Supplementary Materials:
Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 90,000 patients across three healthcare systems

Amanda B Zheutlin, Jessica Dennis, Nicole Restrepo, Peter Straub, Douglas Ruderfer, Victor M Castro, Chia-Yen Chen, H Lester Kirchner, Christopher F Chabris, Lea K Davis, & Jordan W Smoller

Supplementary Methods

An overview of the genotype quality control methods employed across sites was described in the main text (Methods: Quality Control of Genetic Data). Details for each site were listed below.

GHS: The Geisinger patient samples (N = 59,085) were genotyped on the Illumina Human OmniExpress Exome BeadChip (958,497 markers) by the Regeneron Genetics Center. Samples were genotyped in batches of ~20,000-25,000 samples. Individuals and SNPs were filtered based on call rate (<1%), minor allele frequency (<1%), and Hardy Weinberg equilibrium (HWE) ($p < 10^{-7}$). A random individual from pairs of related individuals was removed (pihat > 0.125). Patients with European ancestry determined genetically were retained. Strand alignment, checks, and phasing were performed with SHAPEIT v2; palindromic SNPs were removed during strand alignment. Imputation was performed with IMPUTE2 separately by batch on autosomal chromosomes with the 1000 Genomes Phase 1 reference panel. IMPUTE2 files were converted to hard-call PLINK format using filters for genotype probability (>0.9) and INFO score (>0.7). Batches were merged after imputation in PLINK v1.90. Principal components to include as ancestry covariates were generated in EIGENSOFT.

PHS: 25,540 individuals were genotyped on one of three different Illumina arrays (MEGA, MEGA^{EX}, and MEG BeadChip). Individuals and SNPs were excluded from each array for call rate (<2%), sex errors, and heterozygosity ($|F_{het}| > 0.2$); then all three arrays were merged using only variants present in all three datasets. SNPs that had allele frequencies that differed between chips (>1%) or showed significant array effects ($p < 10^{-6}$) in the merged dataset were excluded. To identify and extract individuals of European-American ancestry, we pruned the initial merged dataset by linkage disequilibrium and randomly removed one individual from pairs with apparent relatedness (pihat > 0.2). We merged this dataset with the 1000 Genomes sample[2] and generated principal components (PCs). 1000 Genomes has labeled individuals within their dataset as belonging to one of five ancestrally distinct super-populations. We used the first four PCs to identify individuals from PHS that clustered with the European 1000 Genomes super-population (N = 19,136). Next, we used 1000 Genomes Phase III reference panel to impute the initial merged dataset (N_{SNP} = 1,345,786) with only unrelated European-American individuals (N = 19,136). We used Eagle v2.3.5 for prephasing and Minimac3 for imputation. Dosage files were converted to hard-call genotypes with PLINK v1.90 using filters for genotyping probability (P >.8), INFO score (>.9), and SNP missingness (<2%). After a final round of quality control using filters for violations of HWE ($p < 10^{-4}$) and minor allele frequency (<1%), our dataset included 19,136 European-American patients and 6,237,592 variants.

VUMC: A subset of BioVU patients (N = 24,262) were genotyped on the Illumina MEGA^{EX} platform of more than 2 million markers. Ruderfer and colleagues described the first phase of quality control elsewhere[1], including filters for SNP and individual call rate (<2%), minor allele frequency (<1%), violations of HWE ($p < 5 \times 10^{-5}$), batch effects ($p < 5 \times 10^{-5}$), heterozygosity ($|F_{het}| > 0.2$), and relatedness (pihat > 0.2). Ancestry principal components were used to identify individuals of European ancestry, and SHAPEIT / IMPUTE2 were used to pre-phase and impute genotypes according to the 1000 Genomes Phase I reference panel. In the second phase of quality control, we converted dosage data to hard genotype calls and excluded variants with certainty less than 0.9 or INFO < 0.95. After these quality control measures, 18,385 individuals remained. In all subsequent analyses, we used genotype batch and the first 10 ancestry principal components calculated by multidimensional scaling in PLINK v1.90 as covariates.

Table S1.
Schizophrenia PRS PheWAS Results by Site

		GHS					PHS					VUMC				
Phecode	**Description**	**OR**	***p***	**Cases**	**Controls**	**Rank**	**OR**	***p***	**Cases**	**Controls**	**Rank**	**OR**	***p***	**Cases**	**Controls**	**Rank**
300	Anxiety, phobic and dissociative disorders	1.095	**5.99×10^{-16}**	16,749	33,986	2	1.205	**2.36×10^{-15}**	3,178	11,718	1	1.045	0.0542	2,631	14,169	74
300.1	Anxiety disorder	1.100	**1.32×10^{-16}**	14,848	36,096	1	1.198	**9.03×10^{-13}**	2,640	12,400	2	1.051	0.0467	2,114	14,817	68
296.1	Bipolar	1.179	**2.98×10^{-8}**	1,573	54,343	8	1.297	**3.27×10^{-7}**	537	15,918	6	1.248	**2.50×10^{-5}**	419	17,739	1
295.1	Schizophrenia	1.644	**2.09×10^{-12}**	267	55,968	3	1.657	**1.22×10^{-5}**	100	16,513	8	NA	NA	67	18,254	NA
295	Schizophrenia and other psychotic disorders	1.291	**5.60×10^{-8}**	607	54,645	9	1.434	**1.85×10^{-7}**	286	15,999	5	1.238	4.78×10^{-4}	304	17,611	2
296	Mood disorders	1.061	**7.78×10^{-8}**	17,496	34,791	10	1.158	**5.75×10^{-11}**	3,529	11,895	3	1.055	0.0099	3,370	13,418	20
316	Substance addiction and disorders	1.181	**5.35×10^{-9}**	1,702	53,445	5	1.204	1.17×10^{-4}	609	15,660	11	1.167	0.0010	532	17,461	4
296.2	Depression	1.052	**4.53×10^{-6}**	16,587	35,701	14	1.145	**7.98×10^{-9}**	3,071	12,366	4	1.048	0.0279	3,025	13,794	40
318	Tobacco use disorder	1.070	**2.30×10^{-8}**	11,910	41,219	7	1.125	0.0012	1,110	14,760	14	1.042	0.1203	1,763	15,394	147
278.11	Morbid obesity	0.930	**4.15×10^{-10}**	13,462	40,275	4	0.977	0.5728	840	15,400	542	0.943	0.0755	1,113	16,648	94
300.9	Posttraumatic stress disorder	1.185	0.0021	440	55,592	45	1.287	1.03×10^{-4}	323	16,199	10	1.243	0.0021	230	18,035	6
296.22	Major depressive disorder	1.074	1.31×10^{-4}	4,167	47,526	21	1.101	0.0063	1,173	15,106	33	1.133	0.0010	837	17,086	3
317	Alcohol-related disorders	1.123	**1.48×10^{-5}**	1,939	53,241	17	1.096	0.0427	686	15,518	83	1.136	0.0044	587	17,421	11
599.3	Dysuria	1.083	**2.27×10^{-6}**	5,488	43,661	13	1.105	0.0766	432	15,331	123	1.096	0.0219	764	16,339	31
300.11	Generalized anxiety disorder	1.094	**1.30×10^{-6}**	4,205	50,206	12	1.089	0.1445	407	15,999	190	1.146	0.0228	325	17,760	33
301	Personality disorders	1.261	**2.53×10^{-7}**	667	55,149	11	1.14	0.1663	151	16,435	211	NA	NA	91	18,202	NA
798	Malaise and fatigue	1.038	0.0019	13,091	31,868	43	1.1	**2.36×10^{-5}**	3,685	10,604	9	1.042	0.0241	5,257	10,123	34
599	Other symptoms/disorders or the urinary system	1.049	**2.76×10^{-5}**	15,429	31,937	18	1.05	0.0591	2,430	12,621	105	1.053	0.0222	2,735	13,605	32
297.1	Suicidal ideation	1.198	0.0036	346	55,388	51	1.466	**7.37×10^{-6}**	188	16,314	7	NA	NA	98	18,127	NA
297	Suicidal ideation or attempt	1.189	2.82×10^{-4}	595	54,922	25	1.139	0.0063	618	14,973	31	1.208	0.0314	146	18,067	47
278.1	Obesity	0.943	**1.45×10^{-8}**	22,597	29,077	6	1.007	0.7791	2,833	12,827	696	0.985	0.5003	2,546	14,694	488

300.12	Agorophobia, social phobia, and panic disorder	1.133	**9.48×10^{-6}**	1,752	53,402	15	1.154	0.0269	330	16,076	59	0.998	0.9829	188	17,990	883
317.1	Alcoholism	1.114	5.93×10^{-4}	1,387	53,812	29	1.107	0.0442	542	15,693	85	1.130	0.0167	444	17,587	29
539	Bariatric surgery	0.918	5.24×10^{-5}	3,238	52,699	19	0.881	0.0518	324	16,243	96	0.945	0.3954	256	18,024	394
532	Dysphagia	1.082	2.64×10^{-4}	3,026	49,917	24	1.072	0.0781	907	15,003	127	1.047	0.1048	1,505	15,833	131
427.9	Palpitations	1.054	0.0054	4,028	47,923	54	1.075	0.0152	1,741	13,518	42	1.057	0.0285	2,085	14,793	43

Phenotypes were listed in rank order of significance from the schizophrenia PRS PheWAS meta-analysis. Individual effects were listed for each site and those that surpassed the site-specific Bonferroni-corrected significance threshold were bolded (GHS, 1263 phecodes, $p < 3.96 \times 10^{-5}$; PHS, 850 phecodes, $p < 5.88 \times 10^{-5}$; VUMC, 897 phecodes, $p < 5.57 \times 10^{-5}$).

Table S2.
Schizophrenia PRS PheWAS Meta-Analysis Conditioned on Schizophrenia Diagnosis

Phecode	Description	Beta	SE	OR	*p*	Total	Cases	Controls
300.1	Anxiety disorder	0.093	1.098	0.010	**1.54×10^{-21}**	82,177	19,316	62,861
300	Anxiety, phobic and dissociative disorders	0.088	1.092	0.009	**8.91×10^{-21}**	81,709	22,222	59,487
278.11	Morbid obesity	-0.071	0.932	0.011	**3.29×10^{-11}**	86,917	15,329	71,588
296	Mood disorders	0.060	1.061	0.009	**5.34×10^{-11}**	83,735	23,933	59,802
316	Substance addiction and disorders	0.145	1.156	0.022	**1.09×10^{-10}**	88,611	2,735	85,876
296.1	Bipolar	0.147	1.159	0.024	**1.23×10^{-9}**	89,721	2,422	87,299
296.2	Depression	0.052	1.054	0.009	**1.53×10^{-8}**	83,786	22,265	61,521
318	Tobacco use disorder	0.060	1.062	0.011	**2.05×10^{-8}**	85,362	14,634	70,728
278.1	Obesity	-0.047	0.954	0.009	**9.02×10^{-8}**	83,790	27,765	56,025
300.11	Generalized anxiety disorder	0.087	1.091	0.017	**4.19×10^{-7}**	88,084	4,878	83,206
599.3	Dysuria	0.074	1.077	0.015	**9.60×10^{-7}**	81,261	6,617	74,644
539	Bariatric surgery	-0.089	0.915	0.019	**3.61×10^{-6}**	89,942	3,798	86,144
317	Alcohol-related disorders	0.096	1.100	0.021	**5.33×10^{-6}**	88,588	3,103	85,485
599	Other symptoms/disorders or the urinary system	0.043	1.044	0.010	**9.03×10^{-6}**	78,034	20,338	57,696
798	Malaise and fatigue	0.041	1.042	0.009	**9.84×10^{-6}**	73,928	21,595	52,333
250.2	Type 2 diabetes	-0.040	0.960	0.009	**1.01×10^{-5}**	87,355	22,411	64,944
250	Diabetes mellitus	-0.040	0.961	0.009	**1.12×10^{-5}**	87,202	23,009	64,193
300.9	Posttraumatic stress disorder	0.164	1.178	0.038	**1.43×10^{-5}**	89,987	922	89,065
300.12	Agorophobia, social phobia, and panic disorder	0.107	1.113	0.025	**1.60×10^{-5}**	88,918	2,230	86,688
296.22	Major depressive disorder	0.067	1.069	0.016	**1.69×10^{-5}**	85,113	5,998	79,115
427.9	Palpitations	0.054	1.056	0.014	6.73×10^{-5}	83,344	7,742	75,602
727.1	Synovitis and tenosynovitis	-0.061	0.941	0.016	1.16×10^{-4}	84,210	5,707	78,503
317.1	Alcoholism	0.093	1.098	0.024	1.17×10^{-4}	88,651	2,287	86,364
532	Dysphagia	0.061	1.062	0.016	1.65×10^{-4}	85,418	5,301	80,117
278	Overweight, obesity and other hyperalimentation	-0.032	0.969	0.009	2.43×10^{-4}	81,497	32,366	49,131

Effects that surpassed the Bonferroni-corrected significance threshold (3.7×10^{-5}) were bolded.

Table S3.
Biobank and Hospital Demographics and Case Prevalence

	GHS		PHS		VUMC	
	Biobank N = 56,926	Hospital N = 1,184,064	Biobank N = 16,684	Hospital N = 1,855,688	Biobank N = 18,358	Hospital N = 2,354,159
Mean age, years (SD)	58.6 (17.3)	46.0 (21.9)	57.7 (16.1)	47.4 (19.1)	60.7 (17.2)	48.6 (23.6)
Females, N (%)	35,111 (59%)	628,592 (53%)	9,003 (54%)	1,074,789 (58%)	9,350 (51%)	1,278,720 (54%)
Median EHR length, days	4,811	1,710	3,910	377	3,568	533
300: Anxiety, phobic and dissociative disorders	16,749 (29%)	144,380 (12%)	3,178 (19%)	163,248 (9%)	2,631 (14%)	123,134 (5%)
300.1: Anxiety disorder	14,848 (26%)	129,330 (11%)	2,640 (16%)	132,354 (7%)	2,114 (12%)	92,584 (4%)
296.1: Bipolar	1,573 (3%)	16,911 (1%)	537 (3%)	23,972 (1%)	419 (2%)	28,826 (1%)
295.1: Schizophrenia	267 (1%)	4,587 (0%)	100 (1%)	7,847 (0%)	67 (0%)	7,799 (0%)
295: Schizophrenia and other psychotic disorders	607 (1%)	8,424 (1%)	286 (2%)	19,733 (1%)	304 (2%)	21,383 (1%)
296: Mood disorders	17,496 (31%)	144,519 (12%)	3,529 (21%)	177,489 (10%)	3,370 (18%)	135,233 (6%)
316: Substance addiction and disorders	1,702 (3%)	12,015 (1%)	609 (4%)	34,340 (2%)	532 (3%)	47,186 (2%)
296.2: Depression	16,587 (29%)	120,139 (10%)	3,071 (18%)	143,464 (8%)	3,025 (16%)	111,064 (5%)
318: Tobacco use disorder	11,910 (21%)	91,858 (8%)	1,110 (7%)	64,813 (4%)	1,763 (10%)	113,105 (5%)
278.11: Morbid obesity	13,462 (24%)	56,636 (5%)	841 (5%)	32,165 (2%)	1,113 (6%)	37,502 (2%)
300.9: Posttraumatic stress disorder	440 (1%)	2,050 (0%)	323 (2%)	13,728 (1%)	230 (1%)	16,433 (1%)
296.22: Major depressive disorder	4,167 (7%)	16,419 (1%)	1,173 (7%)	52,397 (3%)	837 (5%)	45,794 (2%)
317: Alcohol-related disorders	1,939 (3%)	15,519 (1%)	686 (4%)	38,169 (2%)	587 (3%)	41,803 (2%)
599.3: Dysuria	5,488 (10%)	49,048 (4%)	432 (3%)	24,256 (1%)	764 (4%)	55,857 (2%)
300.11: Generalized anxiety disorder	4,205 (7%)	32,541 (3%)	407 (2%)	16,965 (1%)	325 (2%)	17,997 (1%)
301: Personality disorders	667 (1%)	3,168 (0%)	151 (1%)	4,102 (0%)	91 (1%)	6,059 (0%)
798: Malaise and fatigue	13,091 (23%)	101,495 (9%)	3,686 (22%)	139,168 (8%)	5,257 (29%)	199,327 (9%)
599: Other symptoms / disorders of the urinary system	15,429 (27%)	125,269 (11%)	2,430 (15%)	103,605 (6%)	2,735 (15%)	127,729 (5%)
297.1: Suicidal ideation	346 (1%)	1,981 (0%)	188 (1%)	5,955 (0%)	98 (1%)	17,355 (1%)
297: Suicidal ideation or attempt	595 (1%)	2,805 (0%)	618 (4%)	27,048 (2%)	146 (1%)	24,973 (1%)
278.1: Obesity	22,597 (40%)	134,552 (11%)	2,834 (17%)	120,627 (7%)	2,546 (14%)	98,891 (4%)
300.12 Agorophobia, social phobia, and panic disorder	1,752 (3%)	12,763 (1%)	330 (2%)	13,650 (1%)	188 (1%)	11,367 (1%)
317.1: Alcoholism	1,387 (2%)	10,843 (1%)	542 (3%)	29,753 (2%)	444 (2%)	31,849 (1%)
539: Bariatric surgery	3,238 (6%)	10,094 (1%)	324 (2%)	11,116 (1%)	256 (1%)	8,077 (0%)

532: Dysphagia	3,026 (5%)	17,460 (2%)	907 (5%)	38,595 (2%)	1,505 (8%)	54,845 (2%)
427.9: Palpitations	4,028 (7%)	27,386 (2%)	1,743 (10%)	58,027 (3%)	2,085 (11%)	65,812 (3%)

Mean age is the average age of the patients defined by the most recent hospital visit in which they received an ICD-9 code. Hospital values reflect all patients with an EHR visit at age 10 or later.

Figure S1
Case Prevalence by PRS Ventile for Phenome-Wide Significant Phenotypes

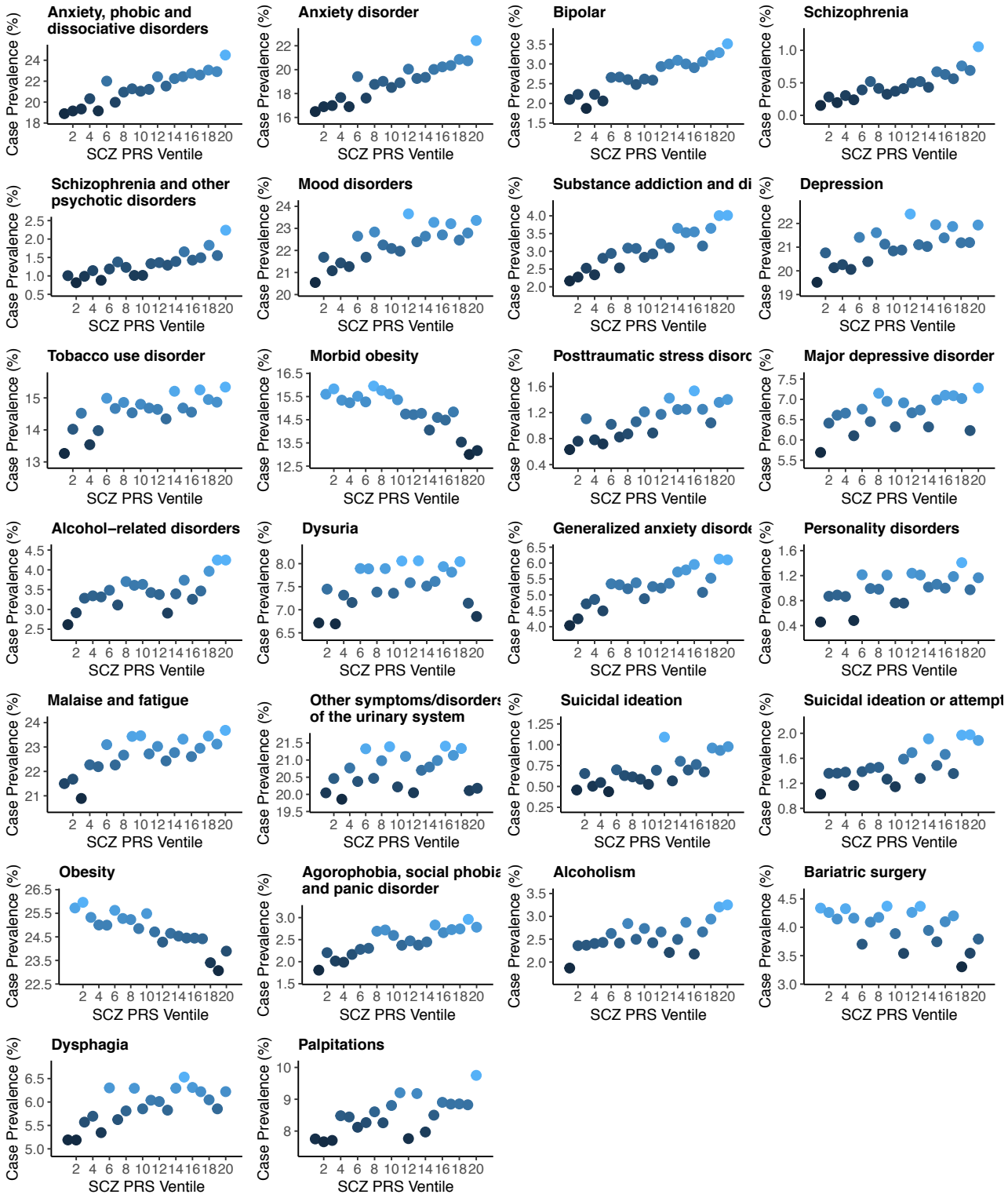
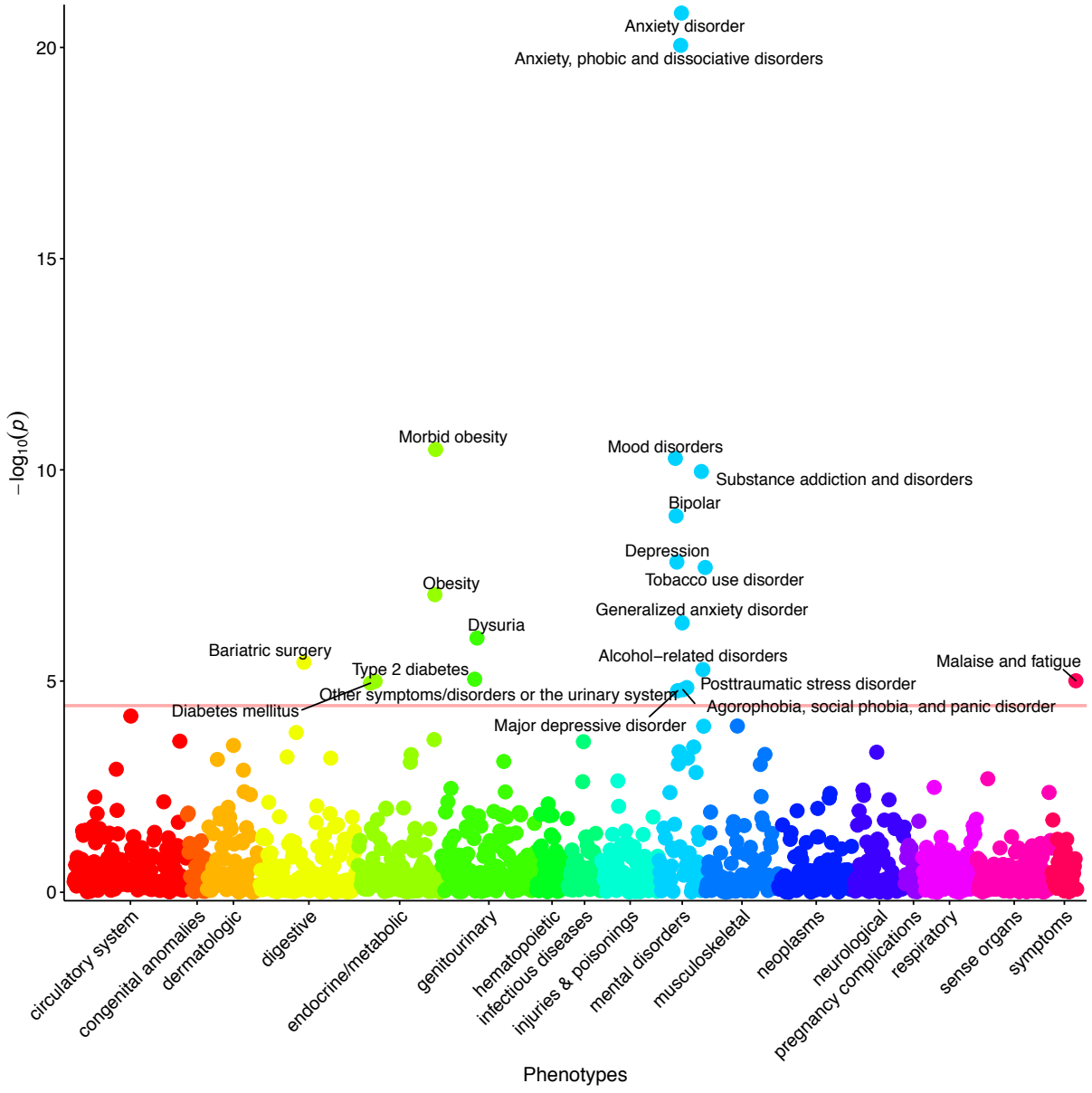


Figure S2.
Schizophrenia PRS PheWAS Meta-Analysis Conditioned on Schizophrenia Diagnosis



Manhattan plot for phenome-wide association analysis with schizophrenia polygenic risk scores conditioned on psychosis-related diagnoses (phecode 295) meta-analyzed across three healthcare systems (1338 phenotypes; 91,980 patients). The *x* axis is phenotype (grouped by broad disease category) and the *y* axis is significance ($-\log_{10} P$; 2-tailed) of association derived by logistic regression including covariates for median age across the medical record, sex, top 10 principal components, and phecode 295. The red line shows phenome-wide level significance (3.7×10^{-5}). All significant effects were positive (i.e., higher polygenic risk scores resulted in higher incidence of the phenotype) with five exceptions: morbid obesity, obesity, bariatric surgery, type 2 diabetes, and diabetes mellitus.

References

1. Ruderfer DM, Walsh CG, Aguirre MW, et al. Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *bioRxiv*. 2018. doi:10.1101/266411.
2. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393.