <sup>-16</sup>	-2.965	8.21x10 <sup>-197</sup>	WB
SFTPD	22/25	-0.347	2.64x10 <sup>-17</sup>	-0.050	5.68x10 <sup>-04</sup>	Heart AA
		-0.749	6.43x10 <sup>-06</sup>			WB
BLMH	2/2	-0.873	3.07x10 <sup>-15</sup>	-0.985	1.48x10 <sup>-58</sup>	Artery Tib.
CCL15	15/16	-0.459	1.02x10 <sup>-14</sup>			Nerve Tib.
PI3	1/1	-0.731	1.02x10 <sup>-08</sup>			WB
AXL	6/6	-0.243	6.03x10 <sup>-06</sup>			Artery Tib.
CCL16	15/15	-0.181	1.28x10 <sup>-05</sup>			Liver
CASP3	5/6	-0.440	1.89x10 <sup>-05</sup>	-0.173	1.57x10 <sup>-06</sup>	Cells fibro.
		0.914	1.83x10 <sup>-05</sup>	0.865	1.51x10 <sup>-07</sup>	WB
TFRC	1/1	-0.404	7.94x10 <sup>-05</sup>			Lung
CXCL16	17/17	-0.225	1.69x10 <sup>-04</sup>	-0.213	1.68x10 <sup>-05</sup>	Thyroid
		-0.097	2.97x10 <sup>-03</sup>	-0.067	1.54x10 <sup>-02</sup>	WB
CDH5	1/1	-0.148	1.92x10 <sup>-02</sup>			Thyroid
CD163	1/1	-0.129	3.66x10 <sup>-02</sup>			Testis

MR results are shown for the best tissue with negative effect estimate and whole blood, if available. We added respective MetaXcan results for comparison.

Neg = Number of significant negative causal estimate across tissue. Tot = Total number of significant MR tests.

<https://doi.org/10.1371/journal.pone.0268815.t004>

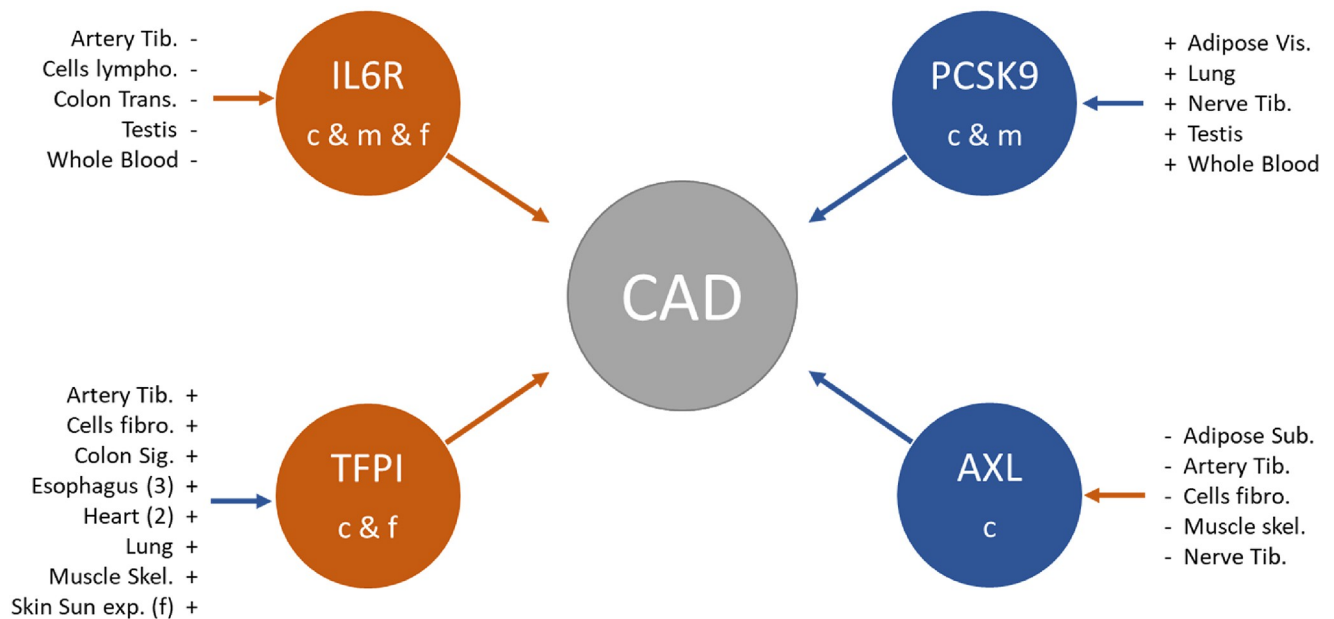
survive multiple testing correction. PCSK9 was also causally linked in males, while in females the PCSK9 instrument did not reach the significance threshold to be included for MR analysis. For AXL, both sex-stratified instruments were above the MR significance threshold and hence excluded.

Finally, we searched for causal chains from GE over PE to CAD for all four proteins with significant causal link to CAD, which were also causally affected by gene expression ( $n = 50$  tests). The total causal effect estimates of GE on CAD were significant in 26 gene-tissue pairs and negative in four of them (*TFPI* in tibial artery and cultured fibroblast cells in all and females). *AXL* was not causally linked to CAD in any tissue. The indirect effect was estimated as product of the  $GE \rightarrow PE$  and  $PE \rightarrow CAD$  effects, which corresponds to the effect of GE on CAD mediated by PE. These indirect effect estimates were significant for all 50 gene-tissue pairs and for 45 of them no significant difference between the total and indirect GE-CAD effect was observed, indicating complete mediation of GE via PE towards CAD. We summarized the results in S12 Table in S2 File and displayed the causal chains in Fig 4.

## Discussion

In this work, we performed a genetic cis-association analysis of 92 cardiovascular biomarkers and found 66 regulations. We used these signals to unravel the relationship of cis-pQTLs and cis-eQTLs by (1) testing for co-localization of signals, (2) analyzing the correlation of genetically regulated GE and PE, and (3) testing for causal effects of the GE/PE associations. Finally, we established causal chains of GE, PE and CAD across tissues.

In our study, we focus on cis-effects rather than a whole-genome hypothesis-free approach. Similar to eQTL analyses, cis-effects tend to be true positives, while trans-effects are often false positives requiring much more stringent false positive control limiting power of this type of analysis. We check the GWAS catalog for known associations, and detected that 37 of our



**Fig 4. Graphical overview of causal networks from gene expression (GE) over protein expression (PE) on the outcome coronary artery disease (CAD).** Orange arrows indicate a negative causal effect of protein level on CAD (IL6R and TFPI) or of gene expression on protein levels, blue arrows denote positive links. The settings are indicated by c (combined), m (males) and f (females). Tissues in which gene expression showed a significant indirect effect on CAD are listed next to their gene. Bold tissues indicate also a significant direct effect.

<https://doi.org/10.1371/journal.pone.0268815.g004>

associations are novel. Moreover, for 38 of the 66 associated proteins, our lead SNP achieved genome-wide significance ( $p < 5 \times 10^{-8}$ ).

As example, *PI3* was a genome-wide significant and novel loci, which was also associated in the sex-stratified analyses. Here, it showed a significant sex-related effect, i.e. the effect estimate in men was twice that of women. We tested our own GE and eQTL data for this interaction but could not confirm this sex-dimorphism at the GE layer. Moreover, we detected a co-localization of eQTLs and pQTLs for men (whole blood), but not for women. In both MetaXcan and MR, the estimates were significant in each sex, but with stronger effect in men (MR in whole blood:  $\beta_{\text{men}} = -0.990$ ,  $\beta_{\text{women}} = -0.481$ ,  $p_{\text{IA}} = 0.002$ ). *PI3* codes for elafin, which has been linked to the inflammatory response in atherosclerosis [50] and myocardial infarction [51]. Most atherosclerotic outcomes show sexual dimorphism as well, with higher risk for men. This makes elafin an interesting target for further studies of sex-dimorphisms in cardiovascular research.

To unravel the relationship between GE and PE, we performed several analyses across different tissues since the origin of plasma PE is not necessarily whole blood. We found that the lead pQTL was in most cases not the best-associated eQTL. We tested pairwise LD between eQTLs and pQTLs and found most of them in low LD ( $r^2 < 0.1$ ) across tissues and proteins. When comparing LD results with our co-localization results, we found as expected that amongst the low LD pairs signals are often predicted as independent. More surprisingly, for the high LD pairs, the ratio between independence and co-localization was balanced, i.e. LD does not ensure co-localization of the signals. For example, both *IL17RA* and *SFTPD* have high LD pairs in 34 and 39 tissues, respectively. *IL17RA* showed co-localization of these signals in all these tissues. In contrast, for *SFTPD*, co-localization was refuted for all these signals.

In general, we observed a good agreement of MetaXcan and MR results although the MetaXcan approach does not show causality *per se* and is also based on different gene-models



compared to the instruments used for MR. In contrast, we observed only a moderate overlap between co-localization results and MetaXcan / MR results. Signals with co-localization but no significant causal estimate could be explained by pleiotropic effects while causality but lack of co-localization could be explained by locus heterogeneity, i.e. different causal e- and pQTLs. However, most of the detected associations were found in for high LD pairs of e- and pQTL (Co-localization: 295 of 313 gene tissue pairs [94.2%]; MetaXcan: 659 of 1,474 [44.7%], MR: 357 of 501 [71.3%]).

Most interestingly, we detected five gene-protein pairs, which consistently showed opposite effect directions of eQTLs and pQTLs, negative correlation of GE and PE (MetaXcan) and negative causal effects (Mendelian Randomization). Those were BLMH, CASP3, CXCL16, IL6R and SFTPD. While BLMH was co-localized in 10 of 12 analyzed tissues, the other four had one co-localizing tissue each, but all other tissues suggested independent signals. Only BLMH was co-localized in most tissues. We discuss the functional relevance of this observation in more detail by a GTEx v8 look-up of gene expression levels by tissue [38]. While the highest rate of BLMH expression occurs in skin tissue, it is also expressed in artery tissues, for which we observed negative association and in tibial artery tissue also negative causal effect. High expression levels in GTEx v8 [38] and negative links in our analyses were also found for CXCL16 in tissues testis, whole blood, and skin, and for IL6R in tissues muscle skeletal, whole blood, esophagus muscularis, and colon transverse. CASP3 had negative links between GE and PE in several tissues, but also four positive links for brain (substantia nigra), liver, pancreas and whole blood tissue. The negative links were detected in tissues with higher GE, e.g. cells cultured fibroblasts. SFTPD is highly expressed in lung tissue, in which we found independent e- and pQTL signals and a positive causal estimates. In other tissues with lower SFTPD expression, negative links were observed, e.g. thyroid. Thus, in the last case, the relationship of GE and PE could be dominated by single tissues showing a positive correlation, while for the other four the negative links happen in tissues with substantial gene expression. This suggests functional relevance of BLMH, CASP3, CXCL16, and IL6R that needs further biological validation.

Consistent negative effects between GE and PE are of particular interest to further studies, since the mechanism behind this is not clear. Explanations for this observation could be (1) tissue-specific protein levels that differ from those measured in whole blood; (2) whole blood acting only as transport compartment to a specific target tissue; (3) upregulation of pathways, in which the protein is further metabolized; (4) post-translational modifications that influence protein degradation; or (5) upregulation of genes in response to increased consumption / degradation of protein.

Among our results, IL6R showed the most pronounced effects, with causal negative effects in whole blood, testis, artery tibial and colon transverse tissues. We speculate that inflammatory conditions consume IL6R resulting in low plasma levels but increased gene-expression to counter the IL6R loss. The negative effect of IL6R on CAD was previously reported by Yuan et al. [52] and could be explained by reduced inflammation. However, the GE effect on CAD was positive, which is mainly mediated by IL6R PE.

In our MR mediation analyses, we found several indirect causal links of tissue-specific GE over PE in whole blood on CAD. Only for TFPI, we found indirect effects from heart-specific tissues (atrial appendage and left ventricle). Although these two tissues might be more specific for CAD, cis-effects are usually shared across tissues, with exception of brain tissues [38]. Hence, other tissues such as muscle and whole blood with larger sample sizes might still detect the true contribution of the gene expression. In addition, we do not know which tissue our measured proteins come from. Therefore, it is also possible that increased GE in CAD-unrelated tissues lead to higher blood protein levels. The proteins can be transported to heart tissues, where they increase the risk for an event. The detected direct effects of *AXL* in four

tissues could be false positives, given the comparison to a non-significant total effect. The direct effect of PCSK9 gene expression occurs in Adipose tissue (visceral omentum), where it is only weakly expressed (TPM = 0.27). This needs further biological validation.

One limitation of this study was its relatively small sample size, and with it, reduced power. We therefore refrained from analyzing trans-pQTLs and focused on cis-effects. Larger studies of meta-analyses are required to resolve this limitation. The observed trans-associations could then be used as independent instruments for a bivariate Mendelian Randomization analysis, checking if high protein levels showed reverse causality on the gene expressions, an issue which could not be addressed by our study.

In conclusion, we discovered several causal links of tissue-specific gene-expression and blood protein levels of cardiovascular biomarkers. Observed negative causal relationships are of interest for further studies to unravel the underlying post-transcriptional or pathway-associated regulatory processes. Finally, we established a causal patho-mechanistic network of GE and PE of IL6R, PCSK9, TFPI, and AXL and coronary artery disease providing possible new therapy targets.

## Supporting information

**S1 File.**  
(PDF)

**S2 File.**  
(XLSX)

## Acknowledgments

We thank all study participants of the LIFE-Adult study whose personal dedication and commitment have made this project possible. LIFE-Adult genotyping (round 3) was done at the Cologne Center for Genomics (CCG, University of Cologne, Peter Nürnberg and Mohammad R. Toliat). For genotype imputation, compute infrastructure provided by ScaDS (Dresden/Leipzig Competence Center for Scalable Data Services and Solutions) at the Leipzig University Computing Centre was used.

## Author Contributions

**Conceptualization:** Markus Scholz.

**Data curation:** Tarcyane Garcia, Stefanie M. Hauck, Agnese Petrera, Kerstin Wirkner, Holger Kirsten.

**Formal analysis:** Janne Pott.

**Funding acquisition:** Markus Loeffler, Annette Peters.

**Investigation:** Janne Pott, Stefanie M. Hauck, Agnese Petrera, Markus Loeffler, Markus Scholz.

**Methodology:** Janne Pott.

**Project administration:** Markus Scholz.

**Resources:** Markus Loeffler, Annette Peters.

**Software:** Holger Kirsten.

**Supervision:** Markus Scholz.

**Visualization:** Janne Pott.

**Writing – original draft:** Janne Pott.

**Writing – review & editing:** Markus Scholz.

## References

1. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015; 47:1121–30. Epub 2015/09/07. <https://doi.org/10.1038/ng.3396> PMID: 26343387.
2. Howson JMM, Zhao W, Barnes DR, Ho W-K, Young R, Paul DS, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet.* 2017; 49:1113–9. Epub 2017/05/22. <https://doi.org/10.1038/ng.3874> PMID: 28530674.
3. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet.* 2017; 49:1392–7. Epub 2017/07/17. <https://doi.org/10.1038/ng.3914> PMID: 28714974.
4. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet.* 2017; 49:1385–91. Epub 2017/07/17. <https://doi.org/10.1038/ng.3913> PMID: 28714975.
5. Verweij N, Eppinga RN, Hagemeijer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Sci Rep.* 2017; 7:2761. Epub 2017/06/05. <https://doi.org/10.1038/s41598-017-03062-8> PMID: 28584231.
6. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res.* 2018; 122:433–43. Epub 2017/12/06. <https://doi.org/10.1161/CIRCRESAHA.117.312086> PMID: 29212778.
7. Chen G, Gharib TG, Huang C-C, Taylor JMG, Misk DE, Kardina SLR, et al. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics.* 2002; 1:304–13. <https://doi.org/10.1074/mcp.m200008-mcp200> PMID: 12096112.
8. Pascal LE, True LD, Campbell DS, Deutsch EW, Risk M, Coleman IM, et al. Correlation of mRNA and protein levels: cell type-specific gene expression of cluster designation antigens in the prostate. *BMC Genomics.* 2008; 9:246. Epub 2008/05/23. <https://doi.org/10.1186/1471-2164-9-246> PMID: 18501003.
9. Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, et al. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet.* 2011; 7:e1001393. Epub 2011/06/09. <https://doi.org/10.1371/journal.pgen.1001393> PMID: 21695224.
10. Yeung ES. Genome-wide correlation between mRNA and protein in a single cell. *Angew Chem Int Ed Engl.* 2011; 50:583–5. <https://doi.org/10.1002/anie.201005969> PMID: 21226136.
11. Haider S, Pal R. Integrated analysis of transcriptomic and proteomic data. *Curr Genomics.* 2013; 14:91–110. <https://doi.org/10.2174/1389202911314020003> PMID: 24082820.
12. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat Commun.* 2018; 9:3268. Epub 2018/08/15. <https://doi.org/10.1038/s41467-018-05512-x> PMID: 30111768.
13. He B, Shi J, Wang X, Jiang H, Zhu H-J. Genome-wide pQTL analysis of protein expression regulatory networks in the human liver. *BMC Biol.* 2020; 18:97. Epub 2020/08/10. <https://doi.org/10.1186/s12915-020-00830-3> PMID: 32778093.
14. Loeffler M, Engel C, Ahnert P, Alfermann D, Arelin K, Baber R, et al. The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health.* 2015; 15:691. Epub 2015/07/22. <https://doi.org/10.1186/s12889-015-1983-z> PMID: 26197779.
15. Assarsson E, Lundberg M, Holmquist G, Björkstén J, Thorsen SB, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One.* 2014; 9:e95192. Epub 2014/04/22. <https://doi.org/10.1371/journal.pone.0095192> PMID: 24755770.
16. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008; 24:1547–8. Epub 2008/05/08. <https://doi.org/10.1093/bioinformatics/btn224> PMID: 18467348.
17. Schmid R, Baum P, Ittrich C, Fundel-Clemens K, Huber W, Brors B, et al. Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3. *BMC Genomics.* 2010; 11:349. Epub 2010/06/02. <https://doi.org/10.1186/1471-2164-11-349> PMID: 20525181.

18. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8:118–27. Epub 2006/04/21. <https://doi.org/10.1093/biostatistics/kxj037> PMID: 16632515.
19. Cohen Freue GV, Hollander Z, Shen E, Zamar RH, Balshaw R, Scherer A, et al. MDQC: a new quality assessment method for microarrays based on quality control reports. *Bioinformatics*. 2007; 23:3162–9. Epub 2007/10/12. <https://doi.org/10.1093/bioinformatics/btm487> PMID: 17933854.
20. Affymetrix. Axiom™ Analysis Suite. UserGuide. 2015. <http://www.affymetrix.com/support/technical/byproduct.affx?product=axiomanalysissuite>.
21. Wang J. An estimator for pairwise relatedness using molecular markers. *Genetics*. 2002; 160:1203–15. <https://doi.org/10.1093/genetics/160.3.1203> PMID: 11901134
22. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015; 526:68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245.
23. Delaneau O, Howie B, Cox AJ, Zagury J-F, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet*. 2013; 93:687–96. <https://doi.org/10.1016/j.ajhg.2013.09.002> PMID: 24094745.
24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; 5:e1000529. Epub 2009/06/19. <https://doi.org/10.1371/journal.pgen.1000529> PMID: 19543373.
25. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4:7. Epub 2015/02/25. <https://doi.org/10.1186/s13742-015-0047-8> PMID: 25722852.
26. Barbeira AN, Vairus L. summary-gwas-imputation. harmonization, liftover, and imputation of summary statistics from GWAS. GitHub repository: IM-Lab; 2020. <https://github.com/hakyimlab/summary-gwas-imputation>.
27. Peterson CB, Bogomolov M, Benjamini Y, Sabatti C. Many Phenotypes Without Many False Discoveries: Error Controlling Strategies for Multitrait Association Studies. *Genet Epidemiol*. 2016; 40:45–56. Epub 2015/12/02. <https://doi.org/10.1002/gepi.21942> PMID: 26626037.
28. Benjamini Yoav, and Hochberg Yosef. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995; 57:289–300. Available from: [www.jstor.org/stable/2346101](http://www.jstor.org/stable/2346101).
29. Benjamini Y, Bogomolov M. Selective inference on multiple families of hypotheses. *J R Stat Soc B*. 2014; 76:297–318. <https://doi.org/10.1111/rssb.12028>
30. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhaj J, et al. Ensembl 2018. *Nucleic Acids Res*. 2018; 46:D754–D761. <https://doi.org/10.1093/nar/gkx1098> PMID: 29155950.
31. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019; 47:D1005–D1012. <https://doi.org/10.1093/nar/gky1120> PMID: 30445434.
32. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature*. 2017; 550:204–13. <https://doi.org/10.1038/nature24277> PMID: 29022597.
33. Joehanes R, Zhang X, Huan T, Yao C, Ying S-X, Nguyen QT, et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol*. 2017; 18:16. Epub 2017/01/25. <https://doi.org/10.1186/s13059-016-1142-6> PMID: 28122634.
34. Kirsten H, Al-Hasani H, Holdt L, Gross A, Beutner F, Krohn K, et al. Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *Hum Mol Genet*. 2015; 24:4746–63. Epub 2015/05/27. <https://doi.org/10.1093/hmg/ddv194> PMID: 26019233.
35. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013; 45:1238–43. Epub 2013/09/08. <https://doi.org/10.1038/ng.2756> PMID: 24013639.
36. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019; 47:D886–D894. <https://doi.org/10.1093/nar/gky1016> PMID: 30371827.
37. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003; 326:219. <https://doi.org/10.1136/bmj.326.7382.219> PMID: 12543843.
38. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020; 369:1318–30. <https://doi.org/10.1126/science.aaz1776> PMID: 32913098.
39. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*.

- 2014; 10:e1004383. Epub 2014/05/15. <https://doi.org/10.1371/journal.pgen.1004383> PMID: 24830394.
40. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.* 2019; 15:e1007889. Epub 2019/01/22. <https://doi.org/10.1371/journal.pgen.1007889> PMID: 30668570.
  41. Barbeira AN, PrediXcan Team. MetaXcan. GitHub repository: IM-Lab; 2021. <https://github.com/hakyimlab/MetaXcan>.
  42. PredictDB Team. GTEx v8 models on eQTL and sQTL. PredictDB: Im Lab, Genetic Medicine, Department of Medicine, The University of Chicago. [cited 21 May 2021]. <https://predictdb.org/post/2021/07/21/gtex-v8-models-on-eqtl-and-sqtl/>.
  43. Burgess S, Daniel RM, Butterworth AS, Thompson SG. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol.* 2015; 44:484–95. Epub 2014/08/22. <https://doi.org/10.1093/ije/dyu176> PMID: 25150977.
  44. Burgess S, Thompson SG. Mendelian Randomization. *Methods for Using Genetic Variants in Causal Estimation.* 1st ed. CRC Press; 2015.
  45. Sun W, Kechris K, Jacobson S, Drummond MB, Hawkins GA, Yang J, et al. Common Genetic Polymorphisms Influence Blood Biomarker Measurements in COPD. *PLoS Genet.* 2016; 12:e1006011. Epub 2016/08/17. <https://doi.org/10.1371/journal.pgen.1006011> PMID: 27532455.
  46. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun.* 2017; 8:14357. Epub 2017/02/27. <https://doi.org/10.1038/ncomms14357> PMID: 28240269.
  47. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* 2017; 13:e1006706. Epub 2017/04/03. <https://doi.org/10.1371/journal.pgen.1006706> PMID: 28369058.
  48. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018; 558:73–9. Epub 2018/06/06. <https://doi.org/10.1038/s41586-018-0175-2> PMID: 29875488.
  49. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science.* 2018; 361:769–73. Epub 2018/08/02. <https://doi.org/10.1126/science.aag1327> PMID: 30072576.
  50. Henriksen PA, Hitt M, Xing Z, Wang J, Haslett C, Riemersma RA, et al. Adenoviral gene delivery of elafin and secretory leukocyte protease inhibitor attenuates NF-kappa B-dependent inflammatory responses of human endothelial cells and macrophages to atherogenic stimuli. *J Immunol.* 2004; 172:4535–44. <https://doi.org/10.4049/jimmunol.172.7.4535> PMID: 15034071.
  51. Shavadia JS, Granger CB, Alemayehu W, Westerhout CM, Povsic TJ, Brener SJ, et al. High-throughput targeted proteomics discovery approach and spontaneous reperfusion in ST-segment elevation myocardial infarction. *Am Heart J.* 2020; 220:137–44. Epub 2019/11/09. <https://doi.org/10.1016/j.ahj.2019.09.015> PMID: 31812755.
  52. Yuan S, Lin A, He Q-Q, Burgess S, Larsson SC. Circulating interleukins in relation to coronary artery disease, atrial fibrillation and ischemic stroke and its subtypes: A two-sample Mendelian randomization study. *Int J Cardiol.* 2020; 313:99–104. Epub 2020/03/21. <https://doi.org/10.1016/j.ijcard.2020.03.053> PMID: 32223966.