

**Cell Genomics, Volume 3**

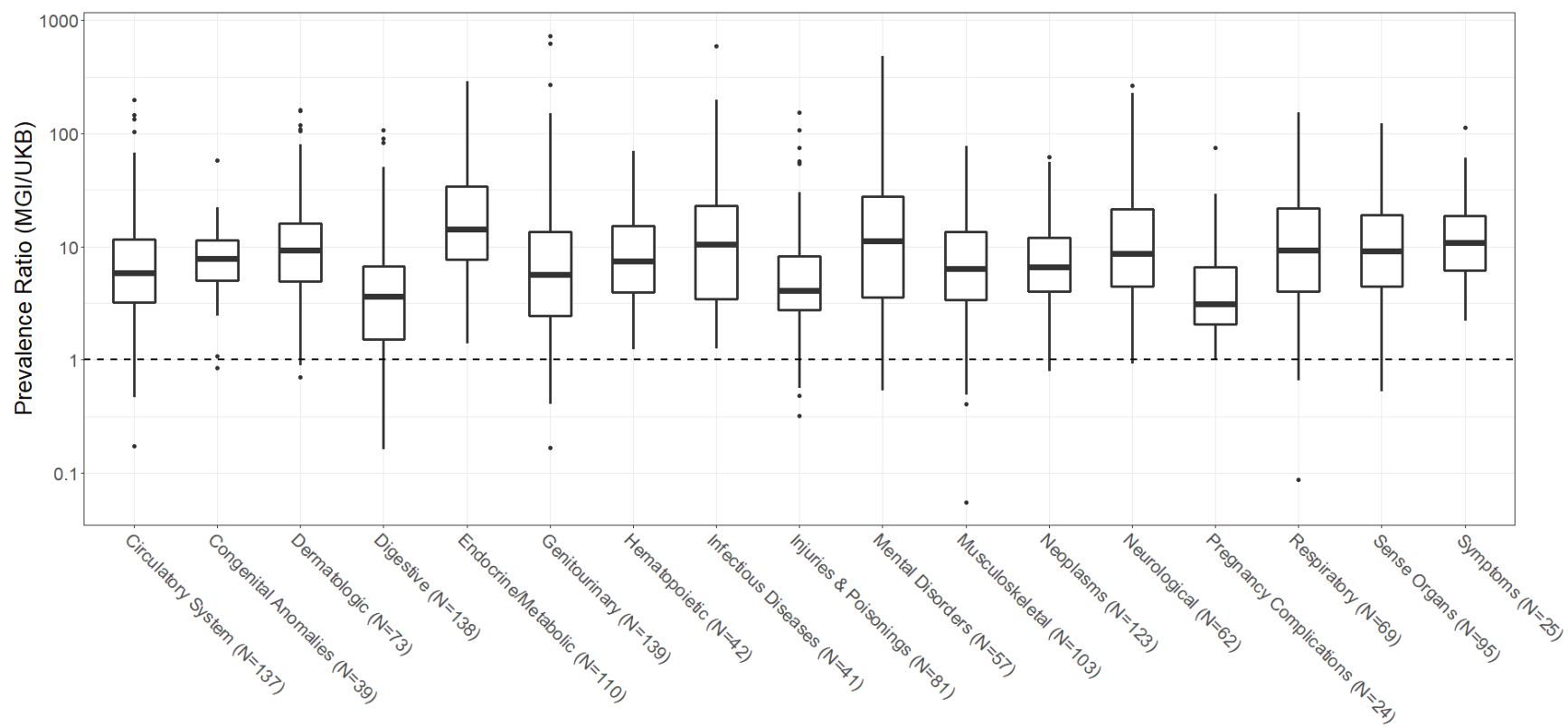
**Supplemental information**

**The Michigan Genomics Initiative: A biobank  
linking genotypes and electronic clinical  
records in Michigan Medicine patients**

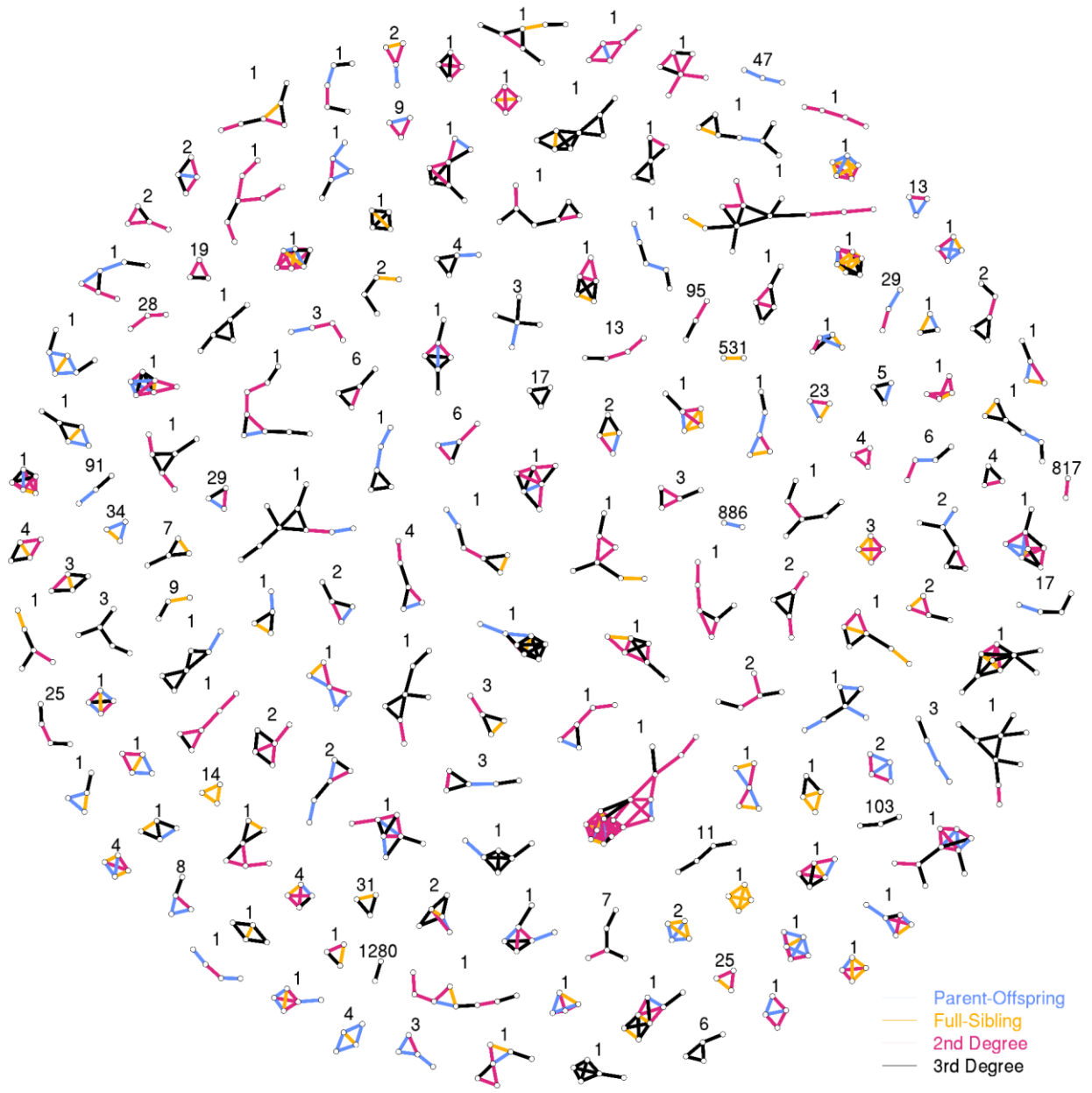
**Matthew Zawistowski, Lars G. Fritsche, Anita Pandit, Brett Vanderwerff, Snehal Patil, Ellen M. Schmidt, Peter VandeHaar, Cristen J. Willer, Chad M. Brummett, Sachin Kheterpal, Xiang Zhou, Michael Boehnke, Gonçalo R. Abecasis, and Sebastian Zöllner**

**Supplementary Material for “The Michigan Genomics Initiative: a biobank linking genotypes and electronic clinical records in Michigan Medicine patients.”**

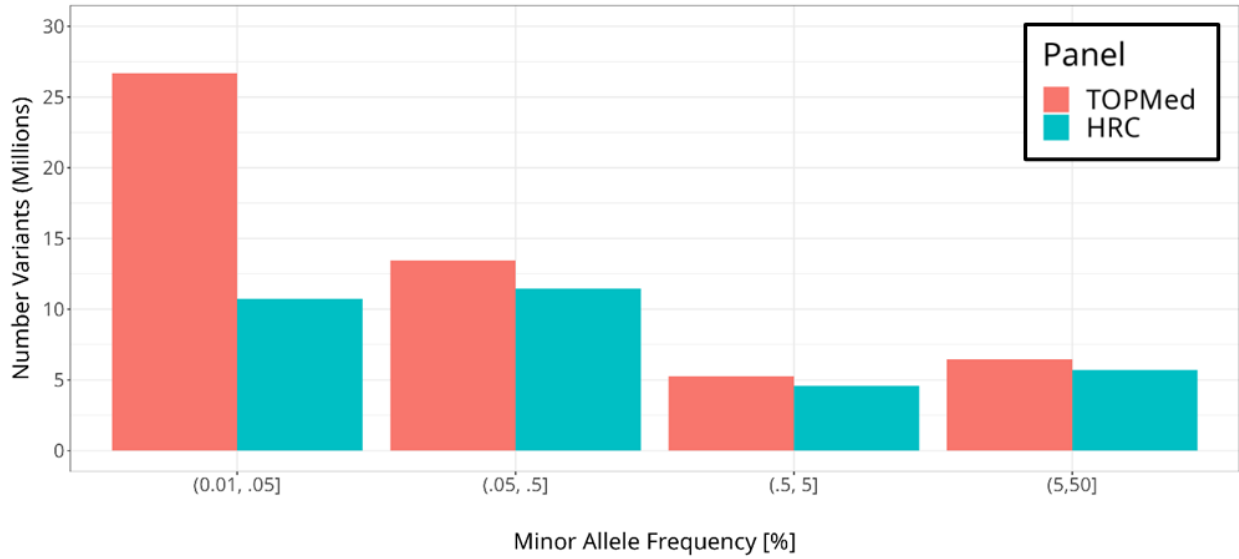
**Figure S1: Ratio of disease prevalence in MGI and UKB by Phecode categories. Related STAR Methods (Clinical Phenotype Data, Phecodes in UK Biobank).**



**Figure S2: Unique genetically-inferred familial configurations involving Parent-Offspring, Full-Sibling, second degree and third degree relationships among MGI participants. The numbers are the observed count for each configuration. Related to STAR Methods (Genetic Quality Control Procedures).**



**Figure S3: Comparison of TopMed and HRC based imputation in MGI. Frequency spectrum for TopMed and HRC imputation. The largest gain in coverage for TopMed is at the lower end of the frequency spectrum. Related to Figure 5(C).**



**Figure S4: Comparison of TopMed and HRC imputation accuracy by MAF. TopMed provides increased accuracy for all MAF>0.01 bins, with greatest improvement for SNPs < 5%. Related to Figure 5(C).**

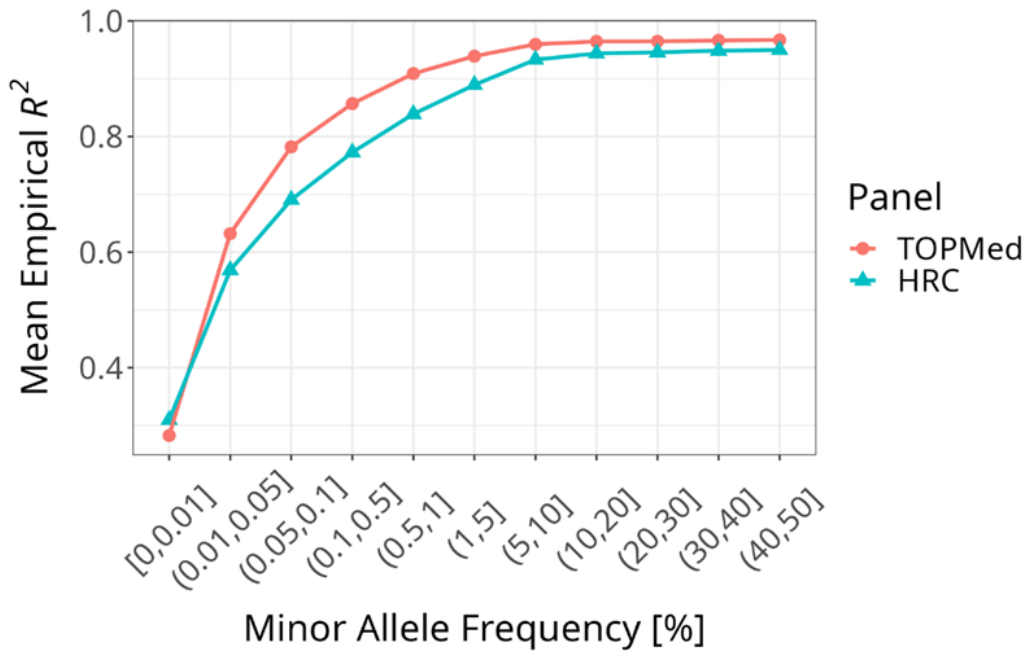
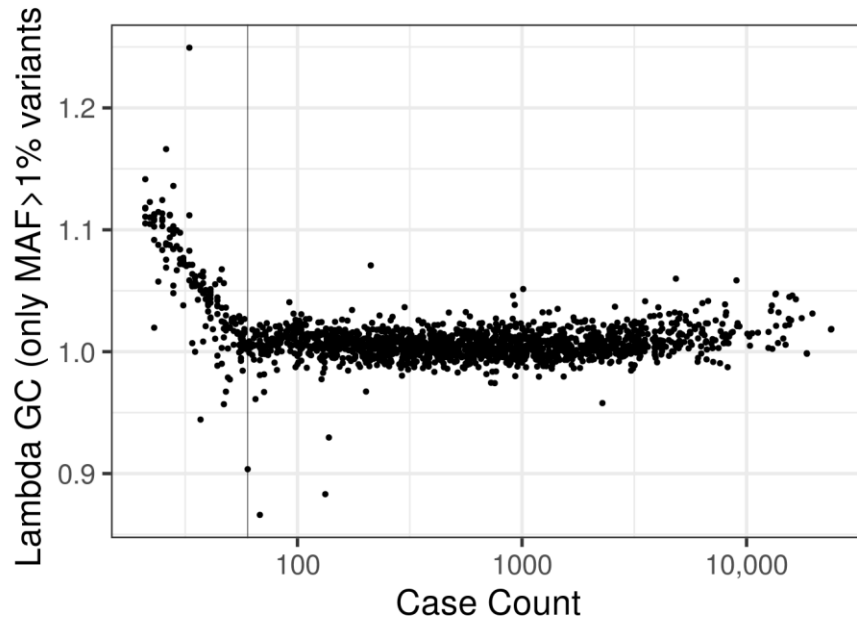
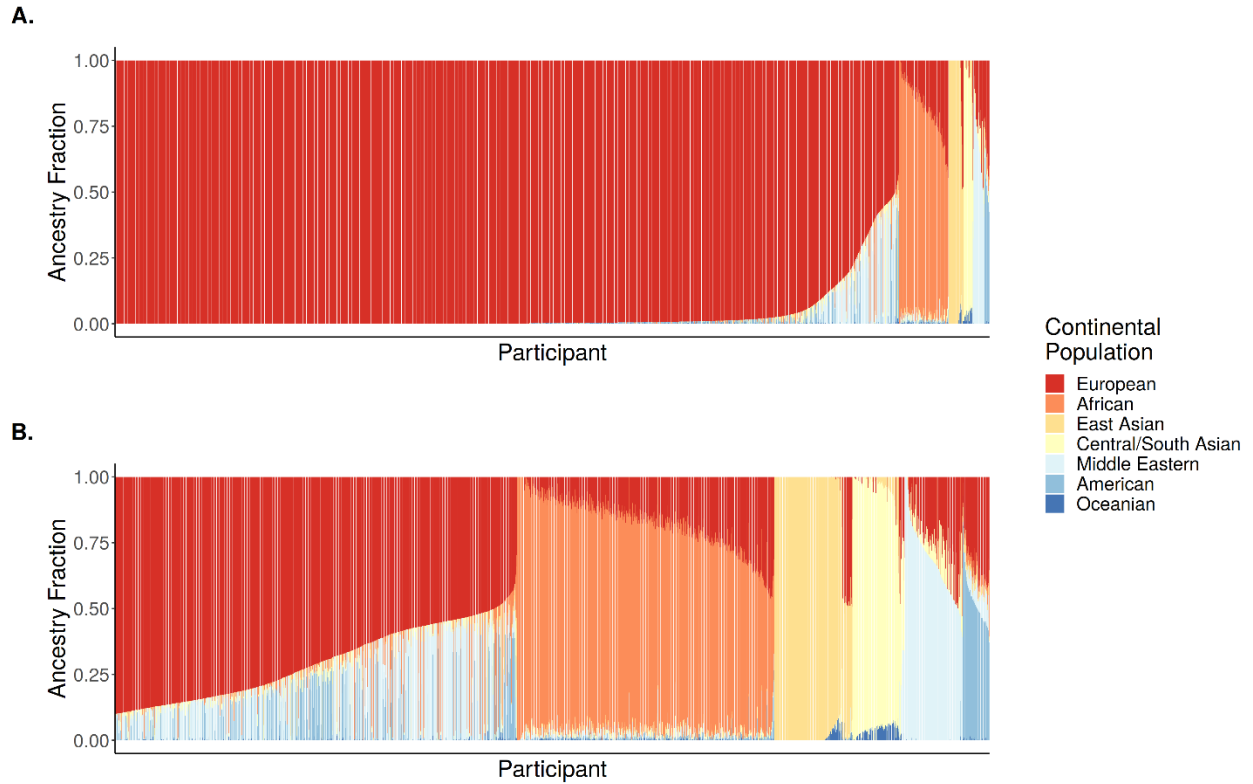


Figure S5: Genomic Control and case count (log10 scale) for 1712 PheCode traits with at least 20 cases. Traits with fewer than N=60 cases (vertical line) showed evidence of severe inflation. We therefore report GWAS results for PheCode traits with at least 60 cases. Related to STAR Methods (Genetic Analysis).



**Figure S6: Ancestry proportions of MGI participants estimated using ADMIXTURE. The top panel (A) shows all MGI participants and the bottom panel (B) shows only participants with <90% estimated EUR ancestry. We created continental ancestry bins for the imputation accuracy analysis by assigning samples whose majority ancestry assignment was at least 25% to the corresponding ancestry bin. Related to STAR Methods (Genetic Quality Control Procedures).**



**Table S1. Demographic and Clinical Outcome table including descriptive statistics for quantitative laboratory measures and select phecode traits. Related to Figure 2B.**

	All	Male	Female
<b>Number of Individuals</b>	57,055	26,732 (47%)	30,323 (53%)
<b>Age, yr (range 18-90+; mean <math>\pm</math> SD)</b>	56.44 $\pm$ 17	58.38 $\pm$ 17	54.74 $\pm$ 16
<b>BMI, kg/m<sup>2</sup></b>	29.87 $\pm$ 7.0	29.69 $\pm$ 6.1	30.03 $\pm$ 7.7
<b>Self-Reported Race, N</b>			
African American	3,223	1,264	1,959
American Indian or Alaska Native	237	112	125
Asian	1,324	601	723
Caucasian	49,605	23,534	26,071
Native Hawaiian or Pacific Islander	43	14	29
Other	1,023	463	560
Patient Refused	200	102	98
Unknown/Missing	1,400	643	757
<b>Self-Reported Ethnicity, N</b>			
Hispanic or Latino	805	337	468
Non-Hispanic or Latino	34,982	16,809	18,173
Patient Refused	155	70	85
Unknown/Missing	21,113	9,517	11,596
<b>Quantitative Measurements (mean <math>\pm</math> SD)</b>			
SBP, mm Hg (N=56,293)	70.52 $\pm$ 7.4	72.45 $\pm$ 7.4	68.83 $\pm$ 7.0
LDL-C, mg/dL (N=24,688)	100.38 $\pm$ 35.7	95.92 $\pm$ 35.2	104.65 $\pm$ 35.7
Albumin (N=43,558)	3.96 $\pm$ 0.6	3.93 $\pm$ 0.6	3.99 $\pm$ 0.6
Creatinine (N=50,255)	1.21 $\pm$ 1.2	1.37 $\pm$ 1.3	1.04 $\pm$ 1.0
Glucose (N=50,327)	123.37 $\pm$ 56.2	127.06 $\pm$ 56.6	119.40 $\pm$ 55.4
Mean corpuscular hemoglobin (N=50,246)	29.8 $\pm$ 2.6	29.98 $\pm$ 2.6	29.60 $\pm$ 2.6
Thyroid stimulating hormone (N=27,753)	3.19 $\pm$ 10.6	3.21 $\pm$ 9.1	3.17 $\pm$ 11.4
<b>Cardiometabolic Phenotypes, Phecode (N cases, % of sample)</b>			
Hypertension (401)	26,223 (46%)	13,858 (52%)	12,365 (41%)
Obesity (278.1)	17,547 (31%)	7,455 (28%)	10,092 (33%)
Cardiac Dysrhythmias (427)	15,896 (28%)	7,853 (29%)	8,043 (27%)
Type 2 Diabetes (250.2)	11,377 (20%)	6,036 (23%)	5,341 (18%)
<b>Neurodegenerative and Neurological Phenotypes (N cases, % of sample)</b>			
Sleep Apnea (327.3)	12,418 (22%)	6,884 (26%)	5,534 (18%)
Epilepsy (345.1)	606 (1%)	275 (1%)	331 (1%)
Parkinson's Disease (332)	399 (0.7%)	272 (1%)	127 (0.4%)
Multiple Sclerosis (335)	415 (0.7%)	109 (0.4%)	306 (1%)
<b>Respiratory and Immunological Phenotypes (N cases, % of sample)</b>			
Asthma (495)	10,135 (18%)	3,446 (13%)	6,689 (22%)
Pneumonia (480)	5,327 (9%)	2,467 (9%)	2,860 (9%)
Rheumatoid Arthritis (714.1)	1,750 (3%)	478 (2%)	1,272 (4%)



Ulcerative Colitis (555.2)	925 (2%)	428 (2%)	497 (2%)
<b>Oncology Phenotypes (N cases, % of sample)</b>			
Skin Cancer (172)	8,307 (15%)	4,580 (17%)	3,727 (12%)
Melanoma (172.11)	3,081 (5%)	1,772 (7%)	1,309 (4%)
Basal Cell Carcinoma (172.21)	3,522 (6%)	1,921 (7%)	1,601 (5%)
Breast Cancer (174)	3,418 (6%)	46 (0.2%)	3,372 (11%)
Prostate Cancer (185)	3,002 (5%)	3,002 (11%)	0 (0%)
Bladder Cancer (189.2)	1,746 (3%)	1,343 (5%)	403 (1%)
<b>Mental Health Phenotypes (N cases, % of sample)</b>			
Depression (296.2)	14,889 (26%)	4,992 (19%)	9,897 (33%)
Anxiety disorders (300)	14,922 (26%)	4,810 (18%)	10,112 (33%)
Bipolar (296.1)	1,134 (2%)	357 (1%)	777 (3%)
Schizophrenia (295.1)	115 (0.2%)	50 (0.1%)	65 (0.2%)

**Table S2. Phecode traits with >10-fold enrichment of case samples in MGI compared to UKB.  
Related STAR Methods (Clinical Phenotype Data, Phecodes in UK Biobank).**

PheCode	Description	Category	MGI Cases	MGI Controls	UKB Cases	UKB Controls	MGI:UKB Case Ratio
611.1	Abnormal mammogram	genitourinary	4,851	42,782	56	400,113	86.6
079.1	Varicella infection	infectious diseases	4,143	39,472	64	401,601	64.7
599.8	Other symptoms involving urinary system	genitourinary	3,888	35,037	62	383,297	62.7
300.4	Dysthymic disorder	mental disorders	2,050	27,399	53	363,984	38.7
792	Abnormal Papanicolaou smear of cervix and cervical HPV	genitourinary	2,788	17,968	97	193,707	28.7
327.4	Insomnia	neurological	4,252	35,287	163	401,998	26.1
278.4	Abnormal weight gain	endocrine/metabolic	1,625	32,094	66	396,288	24.6
415.11	Pulmonary embolism and infarction, acute	circulatory system	1,825	47,402	75	400,268	24.3
338.2	Chronic pain	neurological	9,127	34,356	379	406,436	24.1
110.11	Dermatophytosis of nail	infectious diseases	1,664	44,101	74	404,551	22.5
272.12	Hyperglyceridemia	endocrine/metabolic	1,210	31,466	58	371,432	20.9
272.13	Mixed hyperlipidemia	endocrine/metabolic	3,221	31,466	164	371,432	19.6
313	Pervasive developmental disorders	mental disorders	1,449	49,104	76	406,624	19.1
279.7	Other immunological findings	endocrine/metabolic	4,513	44,339	248	406,707	18.2
613.5	Mastodynia	genitourinary	1,396	47,335	77	405,229	18.1
338	Pain	neurological	13,542	34,356	766	406,436	17.7
690.1	Seborrheic dermatitis	dermatologic	924	43,414	53	400,950	17.4
340.1	Migraine with aura	neurological	3,640	43,699	211	397,071	17.3
464	Acute sinusitis	respiratory	3,174	39,262	196	404,740	16.2
949	Allergies, other	injuries & poisonings	6,371	35,760	396	403,085	16.1
938.2	Chronic dermatitis due to solar radiation	dermatologic	3,992	35,760	251	403,085	15.9
338.1	Acute pain	neurological	6,081	34,356	393	406,436	15.5
246	Other disorders of thyroid	endocrine/metabolic	6,055	39,246	397	389,743	15.3
250.42	Other abnormal glucose	endocrine/metabolic	6,284	36,652	418	387,082	15.0

261.4	Vitamin D deficiency	endocrine/metabolic	6,612	40,977	446	404,730	14.8
456	Chronic venous insufficiency [CVI]	circulatory system	2,633	36,953	183	367,908	14.4
356	Hereditary and idiopathic peripheral neuropathy	neurological	2,114	47,411	152	405,102	13.9
367.2	Astigmatism	sense organs	2,622	43,024	189	404,787	13.9
627.2	Symptomatic menopause	genitourinary	2,534	17,353	183	189,130	13.8
090	Sexually transmitted infections (not HIV or hepatitis)	infectious diseases	953	50,467	71	407,129	13.4
690	Erythematousquamous dermatosis	dermatologic	929	43,414	71	400,950	13.1
790	Nonspecific findings on examination of blood	symptoms	1,930	44,032	149	400,617	13.0
401.3	Other hypertensive complications	circulatory system	1,827	26,620	144	328,788	12.7
840	Sprains and strains	injuries & poisonings	5,944	42,380	469	406,668	12.7
605	Erectile dysfunction [ED]	genitourinary	3,436	16,004	276	167,468	12.4
704.8	Other specified diseases of hair and hair follicles	dermatologic	943	46,497	76	400,627	12.4
476	Allergic rhinitis	respiratory	12,834	31,134	1,052	388,553	12.2
136	Other infectious and parasitic diseases	infectious diseases	2,459	47,630	205	406,672	12.0
277.51	Lipoprotein disorders	endocrine/metabolic	620	48,876	52	405,784	11.9
429.1	Heart transplant/surgery	circulatory system	762	45,087	64	400,869	11.9
573.9	Abnormal serum enzyme levels	digestive	2,452	41,044	209	398,276	11.7
279	Disorders involving the immune mechanism	endocrine/metabolic	2,845	44,339	249	406,707	11.4
483	Acute bronchitis and bronchiolitis	respiratory	2,122	42,165	193	396,438	11.0
505	Other pulmonary inflammation or edema	respiratory	659	43,653	61	395,582	10.8
938	Dermatitis due to solar radiation	dermatologic	4,174	35,760	387	403,085	10.8
527.7	Disturbance of salivary secretion	digestive	667	47,262	62	401,593	10.8
260	Protein-calorie malnutrition	endocrine/metabolic	2,103	40,977	202	404,730	10.4
250.4	Abnormal glucose	endocrine/metabolic	7,061	36,652	681	387,082	10.4