Online Data Supplement

The Undiagnosed Disease Burden Associated with Alpha-1 Antitrypsin Deficiency Genotypes

Tomoko Nakanishi, Vincenzo Forgetta, Tomohiro Handa, Toyohiro Hirai, Vincent Mooser, Mark G. Lathrop, William O.C.M. Cookson, J. Brent Richards

MATERIAL AND METHODS

Genotyping and ancestry assignment

Genotyping microarrays used in UK Biobank are Affymetrix Axiom array (90%) and Affymetrix BiLEVE array (10%), which share more than 95% of SNPs. We removed those who had discordance between inferred gender and submitted gender, outliers based on heterozygosity and missing rates, those with excessive relatives who have more than 10 putative third-degree relatives, and those with putatively carrying sex chromosome configurations that are not either XX or XY using the quality control (QC) metrics provided by UK Biobank[1]. This genotype QC process retained 486,355 individuals. To assign European ancestry, we projected the UK Biobank samples onto the first 20 principal components (PCs) estimated from the 1000 Genome Phase 3 project data[2] (1000G) using FastPCA version 2[3]. Projections used a curated set of 38,511 LDpruned HapMap Release 3 (HapMap3)[4] bi-allelic SNPs that were shared between the 1000G and UK Biobank datasets (minor allele frequency > 1%, minor allele count > 5, genotyping call rate > 95%, Hardy-Weinberg p > 1.0 x 10^{-6}, and regions of extensive LD removed). Expectation Maximization (EM) clustering (implemented in R using "EMCluster" [https://github.com/snoweye/EMCluster]) was used to compute probabilities of cluster membership based on a finite mixture of multivariate Gaussian distributions with unstructured dispersion. Eigenvectors 1, 2 and 5 were used for clustering as they represented the smallest number of eigenvectors that were able to resolve the British 1000G subpopulation (GBR) from other ethnicities. Twelve predefined clusters were chosen for EM clustering as sensitivity analyses suggested that this number provided a good compromise between model fit (as quantified by log likelihood, Bayesian information criterion, and Akaike information criterion) and computational burden. Detail on this method are discussed in our previous report[5]. Here, we defined 458,861 UK Biobank participants as European that clustered together with 1000G European superpopulation (see Figure S1 and Table S1). Out of 458,861 European participants, 458,164 participants with non-missing genotype information of rs28929474 and rs17580 were used for the downstream analyses (Figure S2). 9,243 participants of African-descent, 3,111 participants of Admixed American-descent, 2,475 participants of East Asian, and 11,391 participants of South Asian were also defined along with European ancestry (Table S3).

Definition of the presence of respiratory symptoms

We defined the presence of respiratory symptoms through touchscreen questionnaire responses completed at the assessment centres or online follow-ups and included as symptoms: wheeze or whistling in the chest in last year (Data-Field 2316), shortness of breath walking on level ground (Data-Field 4717), cough on most days (Data-Field 22502) and bringing up phlegm/sputum/mucus on most days (Data-Field 22504). All the participants with missing data or "-3" (Prefer not to answer) were treated as without symptoms, which may decrease the frequency of symptoms.

Spirometry quality control

The quality control of spirometry data was based on the American Thoracic Society and European Respiratory Society (ATS/ERS) criteria[6] of acceptability and reproducibility and referred to the previous paper[7] which also analyzed UK Biobank spirometry data.

Acceptability of blows

At first, 1,241,336 blows were available at baseline visits and we included to be acceptable those with the following values in the Vitalograph spirometer blow quality metrics (Data-Field 3061), namely “blank” (0), “TEST_DURATION (set if the test is less than 6 seconds)” (16), “USER_ACCEPTED” (32) and “TEST_DURATION and USER_ACCEPTED” (48). A total of 790,263 blows were retained.

Next, to ensure the good start of blow, the extrapolated volume at the start of test was calculated using the data points of blow (Data-Field 3066) and was removed if the extrapolated volume was larger than 150ml or 5% of FVC. A total of 788,193 acceptable blows were retained.

Reproducibility of measurements

To meet the criteria for reproducibility, we compared each blow to the other blows and kept them if the difference was within 250ml. Since epidemiological studies are not designed for monitoring the individual patients, this threshold of 250ml, was applied which was confirmed as reliable in previous study by comparing the longitudinal results of same individuals with intervals of years[7]. Finally, 373,693 blows for FEV1 and 362,048 blows for FVC were defined as “acceptable” and “reproduced”.

Best measurements

The best measurements of FEV1 and FVC were defined as the highest measurements from the "acceptable" and "reproduced" measurements. FEV1/FVC was derived from the best measurements of FEV1 and FVC, which means that the best FEV1 and FVC of a single participant are not necessarily derived from the same blow. In total, 354,522 participants from total population have final valid FEV1/FVC information.

Calculation of percentage of predicted FEV1 value

We calculated the percentage of predicted FEV1 value in R v3.5.0 by using "rspiro" package (https://github.com/thlytras/rspiro) which implements the Global Lung Function Initiative 2012 spirometry equations[8]. Age, height, and self – reported ethnicity information was obtained from Data-Field 21003, 50, and 21000, respectively. Table S5 shows the detailed information of the coding for ethnicity used for the calculation of %FEV1 predicted.

Bronchodilator information

Although 2,298 participants who had diagnoses of asthma used inhalers for chest within last hour (Data-Field 3090), we did not remove these participants. This is because UK Biobank dataset does not cover all the medication lists in asthmatic patients and excluding all asthmatic people in the analyses may cause a selection bias.

Statistical analysis

Regression models were fitted to assess the associations of *SERPINA1* genotypes and clinical outcomes compared to PI*MM genotype. All the models were adjusted for age (Data-Field[DF] 21003), sex (DF 31), genotype arrays (DF 22000), assessment centre (DF 54) and the first five principal components (PC) in order to account for population structure.

Definition of smoking status and exposure to smoke or polluted air in household or workplace

Smoking status was defined by the questionnaire - based information as was previously reported[7]. "Never smokers" were people who answered "not smoking at present" and "never smoked in the past", or who answered "not smoking at present, smoked occasionally or just tried once or twice in the past" but didn't have more than 100 episodes of smoking over their lifetime. "Current smokers" were those who smoke at present on most, or all days, or occasionally. "Past smokers" were those who do not smoke at present but smoked on most or all days in the past, who do not smoke at present and smoked occasionally or just tried once or twice in the past but had at least 100 episodes of smoking over their lifetime, or who smoked on most or all days or occasionally with more than 100 episodes of smoking over their lifetime but prefer not to answer whether they smoke at present. Table S4 summarizes the Data-Fields and the codes we curated for smoking status. “Current smokers” and “past smokers” were combined and termed as “ever-smokers” for the downstream regression analyses. We also defined “heavy-smokers” as those with more than 20 pack-years[9] using the Data-Field 20161. With respect to exposure to smoke or polluted air in household or workplace, we selected all individuals with the answers of “at least one household member smokes” (coding “1” or “2”) in smoking/smokers in household (Data-Field 1259) in addition to the individuals with the answers of “Often (-141)” or “Sometimes (-131)” in occupational exposure questions (Workplace very dusty [Data-Field 22609], Workplace full of chemical or other fumes [22610], Workplace had a lot of cigarette smoke from other people smoking Employment history [22611], Worked with materials containing asbestos [22612], Worked with paints, thinners or glues [22613], Worked with pesticides [22614], Workplace had a lot of diesel exhaust [22615]).

Survival analysis

We set the date of attending assessment centre (Data-Field 53) as the starting date and participants with the date of death (Data-Field 40000) were determined to have died while the rest of the participants without death registry information were assumed to be alive until 2016/02/16, which is the last date included in the Data-Field 40000. We used R package “survminer” for this analysis. (http://www.sthda.com/english/rpkgs/survminer/). “Cox.zph” function was used to test the proportional hazard assumption in Cox regression. As a sensitivity analysis, multivariate Cox proportional hazard model adjusted for age was also applied for survival analysis.

Phenome-wide association study

Mapping ICD codings to phecode

We used the information in Data-Field 41203 / 40205 and 41202 / 40204 for ICD-9 and ICD-10, respectively. We mapped ICD -9 codes to phecodes in Phecode Map 1.2 with ICD-9 Codes by the following method: First, we searched by exact matching of the code descriptions, followed by exact match of meaning description, and if the codes in UK Biobank start with the ICD-9 codes in Phecode Map, we also mapped them to the corresponding phecodes. For example, “V0299” in UK Biobank was treated to be same as “V02.9” in Phecode Map. We mapped similarily ICD-10 codes to phecodes by using both Phecode Map 1.2 with ICD-10 Codes (beta) and Phecode Map 1.2 with ICD-10cm Codes (beta)[10]. For the rest of the ICD-10 codes, we also included the match of the first 3 digits (for example, “A561” in UK Biobank was treated to be same as “A56.*” in Phecode Map). 43,928 (84%) out of 52,259 cumulative ICD-9 codings and 2,797,060 (95%) out of cumulative 2,933,010 ICD-10 codings were successfully mapped to phecodes. The non-mapped codings are listed in Table S7 (ICD-9) and Table S8 (ICD-10). Table S8 includes mainly abortions, external causes of injury and poisoning (E) or factors influencing health status and contact with health services (V). For ICD-10 codes, several unmatched codings with prefixes from B to M were manually mapped to the phecodes listed in Table S8. Finally, the majority of phenotypes uncovered that were unmatched had prefixes from O to Z, which were predominantly pregnancy related outcomes, injuries or factors influencing health status.

Selection of phecodes tested

We first converted ICD-9 and ICD-10 codes to phecodes and created the matrix of the phenotypes of the participants using “createPhenotypes.R” function in PheWAS R package[11] with the parameters of min.code.count=1 and add.phecode.exclusions=T. We included the phecodes with at least 20 cases in total and at least 1 case in mutant genotype group in the analyses and then undertook a logistic regression model between the *SERPINA1* genotypes and each phecode, adjusting for age, sex, genotyping array, assessment centre and first five principal components using “glm” function in R. We applied both Bonferroni and Benjamini-Hochberg correction and reported the significant phenotypes ($p < 0.05$).

Sensitivity analyses

We performed two sets of sensitivity analyses. As case-control imbalance could cause high type I error rate, we first performed Firth test using “logistf” function (https://cran.r-project.org/web/packages/logistf/index.html) to all phenotypes, which is known to have as low type I error rate. Second, Hospital Episode Statistics are coded by administrative staff who referred to the patient notes of clinicians, our PheWAS design could be affected by the misclassification. Therefore, we reperformed PheWAS by restricting “cases” to only those with >= 2 entries of codes with the parameters of min.code.count=2 in “createPhenotypes.R” function to reduce errors due to incorrect codings. Here we included all the phecodes with at least one case in total. We acknowledge this approach is less sensitive as we have removed substantial true cases with only one entry. For instance, the number of those with asthma (phecode “495”) dropped from 25,331(6.5%) to 1,775(0.48%) in PI*ZZ and PI*MM individuals when we excluded those with only one entry of the code, which is much fewer than the prevalence of asthma (Table S14). We used same sets of covariates as the main analysis in all sensitivity analysis. Quantile-quantile plots were drawn using “qqman” function.

Polygenic risk score of spirometry data

We first filtered SNPs for stringent quality control metrics, retaining only SNPs with a minor allele frequency > 0.05% and an imputation quality score > 0.3. We then hard-called all post-QC genotyped and imputed variants by PLINK[12]. We used GWAS summary statistics of FEV1/FVC derived from 79,055 European individuals in SpiroMeta consortium[7], in which the authors observed no overlap of samples compared to UK Biobank. Polygenic risk score(PRS) for FEV1/FVC was established by implementing LDpred[13] with LD radius of 400 using only HapMap3 SNPs, which is a recommended setting by the author. We randomly selected 5,000 individuals from the cluster of White British in the previous report[14] and used them as LD reference genotypes. The top-performing PRS with the fraction of causal markers set to be 0.01, was selected among LDpred -inf (LDpred specialized to an infinitesimal prior) and LDpred with the fraction of causal markers set at various proportions: 1, 0.3, 0.1, 0.03, 0.01, 0.003 and 0.001. Although in theory this procedure is susceptible to overfitting, overfitting is considered to have a negligible effect in practice given its large sample size (n=328,638) and the small discrete set of parameter choices. To estimate the interactions between *SERPINA1* genotype and common variants (PRS), we applied multivariate logistic regression for

FEV1/FVC < 0.7 with *SERPINA1* genotypes, PRS (for the sake of interpretation, it was first multiplied by -1 and standardized) and interaction terms between each other.

RESULT

Sensitivity analyses

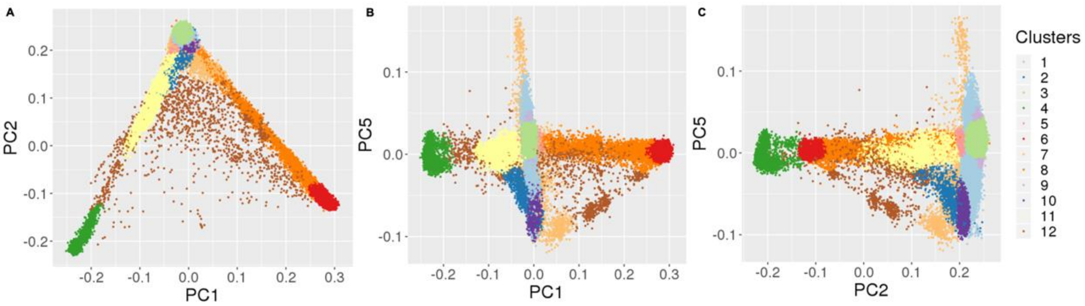
When we included E88.0 in ICD-10 codes for the diagnosis of AATD, there were 294 (0.064%) AATD diagnoses in the total European ancestry population, which is more frequent than the previous estimates[15]. We acknowledge that this likely represents an over-estimate of AATD diagnostic rate, given that this ICD-10 code includes other diseases. Twenty (14%, 95% CI: 9.4% – 21%) out of 140 PI*ZZ individuals were diagnosed as AATD and among 31 PI*ZZ individuals with COPD diagnosis, only 16 (52%, 95% CI: 35% – 68%) were diagnosed as having AATD. We excluded 7,404 participants who have at least one relative identified in UK Biobank and reperformed the association studies. The results were similar to the main results (Table S13). Multivariate Cox proportional hazard model provided similar hazards to all-cause mortality compared to univariate analysis (Table S12).

Phenome-wide association study (PheWAS)

In PheWAS, we tested 1,315 icd.phenotypes for PI*ZZ vs PI*MM individuals (n = 398,564), 1,311 icd.phenotypes for PI*SZ vs PI*ZZ individuals, 1,316 icd.phenotypes for PI*MZ vs PI*ZZ individuals and 1,317 icd.phenotypes for PI*SS vs PI*ZZ individuals (Table S14-S17). PI*ZZ genotype was associated with increased risk of other disorders of metabolism (including AATD) (OR: 124.7, 95%CI: 73.4 – 211.9, p = 3.2 x 10^{-71}), emphysema (OR: 53.2, 95%CI: 31.2 – 90.7, p = 2.2 x 10^{-48}), obstructive chronic bronchitis (OR: 16.4, 95%CI: 8.0 – 33.5, p = 1.5 x 10^{-14}), chronic airway obstruction (OR: 8.4, 95%CI: 4.6 – 15.3, p = 4.3 x 10^{-12}), dependence on respirator or supplemental oxygen (OR: 15.5, 95%CI: 4.8 – 50.1, p = 4.8 x 10^{-6}), cachexia (OR: 69.2, 95%CI: 9.3 – 515.7, p = 3.6 x 10^{-5}), and secondary polycythemia (OR: 19.7, 95%CI: 4.8 – 81.1, p = 3.6 x 10^{-5}) with p < 0.05/1,317, with p < 0.05/1,317, amongst which the first three were validated by two sensitivity analyses we performed (Table S14, Figure S3). PI*SZ was associated with increased risk of lipoprotein disorders (OR: 20.8, 95%CI: 5.0 – 86.0, p = 2.8 x 10^{-5}) and open wound of neck (OR: 11.8, 95%CI: 3.7 – 37.2, p = 2.6 x 10^{-5}) with p < 0.05/1,311, none of which was validated in the sensitivity analysis (Table S15, Figure S4). PI*MZ was associated with increased risk of cholelithiasis (OR: 1.3, 95%CI: 1.2 – 1.5, p = 7.8 x 10^{-10}), calculus of bile duct (OR: 1.6, 95%CI: 1.3 – 1.8, p = 2.8 x 10^{-8}), abnormal results of function study of liver (OR: 1.5, 95%CI: 1.3 – 1.7, p = 1.8 x 10^{-7}), emphysema (OR: 1.6, 95%CI: 1.3 – 1.7, $p = 1.3 \times 10^{-6}$), and decreased risk of coronary atherosclerosis (OR: 0.8, 95%CI: 0.8 – 0.9, $p = 3.5 \times 10^{-6}$) with $p < 0.05/1,311$, all of which were also significant with Firth test (Table S16, Figure S5). No phenotype remained associated with PI*SS after Benjamini-Hochberg correction (Table S17).

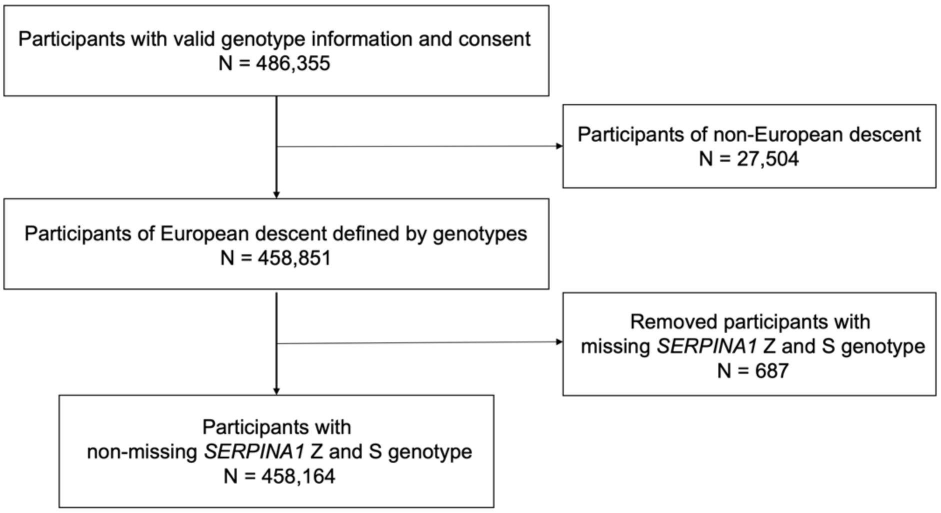
Figures

FigureS1. Bivariate scatterplots of principal components 1, 2 and 5 of the genotypes of UK Biobank samples.



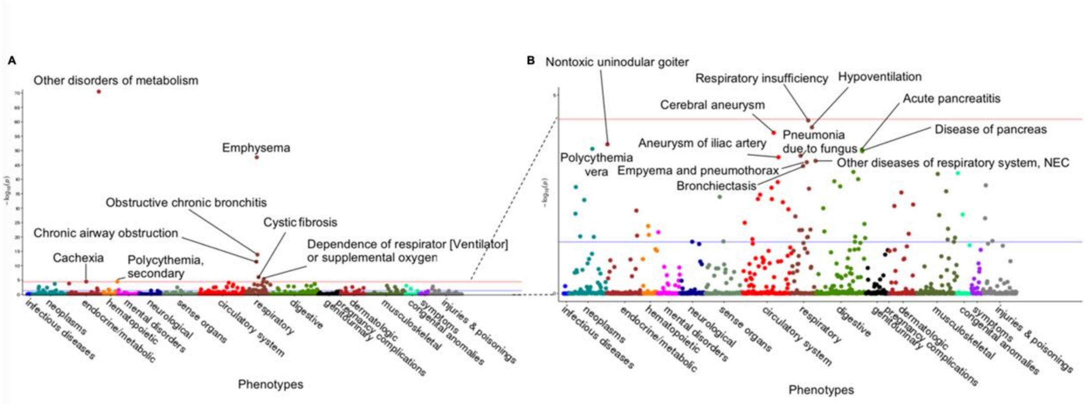
(A) x axis=first principal component(PC) (PC1); y axis=second PC (PC2), (B) x axis=PC1, y axis=fifth PC(PC5), (C) x axis=PC2, y axis=PC5. Clusters 1 – 12 were the clusters identified by clustering method[5] with 1000G individuals[2]. The detailed methods are described above in the online supplement.

Figure S2. Flow diagram of the study participants, inclusion and exclusion for the analysis.


Participants with valid genotype information and consent
N = 486,355
Participants of non-European descent
N = 27,504
Participants of European descent defined by genotypes
N = 458,851
Removed participants with missing *SERPINA1* Z and S genotype
N = 687
Participants with non-missing *SERPINA1* Z and S genotype
N = 458,164

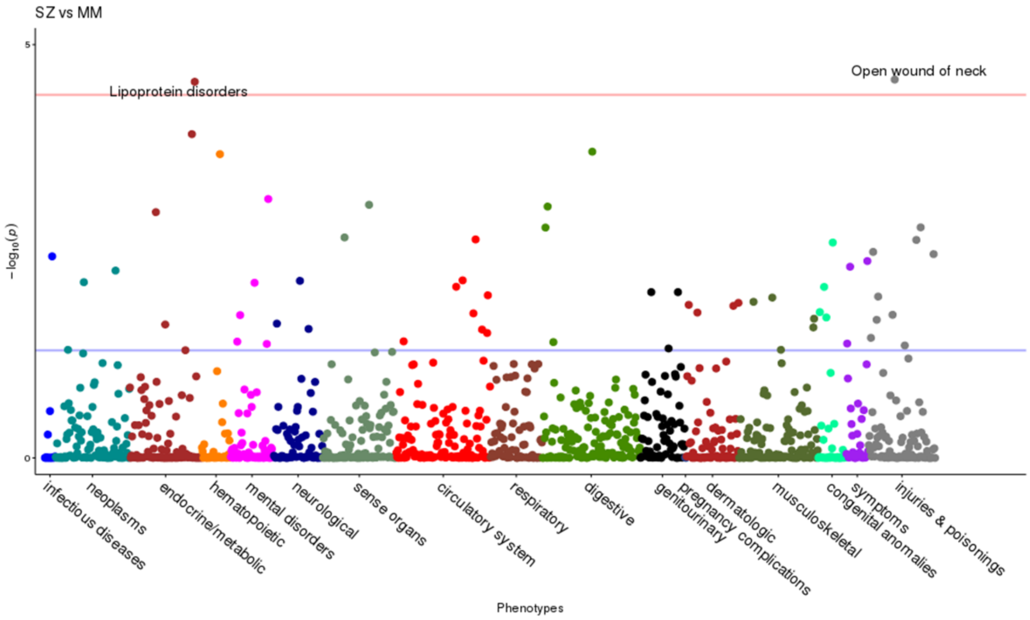
Detailed methods of defining European descent is in the online supplement.

Figure S3. PheWAS Manhattan plot for PI*ZZ vs PI*MM genotypes.



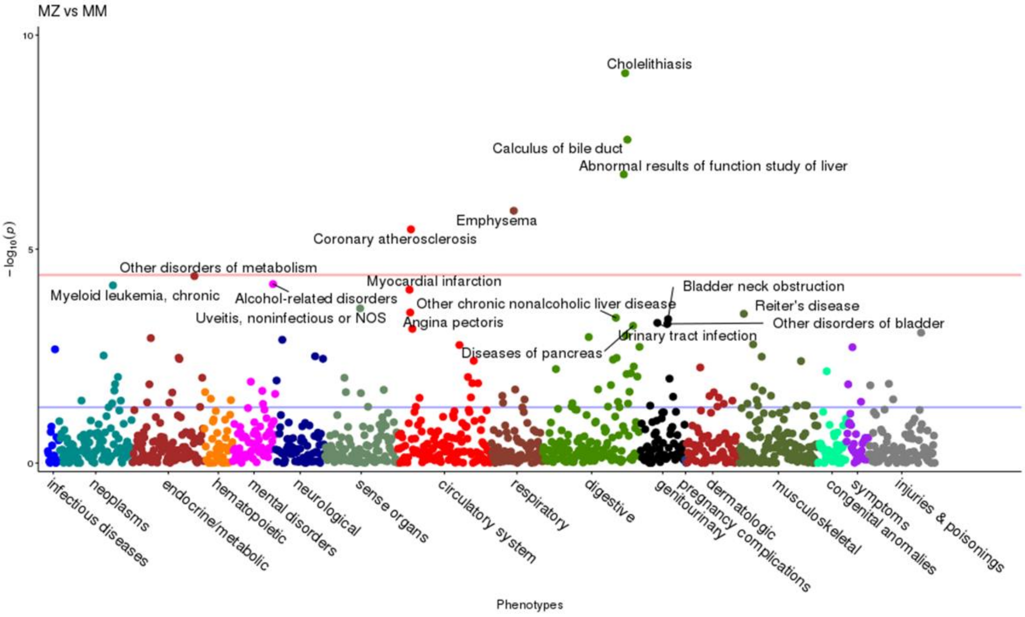
Logistic regression models were adjusted for age, sex, genotyping array, assessment centre and the first five genetic principal components. X axis=each phenotype ordered by phecode; Y axis=-log10(p value) from logistic regression. red line=p-value threshold after Bonferroni corrected value < 0·05; blue line=p-value < 0.05. annotated phenotypes=statistically significant (p < 0.05) after Benjamini-Hochberg correction. (A) Normal view. (B) Enlarged view.

Figure S4. PheWAS Manhattan plot for PI*SZ vs PI*MM.



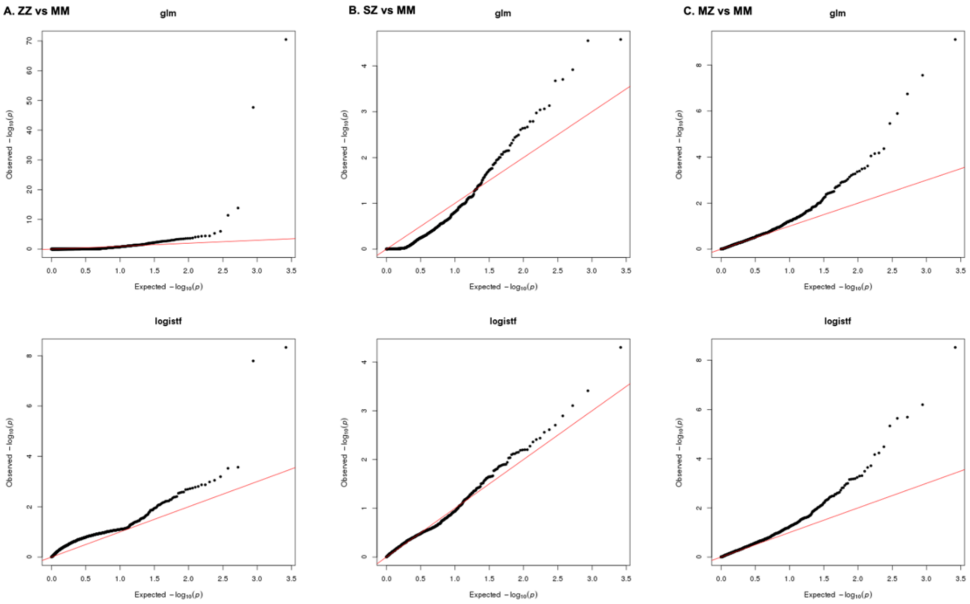
Logistic regression models were adjusted for age, sex, genotyping array, assessment centre and the first five genetic principal components. X axis=each phenotype ordered by phecode; Y axis=-log10(p value) from logistic regression. red line=p-value threshold after Bonferroni corrected value < 0.05; blue line=p-value < 0.05. annotated phenotypes=statistically significant (p < 0.05) after Benjamini-Hochberg correction.

Figure S5. PheWAS Manhattan plot for PI*MZ vs PI*MM.



Logistic regression models were adjusted for age, sex, genotyping array, assessment centre and the first five genetic principal components. X axis=each phenotype ordered by phecode; Y axis=-log10(p value) from logistic regression. red line=p-value threshold after Bonferroni corrected value < 0.05; blue line=p-value < 0.05. annotated phenotypes=statistically significant (p < 0.05) after Benjamini-Hochberg correction.

Figure S6. Quantile-quantile plots of PheWAS.



Quantile-quantile plots for PheWAS. (A) ZZ vs MM. (B) SZ vs MM. (C) MZ vs MM. All the test statistics were calculated with “glm” and “logistf”, respectively.

References

1. (No Title) [Internet]. [cited 2020 Jun 10].Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2017/07/ukb_genetic_file_description.txt.

2. Consortium T 1000 GP, Auton A, Abecasis GR, Altshuler (Co-Chair) DM, Durbin (Co-Chair) RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs (Principal Investigator) RA, et al. A global reference for human genetic variation. *Nature* [Internet] The Author(s); 2015; 526: 68–74Available from: https://doi.org/10.1038/nature15393.

3. Galinsky KJ, Bhatia G, Loh P-R, Georgiev S, Mukherjee S, Patterson NJ, Price AL. Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* 2016; 98: 456–472.

4. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R,

Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; 467: 52–58.

5. Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, Vulpescu NA, Forgetta V, Kleinman A, Mohanty ST, Sergio CM, Quinn J, Nguyen-Yamamoto L, Luco AL, Vijay J, Simon MM, Pramatarova A, Medina-Gomez C, Trajanoska K, Ghirardello EJ, Butterfield NC, Curry KF, Leitch VD, Sparkes PC, Adoum AT, Mannan NS, Komla-Ebri DSK, Pollard AS, Dewhurst HF, Hassall TAD, et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* 2019; 51: 258–266.

6. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CPM, Gustafsson P, Jensen R, Johnson DC, MacIntrye N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wagner J. Standardisation of spirometry. *Eur. Respir. J.* 2005; 26: 319–338.

7. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, Batini C, Fawcett KA, Song K, Sakornsakolpat P, Li X, Boxall R, Reeve NF, Obeidat M, Zhao JH, Wielscher M, Weiss S, Kentistou KA, Cook JP, Sun BB, Zhou J, Hui J, Karrasch S, Imboden M, Harris SE, Marten J, Enroth S, Kerr SM,

Surakka I, Vitart V, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* 2019; 51: 481–493.

8. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MSM, Zheng J, Stocks J. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur. Respir. J.* [Internet] 2012; 40: 1324 LP – 1343Available from: http://erj.ersjournals.com/content/40/6/1324.abstract.

9. Castaldi PJ, DeMeo DL, Kent DM, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, Rennard SI, Stocks JM, Stoller JK, Strange C, Turino G, Sandhaus RA, Griffith JL, Silverman EK. Development of Predictive Models for Airflow Obstruction in Alpha-1-Antitrypsin Deficiency. *Am. J. Epidemiol.* [Internet] 2009; 170: 1005–1013Available from: https://doi.org/10.1093/aje/kwp216.

10. Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, Zhao J, Carroll R, Bastarache L, Denny JC, Theodoratou E, Wei W-Q. Developing and Evaluating Mappings of ICD-10 and ICD-10-CM Codes to PheCodes. *bioRxiv* [Internet] 2019; : 462077Available from:

http://biorxiv.org/content/early/2019/07/03/462077.abstract.

11. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics* [Internet] 2010; 26: 1205–1210Available from: https://doi.org/10.1093/bioinformatics/btq126.

12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* [Internet] 2007/07/25. The American Society of Human Genetics; 2007; 81: 559–575Available from: https://www.ncbi.nlm.nih.gov/pubmed/17701901.

13. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, Hayeck T, Won HH, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H, Pers TH, Agartz I, Agerbo E, Albus M, Alexander M, Amin F, Bacanu SA, Begemann M, Belliveau RA, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* 2015; 97: 576–592.

14. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep phenotyping and genomic data. *Nature* [Internet] 2018; 562: 203–209Available from: https://doi.org/10.1038/s41586-018-0579-z.

15. Silverman EK, Sandhaus RA. Alpha1-Antitrypsin Deficiency. *N. Engl. J. Med.* 2009; 360: 2749–2757.