SUPPLEMENTARY INFORMATION

Single-nucleus transcriptome-wide association study of human brain disorders

Authors: Sanan Venkatesh¹⁻⁵, Zhenyi Wu¹⁻⁵, Marios Anyfantakis¹⁻⁵, Christian Dillard¹⁻⁴, Prashant N.M.¹⁻⁴, David Burstein¹⁻⁶, Deepika Mathur¹⁻⁶, Roman Kosoy¹⁻⁴, Christos Chatzinakos^{7,8}, Bukola Ajanaku¹⁻⁴, Fotis Tsetsos^{1-4,6}, Biao Zeng¹⁻⁴, Aram Hong¹⁻⁴, Clara Casey¹⁻⁴, Marcela Alvia¹⁻⁴, Zhiping Shao¹⁻⁴, Stathis Argyriou¹⁻⁴, Karen Therrien¹⁻⁴, VA Million Veteran Program, PsychAD Consortium, Tim Bigdeli⁷⁻¹⁰, Pavan Auluck¹¹, David A. Bennett¹², Stefano Marenco¹¹, Vahram Haroutunian¹⁻⁵, Kiran Girdhar¹⁻⁴, Jaroslav Bendl¹⁻⁵, Donghoon Lee¹⁻⁴, John F. Fullard¹⁻⁴, Gabriel E. Hoffman¹⁻⁶, Georgios Voloudakis^{1-6,13*}, Panos Roussos^{1-6*}

Affiliations:

¹Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

²Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY USA ³Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁴Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁵Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA

⁶Center for Precision Medicine and Translational Therapeutics, James J. Peters VA Medical Center, Bronx, NY, USA.

⁷Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA

⁸Institute for Genomics in Health (IGH), SUNY Downstate Health Sciences University, Brooklyn, NY, USA

⁹VA New York Harbor Healthcare System, Brooklyn, NY, USA

¹⁰Department of Epidemiology and Biostatistics, School of Public Health, SUNY Downstate Health Sciences University, Brooklyn, NY, USA

¹¹Human Brain Collection Core, National Institute of Mental Health-Intramural Research Program, Bethesda, MD, USA

¹²Rush Alzheimer's Disease Center, Rush University Medical Center.

¹³Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

*co-senior and co-corresponding

TABLE OF CONTENTS

TABLE OF CONTENTS	2
SUPPLEMENTARY FIGURES	3
Supplementary Fig. 1	
Supplementary Fig. 2	4
Supplementary Fig. 3	5
Supplementary Fig. 4	6
Supplementary Fig. 5	7
Supplementary Fig. 6	8
Supplementary Fig. 7	
Supplementary Fig. 8	
Supplementary Fig. 9	
Supplementary Fig. 10	
Supplementary Fig. 11	
Supplementary Fig. 12	
Supplementary Fig. 13	
Supplementary Fig. 14	
Supplementary Fig. 15	
Supplementary Fig. 16	
Supplementary Fig. 17	
Supplementary Fig. 18	
Supplementary Fig. 19	
SUPPLEMENTARY NOTES	
Other Acknowledgments	
SUPPLEMENTARY REFERENCES	41

SUPPLEMENTARY FIGURES

Supplementary Fig. 1



Supplementary Fig. 1 | Comparison of gene-cell-type prediction performance across ancestry-specific snTIMs. Color of each point represents the cell-type (see legend). Comparisons are for EUR vs AFR (A), EUR vs AMR (B) and AFR vs AMR (C) snTIMs. A. 3,819 genes are imputed in EUR and not AFR snTIMs, and vice versa for 2,378 genes. B. 7,824 genes are imputed in EUR and not AMR snTIMs, and vice versa for 1,558 genes. C. 6,795 genes are imputed in AFR and not AMR snTIMs, and vice versa for 1,970 genes. r: Spearman's correlation coefficient.



Supplementary Fig. 2 | Spearman's correlation (A) and Jaccard similarity (B) of cross-ancestry prediction performance at different cell-type levels. Each point represents a pairwise analysis (Spearman's correlation for A and Jaccard index for B, respectively) of R^2_{CV} across all shared imputable genes ("N") between ancestries for the same snTIM. Applicable pairwise ancestry comparisons include EUR-AFR, EUR-AMR and AFR-AMR colored orange, purple and light-blue, respectively. Point shape denotes snTIM level including snBulk, Class and Subclass. Dashed lines represent linear regression lines. **A.** $log_{10}(N)$ is positively associated (p = 1.01×10^{-5}) with Spearman's correlation in the linear regression analysis (p = 1.01×10^{-5} , adjusted $R^2 = 0.21$). **B.** $log_{10}(N)$ is positively associated (p = 1.91×10^{-24}) with the Jaccard index in a linear regression model (p = 3.81×10^{-24} , adjusted $R^2 = 0.679$).



Supplementary Fig. 3 | Expanded heatmap of S-TWAS significant gene-trait associations by cell-type hierarchy level. The number within each cell represents the number of FDR³³-significant gene-trait associations (GTAs) within each snTIM and level aggregate. FDR-correction for class and subclass aggregate categories is performed across all participating models for that hierarchy level. Color denotes the Jaccard index $\left(\frac{|x \cap all|}{|x \cup all|}\right)$ of the specific cell (*x*) against the union of all FDR-significant trait-associated genes across Bulk, snBulk, class and subclass. S-TWAS analysis is limited to EUR ancestry.



Supplementary Fig. 4 | Heatmap of novel GTAs. Number in each cell indicates the number of novel GTAs (as in Fig. 2C) found in each snTIM and level aggregates. Color of each cell indicates the effect size of the fisher's exact test comparing each snTIM to snBulk.



Supplementary Fig. 5 | Expanded gene fine-mapping heatmaps. A, Expanded heatmap of S-snTWAS individual-cell-type gene fine-mapping. The numbers at the top and bottom of each cell indicate the count of fine-mapped genes and loci, respectively. Color indicates the genes per locus for every trait-snTIM combination. Fine-mapping was performed with FOCUS for each cell-type. FOCUS performs analyses at the LD block level rather than the standard 1 mega base pair window from the transcription start site which means that dysregulated GReX for a given gene-trait combination may be driven by more than one locus. In our cell-type-specific snTWAS FOCUS analysis, in most loci one gene was fine-mapped (PIP \ge 0.5), with the exception of a few loci prioritizing more than one gene. Interestingly, in a few instances (e.g. MDD PVM), a gene was fine-mapped in two different LD blocks. **B**, Expanded heatmap of S-snTWAS multiple-cell-type gene fine-mapping. The numbers at the top and bottom of each cell indicate the count of fine-mapped genes and loci, respectively. Color indicates the genes per locus for every trait-snTIM combination. Fine-mapping was performed with FOCUS while jointly considering all cell-types at the class level.











Supplementary Fig. 6 | Multivariate adaptive shrinkage (mash) analysis in S-snTWAS traits. Heatmaps depicting the combinatorial probability that an effect exists in cell-type A and not in cell-type B at the class level. Only results with a $P(Cell_A \cap Cell_B^C) \ge 0.2$ in at least one comparison were retained for visualization. The top 20 results (ordered by the maximum row-wise combinatorial probability) are displayed for BD (A), SCZ (B), AD (C), MS (D), AUD (E), ALS (F), PD (G), Anorexia (H), Migraines (I), MDD (J), ADHD (K), Insomnia (L). Cell type probabilities are estimated with mash. Cells are colored gray when the analysis is not applicable.







- Significant
- Non-significant

Supplementary Fig. 7 | Gene-trait association effect size heterogeneity scatterplots. Gene association effect size heterogeneity (I²) across cell types for BD (**A**), SCZ (**B**), AUD (**C**), Migraines (**D**), MDD (**E**), Insomnia (**F**). Significance denotes the presence of different effects of individual cell types in the analysis (Cochran's Q test). Only genes with significant GTAs in S-snTWAS analysis are visualized. Color corresponds to cell-type (Extended Data Fig. 2). Only traits with a high enough sample size in the MVP (see Methods) were evaluated for heterogeneity, due to need for effect size estimates. AD is visualized in Fig. 3C. ADHD, ALS, Anorexia, MS, and PD are excluded due to few cases in the MVP. PTSD and Anxiety are excluded due to few significant S-snTWAS associations.



Supplementary Fig. 8 | GTA:*BIN1***-AD Forest Plot.** Significant associations are marked by an asterisk. Horizontal bars denote the 95% confidence interval.



D	Vesicle Docking Involved In Exocytosis		*	*	*	*			7	۲		*	*	*	*	*		*	*	*	*	*	*	*	: 7	۲ :	*		*	*			
	Protein N Terminus Binding		*			Г			7	ĸ		*	*	*	*	*	*	*	*	*	*	*	Г	*	: 7	k :	*		*	*			
	Regulation Of Short Term						T				Г		*		*										7	k							
	Intercellular Bridge		*	*	*	*	*	*	< 3	ĸ		*		*	*	*	*	*	*	*	*	*	*	*	T	T	*					*	
ay	Regulation Of Autophagosome Assembly	*	*	*	*	*				-		*	*	*	*	*	*	*	*	*	*	*	*	*			*		*	*			
athw	Regulation Of ER To Golgi														+	+								-						*	-		
Å,	Vesicle Mediated Transport	~	~				4				t	-	^		÷	~	4	- -			- -	4					4				-		
		ملد	<u>بل</u> د	<u>_</u>	-						÷	^	J.	^ 	<u>~</u>	<u>ب</u> لە	^ -		۰۰ ملد	^	<u>_</u>	<u>م</u>					^ _			-	^		
	Goigi Transport Complex	×	×	- X	×	×							× 	×	×	×	×	×	×.		×	- X		×			⊼ .L.		-le	~	_	_	
	Negative Regulation Of Macroautophagy	×	×	×	×	×	-		7	ς	ų.	*	×	×	×	×	×	×	×	×	×	×	×	×			*	-	×	×			
	-log ₁₀ (p)			*				*	5			*		*	*					*	*	*	*	L	L		*				*		
	0.0067	ulk	Z	Z Ш	igo	РС	stro	Ira			DD 1		X6	(B2	<	LB	НC	ST	F	Ì	-2	- 3	Ĩ	-2	Ľ	5 6	99	Ē	ЧN	ЫG	M	MC	РС
	2.8	snB			ō	0	As	Ž		Ĺ	ι	ц Ч		DAR	ź	PVA	Ū -	NS	2-3	<u>––</u>	<u>1-1</u>	T-1	<u>1-1</u>	LI-0	4		Z	Ľ	-9-	~	۵.	ZLI	
	5.4								-	=	Ĺ	-5-1	1P5	١٩٢		Z	ALE	_	N	N L	- S-	- е́	3-6	Z	Z		ш	Z Ш	Ľ2				
	8.5											AN S	A	≤			Ч		Ш	Ш	N.	Z	Z	Ш					Ц				
	27											z	Z				\leq																
																	sn]	ΓΙΜ															
Е	Positive Regulation Of Receptor Recycling		*						*		*									*				*									
	Phospholipase Binding	*	*						*		*				⊢			*		*				*				*					
	Regulation Of Glutamate Secretion		*						*		*							*	*	*				*			Г	*					
	Synaptic Vesicle Endocytosis		*						*		*							*		*				*				*					
vay	Wnt Signalosome										*			*				*									Г						
ath	Nucleoplasm Part							*		*		*																	*		*	*	*
₽.	Protein Phosphatase 2a Binding						*										*																
	Regulation Of Microtubule Based Movement						*										*																
	Receptor Agonist Activity													*				*															
	Regulation Of Oxidative Phosphorylation		*						*		*									*				*				*					
	-log ₁₀ (p)	¥	z	z	0	U	0	al	e	0	z	9	23	۵.	Ш.	Ö	Ŀ	<u> </u>	Ţ	Ņ	က္	Ţ	Ņ	F.	ß	E	۵.	Ċ	Σ	e.	U	υ	0
	0.0031	лВи	_	Ш	Olig	ОР	Astı	Mur	mur	Enc	REL	Ē	ARB	2	VAL	ĠH	SS	-3-1	Ę.	Ļ.	Ė.	Ļ	Ė.	0-0	1 L6	-5-E	N-0	Σ	Ы	aptiv	/LM	SM	۵.
	0.24	S							Ē		5-F	-2-1	AD/	=	۲ ۲	Ъ	\leq	Ľ	<u>-</u> 9-	3-5-	3-5-	3-5-	Ľė	Г Z	ш	N	-2-(Ada	/		
	0.7										AMF	AMF	Ζ		-	M		Ш	ΕN	Z			ЫN	ш		ш	Z						
										N	Z				Z				Ш	Ξ	Ξ					_							
	19										=	_					sn⊺	гім															
F	Regulation Of Intracellular Estrogen			ala										_	-	-						_	_	_	_								
	Receptor Signaling Pathway			<u>х</u>	_		×			_	_			-							_	-	_	-	_			-					
	Regulation of Gilar Cell Prolieration			ر •	_	_	<u> </u>							-	⊢						_	4		-			⊢	⊢			_	_	
	Collular Response To Interloukin 4			^	_		~			_								*					_	+	_		H	-					
/ay	T Helper 1 Type Immune Response			_		-												*				-			-								
athw	Transcription Coactivator Activity	\vdash		_														^		_										┍			+
ä	Mammary Gland Enithelial Cell Proliferation	\vdash		_	_	-												*			_	-											^
	Contractile Actin Eilament Bundle Assembly			_	_	_	_			_	*			*	┝			^			-						F						
	Protein Phosphatase 1 Binding	\vdash		-		_					*			*	⊢						-				_		⊢	-					
	Phosphatase Inhibitor Activity			_	_	_					*			^ *	-						_	+	_	+	_		H				_		
	Thosphatase ministor Activity		_	_		4.5		_	4					Ē.	L		<u> </u>	<u> </u>			~~~	_	<u>.</u>	_				L	_	_	4		
	-log ₁₀ (p)	Bulk	Z	Z Ш	Jligo	DPC	stro	lura	nune	:ndo	ELN	HX6	RB2	VIP	ALB	SHC	SST	3-IT	T-1	1-2	T-3	T-1	1-2	Ģ	L6B	ET	-NP	ВМ	MVc	otive	MC	SMC	РС
	snl			0	0	A	Σ	mm	Ш	I-RI	5-LI	DA	Z	ЪХ	B-0	Z	2-0	-9-	-5-1	-2-1	-2-1	-9-	I L6	Ы	N L5	-9-2		ш	Vdap	Z	0)		
										MΡĒ	MPč	N A		Ζ	VAL		IN L	I N I	L3-	L3-	ĽЗ	I N	Ц		Ш	N LÉ			ব				
										ΓAI	ILA				۲ Z		ш	ш	ЫN	ЫN	Ы	ш				Ē							
	0.93										Ζ	Z				-																	
	14 snTIM																																



Supplementary Fig. 9 | Top 10 pathways enriched within all S-snTWAS traits. Pathways were prioritized by p-value. Hierarchical pruning was performed to deprioritize pathways (see Methods). For each cell-type, the top 2 significant pathways after hierarchical pruning (if any) were prioritized.

We describe the following traits in each panel: BD (**A**), MS (**B**), AUD (**C**), ALS (**D**), PD (**E**), Migraines (**F**), MDD (**G**), ADHD (**H**), Insomnia (**I**). AD and SCZ are described within Extended Data Fig. 5B.



-log₁₀ p-value Threshold

BD Control AD Co ニンンンンンンンン soz \sim ross-Disorder Progressive Thresholding Correlatior MDD ADHD $\begin{array}{c} 100\% \\ -0\%$ log₁₀ N(Top Associations)

Supplementary Fig. 10 | Cross-disorder concordance. A, Fisher's exact test Heatmap of cross-disorder significant gene sharing. *Lower-Left Triangle*: The number in each cell describes the number of shared significant genes (*lower-left triangle*) or pathways (*upper-right triangle*) between each pair of disorders. Fisher's exact test p-values were FDR³³-corrected, and significant values are annotated by asterisks. **B**, Cross-disorder association concordance. Each point represents the percent of concordant gene-cell-type (*Lower-Left Triangle*) or pathway-cell-type (*Upper-Right Triangle*) combinations (same association z-score sign) under the appropriate p-value threshold in both disorders. Each progressive point restricts the p-value threshold 10-fold (i.e. $p \le 1$, $p \le 1 \times 10^{-1}$, $p \le 1 \times 10^{-2}$, ...). Red and blue dashed horizontal lines denote a 0.75 (75%) and 0.25 (25%) sign concordance, respectively. **C**, snTWAS cross-disorder association z-score Pearson's correlation analysis (PTCA). Each line graph tracks the cross-disorder association z-score Pearson's correlation for progressively higher ranked (see Methods) gene-cell-type (*Lower-Left Triangle*) or pathway-cell-type (*Upper-Right Triangle*) combinations. Red and blue dashed horizontal lines denote a 0.75 and 0.25 Pearson's correlation coefficient (r), respectively. Vertical dotted lines denote top 1% of gene-cell-type and pathway-cell-type combinations for the lower-left and upper-right triangles, respectively.

	BD	SCZ	AD	MS	AUD	ALS	РD	Anorexia	Migraines	MDD	ADHD	Insomnia	1
BD	1.00	0.69	0.02	0.06	0.08	-0.07	0.06	0.23	0.01	0.46	0.23	0.11	
SCZ	0.69	1.00	0.07	0.07	0.02	0.06	0.00	0.22	-0.03	0.33	0.18	0.02	- 0.8
AD	0.02	0.07	1.00	-0.03	0.00	0.23	0.21	0.07	0.08	-0.04	-0.10	-0.02	- 0.6
MS	0.06	0.07	-0.03	1.00	-0.07	0.03	0.03	0.02	0.00	0.13	0.01	0.06	- 0.4
AUD	0.08	0.02	0.00	-0.07	1.00	0.03	0.05	0.01	-0.29	-0.05	-0.07	0.02	- 0.2
ALS	-0.07	0.06	0.23	0.03	0.03	1.00	0.13	0.15	-0.02	-0.05	0.00	0.02	
PD	0.06	0.00	0.21	0.03	0.05	0.13	1.00	0.12	0.01	-0.01	-0.11	0.01	
Anorexia	0.23	0.22	0.07	0.02	0.01	0.15	0.12	1.00	0.01	0.29	0.10	0.09	0.2
Migraines	0.01	-0.03	0.08	0.00	-0.29	-0.02	0.01	0.01	1.00	0.20	0.13	0.13	0.4
MDD	0.46	0.33	-0.04	0.13	-0.05	-0.05	-0.01	0.29	0.20	1.00	0.66	0.50	0.6
ADHD	0.23	0.18	-0.10	0.01	-0.07	0.00	-0.11	0.10	0.13	0.66	1.00	0.40	0.8
Insomnia	0.11	0.02	-0.02	0.06	0.02	0.02	0.01	0.09	0.13	0.50	0.40	1.00	1

Supplementary Fig. 11 | LD Score Regression bivariate heritability analysis. Heatmap values range from -1 to 1 to indicate heritability correlation. Red indicates positive correlation, violet indicates negative correlation.



Supplementary Fig. 12 | I-snTWAS GTA Heatmap for EUR (A), AFR (B) and AMR (C) ancestries. The number within each cell represents the number of FDR³³-significant gene-trait associations for the respective snTIM. "Class" and "Subclass" represent the number of FDR-significant genes among all class- and subclass-level snTIMs, respectively.



Supplementary Fig. 13 | Overlap of cross-ancestry fine-mapping. A, Overlap of TWAS and fine-mapping in single ancestry and multi-ancestry fine-mapping across the 9 I-snTWAS traits. Annotated numbers in the barplot indicate the total number of GTAs in each category (i.e. there are 550 fine-mapped GTAs in EUR and 1,636 I-snTWAS significant GTAs). Y-axis indicates the population in which analysis was performed. "MA" indicates the MA-FOCUS bi-ancestral analysis and the union of EUR and AFR I-snTWAS significant trait-gene-cell-type combinations for fine-mapping and TWAS, respectively. B, Overlap of fine-mapped GTAs (FOCUS for EUR and AFR; MA-FOCUS

for bi-ancestral EUR and AFR). **C**, Distribution of PIP values (FOCUS for EUR and AFR, MA-FOCUS for bi-ancestral EUR and AFR) amongst I-snTWAS significant trait-gene-cell-type combinations for each ancestry ("MA" indicates the union of EUR and AFR I-snTWAS significant trait-gene-cell-type combinations). Square points indicate the median values of each distribution and circular points indicate the mean values of each distribution.



Scaled TWAS Z







Supplementary Fig. 14 | MAFOCUS plots for NPD/NDDs. Bi-ancestral (EUR and AFR) fine-mapping of BD (**A**), SCZ (**B**), AD (**C**), Insomnia (**D**), MDD (**E**), Migraines (**F**), Anxiety (**G**), PTSD (**H**). The top 10 trait-associated genes in the AFR I-snTWAS analysis are visualized across class level snTIMs. Z-scores are scaled across all genes within each population. AUD is visualized in Fig. 5E.



Supplementary Fig. 15 | Conservation of cross-ancestry concordance of GReX-PheWAS associations among top trait-gene-cell-type combinations. PheWAS was performed for all genes found to have a significant (Bonferroni-adjusted p-value ≤ 0.05) association in the I-snTWAS analysis spanning all snTIMs and ancestries (EUR, AFR and AMR). For each pairwise ancestry combination, trait-specific PheWAS associations were binned (x-axis) based on the maximum p-value of the shared gene-cell-type combinations (n=1,757, 880 and 681 in EUR-AFR, EUR-AMR, and AFR-AMR, respectively); only phecodes with at least 500 cases and 500 controls were considered. Consequently, the percentage of associations with concordant effect size signs for each ancestry pairwise combination is visualized among increasing p-value significance thresholds. Shaded region around each line represents the 95% confidence interval. Size of the point represents the number of gene-cell-type combinations corresponding to 2% of the applicable gene-cell-type combinations (36, 18 and 14 for EUR-AFR, EUR-AMR and AFR-AMR, respectively). Please note that we are underpowered for the AFR-AMR comparison.



Supplementary Fig. 16 | *CELF1* sn-eQTLs for inhibitory and excitatory neurons. Sn-eQTLs for *CELF1* across 2 cell-types. Cell-types were chosen based on the differing association of cell-type specific *CELF1* GReX with various phenotypes. Labeled points indicate SNPs in which there is a significant (FDR³³ \leq 0.05; threshold for significance is indicated by horizontal dashed lines) effect in one cell-type but not the other. No SNPs contained significant effects in opposite directions for the two cell-types. Plotted SNPs were selected using a superset of SNP predictors in both *CELF1* snTIMs. Upward arrow indicates positive effect size and downward arrow indicates negative effect size. The vertical dotted line divides the LD blocks (utilizing EUR LD blocks as previously described¹¹³).



「」 「」 「」 「」 」 「」 「」 「」 「」 「」





Supplementary Fig. 17 | Clustering of top PheWAS associations in S-snTWAS. PheWAS for the top 20 gene-cell-type combinations ranked from the S-snTWAS for each trait; each gene is represented by the cell-type with the lowest association p-value. In addition to each trait (mapped to the bolded italicized phecode), the top 20 (if at least 20 distinct phecodes are significant amongst the op 20 gene-cell-type combinations) phecodes among all phecode categories ranked by association p value are visualized. Ward's hierarchical agglomerative clustering was performed with Ward's criterion preservation. We describe the following traits in each panel: BD (**A**), SCZ (**B**), AD (**C**), Migraines (**D**), MDD (**E**), Insomnia (**F**). AUD is described within fig. 6C. ADHD, ALS, Anorexia, MS, and PD are excluded due to few cases in the MVP. PTSD and Anxiety are excluded due to few significant S-snTWAS associations.



Supplementary Fig. 18 | Comparison of batch effect correction methods. To assess the impact of different batch effect correction methods on overall power, we utilized the FACS-MG cohort. "rePEER" refers to the method to perform PEER on each batch separately and again altogether after aggregation of all PEER-residualized expression data. "PEER" refers to the method to just perform PEER on all expression data aggregated across all batches. "PEER+PC1:3" refers to the method to perform PEER on all expression data aggregated across all batches with genotyping PCs 1, 2, and 3 supplied to the PEER algorithm. Each point represents the number of significant eQTLs

(FDR³³-adjusted p-value \leq 0.05) found at each selected number of PEER factors used in each method (using the second round of PEER for the "rePEER" method).



Supplementary Fig. 19 | GWAS imputation accuracy. We randomly selected SNPs for each GWAS (Table S17) (mean number of selected SNPs was 23,018 with a standard deviation of 14,334). For all selected SNPs, we performed GWAS imputation as if the SNP was missing so that we could observe the accuracy of GWAS imputation on missing data. GWAS imputation is described for each included trait: BD (A), SCZ (B), AD (C), MS (D), AUD (E), ALS (F), PD (G), Anorexia (H), Migraines (I), MDD (J), ADHD (K), Insomnia (L), Anxiety (M), PTSD (N). Every plot is annotated with the Pearson's correlation coefficient (r) and p-value. Color of the point indicates the imputation r^2 . Only missing SNPs with imputation $r^2 \ge 0.7$ were retained for further analysis. Black diagonal line is a reference line with a slope of 1.

SUPPLEMENTARY NOTES

Other Acknowledgments

VA Million Veteran Program: Core Acknowledgements for Publications May 2024

MVP Program Office

Sumitra Muralidhar, Ph.D., Program Director¹ Jennifer Moser, Ph.D., Associate Director, Scientific Programs¹ Jennifer E. Deen, B.S., Associate Director, Cohort & Public Relations¹

MVP Executive Committee

Co-Chair: Philip S. Tsao, Ph.D.² Co-Chair: Sumitra Muralidhar, Ph.D.¹ J. Michael Gaziano, M.D., M.P.H.³ Elizabeth Hauser, Ph.D.⁴ Amy Kilbourne, Ph.D., M.P.H.⁵ Michael Matheny, M.D., M.S., M.P.H.⁶ Dave Oslin, M.D.⁷ Deepak Voora, MD⁴

MVP Co-Principal Investigators

J. Michael Gaziano, M.D., M.P.H.³ Philip S. Tsao, Ph.D.²

MVP Core Operations

Jessica V. Brewer, M.P.H., Director, MVP Cohort Operations³ Mary T. Brophy M.D., M.P.H., Director, VA Central Biorepository³ Kelly Cho, M.P.H, Ph.D., Director, MVP Phenomics³ Lori Churby, B.S., Director, MVP Regulatory Affairs² Scott L. DuVall, Ph.D., Director, VA Informatics and Computing Infrastructure (VINCI)⁸ Saiju Pyarajan Ph.D., Director, Data and Computational Sciences³ Robert Ringer, Pharm.D., Director, VA Albuquerque Central Biorepository⁹ Luis E. Selva, Ph.D., Director, MVP Biorepository Coordination³ Shahpoor (Alex) Shayan, M.S., Director, MVP PRE Informatics³ Brady Stephens, M.S., Principal Investigator, MVP Information Center¹⁰ Stacey B. Whitbourne, Ph.D., Director, MVP Cohort Development and Management³

MVP Publications and Presentations Committee

Co-Chair: Themistocles L. Assimes, M.D., Ph. D² Co-Chair: Adriana Hung, M.D.; M.P.H⁶ Co-Chair: Henry Kranzler, M.D.⁷

MVP Affilations

1: US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

2: VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

3: VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

4: Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705

5: VA HSR&D, 2215 Fuller Road, Ann Arbor, MI 48105

6: VA Tennessee Valley Healthcare System, 1310 24th Ave. South, Nashville, TN 37212

7: Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104

8: VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT 84148

9: New Mexico VA Health Care System, 1501 San Pedro Drive SE, Albuquerque, NM 87108

10: Canandaigua VA Medical Center, 400 Fort Hill Avenue, Canandaigua, NY 14424

PsychAD Acknowledgements

PsychAD Consortium Authors

Aram Hong (1, 4, 6, 7); Athan Z. Li (10, 12); Biao Zeng (1, 4, 6, 7); Chenfeng He (9, 12); Chirag Gupta (9, 12); Christian Porras (1, 4, 6, 7); Clara Casey (1, 4, 6, 7); Colleen A. McClung (18); Collin Spencer (1, 4, 6, 7); Daifeng Wang (9, 10, 12); David A. Bennett (19); David Burstein (1, 2, 4, 6, 7, 8); Deepika Mathur (1, 4, 6, 7); Donghoon Lee (1, 4, 6, 7); Fotios Tsetsos (1, 2, 4, 6, 7); Gabriel E. Hoffman (1, 2, 4, 6, 7, 8); Genadi Ryan (13, 17); Georgios Voloudakis (1, 2, 3, 4, 6, 7, 8); Hui Yang (1, 4, 6, 7); Jaroslav Bendl (1, 4, 6, 7); Jerome J. Choi (11, 12); John F. Fullard (1, 4, 6, 7); Kalpana H. Arachchilage (9, 12); Karen Therrien (1, 4, 6, 7); Kiran Girdhar (1, 4, 6, 7); Lars J. Jensen (21); Lisa L. Barnes (19); Logan C. Dumitrescu (22, 23); Lyra Sheu (1, 4, 6, 7); Madeline R. Scott (18); Marcela Alvia (1, 4, 6, 7); Marios Anyfantakis (1, 4, 6, 7); Maxim Signaevsky (6, 7); Mikaela Koutrouli (1, 4, 6, 7, 21); Milos Pjanic (1, 4, 6, 7); Monika Ahirwar (13, 17); Nicolas Y. Masse (1, 4, 6, 7); Noah Cohen Kalafut (10, 12); Panos Roussos (1, 2, 4, 6, 7, 8); Pavan K. Auluck (20); Pavel Katsel (6); Pengfei Dong (1, 4, 6, 7); Pramod B. Chandrashekar (9, 12); Prashant N.M. (1, 4, 6, 7); Rachel Bercovitch (1, 4, 6, 7); Roman Kosoy (1, 4, 6, 7); Sanan Venkatesh (1, 2, 4, 6, 7); Saniya Khullar (9, 12); Sayali A. Alatkar (10, 12); Seon Kinrot (1, 4, 6, 7); Stathis Argyriou (1, 4, 6, 7); Stefano Marenco (20); Steven Finkbeiner (13, 14, 15, 16, 17); Steven P. Kleopoulos (1, 4, 6, 7); Tereza Clarence (1, 4, 6, 7); Timothy J. Hohman (22, 23); Ting Jin (9, 12); Vahram Haroutunian (5, 6, 7, 8); Vivek G. Ramaswamy (13, 17); Xiang Huang (12); Xinyi Wang (1, 4, 6, 7); Zhenyi Wu (1, 4, 6, 7); Zhiping Shao (1, 4, 6, 7)

PsychAD Consortium Affiliations

1: Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA 2: Center for Precision Medicine and Translational Therapeutics, James J. Peters VA Medical Center, Bronx, NY, USA

3: Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

4: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

5: Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

6: Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

7: Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

8: Mental Illness Research, Education and Clinical Center VISN2, James J. Peters VA Medical Center, Bronx, NY, USA

9: Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA

10: Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, USA

11: Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI, USA

12: Waisman Center, University of Wisconsin-Madison, Madison, WI, USA

13: Center for Systems and Therapeutics, Gladstone Institutes, San Francisco, CA, USA

14: Department of Neurology, University of California San Francisco, San Francisco, CA, USA

15: Department of Physiology, University of California San Francisco, San Francisco, CA, USA

16: Neuroscience and Biomedical Sciences Graduate Programs, University of California San Francisco, San Francisco, CA, USA

17: Taube/Koret Center for Neurodegenerative Disease Research, Gladstone Institutes, San Francisco, CA, USA

18: Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA19: Rush Alzheimer's Disease Center and Department of Neurological Sciences, Rush UniversityMedical Center, Chicago, IL, USA

20: Human Brain Collection Core, National Institute of Mental Health-Intramural Research Program, Bethesda, MD, USA

21: Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

22: Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA 23: Vanderbilt Memory & Alzheimer's Center, Vanderbilt University Medical Center, Nashville, TN, USA