SUPPLEMENTARY METHODS

Quality control of the GSA data for n=635623 variants was performed stratified by ethnicity category (African American, European American, Hispanic American, Other). We performed sample level quality control based on a number of steps including gender discordance, palindromic SNPs, low coverage, contamination, duplicates, discordance with genotyped data; one individual from every pair of individuals with closer than second degree relatedness was removed. Individuals with an ethnicity-specific heterozygosity rate that surpassed +/- 6 standard deviations of the population-specific mean, along with individuals with a call rate of <95% were removed (N=684 participants in total). N=84 individuals were then removed for exhibiting persistent discordance between EHR recorded and genetic sex. An additional N=4 individuals with phenotypically indeterminate sex were also excluded. A further N=102 duplicate individuals were also excluded from downstream analysis. Finally, one of each pair of 28 duplicates were excluded. In total 31,911 passed sample level QC for downstream analysis. Sites with a call rate below 95% were excluded (n=19253), along with sites that were seen to significantly violate Hardy-Weinberg equilibrium (HWE) when calculated stratified by ancestry. HWE thresholds varied by ethnicity, specifically we set a threshold of $p < 1e\text{-}5$ in African American and European American, or $p < 1e\text{-}13$ in Hispanic American (n=11503 SNPs in total). Variants with a minor allele frequency < 5% were removed.

Imputation of Genotyping Data

All BioMe individuals who were successfully genotyped on GSA chip were subsequently imputed into the 1000 Genomes Phase 3 data release. Genotype data which passed

the above quality control filters was phased with SHAPEIT2,1 and imputed to 1000 Genomes Phase 3 reference data using IMPUTE version 2.3.2.2.

Segments of the genome, which were known to harbor gross chromosomal anomalies, were filtered out of the final genotype probabilities files. Imputed sites were excluded if the IMPUTE info score was less than 0.4. The mean IMPUTE info score was 0.873.

Genetic Ancestry

Following imputation, principal component analysis was performed on the imputed genotyping data. In all subsequent statistical analyses, the top 10 genetic principal components (PCs) were included as covariates unless otherwise indicated.

Supplementary Tables

Supplementary Table 1. Characteristics of Cases and Controls in UK BioBank GWAS Cohort

	Case (N=5019)	**Control (N= 352641)**	**P value**
Male	3522 (70%)	161755 (46%)	<0.001
Age, Mean (SD)	58.7 (7.5)	56.8 (8.0)	<0.001
Body Mass Index, Mean (SD)	29.0 (5.0)	27.4 (4.7)	<0.001
Clinical Comorbidities			
Hypertension, n (%)	2087 (70%)	98451 (46%)	<0.001
Type 2 Diabetes (%)	898 (18%)	23773 (7%)	<0.001
Obesity, n (%)	1729 (36%)	81651 (24%)	<0.001
Gout, n (%)	173 (3%)	3522 (1%)	<0.001

Supplementary Table 2. Characteristics of Cases and Controls in PMBB

	Cases (n=811)	**Controls (n=14585)**	**P**
Male	522 (64%)	8102 (56%)	<0.001
Age, Mean (SD)	68.0 (13.1)	65.8 (16.2)	0.002
Race, n (%)			
African American	223 (27%)	5200 (36%)	<0.001
European American	588 (73%)	9385 (64%)	<0.001
Body Mass Index, Mean (SD)	30.0 (6.78)	29.8 (7.14)	0.24
Clinical Comorbidities			
Hypertension, n (%)	661 (82%)	3924 (27%)	<0.001
Type 2 Diabetes (%)	309 (38%)	4408 (30%)	<0.001
Obesity, n (%)	376 (46%)	6560 (45%)	0.46
Gout, n (%)	151 (19%)	1472 (10%)	<0.001

Supplementary Table 1. Association with urinary tract stone diagnosis in Bio*Me* adjusted for age, sex, BMI, and 10 genetic principal components for different values of LDPred rho.