SUPPLEMENTARY INFORMATION

Supplemental Figure 1. Distribution of standardized, adjusted PRS by release batch for six diseases in MGBB.

Supplemental Figure 2. Distribution of standardized, adjusted PRS by age for six diseases in MGBB.

Supplemental Figure 3. Distribution of adjusted PRS by sex for four diseases in MGBB.

Supplemental Figure 4. Correlation between standardized, adjusted PRS and odds of disease in reported white MGBB participants.

Supplemental Figure 5. Correlation between standardized, adjusted PRS and odds of disease in reported Black MGBB participants.

Supplemental Figure 6. Correlation between standardized, adjusted PRS and odds of disease in reported Asian MGBB participants.

Supplemental Figure 7. Correlation between standardized, adjusted PRS and odds of disease in MGBB participants of unknown or other reported race.

Supplemental Figure 8. Difference in standardized, adjusted PRS between WGS and imputed genotyping arrays for 22 individual samples.

Supplemental Table 1. Disease phenotypes and counts in MGB Biobank.

Supplemental Table 2: Counts of constituent SNPs from PRS base files and from 4 genotype arrays after imputation for 6 diseases.

Supplemental Table 3. PRS thresholds corresponding to published OR>2.

Supplemental Table 4. Prevalence and disease associations of high-risk PRS for 6 diseases in MGB Biobank, before adjustment for population structure.

Supplemental Table 5. Analytical performance of genotyping arrays and imputed data for three Genome in a Bottle samples.

Supplemental Table 6. Analytical performance of genotyping arrays and imputed data for 22 samples with WGS.

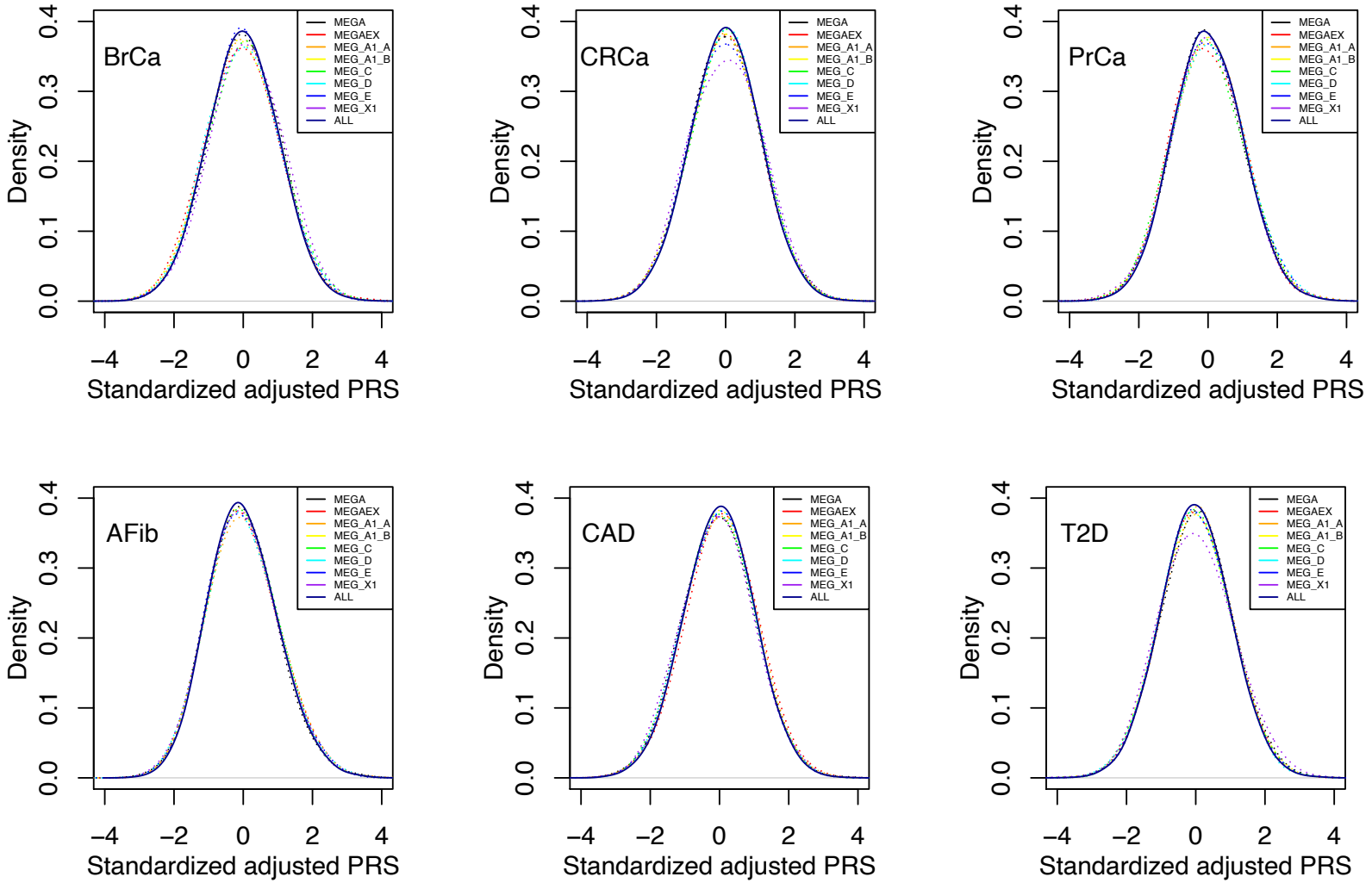
Supplemental Table 7. Standardized, adjusted PRS scores of three Genome in a Bottle samples compared between genome sequencing and genotyping arrays.

Supplemental Table 8. Comparison of standardized, adjusted PRS between WGS and genotyping arrays for 22 samples.

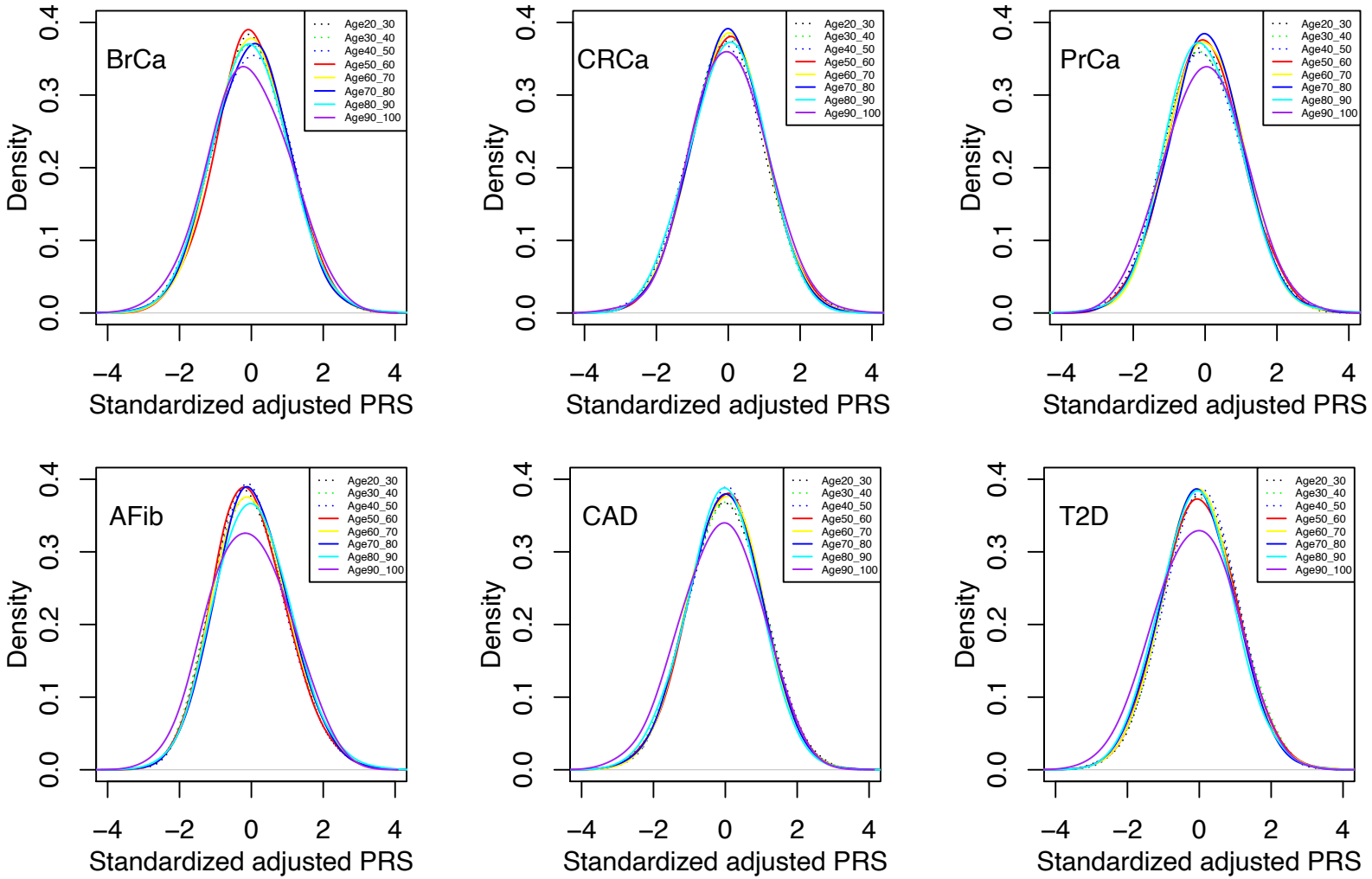
Supplemental Table 9. Comparison between standardized, adjusted PRS scores from MGBB data and prospective assay for diseases in nine samples with high risk.

Supplemental File 1. Supplemental methods

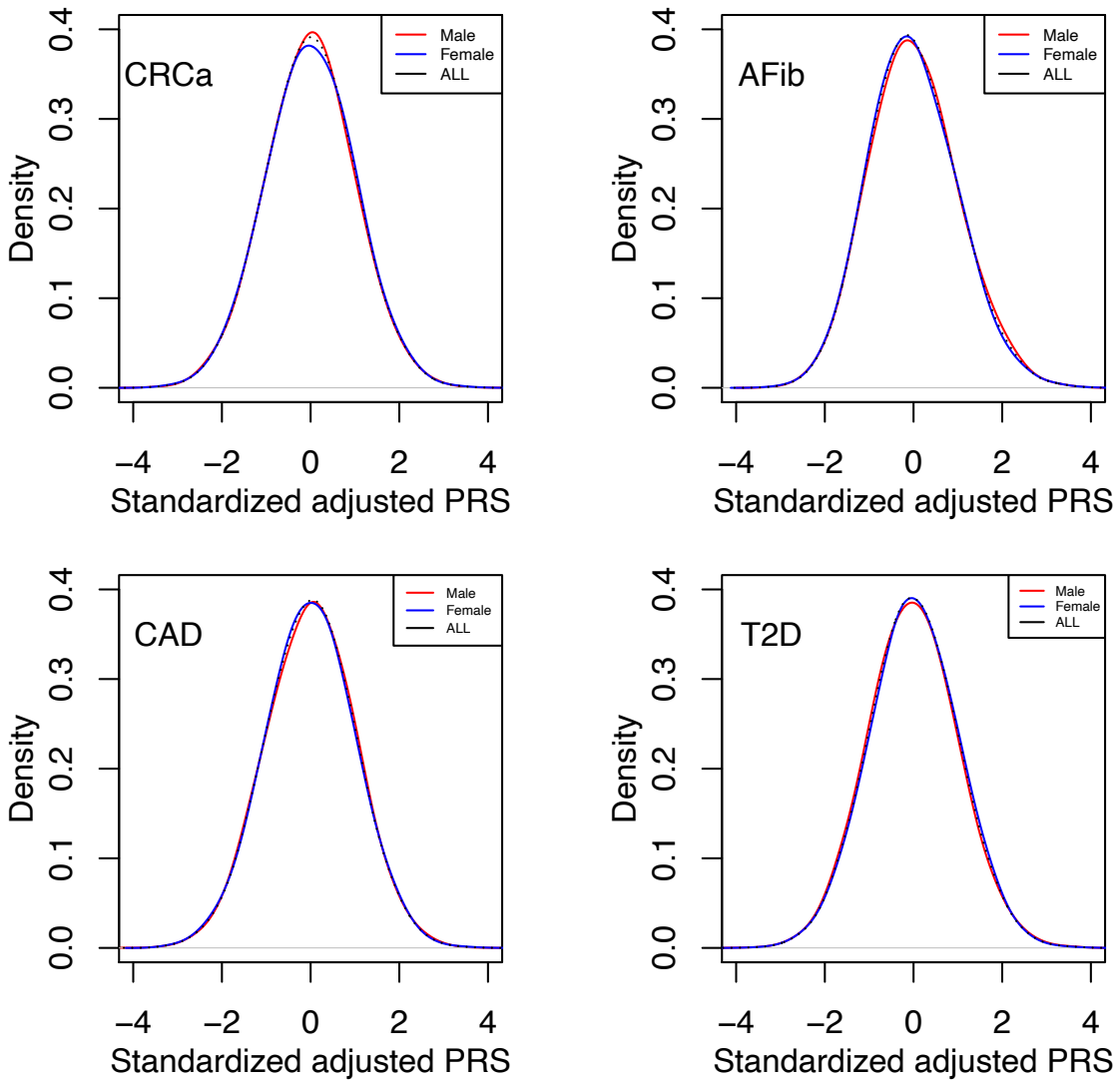
Supplemental File 2. Example of PRS report



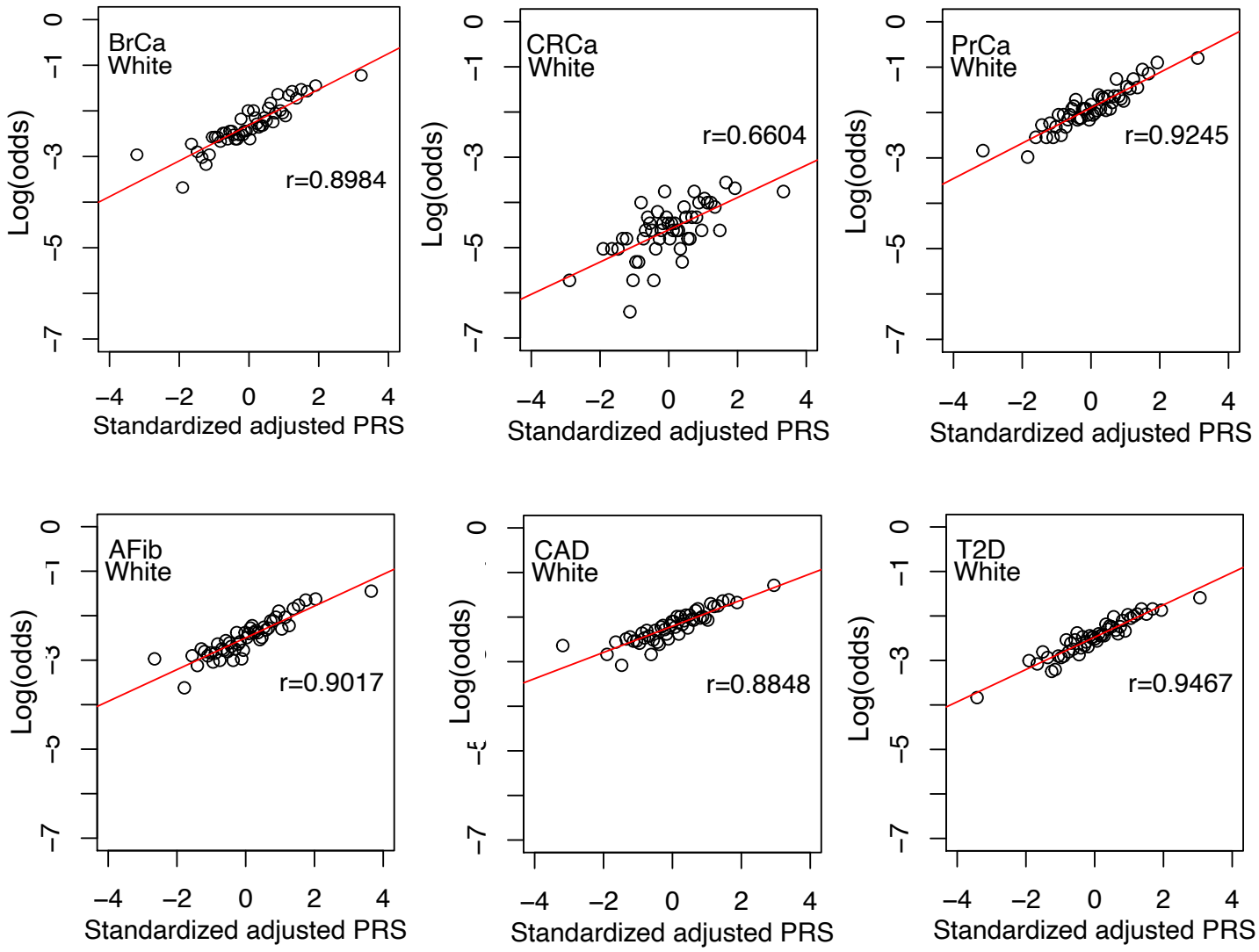
Supplemental Figure 1. Distribution of standardized, adjusted PRS by release batch for six diseases in MGBB. Standardized, adjusted PRS ($PRS_{std\text{-}adj}$) plotted by eight batches of three versions of Illumina genotyping arrays (MEG, MEGA, MEGAEX) used to analyze data from up to 36,423 MGBB participants. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MEG, Multi-Ethnic Global; MEGA, Multi-Ethnic Genotyping Array; MEGAEX, Expanded Multi-Ethnic Genotyping Array; MGBB, Mass General Brigham Biobank; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.



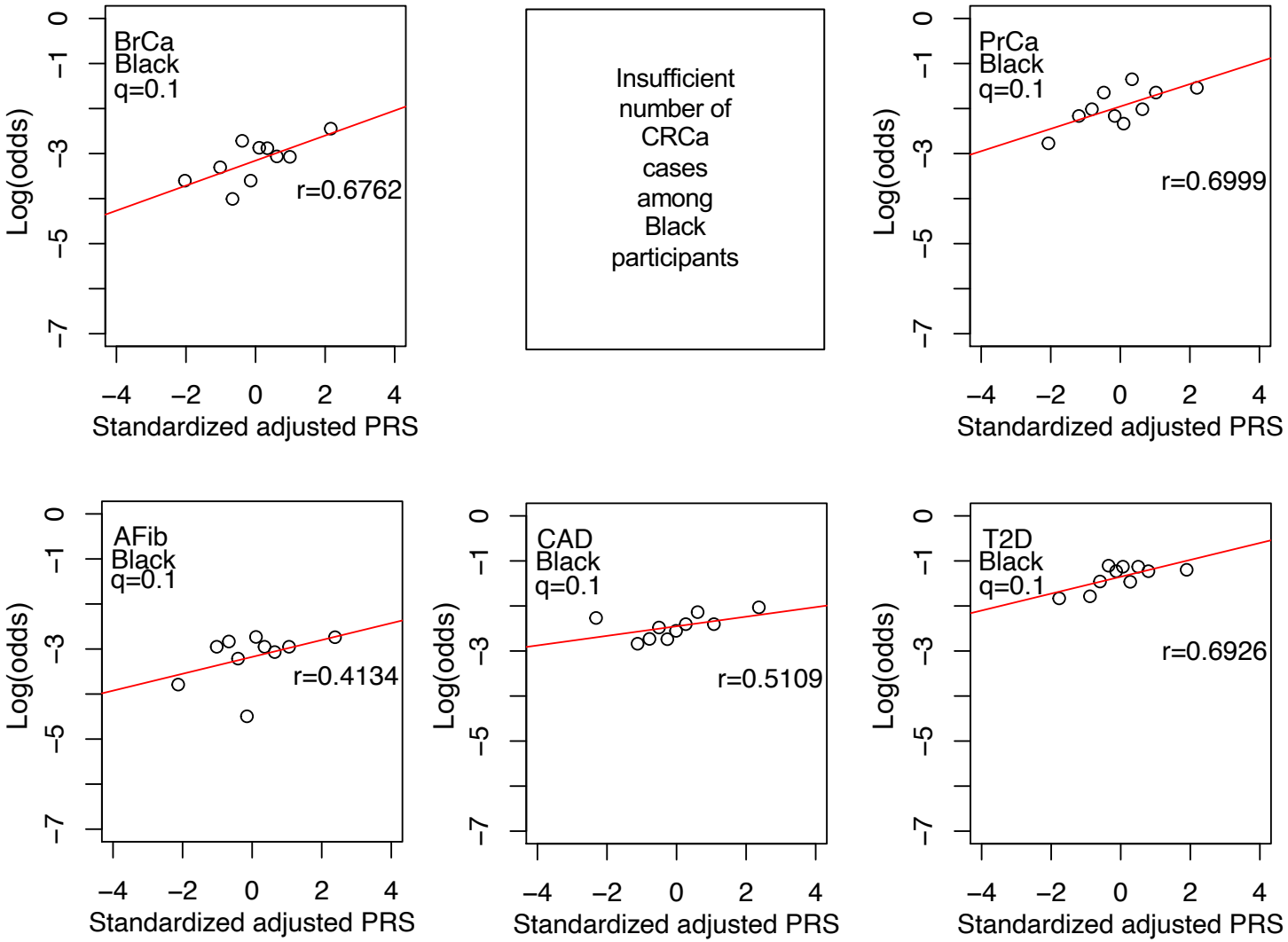
Supplemental Figure 2. Distribution of standardized, adjusted PRS by age for six diseases in MGBB. Standardized, adjusted PRS ($PRS_{std\text{-}adj}$) plotted by decade of age among up to 36,423 MGBB participants. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.



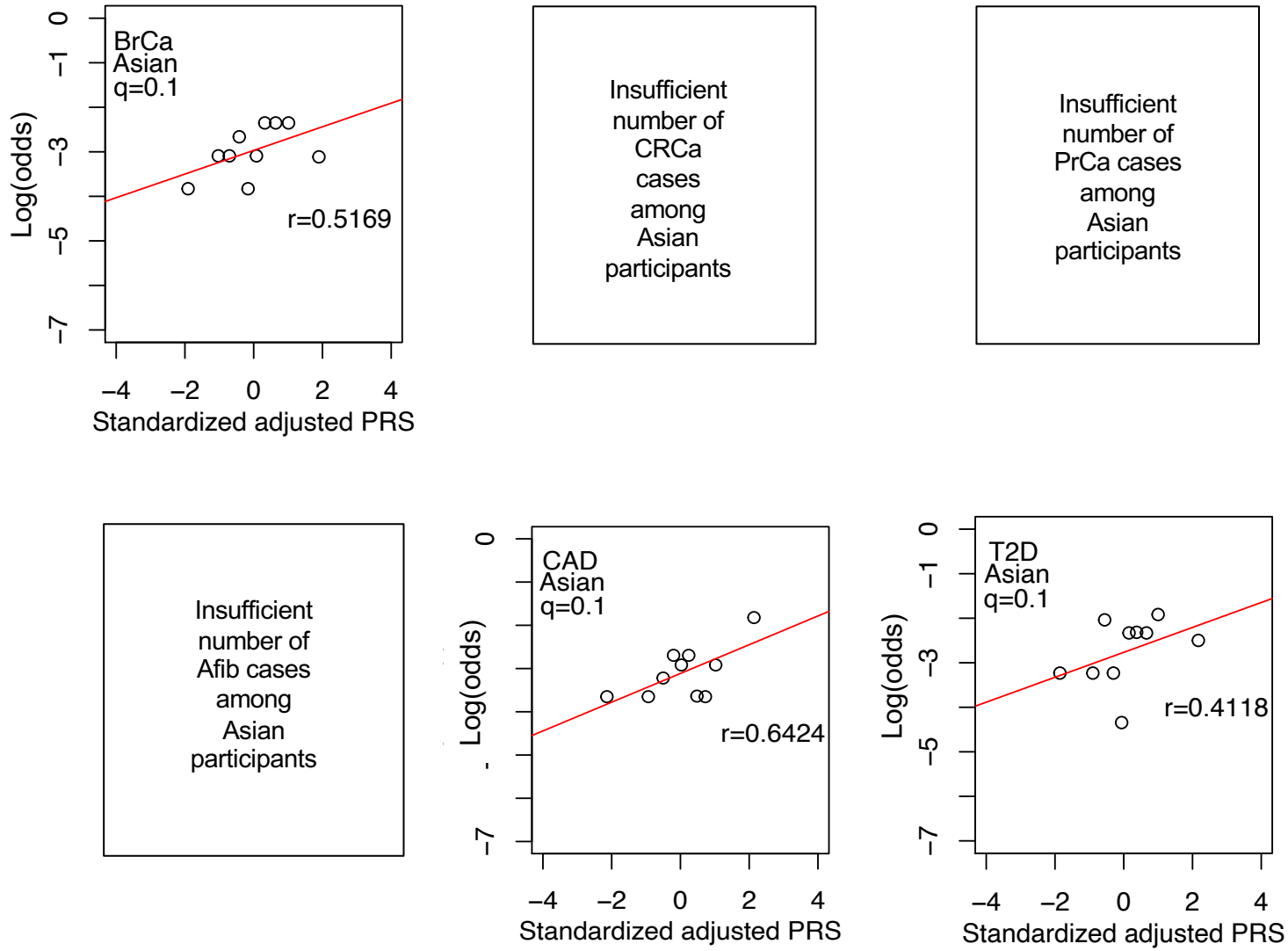
Supplemental Figure 3. Distribution of adjusted PRS by sex for four diseases in MGBB. Standardized, adjusted PRS ($PRS_{std\text{-}adj}$) plotted by sex among 16,704 male and 19,719 female MGBB participants. Abbreviations: AFib, atrial fibrillation; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; PRS, polygenic risk score; T2D, type 2 diabetes.



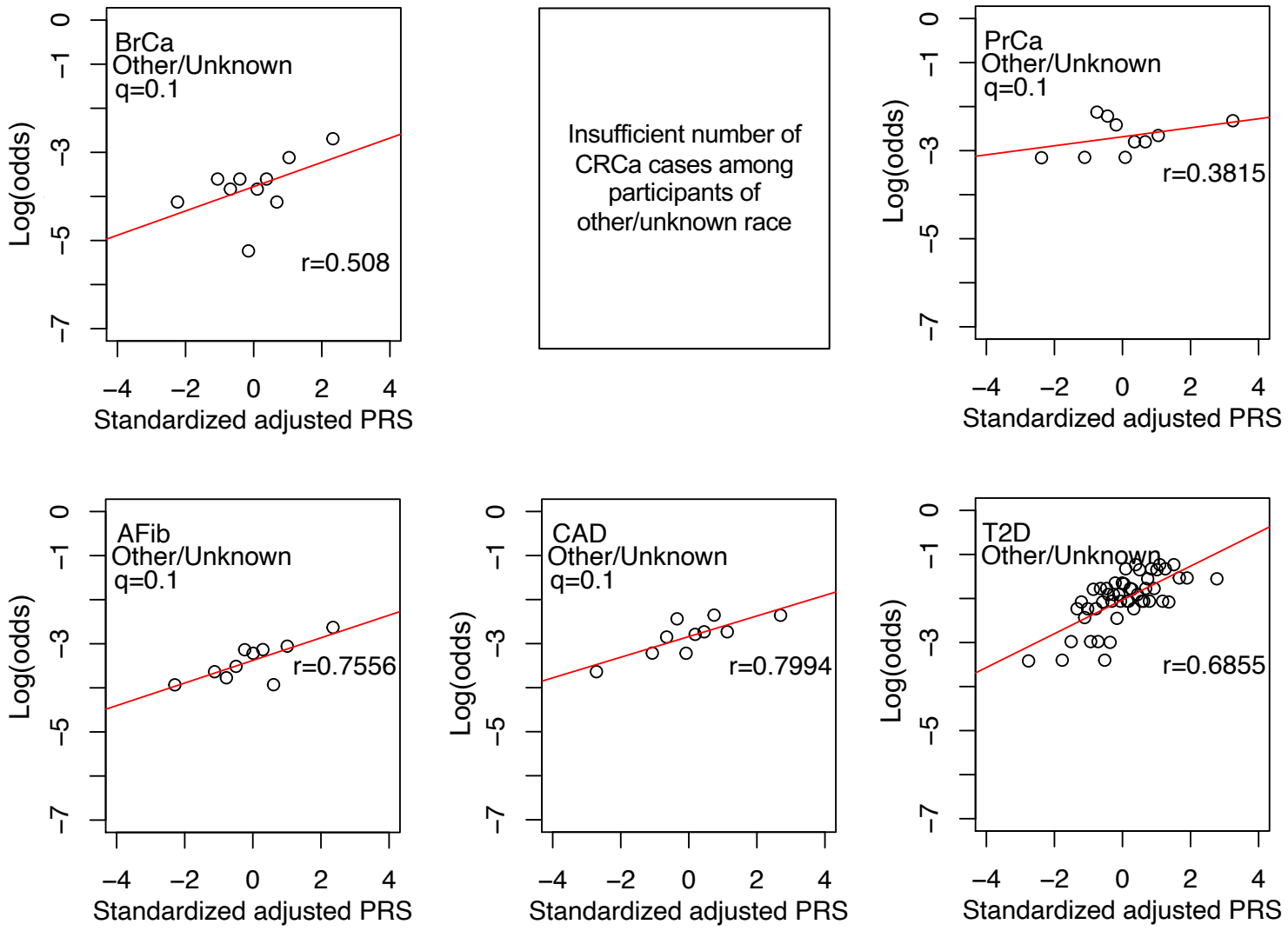
Supplemental Figure 4. Correlation between standardized, adjusted PRS and odds of disease in reported white MGBB participants. Plots show log(odds) of each of six diseases versus quantile (n=50) of standardized population structure-adjusted PRS ($PRS_{std\text{-}adj}$) among up to 30,716 MGBB participants of reported white race. The correlation coefficient, r, is shown in each panel. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; OR, odds ratio; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.



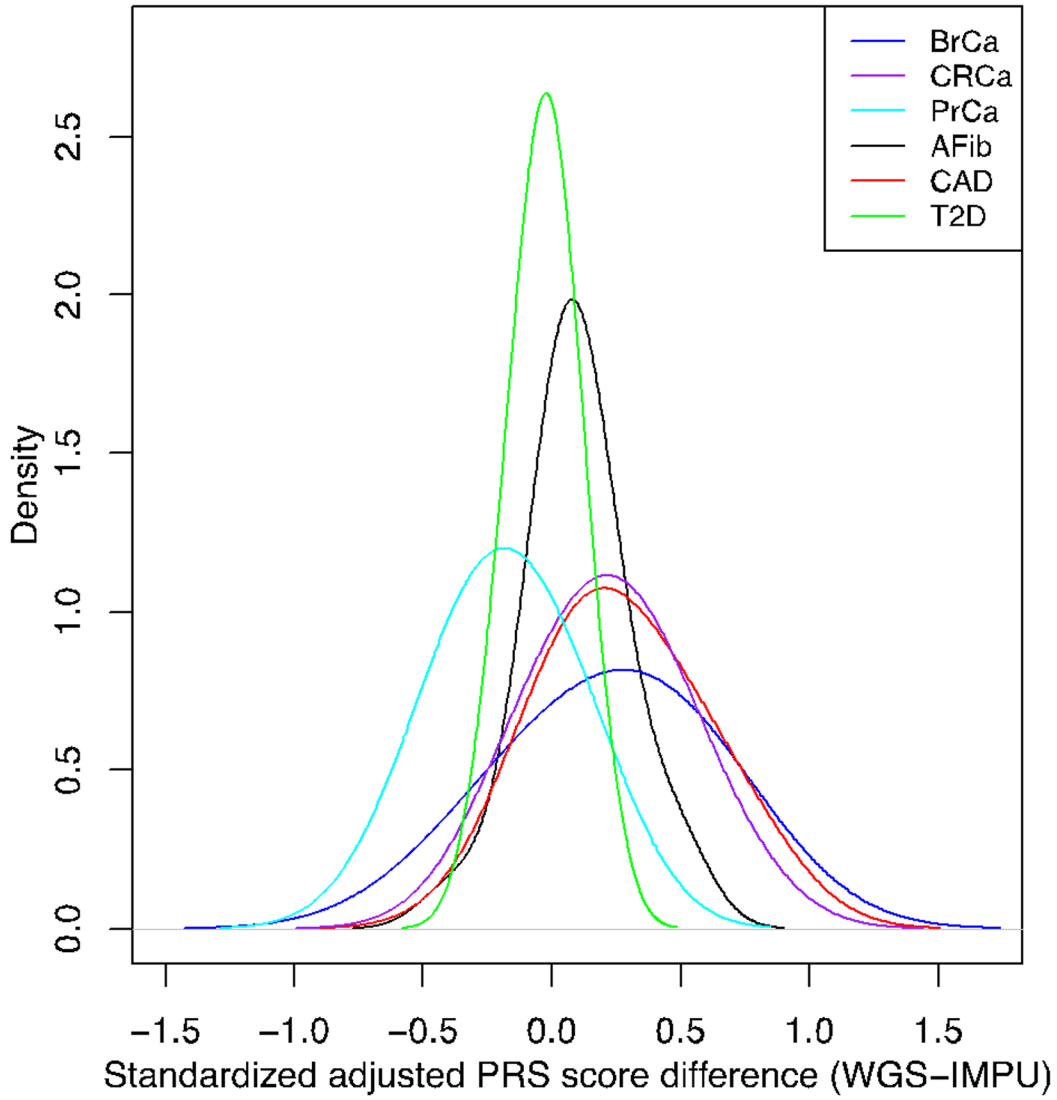
Supplemental Figure 5. Correlation between standardized, adjusted PRS and odds of disease in reported Black MGBB participants. Plots show log(odds) of each of six diseases versus quantile (n=10) of standardized population structure-adjusted PRS ($PRS_{std\text{-}adj}$) among up to 1,807 MGBB participants of reported Black race. Results not reported for CRCa due to 0 CRCa cases in at least one quantile. The correlation coefficient, r, is shown in each panel. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; OR, odds ratio; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.



Supplemental Figure 6. Correlation between standardized, adjusted PRS and odds of disease in reported Asian MGBB participants. Plots show log(odds) of each of six diseases versus quantile (n=10) of standardized population structure-adjusted PRS (PRSstd-adj) among up to 786 MGBB participants of reported Asian race. Results not reported for CRCa, PrCa, or AFib due to 0 cases in at least one quantile. The correlation coefficient, r, is shown in each panel. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; OR, odds ratio; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.



Supplemental Figure 7. Correlation between standardized, adjusted PRS and odds of disease in MGBB participants of unknown or other reported race. Plots show log(odds) of each of six diseases versus quantile (n=50 for T2D, n=10 for all other disease) of standardized population structure-adjusted PRS (PRSstd-adj) among up to 3,113 MGBB participants of unknown or other reported race. Results not reported for CRCa due to 0 cases in at least one quantile (n=10). The correlation coefficient, r, is shown in each panel. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; OR, odds ratio; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.



Supplemental Figure 8. Difference in standardized, adjusted PRS between WGS and imputed genotyping arrays for 22 individual samples. The $PRS_{std\text{-}adj}$ of 22 samples obtained from WGS and from imputed genotyping arrays are subtracted, and the distribution of the difference of the scores is plotted for each disease. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; IMPU, imputed genotype data; MGBB, Mass General Brigham Biobank; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes; WGS, whole genome sequencing.

Supplemental Table 1. Disease phenotypes and counts in MGB Biobank

Disease	PPV (%)	Controls	Cases	Total
BrCa	95	17949	1770	19719
CRCa	100	36031	392	36423
PrCa	100	13783	2919	16702
AFib	94	33691	2732	36423
CAD	95	32869	3554	36423
T2D	95	33060	3363	36423

Positive predictive values (PPV) for computed disease phenotypes in MGBB, as reported previously (see main text). Also shown are the counts of controls and cases, limited to women and men for BrCa and PrCa, respectively. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGB, Mass General Brigham Biobank; PrCa, prostate cancer; T2D, type 2 diabetes.

Supplemental Table 2. Counts of constituent SNPs from PRS base files and from 4 genotype arrays after imputation

Chip	BrCa	CRCA	PrCa	AFib	CAD	T2D
Base file	3820	81	147	6730541	6630150	6917436
MEGA	3818	80	140	6730205	6630033	6917170
MEGAEX	3818	80	140	6730106	6630095	6917118
MEG	3819	80	140	6730145	6630016	6917080
GDA_EA	3818	80	140	6730541	6630150	6917436
GDA	3818	80	140	6730541	6630150	6917436

Comparison of the PRS model for each disease and the sites available from imputation data among different arrays. Abbreviations: MEGA, Multi-Ethnic Genotyping Array; MEGAEX, Expanded Multi-Ethnic Genotyping Array; MEG, Multi-Ethnic Global array; GDA_EA, Global Diversity Array, Early Access; GDA, Global Diversity Array; BrCa, breast cancer; CRCa, colorectal cancer; PrCa, prostate cancer; AFib, atrial fibrillation; CAD, coronary artery; T2D, type 2 diabetes mellitus.

Supplemental Table 3. PRS thresholds corresponding to published OR>2

Disease	Change in OR per SD change in PRS	Z cutoff for OR>2	Source	Mean (SD) PRS_{raw} in MGBB	Mean (SD) PRS_{adj} in MGBB
BrCa	1.66	1.37	Table 2, Mavaddat 2019	-0.3994271 (0.4717565)	0.01712425 (0.4607471)
CRCa	1.54	1.61	Supplemental Table 16, Huyghe 2019	6.132968 (0.4337869)	0.001579539 (0.424732)
PrCa	1.86	1.12	Supplemental Table 18, Schumacher 2018	11.53247 (0.6906457)	0.03554762 (0.6770496)
AFib	1.63	1.42	Supplemental Table 2, Khera 2018	32.41001 (0.1241203)	0.003250882 (0.1056579)
CAD	1.72	1.28	Supplemental Table 1, Khera 2018	18.06105 (0.09994096)	0.002804128 (0.09180318)
T2D	1.65	1.38	Supplemental Table 3, Khera 2018	55.70348 (0.1893828)	0.003617606 (0.1123435)

Also included are the mean and standard deviation (SD) of raw (PRS_{raw}) and adjusted PRS (PRS_{adj}) as determined in MGBB and used to standardize the PRS. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.

Supplemental Table 4. Prevalence and disease associations of high-risk PRS for 6 diseases in MGB Biobank overall and by reported race, before adjustment for population structure

Disease	High risk (%)	OR Overall	OR White	OR Black	OR Asian	OR Other/Unknown
BrCa	**8.6**	**1.99** [1.72, 2.30] (252/1435, 1461/16571)	**2.42** [2.07, 2.83] (223/927, 1365/13724)	**1.39** [0.73, 2.67] (13/217, 37/860)	**0.84** [0.24, 2.90] (3/61, 22/377)	**2.46** [1.29, 4.70] (13/230, 37/1610)
CRCa	**5.3**	**2.45** [1.8, 3.34] (47/1896, 345/34134)	**2.30** [1.67, 3.17] (44/1768, 309/28595)	**4.15** [0.53, 32.41] (1/25, 17/1764)	**0** [0, NaN] (0/13, 7/766)	**5.57** [1.23, 25.26] (2/90, 12/3009)
PrCa	**13.2**	**2.13** [1.91, 2.39] (486/1711, 1705/12800)	**2.30** [2.04, 2.59] (431/1322, 1581/11143)	**1.71** [1.09, 2.69] (42/206, 46/386)	**0** [0, NaN] (0/19, 8/296)	**1.10** [0.6, 2.04] (13/164, 70/975)
AFib	**8.4**	**1.53** [1.36, 1.73] (326/2738, 2406/30952)	**2.59** [2.25, 2.99] (251/1153, 2272/27040)	**1.26** [0.80, 1.99] (45/871, 35/856)	**1.49** [0.60, 3.69] (12/384, 8/382)	**1.60** [0.95, 2.69] (18/330, 91/2674)
CAD	**9.2**	**1.88** [1.7, 2.07] (532/2820, 3021/30049)	**1.94** [1.74, 2.15] (494/2387, 2689/25146)	**2.47** [0.70, 8.69] (3/14, 143/1647)	**2.12** [1.11, 4.05] (17/193, 23/553)	**1.30** [0.78, 2.15] (18/226, 166/2703)
T2D	**9.6**	**2.53** [2.31, 2.78] (649/2850, 2714/30209)	**2.65** [2.11, 3.31] (97/419, 2429/27771)	**2.58** [1.6, 4.15] (356/1250, 20/181)	**1.01** [0.55, 1.87] (41/533, 15/197)	**1.97** [1.58, 2.45] (155/648, 250/2060)

High-risk PRS, defined here as an unadjusted PRS (PRS_{raw}) associated with OR>2 for disease in the original publication. OR shown are the observed OR [95% CI] among up to 36,423 MGBB participants in the overall cohort and by reported race. Data below each OR take the format (n cases with high-risk PRS / n controls with high-risk PRS, n cases without high-risk PRS / n controls without high-risk PRS). Second column shows the proportion of MGBB participants exceeding the literature-derived OR>2 threshold for each disease. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; MGBB, Mass General Brigham Biobank; NaN, not a number; OR, odds ratio; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes.

Supplemental Table 5. Analytical performance of genotyping arrays and imputed data for three Genome in a Bottle samples

		Genotyped data				Imputed data			
		SNV		Indel		SNV		Indel	
Sample	ROI	Sens (%)	PPV (%)	Sens (%)	PPV (%)	Sens (%)	PPV (%)	Sens (%)	PPV (%)
NA12878	NIST_HC	99.42	99.39	99.58	94.54	98.30	98.71	91.35	93.04
	ACMG59	95.83	63.45	#	#	N/A	N/A	N/A	N/A
NA24385	NIST_HC	99.45	99.42	98.21	97.40	96.50	97.25	89.75	92.28
	ACMG59	96.84	63.01	#	#	N/A	N/A	N/A	N/A
NA24631	NIST_HC	99.25	99.27	98.06	93.79	96.36	97.21	89.52	92.14
	ACMG59	95.83	51.98	#	#	N/A	N/A	N/A	N/A

The sensitivity (Sens) and positive predictive value (PPV) of three Genome in a Bottle (GIAB) Samples for both genotyping data and variants imputed from the genotyping array using 1000 Genomes Project phase 3 data. Metrics are shown for single nucleotide variants (SNVs) and insertion/deletion events (indels) separately. Values are calculated within two regions of interests (ROI): 1) NIST: high confidence regions as determined by the GIAB consortium, and 2) ACMG SF v2.0: the NIST region subsetted to the 59 genes screened for secondary findings. Secondary findings were analyzed only in genotyping data, hence not applicable (N/A) for imputed data. Abbreviations: ACMG, American College of Medical Genetics and Genomics; NIST, National Institute of Standards and Technology; SF, secondary findings.

Supplemental Table 6. Analytical performance of genotyping arrays and imputed data for 22 samples with WGS

	Genotyped data				Imputed data			
	SNV		Indel		SNV		Indel	
	Sens (%)	PPV (%)	Sens (%)	PPV (%)	Sens (%)	PPV (%)	Sens (%)	PPV (%)
Mean	99.28	98.60	96.99	92.50	95.87	93.06	88.59	87.42
SD	0.15	0.47	1.58	3.02	1.11	2.98	1.55	3.65

The mean and standard deviation (SD) for sensitivity (Sens) and positive predictive value (PPV) of 22 samples for both genotyping data and variants imputed from the genotyping array using 1000 Genomes Project phase 3 data. Truth defined as variant calls from whole genome sequencing (WGS). Metrics are shown for single nucleotide variants (SNVs) and insertion/deletion events (indels) separately.

Supplemental Table 7. Standardized, adjusted PRS scores of three Genome in a Bottle samples compared between genome sequencing and genotyping arrays

Sample name	Sample number	Method	BrCa	CRCa	PrCa	AFib	CAD	T2D
NA12878	6	WGS	0.0046	0.1936	N/A	-1.5652	-0.7087	-0.1282
	2	MEG	-0.2425	0.0527	N/A	-1.5501	-0.8127	-0.0756
	9	GDA_EA	-0.3183	0.0567	N/A	-1.5634	-0.5685	-0.0002
	2	GDA	-0.4407	0.0529	N/A	-1.5738	-0.7613	0.0339
NA24631	3	WGS	N/A	-0.0335	0.5378	-0.1916	1.2640	-1.4254
	2	MEG	N/A	0.0066	0.7879	-0.5102	1.2694	-1.3991
	6	GDA_EA	N/A	0.0499	0.5433	-0.5026	0.6418	-0.4139
	1	GDA	N/A	0.0503	0.5451	-1.4717	1.2727	-1.2834
NA24385	3	WGS	N/A	1.1379	-0.0092	-1.3015	0.2716	0.4582
	2	MEG	N/A	1.1379	-0.2275	-1.1228	0.2453	0.6089
	6	GDA_EA	N/A	0.9535	-0.2027	-1.5265	-0.0329	0.6133
	1	GDA	N/A	1.0942	0.0839	-1.4717	0.2333	0.7062

Average standardized, adjusted PRS ($PRS_{std\text{-}adj}$) for 3 Genome in a Bottle (GIAB) samples calculated from whole genome sequencing (WGS) and imputed data from 3 different genotyping arrays. Abbreviations: MEG, Multi-Ethnic Global array; GDA_EA, Global Diversity Array - Early Access; GDA, Global Diversity Array; BrCa, breast cancer; CRCa, colorectal cancer; PrCa, prostate cancer; AFib, atrial fibrillation; CAD, coronary artery; T2D, type 2 diabetes mellitus.

Supplemental Table 8. Comparison of standardized, adjusted PRS between WGS and genotyping arrays for 22 samples

		BrCa		CRCa		PrCa		AF		CAD		T2D	
Sample	Race	WGS	IMPU	WGS	IMPU	WGS	IMPU	WGS	IMPU	WGS	IMPU	WGS	IMPU
1	U	-1.8867	-1.9602	0.5725	0.0541	NA	NA	-0.4783	-0.7542	-0.4876	-0.7174	-0.6895	-0.5798
2	U	-0.7455	-0.8262	-0.7147	-0.9928	NA	NA	0.1293	0.0401	**2.0022**	**1.4268**	-0.8434	-0.7031
3	W	-0.5974	-0.2353	-0.5253	-0.8340	NA	NA	0.8934	0.8328	0.4514	0.1264	-0.7919	-0.5170
4	W	1.2894	0.6875	-0.2244	-0.2234	NA	NA	-1.1754	-1.2732	-0.0191	-0.3584	1.2915	1.2000
5	U	NA	NA	-0.2838	-0.4996	-1.7625	-1.5304	-0.5714	-1.0795	0.5459	0.4484	-0.0951	0.1091
6	W	NA	NA	-1.5450	-1.9562	-0.9071	-0.5058	**1.9406**	**1.7445**	1.1350	0.4059	0.7059	0.7560
7	W	NA	NA	1.2426	1.4942	-1.0596	-1.1552	1.1059	0.8242	-1.8147	-1.8903	-2.1168	-2.1413
8	W	-0.0807	0.0807	0.4206	0.5619	NA	NA	**2.5232**	**2.3368**	-0.3928	-0.5078	-1.1037	-0.9188
9	B	NA	NA	0.9960	0.7864	-1.0988	-1.1681	-0.0085	-0.1192	-0.6908	-0.8890	-0.2007	-0.2609
10	W	NA	NA	0.9896	0.8868	-0.4751	-0.5790	0.0136	-0.0659	-0.8545	-1.3623	-0.3380	-0.5170
11	B	**2.0035**	**1.6508**	**1.7024**	**1.4845**	NA	NA	1.1716	1.1790	-1.7253	-1.8242	0.5494	0.5569
12	W	NA	NA	0.9835	0.5430	-1.4750	-1.3309	-0.2462	-0.2993	1.0579	0.6188	-0.4822	-0.4927
13	B	0.8077	0.4828	0.5556	0.3867	NA	NA	0.4374	0.8153	1.1033	0.5647	-0.1753	-0.2103
14	B	-1.3930	-1.6669	0.9154	0.8493	NA	NA	-0.8602	-0.7490	-0.3770	-1.0892	-0.4187	-0.3697
15	B	NA	NA	0.5894	0.2448	-0.3288	-0.0110	0.3465	0.2209	-1.1982	-1.2453	1.3053	1.3296
16	B	NA	NA	0.6200	0.6782	-1.3133	-1.0626	0.5053	0.4669	-1.9614	-1.8994	1.0686	1.1810
17	B	-1.9561	-1.8142	0.3151	-0.3847	NA	NA	0.9165	1.0213	1.0527	0.8562	0.4446	0.5467
18	B	0.3158	-0.1079	-1.0174	-1.4023	NA	NA	-1.5555	-1.7732	-1.7501	-1.8100	0.6112	0.6805
19	B	-0.3796	-0.9207	**2.1503**	**1.8606**	NA	NA	-0.2506	-0.2548	-1.0279	-0.8536	0.8178	0.7486
20	W	NA	NA	-0.5118	-0.4171	-0.3109	0.2335	0.9822	0.5202	0.2268	-0.0132	0.7244	0.7369
21	B	0.2354	-0.4416	1.0415	0.3766	NA	NA	-0.1370	-0.1748	0.6289	0.5646	0.0659	0.0414
22	B	NA	NA	-0.7704	-0.7109	-1.2318	-1.0596	-1.8894	-1.8348	**1.4781**	**0.7385**	-0.1251	-0.2995

Standardized, adjusted PRS ($PRS_{std\text{-}adj}$) are shown for each of 6 diseases and 22 samples with both whole genome sequencing (WGS) and imputed genotyping arrays (IMPU). Green indicates a concordant categorization above the high-risk threshold from both WGS and IMPU for that sample and disease. Yellow indicates a discordant categorization around the high-risk threshold for the sample and disease. All other cells indicate PRS with concordant categorization below the high-risk threshold for that sample and disease. NA indicates cells where the PRS was not calculated for that disease-sex combination. Abbreviations: AFib, atrial fibrillation; B, reported Black race; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; PrCa, prostate cancer; T2D, type 2 diabetes mellitus; U, unknown race; W, reported white race.

Supplemental Table 9. Comparison between standardized, adjusted PRS scores from MGBB data and prospective assay for diseases in nine samples with high risk

Sample	Disease	MEG	GDA
1	CRCa	1.9337	1.8773
2	CAD	2.1961	2.1347
3	PrCa	1.0736	1.1675
4	CRCa	2.7145	2.7161
5	AF	1.6851	1.6435
6	T2D	2.4144	2.2023
7	CAD	4.0646	4.0855
8	PrCa	4.1363	4.1380
9	BrCa	4.0142	4.4274
	CAD	1.7722	1.8204

Shown are 9 MGBB samples determined to be at high risk for a total of 10 diseases using MEG array data. Standardized, adjusted PRS ($PRS_{std\text{-}adj}$) are shown for both the MGBB data and for the same samples run using the GDA array in the prospective assay. Abbreviations: AFib, atrial fibrillation; BrCa, breast cancer; CAD, coronary artery disease; CRCa, colorectal cancer; GDA, Global Diversity Array; MEG, Multi-Ethnic Global array; PrCa, prostate cancer; PRS, polygenic risk score; T2D, type 2 diabetes mellitus.

SUPPLEMENTAL METHODS

PRS calculation for clinical assay for individual samples

Unadjusted PRS (PRS_{raw}) were calculated for each individual sample as described for the Mass General Brigham Biobank (MGBB) cohort in the main text. To determine the adjusted PRS (PRS_{adj}) for each individual, the eigen variable, eigenvalue, and frequency output from the MGBB principal components (PC) analysis were used to project each new individual sample onto the MGBB PCs, using the following command in PLINK v2.0a:

```
plink2 --pfile individual_data --read-freq ref_pcs.acount --score
ref_pcs.eigenvec.allele 2 5 header-read no-mean-imputation variance-
standardize --score-col-nums 6-15 --out new_projection
```

The resulting projected PCs were then scaled to match the MGBB PCs by taking the square root of the eigenvalue and then multiplying by 2. The scaled PCs were fitted into the linear model for each disease developed in the MGBB data to obtain PRS_{pred}:

BrCa: $PRS_{pred} = 17.609341*PC1 - 4.146935*PC2 + 5.335144*PC3 + 3.833931*PC4 - 0.421679$

CRCa: $PRS_{pred} = -13.659121*PC1 + 6.411109*PC2 - 2.483703*PC3 - 6.869127*PC4 + 6.131384$

PrCa: $PRS_{pred} = 23.441147*PC1 + 13.724771 *PC2 - 9.528270 *PC3 + 4.118756 *PC4 + 11.506243$

AFib: $PRS_{pred} = 9.6269881*PC1 - 3.2878238*PC2 - 6.6519006*PC3 - 3.0149108*PC4 + 32.4067610$

CAD: $PRS_{pred} = -6.1974327*PC1 - 3.6757094*PC2 - 1.3488677*PC3 - 1.3490566*PC4 + 18.0582457$

T2D: $PRS_{pred} = 26.4700782*PC1 - 7.4283370*PC2 + 9.3782116*PC3 + 1.6994457*PC4 + 55.6998719$

PRS_{adj} was then calculated as the difference between PRS_{raw} and PRS_{pred}, as described in the main manuscript. Standardized, adjusted PRS values ($PRS_{std\text{-}adj}$) were calculated using the mean and standard deviation of PRS_{adj} in MGBB and compared against the OR>2 threshold as determined from the original publications (see Methods).

Code used to adjust the PRS for population structure are available for download here: [*URL will be provided prior to publication*].

PARTNERS HEALTHCARE | PERSONALIZED MEDICINE
MASSACHUSETTS GENERAL HOSPITAL | BRIGHAM AND WOMEN'S HOSPITAL

LABORATORY FOR MOLECULAR MEDICINE
65 Landsdowne St, Cambridge, MA 02139
Phone: (617) 768-8500 / Fax: (617) 768-8513
http://www.partners.org/personalizedmedicine/lmm

Name: LAST, FIRST	**MRN:** XXXXX	**LMM Accession ID:** PM-19-X00000
DOB: MM/DD/XXXX	**Referring Facility:** XXXX	**Specimen:** Blood, peripheral
Sex: Female/Male	**Referring Physician:** XXXX	**Received:** MM/DD/XXXX
Family #: F00000		**Page:** 1 of X
Test Performed: GenoVA Polygenic Risk Assessment	**Test Codes:**	

RESULTS SUMMARY*

HIGH POLYGENIC DISEASE RISK: Genotyping indicated an increased polygenic risk for developing colorectal cancer. Result details are provided below.

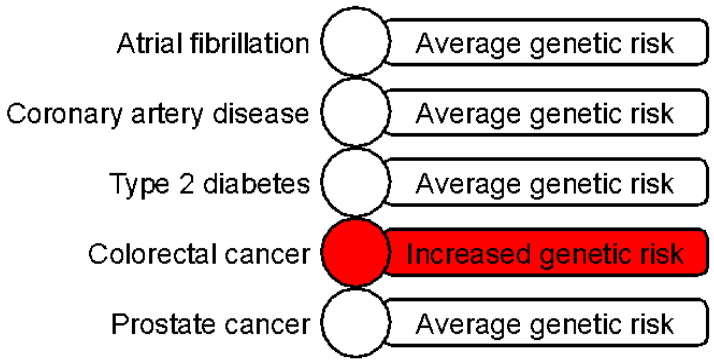
***Polygenic risk calculated using data from predominantly European ancestry individuals. Results are known to be less accurate for non-European ancestry individuals. See details below.**

DETAILED GENOMIC RESULTS

A. POLYGENIC DISEASE RISK

Polygenic risk describes the chance of developing certain health conditions based on a large number of genetic variants across the genome. This test assessed the risk for developing the following conditions: atrial fibrillation, coronary artery disease, type 2 diabetes, colorectal cancer, and prostate cancer (for patients with a prostate) or breast cancer (for patients born with female breast tissue).

This test identified an **increased polygenic risk for colorectal cancer** (see methodology for complete description of the analysis). It **did NOT** indicate increased polygenic risk for the remaining conditions.



Diseases WITH an increased polygenic risk

Disease	This patient's result	General disease prevalence
Colorectal cancer	Increased polygenic risk	1 in 24 people
RISK INTERPRETATION: The patient's calculated polygenic risk score, derived from 80 loci, has been associated with an INCREASED risk for colorectal cancer, defined here as greater than 2-fold risk. Individuals with similar polygenic risk scores have been shown to have an increased risk for colorectal cancer above baseline. Values of this polygenic risk score that fall among the top 5% were associated with a greater than 2-fold greater risk of developing colorectal cancer among >125,000 individuals of European ancestry when compared to the average individual (Huyghe 2019); similar association was observed among 21,630 individuals of East Asian ancestry (Schmit 2018).		
DISEASE INFORMATION: Colorectal cancer is uncontrolled growth in the colon or rectum. It begins as polyps that may turn into cancer over time. Colorectal cancer is the 3rd most common cancer in the United States, and is the 2nd leading cause of cancer death. Colorectal cancer is most often found in people 50 years or older. Symptoms include blood in the stool, stomach pain, aches or cramps, and unintentional weight loss, but many patients have no symptoms (adapted from Centers for Disease Control and Prevention https://www.cdc.gov/cancer/colorectal/, SEER https://seer.cancer.gov/statfacts/html/colorect.html and Colorectal Cancer Alliance https://www.ccalliance.org/colorectal-cancer-information/what-is-colorectal-cancer).		

LABORATORY FOR MOLECULAR MEDICINE
65 Landsdowne St, Cambridge, MA 02139
Phone: (617) 768-8500 / Fax: (617) 768-8513
http://www.partners.org/personalizedmedicine/lmm

Diseases WITHOUT an increased polygenic risk

Disease	This patient's result	General disease prevalence
Atrial fibrillation	Average polygenic risk	Lifetime risk of 1 in 3 to 1 in 5
RISK INTERPRETATION: The patient's calculated polygenic risk score, derived from 6730143 loci, has NOT been associated with high risk for atrial fibrillation, defined here as greater than 2-fold risk. For the majority of individuals with polygenic risk scores similar to this, risk for atrial fibrillation is not likely to be increased above baseline. On the other hand, values of this polygenic risk score that fall among the top 7% were associated with a greater than 2-fold greater risk of developing atrial fibrillation among >400,000 British volunteers of European ancestry when compared to the average individual (Khera 2018).		
DISEASE INFORMATION: Atrial fibrillation (AFib) is an inherited abnormality of the heart's normal rhythm due to episodes of uncoordinated electrical activity in the heart's upper chambers. Symptoms include dizziness, chest pain, sensations of fluttering or chest palpitations, shortness of breath and fainting. AFib and complications associated with the condition (such as stroke and heart failure) may occur at any age, but risk of developing symptoms increases with age (adapted from Genetics Home Reference https://ghr.nlm.nih.gov/condition/familial-atrial-fibrillation and Centers for Disease Control and Prevention https://www.cdc.gov/heartdisease/atrial_fibrillation.htm).		

Disease	This patient's result	General disease prevalence
Coronary artery disease	Average polygenic risk	1 in 5 men aged 60-79 1 in 8 women aged 60-79
RISK INTERPRETATION: The patient's calculated polygenic risk score, derived from 6630016 loci, has NOT been associated with high risk for coronary artery disease, defined here as greater than 2-fold risk. For the majority of individuals with polygenic risk scores similar to this, risk for coronary artery disease is not likely to be increased above baseline. On the other hand, values of this polygenic risk score that fall among the top 10% were associated with a greater than 2-fold greater risk of developing coronary artery disease among >400,000 British volunteers of European ancestry when compared to the average individual (Khera 2018). Having an ancestry-adjusted score in the top 5th percentile has also been associated with an odds ratio of early myocardial infarction (before age 55) of 5.09, 2.02, 3.38, and 3.33 in people of white, black, Hispanic, and Asian ancestry, respectively (Khera 2019).		
DISEASE INFORMATION: Coronary artery disease (CAD) is the most common type of heart disease in the United States, caused by plaque buildup in the walls of the coronary arteries, which supply blood to the heart. Risk of developing CAD increases with age. Symptoms of CAD include chest pain (angina), weakness, light-headedness, nausea, pain or discomfort in the arms or shoulder, shortness of breath, and heart attack (adapted from Centers for Disease Control and Prevention https://www.cdc.gov/heartdisease/coronary_ad.htm).		

Disease	This patient's result	General disease prevalence
Prostate cancer	Average polygenic risk	1 in 8 men
RISK INTERPRETATION: The patient's calculated polygenic risk score, derived from 137 loci, has NOT been associated with high risk for prostate cancer, defined here as greater than 2-fold risk. For the majority of individuals with polygenic risk scores similar to this, risk for prostate cancer is not likely to be increased above baseline. On the other hand, values of this polygenic risk score that fall among the top 13% were associated with a greater than 2-fold greater risk of developing prostate cancer among >140,000 men of European ancestry when compared to the average individual (Schumacher 2018).		
DISEASE INFORMATION: Prostate cancer is characterized by abnormal cell growth in the prostate. All men are at risk for prostate cancer; however, the risks for developing and dying from prostate cancer increase with age, with the highest incidence being observed in men ≥ 65. Prevalence of prostate cancer varies among different ethnic and racial groups, with the highest prevalence observed in African-American men. Prostate cancer is often asymptomatic in its early stages but can be associated with bone pain if it spreads to other parts of the body (adapted from Centers for Disease Control and Prevention www.cdc.gov/cancer/prostate/basic_info/index.htm and www.ncbi.nlm.nih.gov/pmc/articles/PMC6497009/).		

Disease	This patient's result	General disease prevalence
Type 2 diabetes	Average polygenic risk	More than 1 in 5 people >65 years old
RISK INTERPRETATION: The patient's calculated polygenic risk score, derived from 6917079 loci, has NOT been associated with high risk for type 2 diabetes, defined here as greater than 2-fold risk. For the majority of individuals with polygenic risk scores similar to this, risk for type 2 diabetes is not likely to be increased above baseline. On the other hand, values of this polygenic risk score that fall among the top 8% were associated with a greater than 2-fold greater risk of developing type 2 diabetes among >400,000 British volunteers of European ancestry when compared to the average individual (Khera 2018).		
DISEASE INFORMATION: Type 2 diabetes (T2D) is characterized by high blood sugar levels due to abnormal insulin processing. T2D typically occurs during middle or late adulthood and progresses over time. Symptoms include polyuria, polydipsia, fatigue, blurred vision, neuropathy, and weight loss (adapted from Genetics Home Reference https://ghr.nlm.nih.gov/condition/type-2-diabetes).		

[FOR FEMALES]

Disease	This patient's result	General disease prevalence

Breast cancer	Average polygenic risk	1 in 8 women
RISK INTERPRETATION: The patient's calculated polygenic risk score, derived from 3820 loci, has NOT been associated with high risk for breast cancer, defined here as greater than 2-fold risk. For the majority of individuals with polygenic risk scores similar to this, risk for breast cancer is not likely to be increased above baseline. On the other hand, values of this polygenic risk score that fall among the top 8% were associated with a greater than 2-fold greater risk of developing breast cancer among >380,000 women of European ancestry when compared to the average individual (Mavaddat 2019).		
DISEASE INFORMATION: Breast cancer is a disease characterized by the multiplication of abnormal breast cells to form a tumor. Breast cancer can develop in both men and women, but is much more common in females. Breast cancer is the second most common cancer in women, and about 1 in 8 women in the United States will develop breast cancer during her lifetime. Symptoms include a lump or thickening in or near the breast, changes in breast size or shape, nipple discharge, tenderness, inversion of nipples, and skin irritation, dimpling, or scaliness. Many women with breast cancer have no symptoms early in the disease (adapted from Genetic Home Reference https://ghr.nlm.nih.gov/condition/breast-cancer).		

Limitations: The summary risk assessments above are based on combining individual risk allele data in ways that may not always apply to each individual patient. These results should be interpreted in the context of the patient's medical evaluation, family history, and racial/ethnic background. The performance of polygenic risk scores is typically more robust across European ancestries than non-European ancestries.

B. ACTIONABLE MONOGENIC DISEASE FINDINGS

This genotyping test did NOT identify any of the tested medically significant variants in a list of 59 actionable genes (See methodology). Pathogenic variants in these genes are associated with an increased risk of developing certain conditions, including some cancers, heart diseases, and other disorders. NOTE, this assay did not interrogate all possible pathogenic variation within these 59 genes, and thus a negative result does not reflect an absence of a disease-causing variant in one or more of these genes in this individual.

Limitations: As a genotyping assay, this test cannot detect variants which are not specifically targeted on the array and thus should not be considered a screening assay for these genes. Specific types of genetic variation, such as triplet repeat expansions, structural variation, and copy number events, or sequence variants not included in the list of genotyped variants are currently not reliably detected by this assay. Variant interpretation may change over time if more information becomes available. Please note that the presence of pathogenic variants in variants not part of the assay, genes not analyzed, or regions not captured by filtering strategies cannot be fully excluded.

RECOMMENDATIONS

These results should be interpreted in the context of this individual's personal medical history and family history. [IF ACTIONABLE MONOGENIC DISEASE FINDING: Genetic counseling is recommended for this individual and their relatives. Familial variant testing for the *GENE* variant is available if desired.]

METHODOLOGY

Genotyping:
Genomic DNA was genotyped using the Illumina Global Diversity Array and iScan instrumentation by the Clinical Research Sequencing Platform of the Broad Institute. Illumina's Autocall genotype calling software (IAAP) determines genotyping results in GTC format from image data scanned on the Illumina genotyping array. Samples were required to meet or exceed the 98% call rate specification by Autocall in order to be considered passing. The GTC file generated by Autocall is converted to a VCF file using internally validated software. The VCF file from the GDA is imputed using 1000 Genomes as the reference population to determine the likely allele state at sites across the genome using an imputation pipeline that includes quality control, phasing (using EAGLE) and imputation (using minimac4).

Polygenic Disease Risk:
This test assessed the polygenic risk for developing the following six conditions as previously identified: breast cancer (in females), prostate cancer (in males), colorectal cancer, atrial fibrillation, coronary artery disease, and type 2 diabetes (Huyghe et al., 2019; Khera et al., 2018; Mavaddat., 2018; Schumacher et al., 2018). For each condition, a population-standardized PRS is computed as the sum of the patient's risk alleles across multiple SNPs weighted by the SNP-specific effects reported in large genome-wide association studies. An individual is considered to be at an increased PRS if their standardized PRS is above the threshold score for an OR>2 as compared to the median PRS value [computed as ln(2)/ln(change in OR per SD change in PRS)].

Actionable Monogenic Disease Findings:
The American College of Medical Genetics and Genomics (ACMG) recommends reporting variants discovered in certain genes when discovered as secondary findings in a genomics assay. These variants cause conditions that are considered actionable, meaning there are specific guidelines available to monitor and/or treat these conditions. If identified, only those variants recommended by ACMG are reported. Note, this test will miss pathogenic or likely pathogenic variants not included in the genotyping array. Variants present on the array in these 59 genes have been filtered to include: (1) variants identified by our laboratory to be pathogenic or likely pathogenic; (2) variants classified as disease causing mutations in public

databases that have a minor allele frequency <5.0% in the Genome Aggregation Database (gnomAD, http://gnomadexac.broadinstitute.org/); and (3) nonsense, frameshift, and +/-1,2 splice-site variants in disease-associated genes with a minor allele frequency ≤1.0% in gnomAD. The evidence for phenotype-causality has been evaluated for each variant identified from the filtering strategies listed above and variants are classified based on ACMG/AMP criteria (Richards et al. 2015) with ClinGen rule specifications (https://www.clinicalgenome.org/working-groups/sequence-variant-interpretation/). Variants are reported according to HGVS nomenclature (http://varnomen.hgvs.org/). Only those variants with evidence for causing or contributing to disease are reported. All disease-associated variants on this report are confirmed via Sanger sequencing or another orthogonal technology. Please contact the laboratory for additional information or for a complete list of variants and genes analyzed.

The initial genotyping component of this test was performed by the Clinical Research Sequencing Platform of the Broad Institute (320 Charles St, Cambridge, MA 02141; CLIA#22D2055652), and the Sanger confirmation, interpretive algorithms and clinical reports were generated by the Laboratory for Molecular Medicine at Partners Healthcare Personalized Medicine (LMM, 65 Landsdowne St, Cambridge, MA 02139; 617-768-8500; CLIA#22D1005307). This test has not been cleared or approved by the U.S. Food and Drug Administration (FDA). The FDA has determined that such clearance or approval is not necessary.

REFERENCES

Huyghe, J. R., Bien, S. A., Harrison, T. A., Kang, H. M., Chen, S., Schmit, S. L., . . . Peters, U. (2019). Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet, 51*(1), 76-87. doi:10.1038/s41588-018-0286-6

Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P., . . . Miller, D. T. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med, 19*(2), 249-255. doi:10.1038/gim.2016.190

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., . . . Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet, 50*(9), 1219-1224. doi:10.1038/s41588-018-0183-z

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., . . . Easton, D. F. (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet, 104*(1), 21-34. doi:10.1016/j.ajhg.2018.11.002

Schumacher, F. R., Al Olama, A. A., Berndt, S. I., Benlloch, S., Ahmed, M., Saunders, E. J., . . . Mechanisms in Oncology /Elucidating Loci Involved in Prostate Cancer Susceptibility, C. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet, 50*(7), 928-936. doi:10.1038/s41588-018-0142-8

Report Preparation by:
Final Report by: