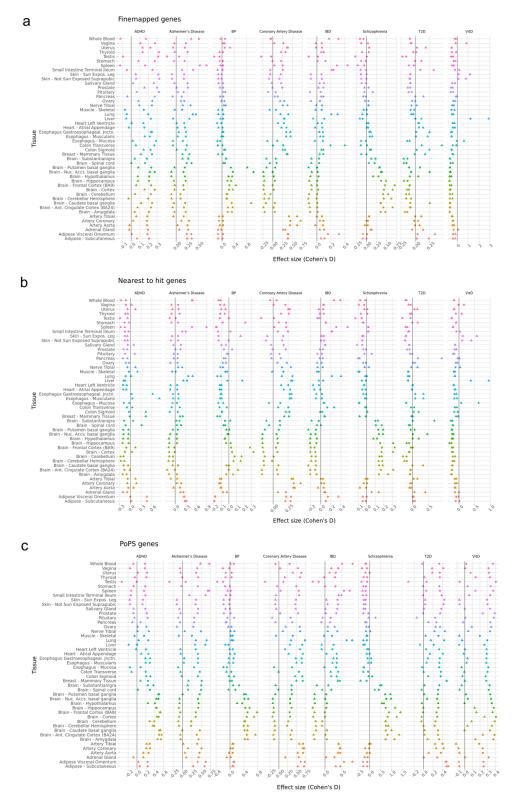
Supplemental Material for:

The Gene Expression Landscape of Disease Genes

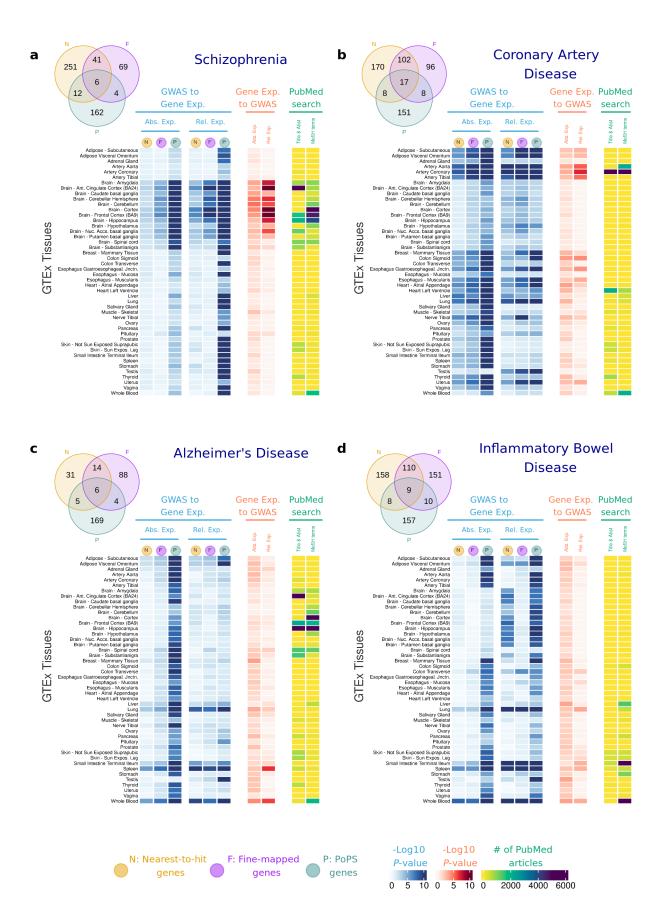
Judit García-González¹, Saul García González^{1,2}, Lathan Liou¹, Paul F. O'Reilly¹,

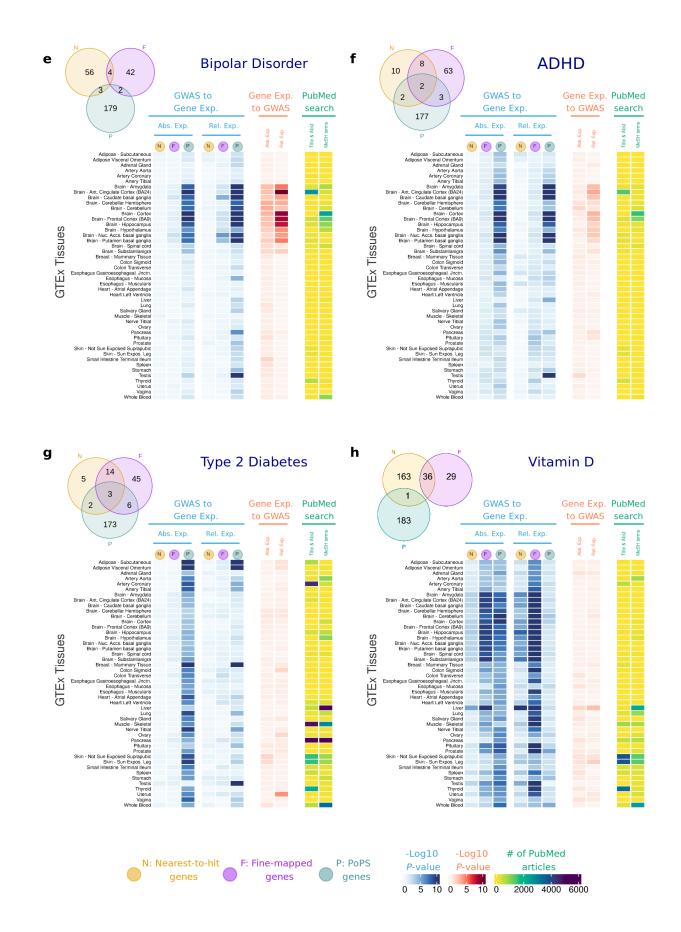
¹Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA ²Center for Excellence in Youth Education, Icahn School of Medicine, Mount Sinai, New York City, NY 10029, USA



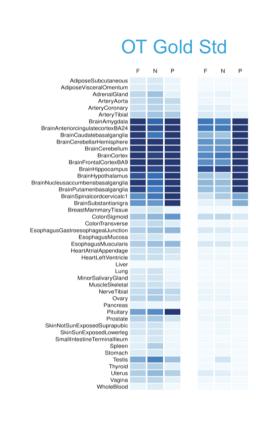


Supplemental Figure 1. Effect size (Cohen's D) across tissues for a) fine-mapped genes, b) Nearest to hit genes and c), PoPs genes. Effect sizes are shown across all the GTEx tissues, and all the diseases investigated. Absolute and relative expression are indicated as circles (absolute expression) and triangles (relative expression).





Supplemental Figure 2: Heatmap showing results of the association between gene expression in each GTEx tissue and a, Schizophrenia; b, Alzheimer's Disease; c, Coronary Artery Disease; d, Inflammatory Bowel Disease; e, Bipolar Disorder; f, ADHD; g, Type 2 Diabetes; h, Vitamin D. In blue, results showing the P-value of the <u>Anderson-Darling test</u>, (as opposed to the One side T-test shown in Main Figure 2). In red, results showing the P-value for enrichment of GWAS signal across the set of genes with highest and most specific expression for each tissue. In yellow, results for the Literature Search using PubMed. F, Fine-mapped genes; N, Nearest-to-hit genes, P, Polygenic Priority Scores genes; TiAb, Title and Abstract are used in the PubMed Search; MeSH, MeSH terms are used in the PubMed Search.



Schizophrenia

OT Gen Evd

F N

N P

F

AdiposeSubcutaneous

AdrenalGland

ArteryAorta

ArteryTibial

BrainCortex

ArteryCoronary

BrainAmygdala

BrainCerebellun

BrainFrontalCortexBA9

BrainPutamenbasalganglia

BrainSpinalcordcervicalc1

BrainHippocampus BrainHypothalamus

BrainSubstantianigra

ColonTransverse

BreastMammaryTissue ColonSigmoid

EsophagusMucosa EsophagusMuscularis

HeartAtrialAppendage

HeartLeftVentricle

MinorSalivaryGland

SkinNotSunExposedSuprapubic

SkinSunExposedLowerleg SmallIntestineTerminallleum

MuscleSkeletal

NerveTibial

Liver

Lung

Ovary

Pancreas Pituitary

Prostate

Spleen

Thyroid

Uterus Vagina

WholeBlood

Testis

0

Stomach

AdiposeVisceralOmentum

BrainAnteriorcingulatecortexBA24 BrainCaudatebasalganglia

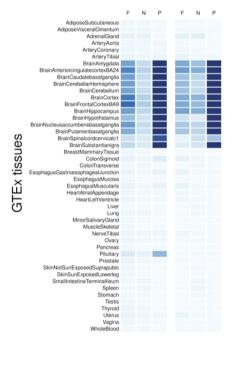
BrainNucleusaccumbensbasalganglia

EsophagusGastroesophagealJunction

BrainCerebellarHemisphere

P

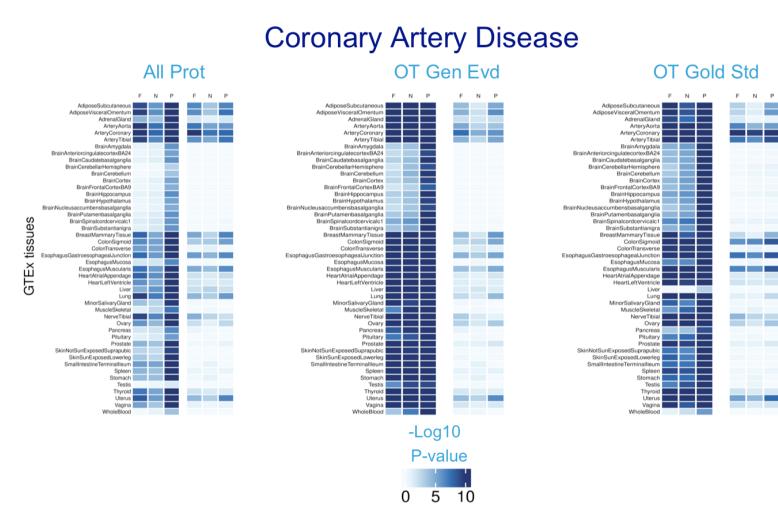
All Prot



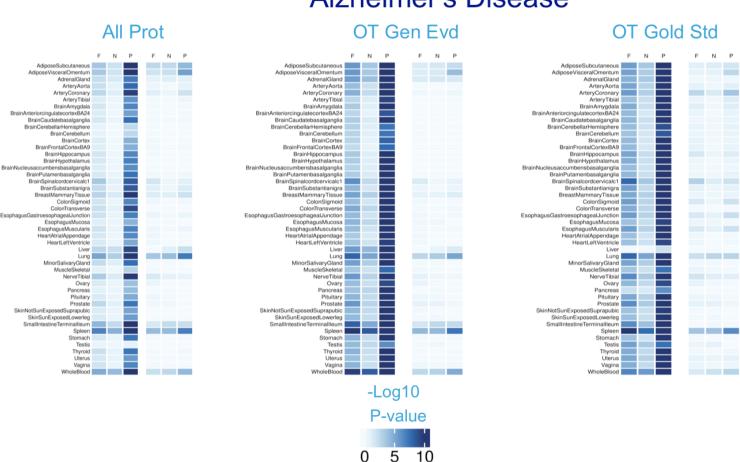
Supplemental Figure 3: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for Schizophrenia. For each disease, three panels are shown showing <u>three different control groups</u> used in the T-test: (i) *"All Prot"*: all protein-coding genes, (ii) *"OT Gen Evd"*: genes prioritized via Open Targets Genetics evidence and (iii) *"OT Gold Std"*: Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for *Absolute* gene expression, and the second three columns represent *Specific* gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.

-Log10 P-value

5



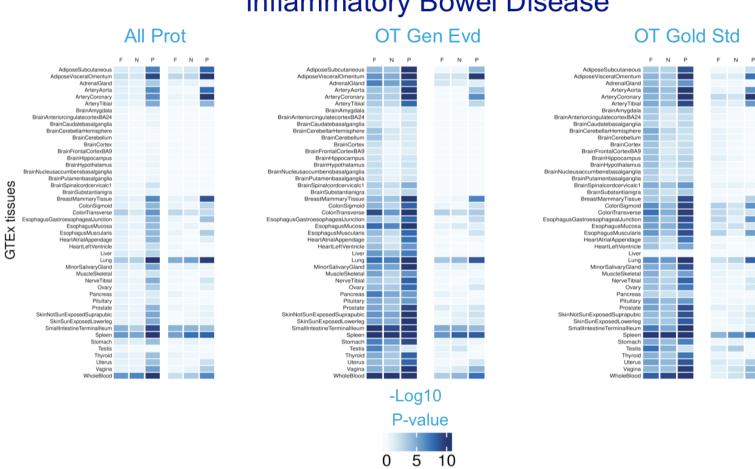
Supplemental Figure 4: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for Coronary Artery Disease. For each disease, three panels are shown showing three different control groups used in the T-test: (i) "All Prot": all protein-coding genes, (ii) "OT Gen Evd": genes prioritized via Open Targets Genetics evidence and (iii) "OT Gold Std": Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for Absolute gene expression, and the second three columns represent Specific gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.



GTEx tissues

Alzheimer's Disease

Supplemental Figure 5: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for Alzheimer's Disease. For each disease, three panels are shown showing <u>three different control groups</u> used in the T-test: (i) *"All Prot"*: all protein-coding genes, (ii) *"OT Gen Evd"*: genes prioritized via Open Targets Genetics evidence and (iii) *"OT Gold Std"*: Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for *Absolute* gene expression, and the second three columns represent *Specific* gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.



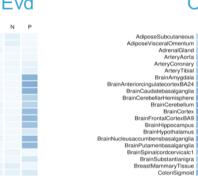
Inflammatory Bowel Disease

Supplemental Figure 6: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for Inflammatory Bowel Disease. For each disease, three panels are shown showing three different control groups used in the T-test: (i) "All Prot": all protein-coding genes, (ii) "OT Gen Evd": genes prioritized via Open Targets Genetics evidence and (iii) "OT Gold Std". Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for Absolute gene expression, and the second three columns represent Specific gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.



F N P

F N P



ColonTransverse EsophagusGastroesophagealJunction

EsophagusMucosa EsophagusMuscularis

HeartLeftVentricle

MinorSalivaryGland MuscleSkeletal

NerveTibial

Pancreas

Pituitary

Prostate

Spleen Stomach

Testis

Thyroid

WholeBlood

Uterus

Vagina

Liver

Lung

Ovary

HeartAtrialAppendage

SkinNotSunExposedSuprapubic

SkinSunExposedLowerleg SmallIntestineTerminallleum

ADHD

F N Р

AdiposeSubcutaneous

AdrenalGland

ArteryCoronary

ArteryTibial BrainAmygdala

BrainCerebellum

BrainFrontalCortexBA9

BrainPutamenbasalganglia

BrainSpinalcordcervicalc1

BrainHippocampus BrainHypothalamus

BrainSubstantianigra

EsophagusMucosa EsophagusMuscularis

HeartLeftVentricle

MinorSalivaryGland

SkinNotSunExposedSuprapubic

SkinSunExposedLowerleg SmallIntestineTerminallleum

MuscleSkeletal

NerveTibial

Pancreas Pituitary

Prostate

Spleen

Testis

Thyroid

Uterus

WholeBlood

Vagina

Stomach

Liver

Lung

Ovary

-Log10 P-value

5 10

HeartAtrialAppendage

ColonTransverse

BreastMammaryTissue ColonSigmoid

BrainCortex

ArteryAorta

AdiposeVisceralOmentum

BrainAnteriorcingulatecortexBA24 BrainCaudatebasalganglia

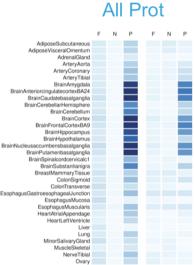
BrainNucleusaccumbensbasalganglia

EsophagusGastroesophagealJunction

BrainCerebellarHemisphere

OT Gen Evd

F



Pancreas

Prostate

SkinNotSunExposedSuprapubic

SkinSunExposedLowerleg SmallIntestineTerminallleum

Pituitary

Spleen

Testis

Thyroid

Uterus

Vagina WholeBlood

Stomach

GTEx tissues

Supplemental Figure 7: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for ADHD. For each disease, three panels are shown showing three different control groups used in the T-test: (i) "All Prot": all protein-coding genes, (ii) "OT Gen Evd": genes prioritized via Open Targets Genetics evidence and (iii) "OT Gold Std": Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for Absolute gene expression, and the second three columns represent Specific gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.

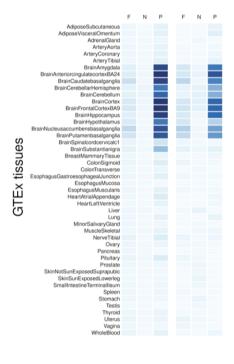


Bipolar Disorder

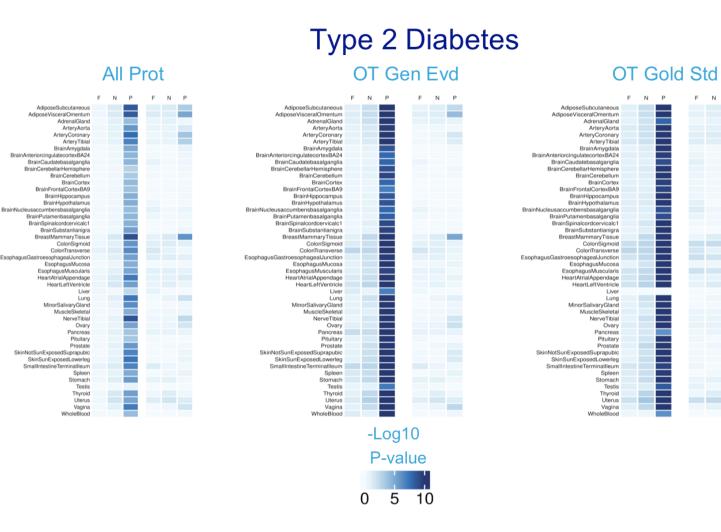
OT Gen Evd



All Prot



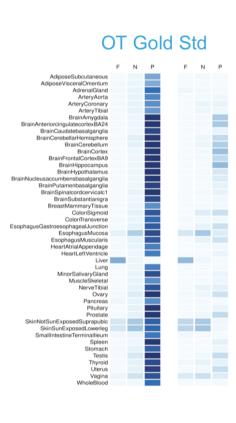
Supplemental Figure 8: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for Bipolar Disorder. For each disease, three panels are shown showing three different control groups used in the T-test: (i) "All Prot": all protein-coding genes, (ii) "OT Gen Evd": genes prioritized via Open Targets Genetics evidence and (iii) "OT Gold Std": Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for Absolute gene expression, and the second three columns represent Specific gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.



GTEx tissues

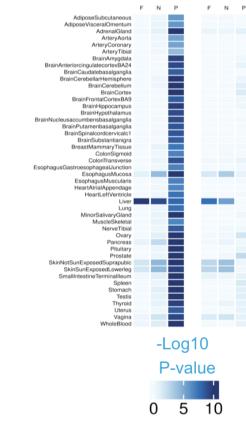
P

Supplemental Figure 9: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for Type 2 Diabetes. For each disease, three panels are shown showing three different control groups used in the T-test: (i) "All Prot": all protein-coding genes, (ii) "OT Gen Evd": genes prioritized via Open Targets Genetics evidence and (iii) "OT Gold Std": Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for Absolute gene expression, and the second three columns represent Specific gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.

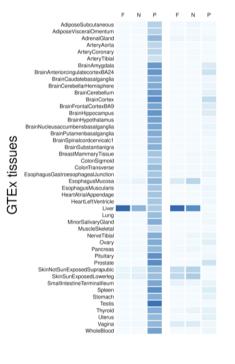


Vitamin D

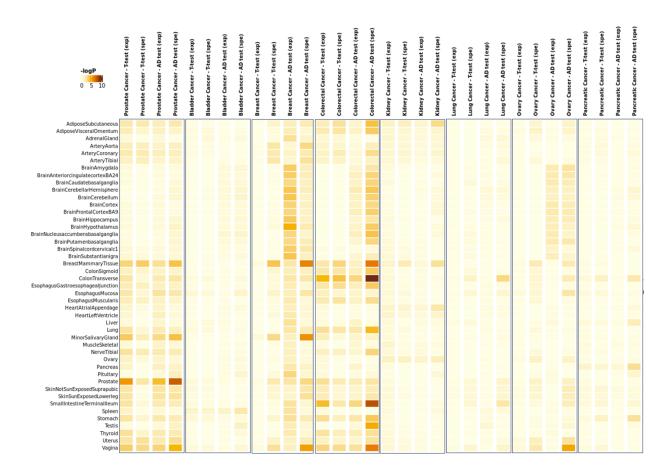
OT Gen Evd



All Prot

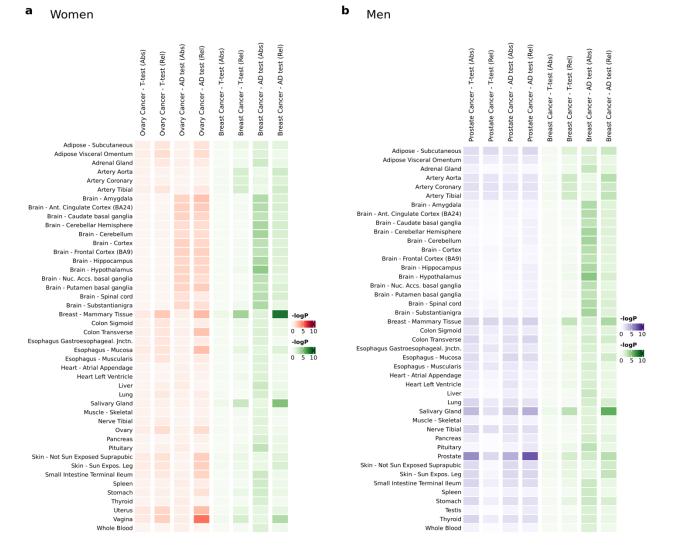


Supplemental Figure 10: Heatmaps showing T-test results of the association between gene expression in each GTEx tissue for Vitamin D. For each disease, three panels are shown showing three different control groups used in the T-test: (i) "All Prot": all protein-coding genes, (ii) "OT Gen Evd": genes prioritized via Open Targets Genetics evidence and (iii) "OT Gold Std": Open Targets gold-standard evidence genes. Within each panel, the three first columns represent results for Absolute gene expression, and the second three columns represent Specific gene expression. F: Fine-mapped genes, N: Nearest-to-hit genes, P: PoPS genes.

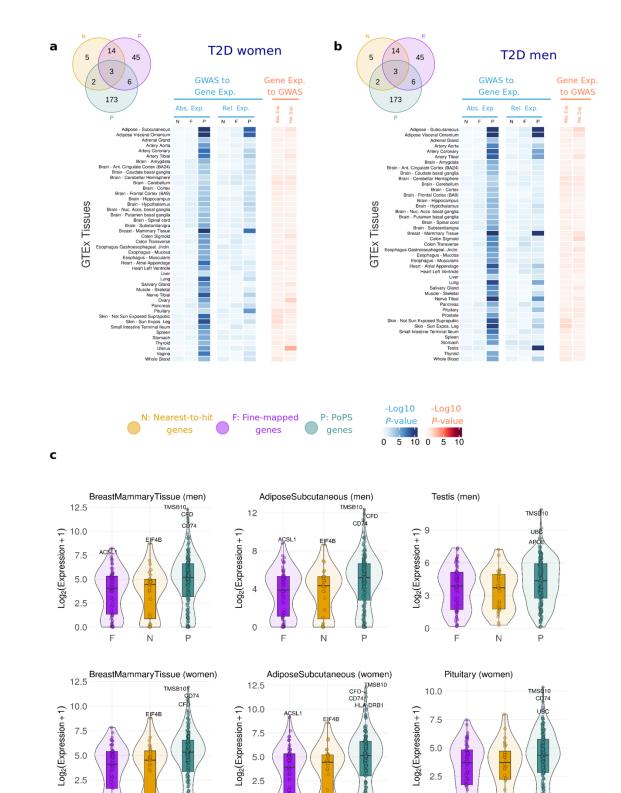


Supplemental Figure 11: Heatmaps showing T-tests and Anderson-Darling (AD test) results of the association between gene expression in each GTEx tissue. exp, Absolute gene expression was measured; spe, Specific (relative) gene expression was measured; AD, Anderson-Darling test.

I



Supplemental Figure 12: Sex-stratified analyses for three cancer traits: ovary cancer, breast cancer, and prostate cancer. Heatmaps show *P*-values for the T-tests and Anderson-Darling (AD test) results of the association between absolute (Abs) and Relative (Rel) gene expression in each GTEx tissue.



Supplemental Figure 13: Sex-stratified analyses for T2D.

N

Ρ

0.0

Ē

0.0

Ē

N

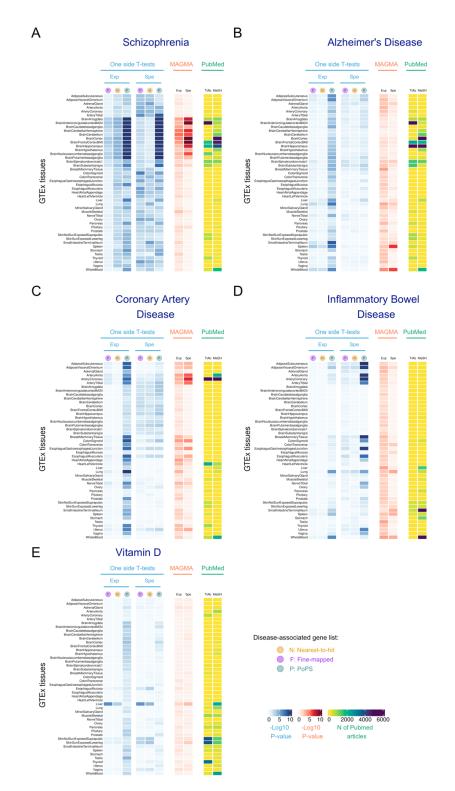
P

Ρ

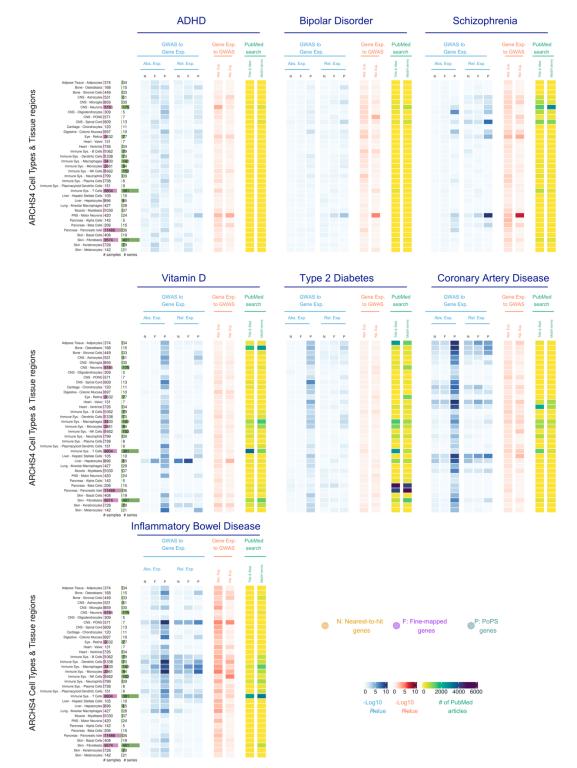
Ň

0.0

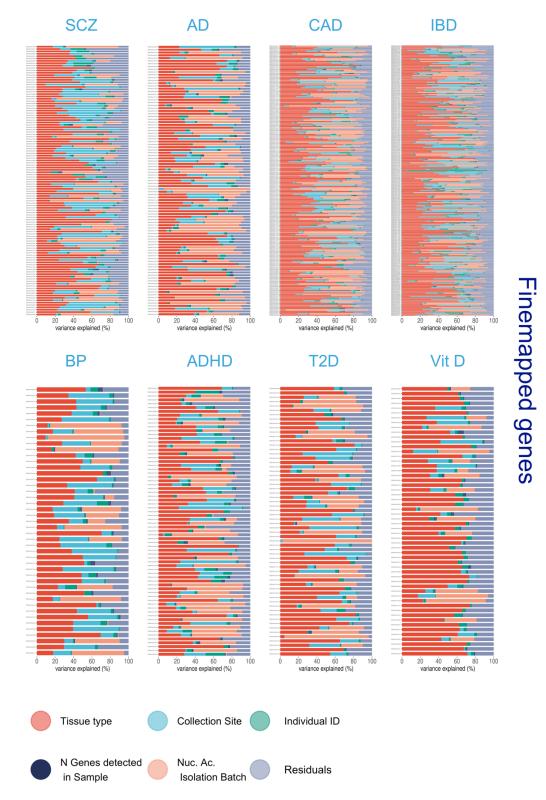
Ē



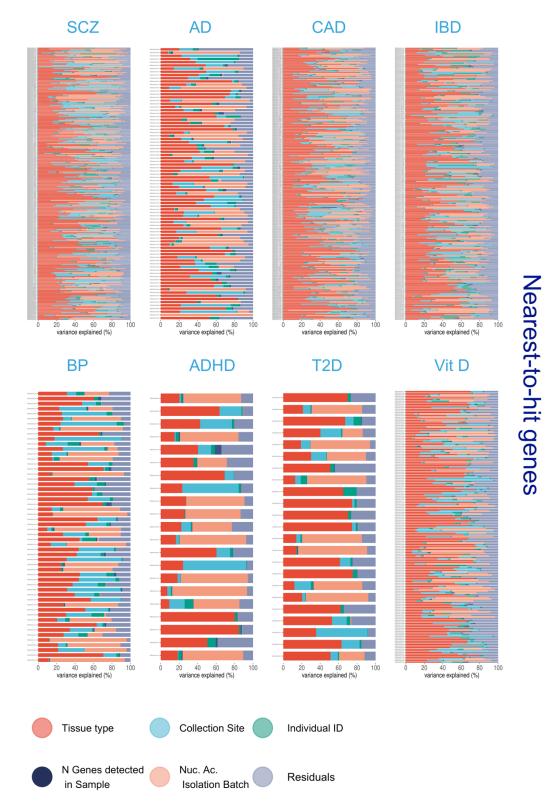
Supplemental Figure 14. Heatmap showing results of the sensitivity analysis where association between gene expression in each GTEx tissue was performed after removal of the top 10% genes most expressed or most specifically expressed in relevant tissues.



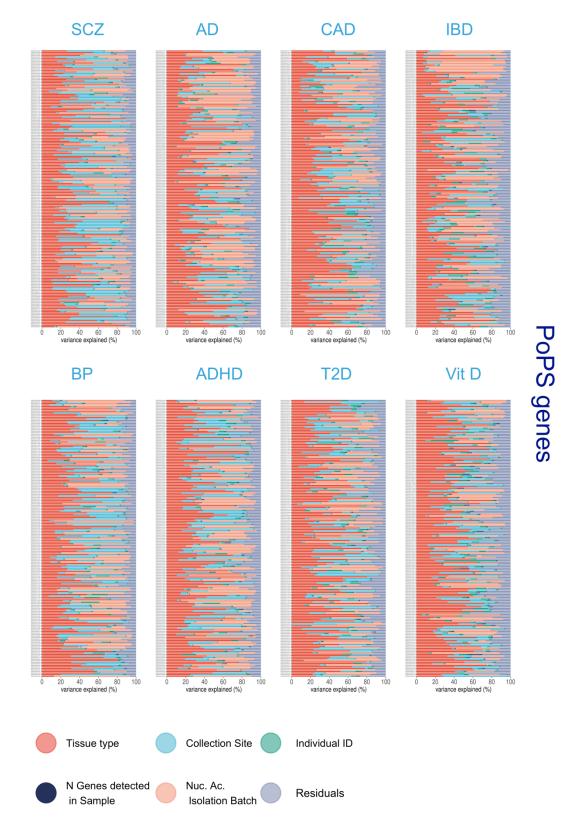
Supplemental Figure 15: Heatmap showing results of the association between gene expression in each cell-type or tissue region and ADHD, Bipolar Disorder, Schizophrenia, Vitamin D, and Type 2 Diabetes using the ARCHS4 dataset. In blue, results showing the Log₁₀ *P*-value for a one-side T-tests. In red, results showing the Log₁₀ *P*-value for enrichment of GWAS signal across the set of genes with highest absolute and relative expression for each tissue. In yellow, results for the Literature Search using PubMed. Abs. Exp, Absolute expression; Spec. Exp, specific expression; F, Fine-mapped genes; N, Nearest-to-hit genes, P, Polygenic Priority Scores genes; Title & Abst, Title and Abstract are used in the PubMed Search; MeSH, MeSH terms are used in the PubMed Search.



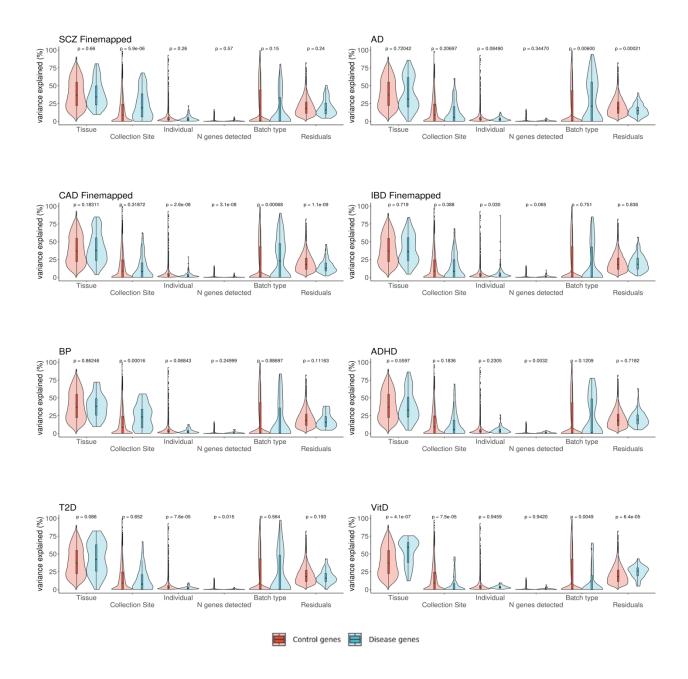
Supplemental Figure 16. Factors explaining gene expression variance in disease-associated genes prioritized via fine-mapping of GWAS hits. Each percentage bar (each row) corresponds to an individual gene. Panels are split by prioritization method and trait investigated.



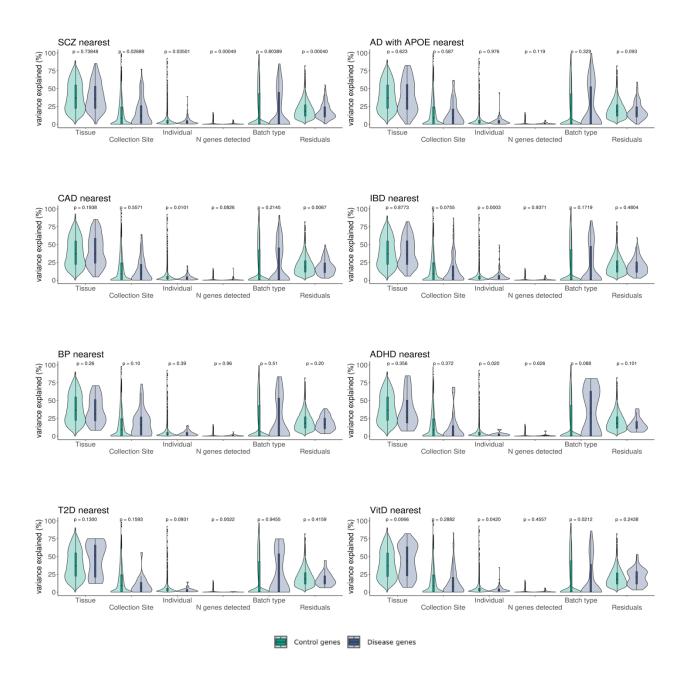
Supplemental Figure 17. Factors explaining gene expression variance in disease-associated genes prioritized using the nearest to GWAS-hit approach. Each percentage bar (each row) corresponds to an individual gene. Panels are split by prioritization method and trait investigated.



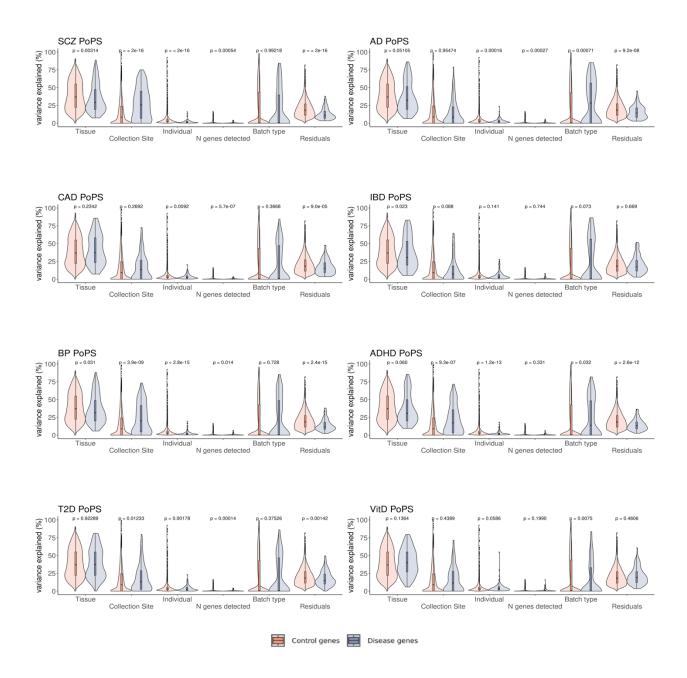
Supplemental Figure 18. Factors explaining gene expression variance in disease-associated genes using the PoPS method. Each percentage bar (each row) corresponds to an individual gene. Panels are split by prioritization method and trait investigated.



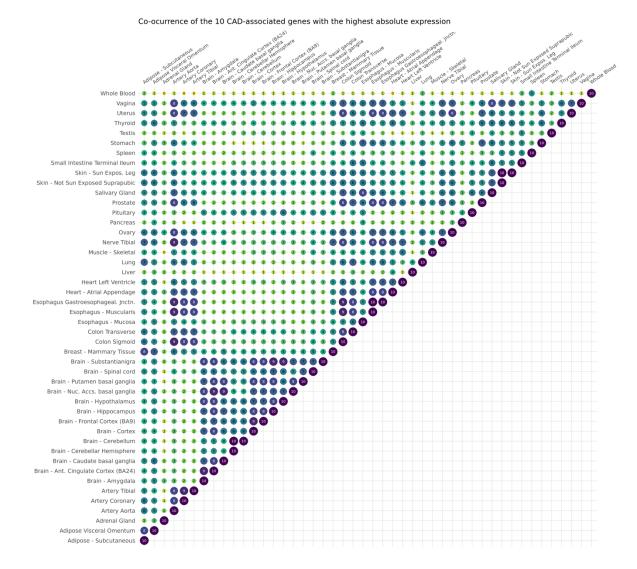
Supplemental Figure 19. Factors explaining gene expression in disease-associated genes via *Finemapping* vs other proteincoding genes. Every data point composing the violin plots represents the proportion of variance explained by a predictor (Tissue, Collection Site, Individual N genes detected, Batch Type) on the gene expression of an individual gene. This is calculated by fitting a linear mixed model for each gene. The results are color-coded to compare the variance explained by each factor in fine-mapped genes vs control genes.



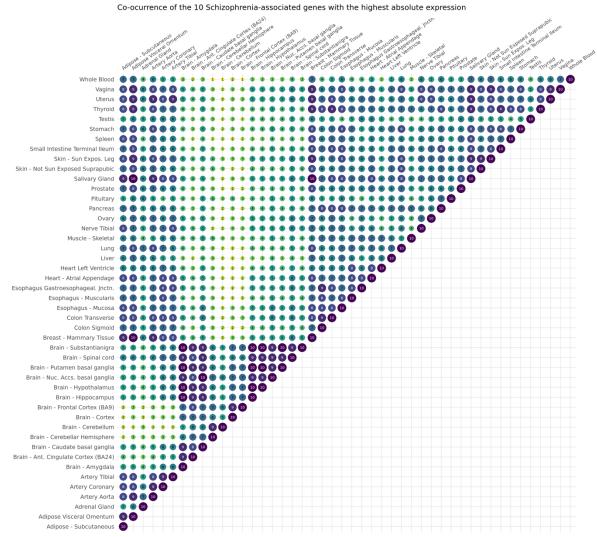
Supplemental Figure 20. Factors explaining gene expression in disease-associated genes via *nearest genomic location* vs other protein-coding genes. Every data point composing the violin plots represents the proportion of variance explained by a predictor (Tissue, Collection Site, Individual N genes detected, Batch Type) on the gene expression of an individual gene. This is calculated by fitting a linear mixed model for each gene. The results are color-coded to compare the variance explained by each factor in *nearest-to-GWAS-hit* genes vs control genes.



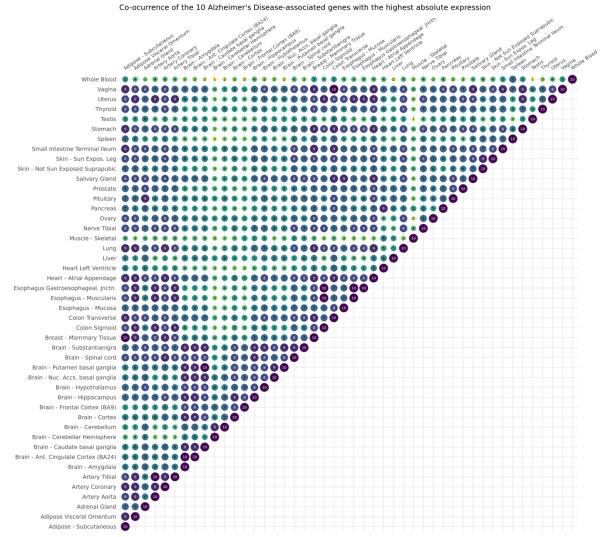
Supplemental Figure 21. Factors explaining gene expression in disease-associated genes via the PoPS prioritization method vs other protein-coding genes. Every data point composing the violin plots represents the proportion of variance explained by a predictor (Tissue, Collection Site, Individual N genes detected, Batch Type) on the gene expression of an individual gene. This is calculated by fitting a linear mixed model for each gene. The results are color-coded to compare the variance explained by each factor in *PoPS* genes vs control genes.

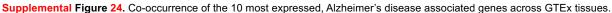


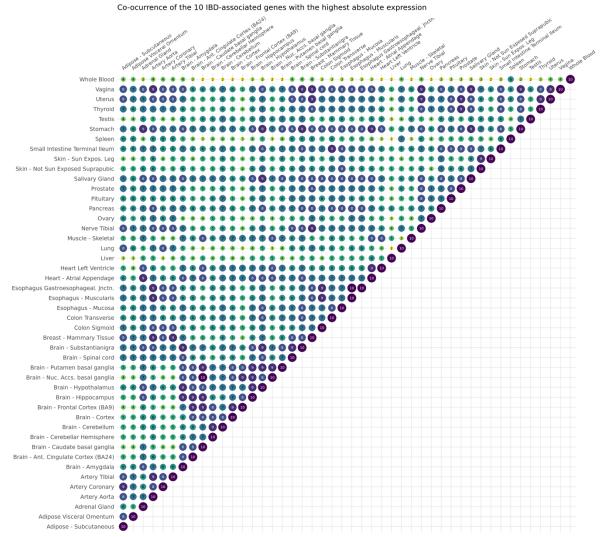
Supplemental Figure 22. Co-occurrence of the 10 CAD most expressed, CAD associated genes across GTEx tissues.











Supplemental Figure 25. Co-occurrence of the 10 most expressed, Inflammatory Bowel Disease associated genes across GTEx tissues.

