The American Journal of Human Genetics, Volume *108*

Supplemental data

Genetic control of the human brain proteome

Chloe Robins, Yue Liu, Wen Fan, Duc M. Duong, Jacob Meigs, Nadia V. Harerimana, Ekaterina S. Gerasimov, Eric B. Dammer, David J. Cutler, Thomas G. Beach, Eric M. Reiman, Philip L. De Jager, David A. Bennett, James J. Lah, Aliza P. Wingo, Allan I. Levey, Nicholas T. Seyfried, and Thomas S. Wingo

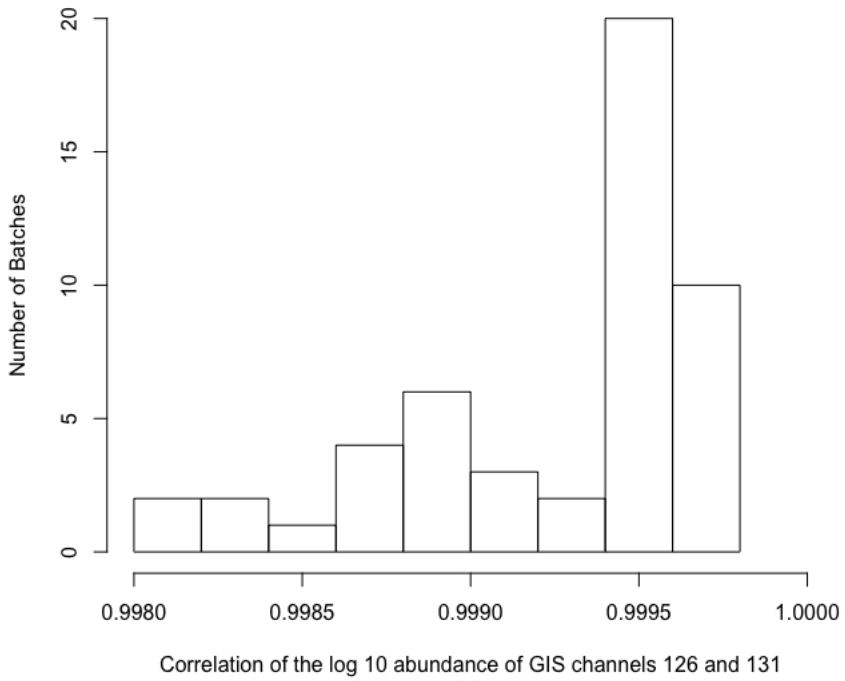

Figure S1. Distribution of the batch-specific correlation of GIS channels. Each TMT proteomic experiment, or batch, contains two GIS channels (126 and 131). Here we show the distribution of correlations of proteomic measurements between the batch-specific GIS channels.

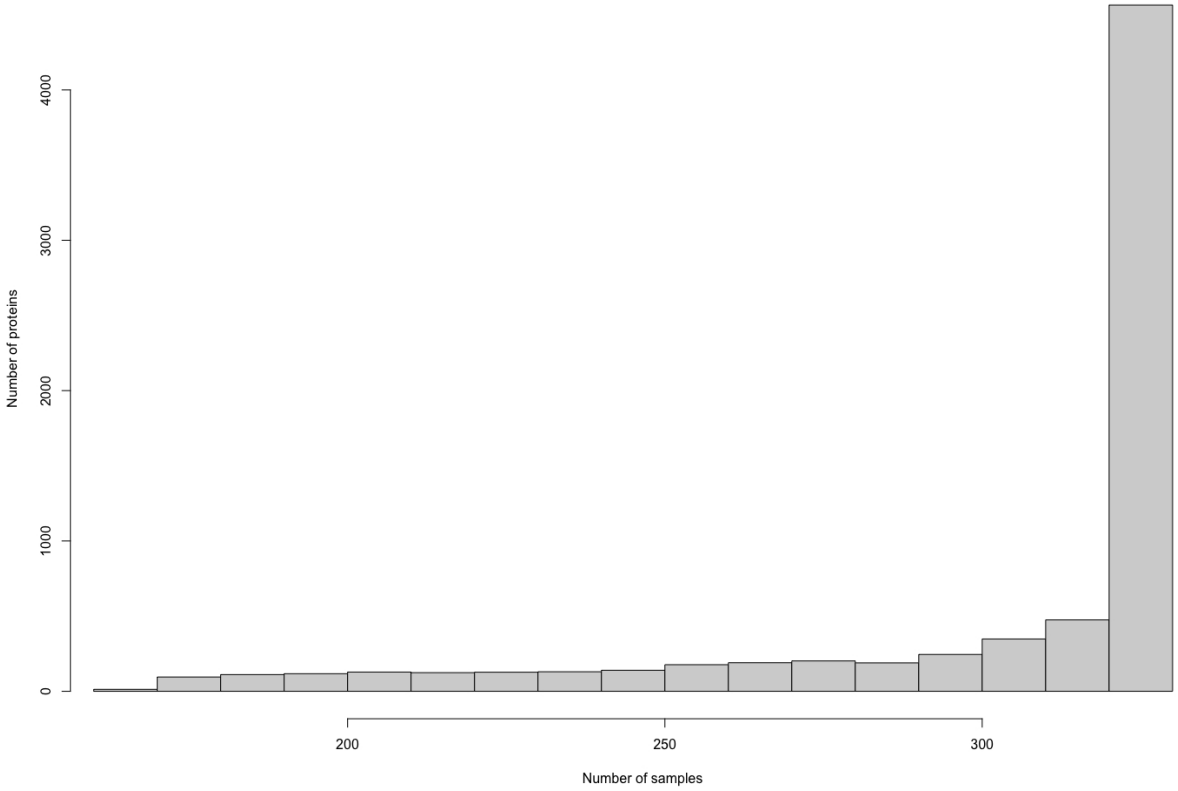

Figure S2. Distribution of the sample sizes used in the pQTL analyses. Due to the highly batch-specific nature of protein measurement in TMT proteomic experiments, each measured protein has a different sample size. This histogram shows the distribution of sample sizes across tested proteins.

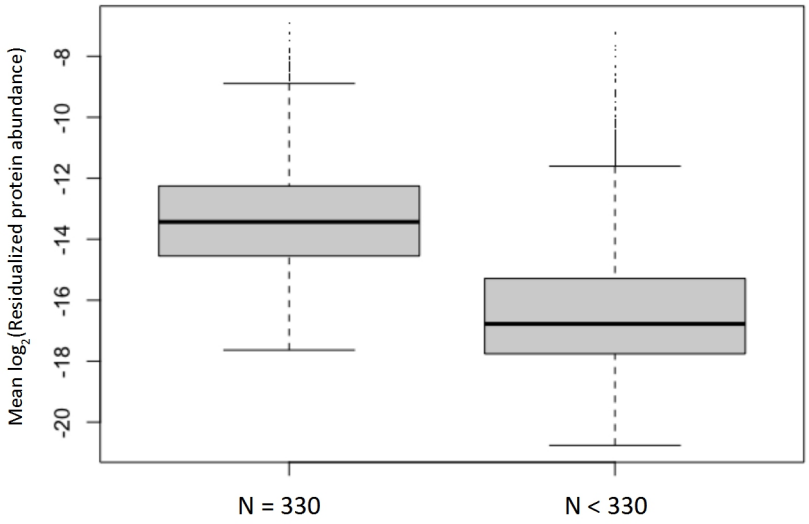

Figure S3. Comparison of the distribution of average protein abundances for proteins measured in all 330 participants (N = 3,843 proteins) vs. those with missing data (N = 4,173 proteins). The difference in distribution is significant by Kruskal-Wallis test (χ^2= 3378.7, p < $2.2x10^{-16}$).

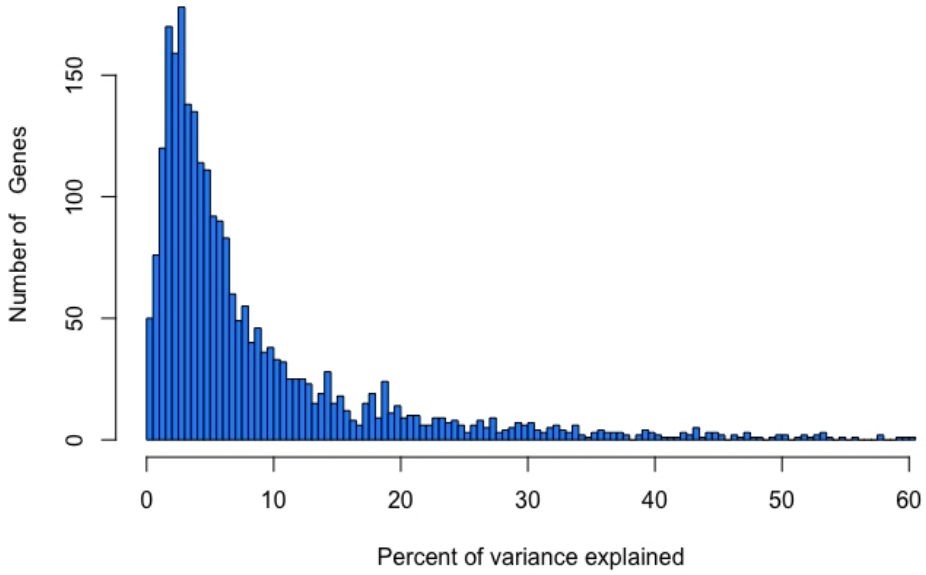

Figure S4. Percentage of variance in protein abundance explained by genotype. For the 2,474 genes with a genetic variant that significantly predicts protein abundance, we used stepwise linear regression to identify all independent pQTLs and assess the proportion of variance in protein abundance explained. The median and mean percentage of variance in protein abundance explained by pQTLs is 4.9% and 8.5% respectively.

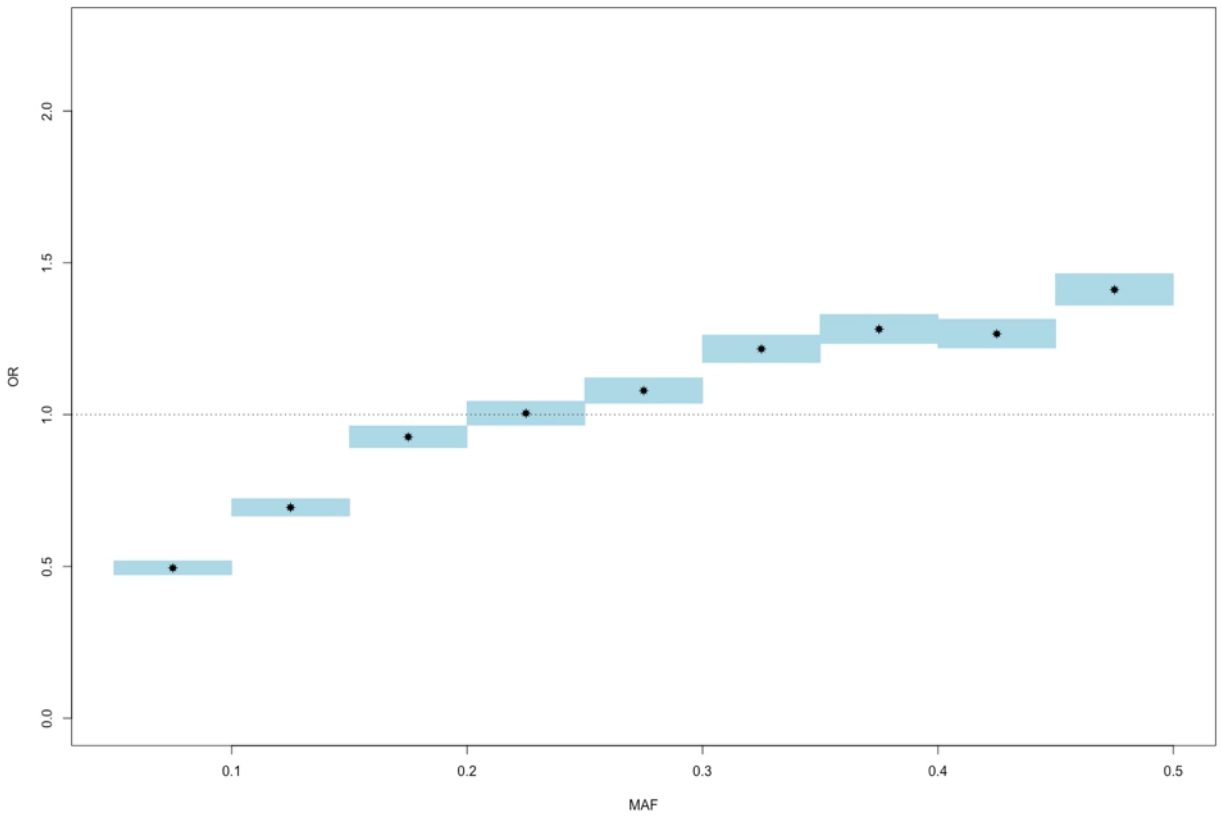

Figure S5. Enrichment of pQTL identification by MAF. Each blue rectangle represents the results of a Fisher's exact test. Each test compared the set of SNVs with a MAF within the range delineated by the blue rectangle and the set of SNVs identified as a pQTL. The height of the dot in the center of each rectangle shows the odds ratio estimate, while the estimate's 95% confidence interval is shown as the height of the rectangle. Tests with blue rectangles below the horizontal dashed line show significant depletion of pQTLs in SNVs with MAFs within the denoted range. Tests with blue rectangles above the horizontal dashed line show significant enrichment of pQTLs in SNVs with MAFs within the denoted range. Only proteins with complete data were considered for this analysis.

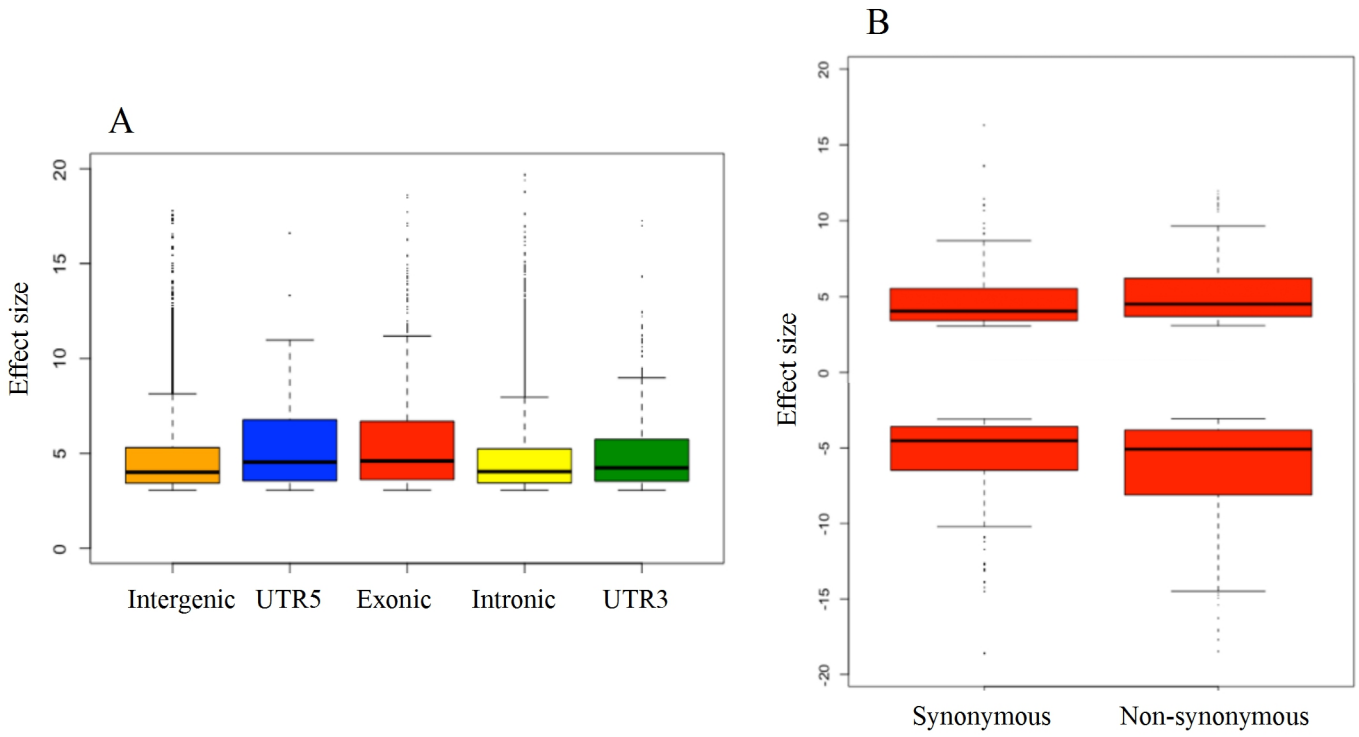

Figure S6. Effect size of pQTLs by genomic annotation. (A) Boxplots showing the distribution of pQTL effect sized by genic location. The shown effect size is the absolute value of the pQTL t-statistic. (B) Boxplots showing the distribution of positive and negative exonic pQTL effects for synonymous and non-synonymous variation.

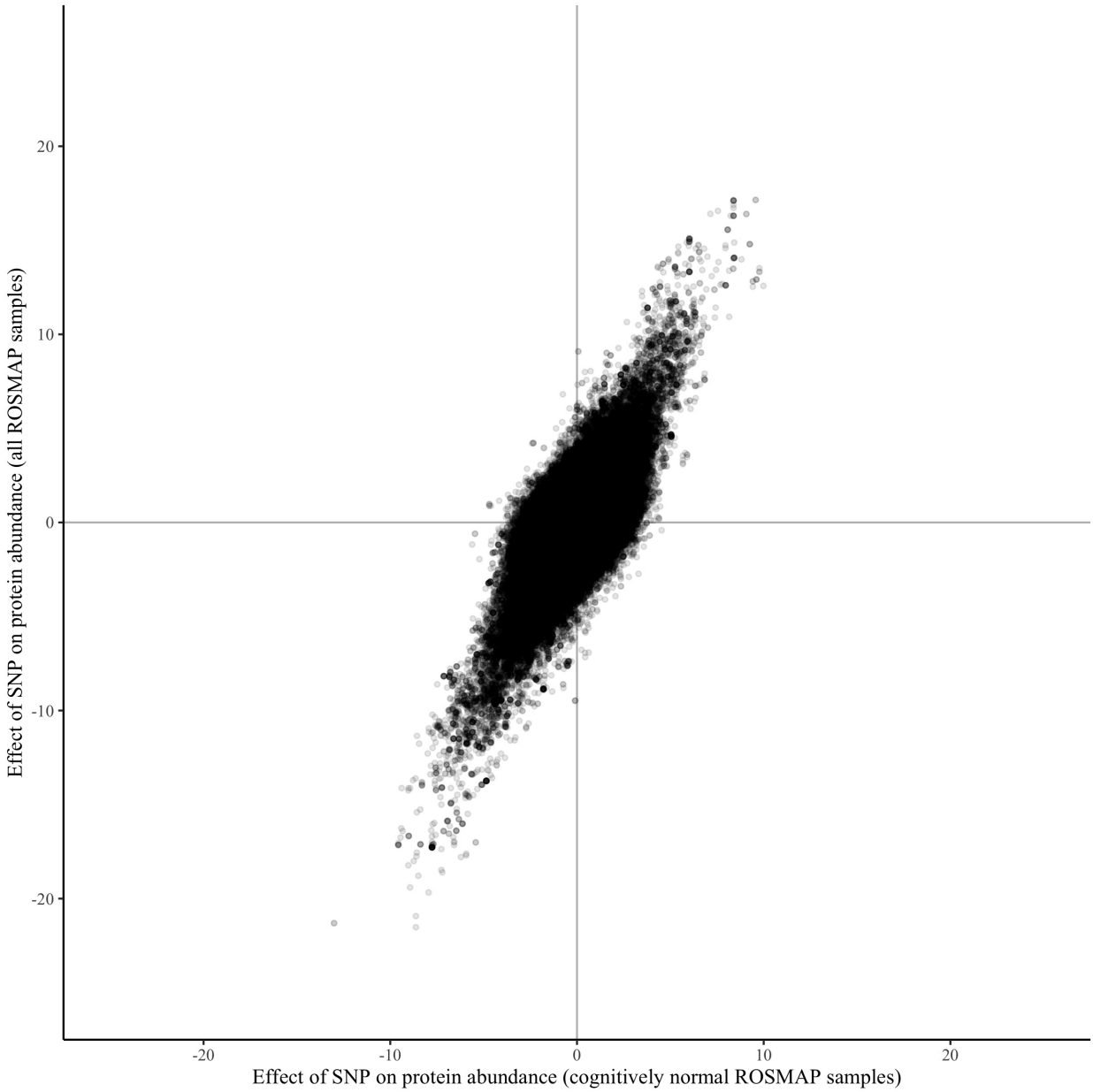

Figure S7. Comparison of pQTL effects estimated using samples with no cognitive impairment vs. all samples. Each point represents a test of a SNV against the protein expression of a single gene. The y-axis shows the effect of a SNV on protein abundance estimated by the main pQTL analysis that used 330 samples and adjusted for clinical diagnosis at death. The x-axis shows the effect of a SNV on protein abundance estimated by a pQTL analysis that used a subset of 139 samples with a clinical diagnosis of no cognitive impairment (NCI) at death. The shown effects are t-statistics. A total of 776,507 tests were performed in both analyses and were plotted here. The correlation between all estimated effects is 0.62 ($p<2.2x10^{-16}$), while the correlation between the estimated effects at sites identified as pQTLs in the main analysis is 0.92 ($p<2.2x10^{-16}$, 37,569 tests at FDR < 0.05).

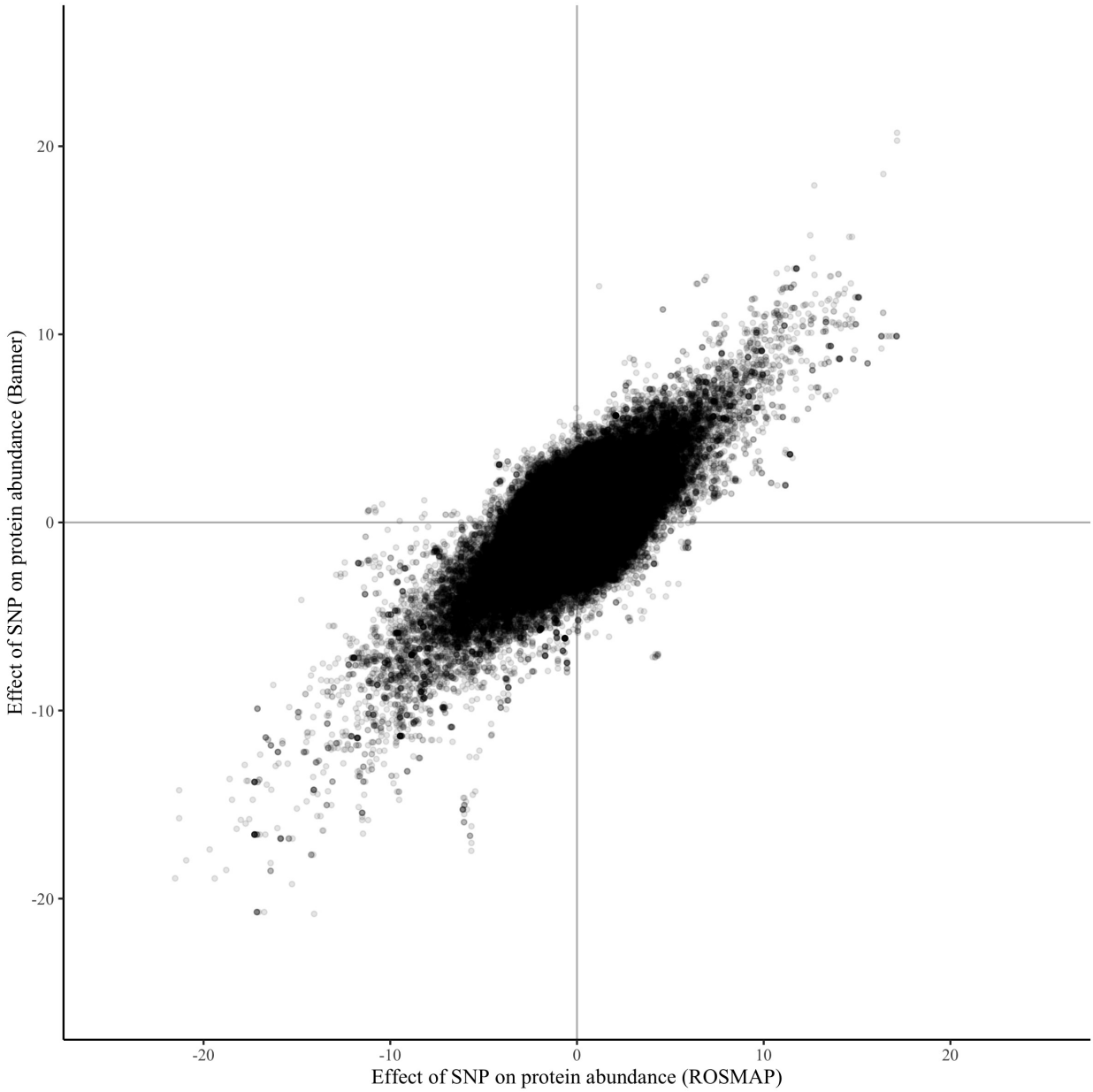

Figure S8. **Comparison of pQTL effects estimated using Banner vs. ROS/MAP samples.** Each point represents a test of a SNV against the protein expression of a single gene. The y-axis shows the effect of a SNV on protein abundance estimated by the Banner pQTL analysis, while the x-axis shows the effect of a SNV on protein abundance estimated by the ROS/MAP pQTL analysis. The shown effects are t-statistics. A total of 591,720 tests were performed in both analyses and were plotted here. The correlation between all estimated effects is 0.57 ($p<2.2x10^{-16}$), while the correlation between the estimated effects at sites identified as pQTLs in the ROS/MAP analysis is 0.90 ($p<2.2x10^{-16}$, 32,679 tests at FDR < 0.05).

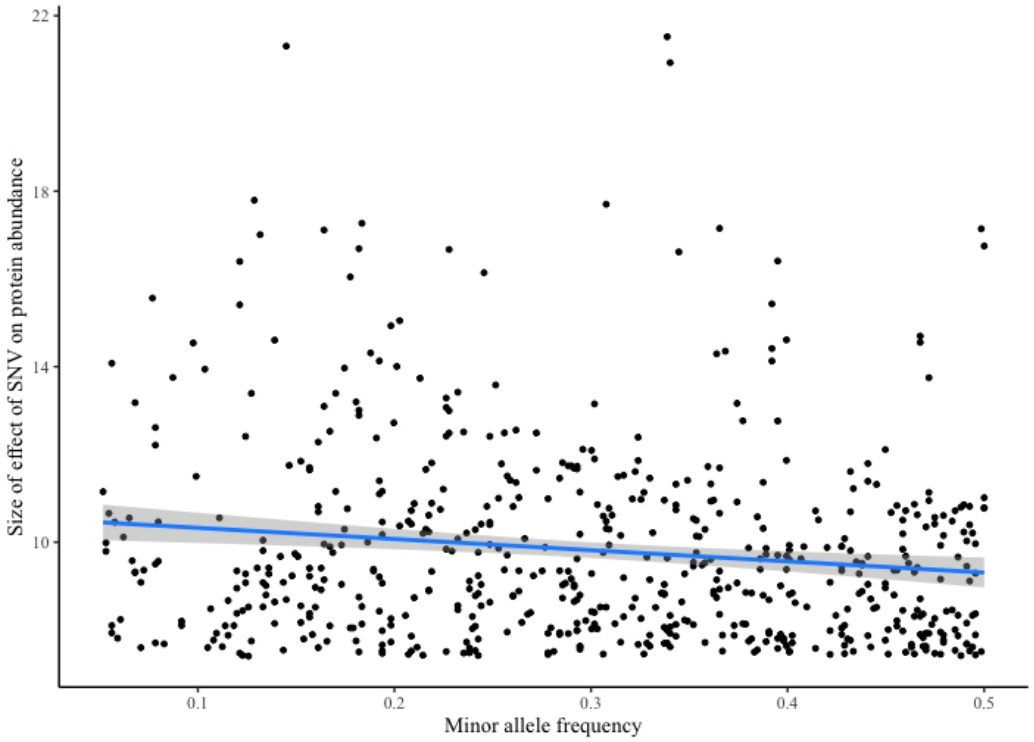

Figure S9. Relationship between pQTL effect size and minor allele frequency (MAF). For this analysis, we considered only independent pQTLs with effects sizes (absolute value of pQTL t-statistic) in the top 10%. The relationship between effect size and MAF was estimated based on a linear regression that modeled the absolute value of the pQTL t-statistic as a function of MAF. We found an increase in MAF to be associated with a decrease in the size of the genetic effect on protein (β =- 2.5479, p = 0.000634).

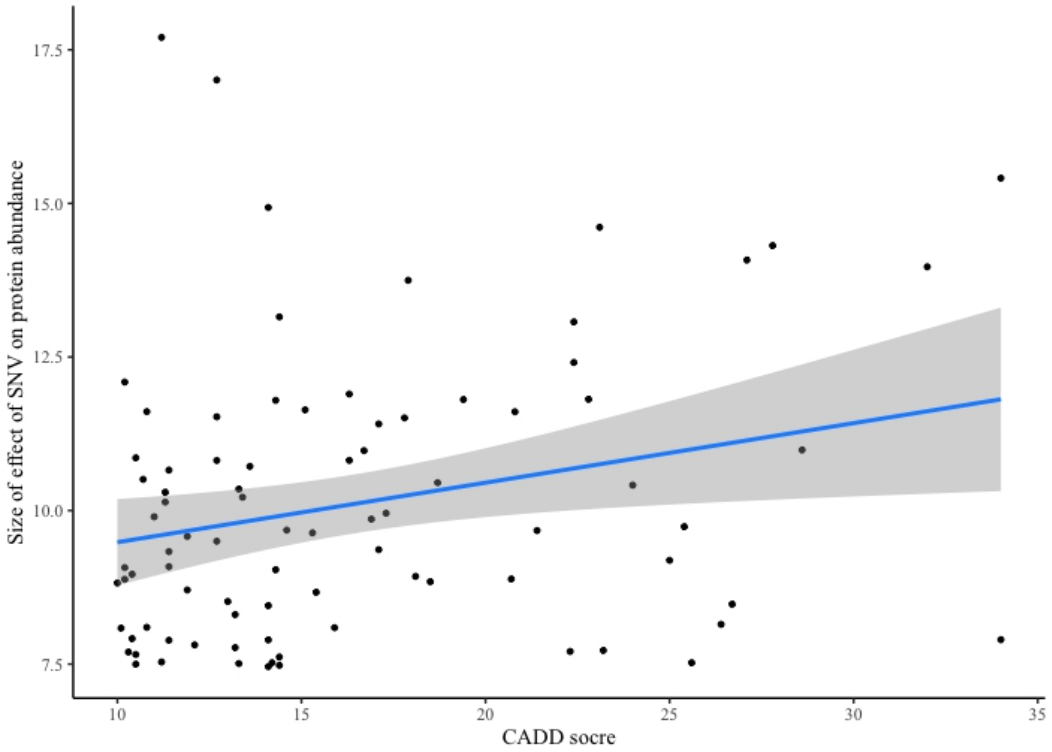

Figure S10. Relationship between pQTL effect size and CADD score. For this analysis, we considered only independent pQTLs with effect sizes (absolute value of pQTL t-statistic) in the top 10% and a CADD score greater than 10. Variants with a CADD score above 10 are predicted to be in the top 10% of deleterious variants in the human genome. The relationship between effect size and CADD score was estimated based on a linear regression that modeled the absolute value of the pQTL t-statistic as a function of CADD score. We found an increase in CADD score to be associated with an increase in the size of the genetic effect on protein (β= 0.09692 , p = 0.0186).

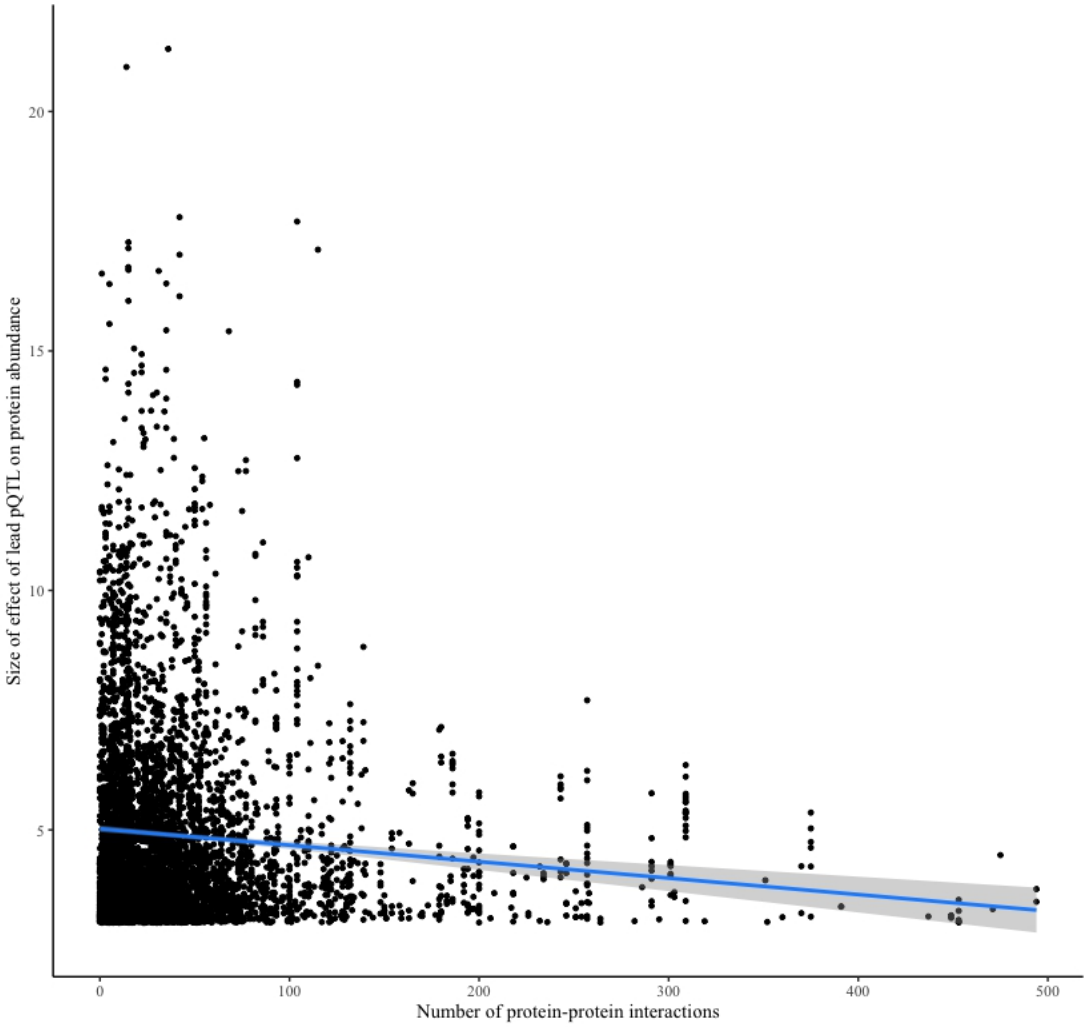

Figure S11. Relationship between the effect size of the lead pQTL and the number of protein-protein interactions. This analysis considered lead pQTLs for proteins with less than 500 protein-protein interactions. The relationship between effect size and number of protein-protein interactions was estimated based on a linear regression that modeled the absolute value of the pQTL t-statistic as a function of the number of protein-protein interactions. We found an increase in the number of protein-protein interactions to be associated with a decrease in the size of the genetic effect on protein (β = -0.001365, p = 1.09e-05).

Table S1. Demographics of analyzed subjects.

	ROS/MAP		Banner BBDP
Characteristic	**Subjects with protein and genotype data**	**Subjects with mRNA, protein, and genotype data**	**Subjects with protein and genotype data**
Sample Size	330	173	149
Female sex (%)	69%	69%	56%
Age at death [years] (median, range)	89 [71 – 106.5]	89 [71 – 106.5]	86 [66 – 103]
Clinical diagnosis of dementia (N, %)			
No cognitive impairment	139 (42%)	78 (45%)	64 (43%)
Mild cognitive impairment	90 (27%)	53 (31%)	20 (13%)
Alzheimer's disease	101 (31%)	42 (24%)	65 (44%)

Table S2. Enrichment of genomic annotations among pQTLs. Enrichments were evaluated with Fisher's exact tests. With the exception of the synonymous and non-synonymous annotations, the background for every test was the set of all SNVs tested in our pQTL study. The background for the synonymous and non-synonymous annotations was the set of all tested exonic SNVs.

Annotation	**# SNVs**	**pQTL enrichment**		
		OR	**95% CI** Lower limit, upper limit	**P**
UTR3	6,654	1.85	1.70, 2.01	1.8e-42
Exonic	5,930	2.44	2.26, 2.64	5.3e-91
synonymous	*3,725*	*0.51*	*0.45, 0.59*	*5.27e-22*
non-synonymous	*2,172*	*1.96*	*1.71, 2.25*	*1.08e-22*
Intronic	218,202	0.88	0.86, 0.91	5.3e-20
UTR5	580	1.93	1.45, 2.52	8.4e-6
Intergenic	177,421	0.75	0.73, 0.77	6.1e-130

Table S3. Large GWASs of brain diseases used to assess the enrichment of disease variants among pQTLs. Only GWAS result from individuals of European descent were analyzed. For each GWAS we used a significance threshold of $5x10^{-8}$ to identify disease-associated variants within 100 kb of genes with proteomic data. Enrichment was assessed for each disease individually using Fischer exact tests.

Brain disease	Study	N	# of disease-associated variants	# of overlapping pQTLs	Enrichment	
					OR	p-value
Alzheimer's disease	Jansen *et al*. 2017	455,258	219	16	1.01	0.90
Parkinson's disease	Nalls *et al*. 2019	471,013	218	83	5.82	4.04e-31
Schizophrenia	Lam *et al*. 2019	154,192	778	142	2.61	4.86e-21
Neuroticism	Nagel *et al*. 2018	449,484	894	182	3.07	9.35e-34

Table S4. Comparison of pQTL identification using the ROS/MAP and Banner BBDP cohorts.

Cohort	Sample size	Number of tested SNVs	Number of tested genes	Number of pQTLs	Number of pQTL genes
ROSMAP	163-330	501,414	7,376	35,601	2,474
Banner BBDP	75-149	460,954	6,526	23,945	1,803
Overlap		429,083	5,712	14,752	1,129

Table S5. List of genes with mRNA-mediated and mRNA-independent pQTLs. Genes in bold are associated with the GO term "neuron apoptotic process". Genes in italic are associated with the GO term "transepithelial transport"

Chr	Genes with mRNA-mediated pQTLs	Genes with mRNA-independent pQTLs
1	RPA2, PADI2, AGL, CCBL2, KYAT3, DBT, SLC25A24, GSTM5, GSTM3, PTGFRN, S100A13, TDRKH, S100A4, TSTD1, DARS2, COA6, NTPCR	ARID1A, ENO1, NASP, ACOT7, SH3GLB1, USP24, BOLA1, CA14, LYSMD1, PSMB4, FDPS, CDC73, CACNA1E, GLUL, TROVE2, CNTN2, IARS2, CAPN2, CCSAP, NID1
2	DPYSL5, RETSAT, GALM, CAPG, PLCL1, ATIC, PPIL3, IDH1, SCRN3	BRE; BABAM2, HS1BP3, MRPL53, TGOLN2, INPP4A, LONRF2, CNTNAP5, TMEFF2, ABCB6, DOCK10
3	PLSCR4, ATG7, MYLK, LARS2, CHL1, LZTFL1	APPL1, CPOX, TF, CDV3, ADCY5, IQSEC1, TFRC
4	DGKQ, TBC1D1, PGM2, GUF1, GPRIN3, SPARCL1, HSD17B11, SCRG1, MMAA	PAICS, KIT
5	SGTB, ERAP1, DIAPH1, TBC1D9B, RUFY1	SLC1A3, SLC12A2, PPIP5K2, HINT1
6	ECI2, HDDC2, SIRT5, GOPC, RWDD1, CAP2, AKAP12, ACAT2, BPHL	ME1, RIMS1, RAB23
7	AMPH, EGFR, ABHD11, PDIA4, ABCB8	GARS, PMPCB, AGFG2, CCDC132; VPS50, SLC25A13, SSBP1, MKRN1
8	LY6H, ADHFE1, SNTB1	OXR1, RALYL, TATDN1, KHDRBS3, ATP6V1B2, GPT
9	AK3, ACO1, NUDT2, GLIPR2, PHYHD1, PTGR1, AIF1L, CCBL1; KYAT1, HDHD3	PSIP1, GBA2
10	SNCG, ANXA11, COX15, PRTFDC1, SFXN3, PRKG1	SEC24C, FAM175B; ABRAXAS2
11	AMPD3, SLC17A6, LRP4, HSD17B12, AAMDC, ASRGL1, C11orf54, MADD, SNX32	CEND1, TPP1, NUCB2, SPON1, SLC1A2, CAPRIN1, CTNND1, INPPL1, CFL1, ZBTB16, SIK3, MCAM, C2CD2L, DCPS
12	CPM, CORO1C, UHRF1BP1L, ESYT1, NT5DC3, ARHGDIB, PIP4K2C, CSRP2, MGST1	ISCU, CS, ANO6, RPAP3, NUAK1, CIT, CALCOCO1
13	CAB39L	DOCK9
14	L3HYPDH, PTGR2, DAAM1, ACOT1, STXBP6, STON2, INF2, ACOT2	HNRNPC, GPHN, COQ6, RTN1, ACYP1, VIPAS39, CDC42BPB
15	FAM82A2; RMDN3, RLBP1, LACTB, RGMA	SQRDL; SQOR, ULK3, SCAMP5
16	LPCAT2, BAIAP3, SULT1A1, NECAB2	COG7, LCMT1, NAE1, ITGAM, SLC9A3R2
17	ASPA, C1QBP, TRPV2, C17orf59; BORCS6, SHMT1, WBP2, TRIM25, FDXR, ACSF2, SEPT9, SPATA20	CAMKK1, TXNDC17, VAT1, DHRS11, SEPT4, GHDC, FLOT2, ACACA, AARSD1; PTGES3L-AARSD1, ACTG1
18		LMAN1
19	PLIN4, LONP1, ATP13A1, PEPD, ALDH16A1	SH3GL1, BRD4, MAP1S, MEGF8, UBE2M
20	CPNE1, TGM2, ITPA	PLCG1, AHCY, PHACTR3, ARFGAP1, RPS21, RPN2, GSS
21	JAM2, PCP4	
22	ARVCF, APOL2, PACSIN2, SYN3	AIFM3