

Supplementary Text: Genetically informed precision drug repurposing for lung function and implications for respiratory infection

CONTENTS

Supplementary Note	2
Supplementary Methods	8
Supplementary Figure 1	13
References	14

Supplementary Note

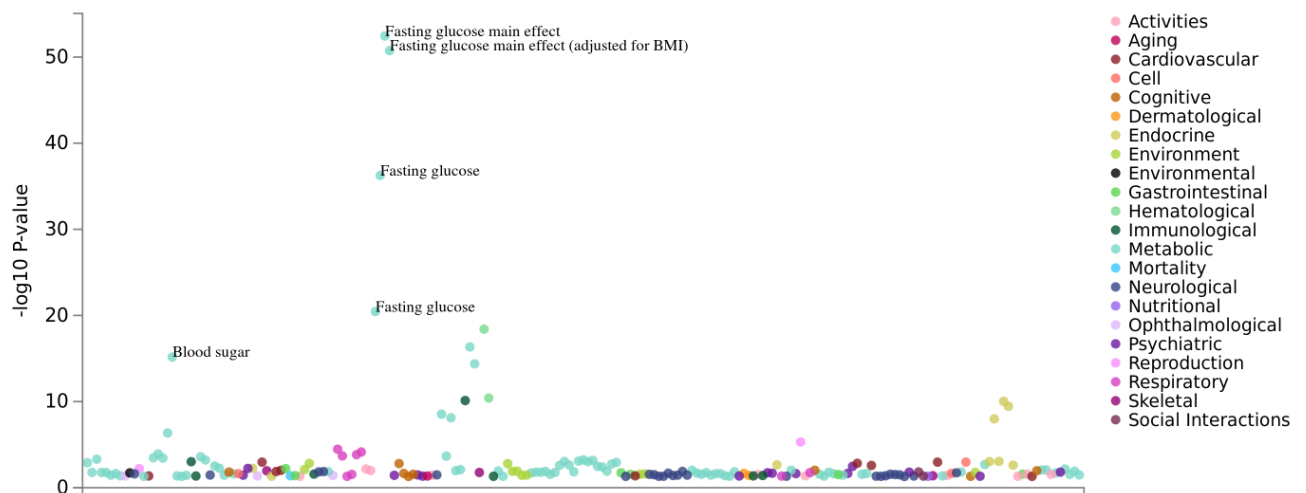
Description of fasting glucose instrumental variables which may be potential outliers

We utilised two-sample Mendelian Randomisation to investigate the effect of fasting glucose on lung function. In these analyses, a ‘leave-one-out’ methodology was applied whereby a single fasting glucose instrumental variable (IV) was sequentially excluded and the causal estimate recalculated (inverse-variance weighted [IVW] estimator with multiplicative random effects, Supplementary Table 7)¹. The IVW estimate was relatively analogous regardless of which IV was removed, however, there were six instances where the causal estimate was no longer statistically significant ($P < 0.05$) upon the exclusion of the IV. There were two such SNPs in the fasting glucose \rightarrow FEV₁ model, and four in the fasting glucose \rightarrow FVC model (two SNPs were shared for both). We investigated whether these outlier SNPs represent either, a) horizontal pleiotropy, or b) a biologically relevant IV with a large effect size. This was undertaken by firstly annotating the SNP to a gene based on genomic position, and then performing a phenome-wide association study (PheWAS) of each SNP using the GWAS atlas database². The five IVs are described below, in all instances we define the PheWAS significance threshold as $P < 1 \times 10^{-5}$, which relates to the number of GWAS considered in the GWAS atlas database.

FEV₁ and FVC: rs6975024

The IVW estimate loses statistical significance upon the removal of this IV - FEV₁: $\beta_{IVW} = -0.078$, $SE = 0.046$, $P = 0.09$; FVC: $\beta_{IVW} = -0.071$, $SE = 0.044$, $P = 0.11$, however, the attenuation in effect size remains relatively small. This variant lies in an intron of the *GCK* (glucokinase) gene. *GCK* is a vital component of glycaemic regulation as it serves as a

pancreatic glucose sensor for insulin secretin and phosphorylates glucose to glucose-6-phosphate, the first stage in glycolysis^{3,4}. The relevance of this SNP to blood glucose is supported by the phenotypes with which this SNP has been associated below phenome-wide significance (SN Figure 1, Supplementary Table 9). Specifically, all of the phenotypes which displayed association with the SNP were glycaemic (fasting glucose, HbA1c, blood sugar, type 2 diabetes, HOMA-B), with only one exception (Birth weight). In summary, this SNP is unlikely to represent pleiotropy due to its plausible involvement in glycaemic biology and lack of association with obvious exposure-outcome confounders.

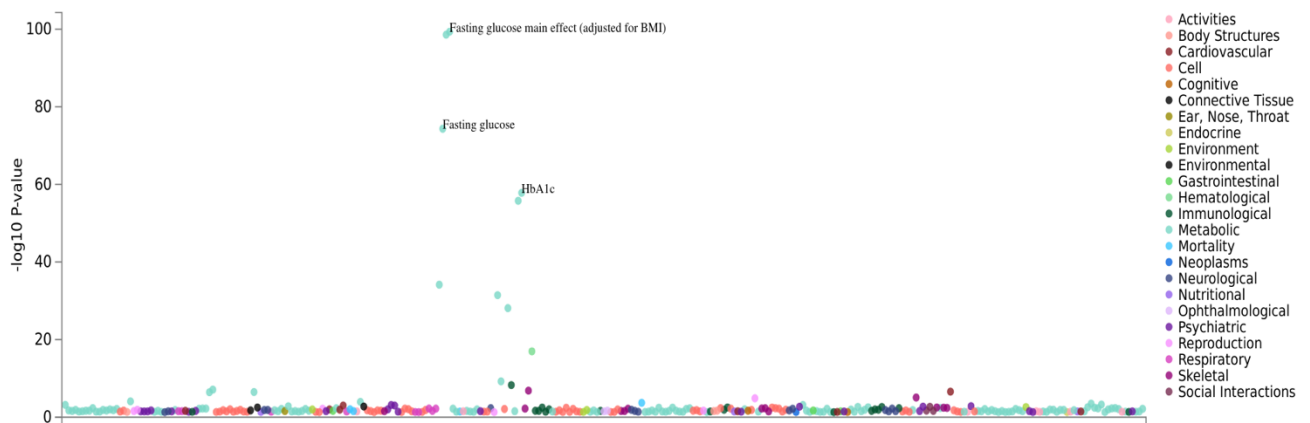


Supplementary Note Figure 1: A phenome-wide association study of rs6975024 using the GWAS atlas database

FVC: rs560887

The causal estimate of fasting glucose on FVC is no longer nominally significant, although similar in magnitude, when this IV is removed: $\beta_{IVW} = -0.084$, $SE = 0.049$, $P = 0.08$. Using its physical genomic location, this SNP is mapped to an intron of *G6PC2*, which encodes the Glucose-6-Phosphatase Catalytic Subunit 2 protein. This gene is a postulated negative regulator of basal glucose-stimulated insulin secretion and is involved in glucose-6-phosphate

hydrolysis^{5,6}. Data from the rs560887 PheWAS supports its biological saliency in glycaemic homeostasis as only glycaemic traits (fasting glucose, HbA1c etc.) were associated with this SNP, with only height the exception (SN Figure 2, Supplementary Table 12). Height could act as a confounder due to its relationship with lung function, however, this SNP has a plausible direct effect on blood glucose, and thus, it does not seem likely that height would mediate the effect of this SNP on lung function.



Supplementary Note Figure 2: A phenome-wide association study of rs560887 using the GWAS atlas database.

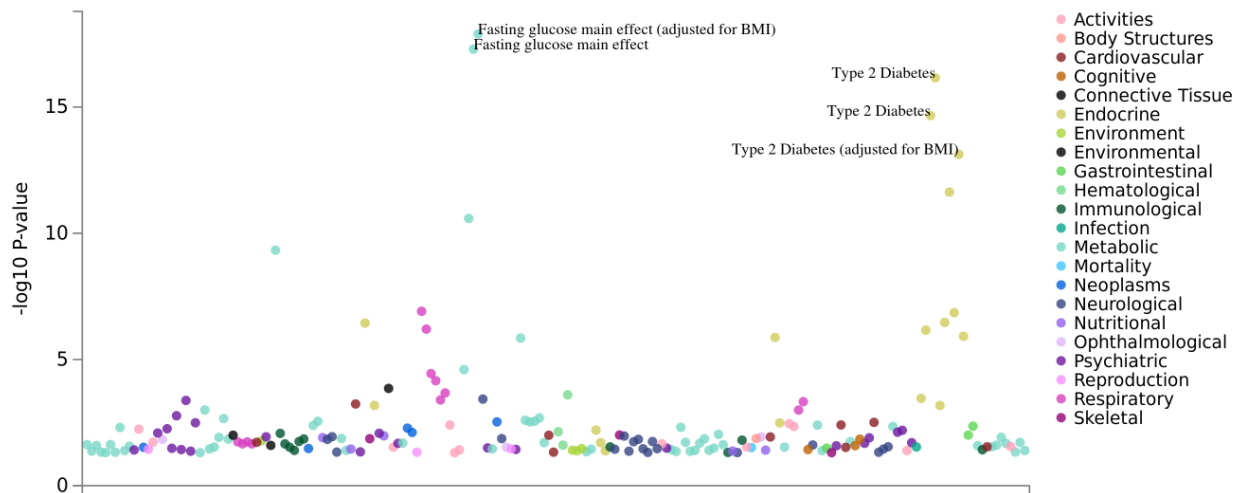
FEV₁ and FVC: rs17168486

The removal of rs17168486 resulted in a non-significant causal estimate of fasting glucose on FVC, although the effect size difference was not large: $\beta_{IVW} = -0.078$, $SE = 0.041$, $P = 0.06$.

In the FEV₁ model, the removal of this SNP retained a causal estimate with borderline statistical significance when rounding to three decimal places: $\beta_{IVW} = -0.082$, $SE = 0.041$, $P = 0.049$. This SNP lies within an intron of the gene which encodes the Diacylglycerol Kinase Beta protein (*DGKB*). Diacylglycerol has important functionality within pancreatic beta cells, with their cellular levels regulated by diacylglycerol kinases such as *DGKB*⁷. Dysregulation of diacylglycerol kinases is associated with downregulation of insulin secretion^{8,9}.

Furthermore, the PheWAS results supported its biological saliency with FEV₁ the only non-

glycaemic or diabetes trait associated using the phenome-wide significance threshold (SN Figure 3, Supplementary Table 8). The association with FEV₁ does not likely constitute a direct effect on lung function, which violate the assumptions of MR, rather it is logical its effect on FEV₁ is through fasting glucose.



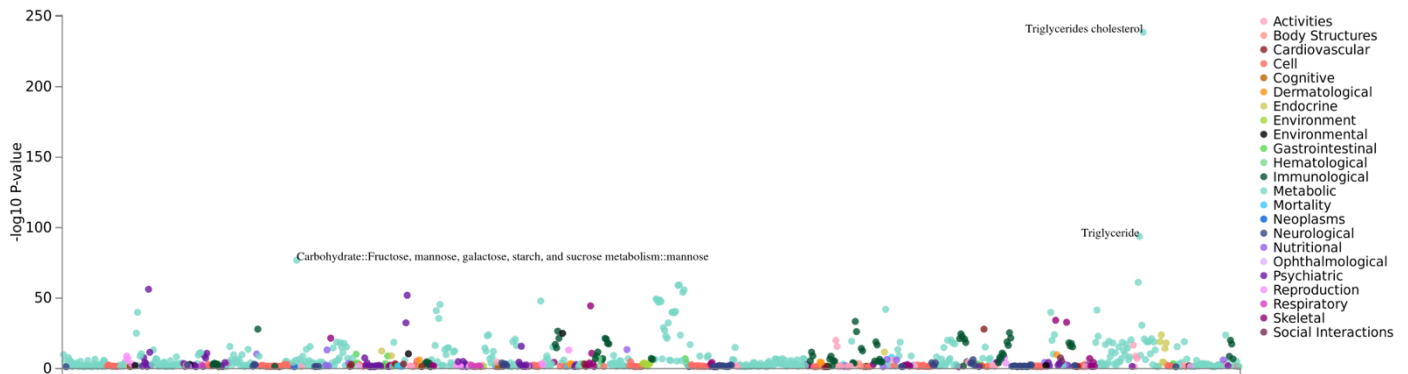
Supplementary Note Figure 3: A phenome-wide association study of rs17168486 using the GWAS atlas database.

FVC: rs1260326

The removal of this SNP produces a minor reduction in the IVW causal estimate, with the estimate borderline nominally significant after excluding rs1260326 - $\beta_{IVW} = -0.083$, SE = 0.043, $P = 0.052$. Using its physical position, this IV is a missense variant mapped to the glucokinase regulator (*GCKR*) gene. As its name suggests, this gene is the key regulator of glucokinase (of which the significance for glycaemic regulation was described above).

Interestingly, the PheWAS of this SNP was associated with a number of lipids, metabolites and haematological traits, beyond its association with glycaemic GWAS (SN Figure 4, Supplementary Table 11). This gene has a number of associations, although they are logically related to glycaemic dysregulation. As a result, we conclude that this IV likely acts through

glucose, for instance, glycolysis enhanced by *GCKR* is postulated to lead to hepatic *de novo* lipogenesis¹⁰.



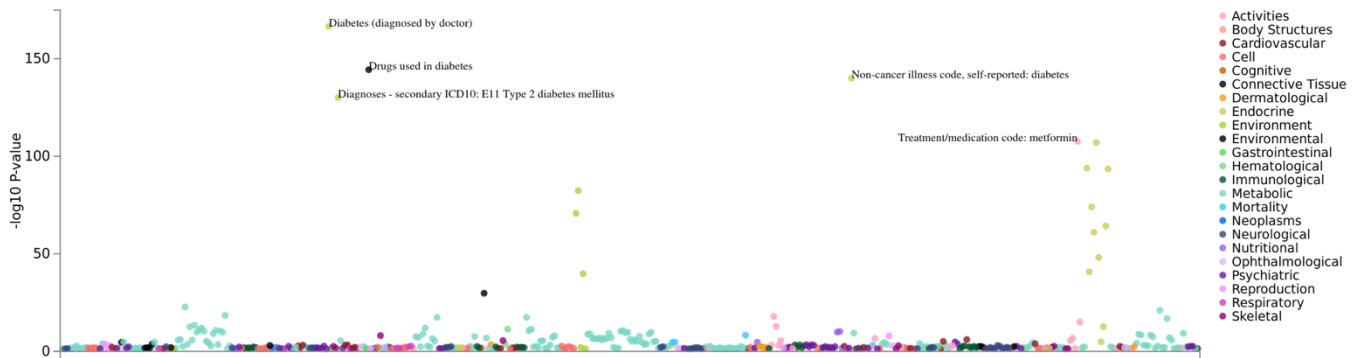
Supplementary Note Figure 4: A phenome-wide association study of rs1260326 using the GWAS atlas database.

FVC: rs7903146

Exclusion of rs7903146 resulted in a small decrease in the causal estimate of fasting glucose to FVC which was nominally no longer significant: $\beta_{IVW} = -0.081$, $SE = 0.042$, $P = 0.052$.

This SNP is located with the largest common variant effect size diabetes risk gene *TCF7L2*, which plays an important role in glucose homeostasis^{11,12}. The PheWAS of this SNP also demonstrated a number of anthropometric traits related to adiposity, for example, BMI, waist-to-hip ratio, and fat-free mass, associated with this SNP, although the relationship with diabetes and glycaemic traits was strong (SN Figure 5, Supplementary Table 9). It is possible that this SNP may also act through these confounding anthropometric phenotypes, particularly as *TCF7L2* is downstream of the pleiotropic WNT/ β -catenin pathways. However, given its characterised relationship with glycaemic function and the small effect of removing this IV on the IVW estimate, we do not believe that this provides significant evidence against

our conclusion of causal effect of fasting glucose on FVC. Further study is needed to refine the biological consequences of this IV SNP.



Supplementary Note Figure 5: A phenome-wide association study of rs7903146 using the GWAS atlas database.

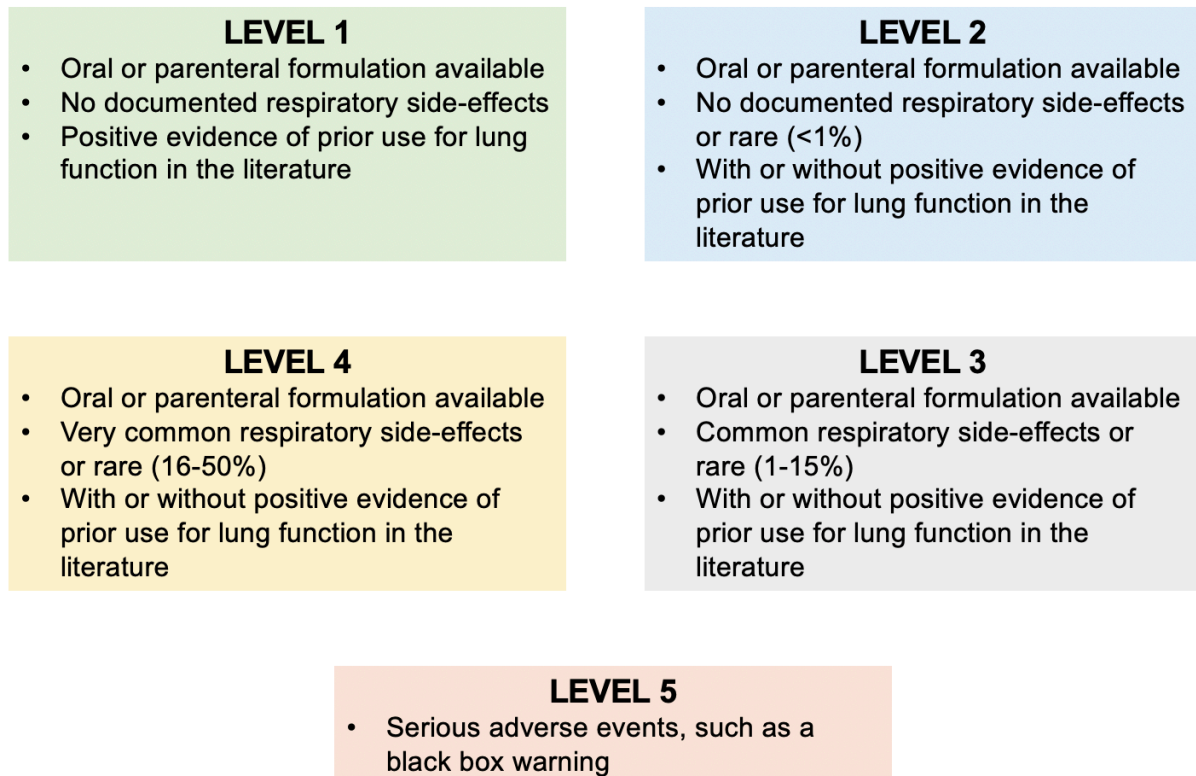
Supplementary Methods

Expert prioritisation of potential drug repurposing candidates from the *pharmagenic enrichment score* approach

A list of 60 FDA approved DrugBank compounds which were over-represented targets in a PES candidate gene-set was reviewed by a pharmacist to prioritise potential useful compounds for lung function. A total of 8 topical compounds were excluded. The remaining 55 oral and/or parenteral compounds were investigated for lung function related adverse events (including all of dyspnea, abnormal breath sounds, decreased respiratory rate, orthopnea, shallow breathing, respiratory distress, respiratory depression or any other related term), other alarming adverse events, important precautions, black-box warnings or any contraindication that might prohibit the drug use in our study population. These data were reviewed for each compound using the following databases: drugs.com, Medscape, SIDER v4.1 and the summaries of each product's characteristics. We also searched for articles that discussed either an improvement or worsening in the lung functions for each compound along with the allowed paediatric age use.

The drugs were then categorised into one of five categories (SM Figure 1, Supplementary Table 18). Level 1 was assigned for an oral or parenteral formulation, with no documented respiratory side effects, and with positive evidence of prior use for lung function in the literature. Level 2 was assigned for an oral or parenteral formulation, with no documented or rare (<1%) respiratory side effects, and with/without positive evidence but no negative evidence of prior use for lung function in the literature. Level 3 was assigned to an oral or parenteral formulation, with common (1-15%) respiratory side effects, and with/without positive evidence but no negative evidence of prior use for lung function in the literature. Level 4 compounds were those oral or parenteral formulations with very common (16-50%)

respiratory side effects or other alarming adverse effects unrelated to respiratory function, without positive evidence but with/without negative evidence of prior use for lung function in the literature. Finally, level 5 was assigned when the drug was associated with a serious adverse event (including a black-box warning or an absolute contraindication).

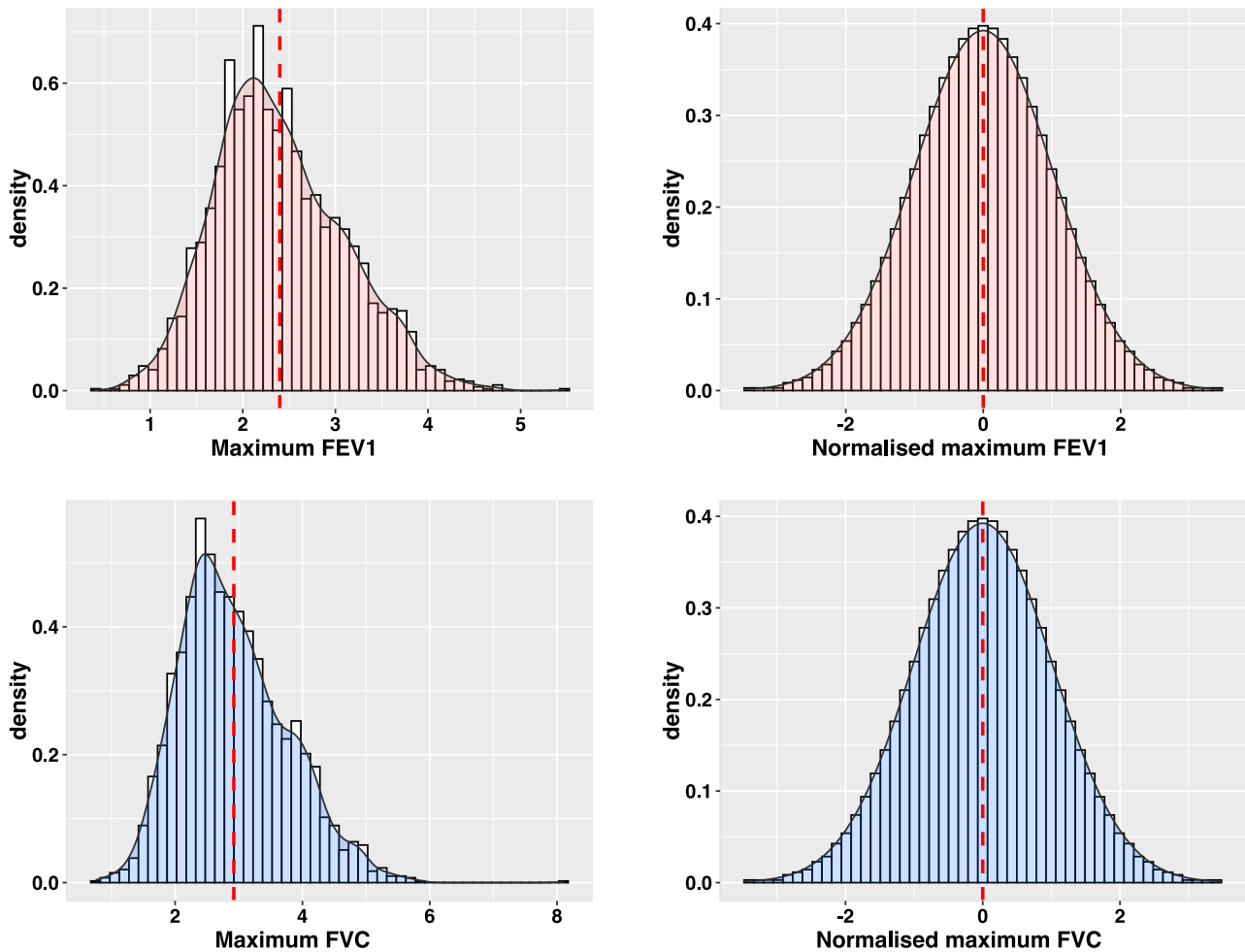


Supplementary Methods Figure 1: Schematic for prioritisation of drug repurposing candidates

Processing the phenotype and genotype data from the Hunter Community Study cohort

The Hunter Community Study (HCS) data were collected from community-dwelling individuals who resided in Newcastle, New South Wales, Australia (second largest city in the state of New South Wales on the east coast of the continent). Individuals were all aged between 55-85 at baseline and were recruited from the electoral roll using a modified Dillman recruiting strategy as described in detail by the original HCS publication¹³. There were 2480 individuals with a maximum (from four attempts) FEV₁ and FVC spirometry measure recorded who also had the following data fields present: baseline age, sex, smoking status ('ever vs never'), height, self-reported asthma, and self-reported bronchitis/emphysema. The sex composition of this HCS subset was approximately even (51.41% female), with a mean age of 66.07 (s.d. = 7.46). Considering the raw spirometry indices, the mean FEV₁ and FVC values were 2.41 (s.d. = 0.69) and 2.93 (s.d. = 0.84), respectively.

We constructed linear regression models for both pulmonary function measures covaried for sex, age, age², height, height², smoking status, self-reported asthma, and self-reported bronchitis/emphysema - both models were highly predictive (FEV₁: adjusted $R^2 = 0.58$, FVC: adjusted $R^2 = 0.61$). The residuals of these models were subjected to inverse-rank normalisation (Blom transformation), and thus, are in standard deviation units (SM Figure 2). The transformed residuals of these models represent greater than covariate predicted lung function (positive values), and less than covariate predicted lung function (negative values).

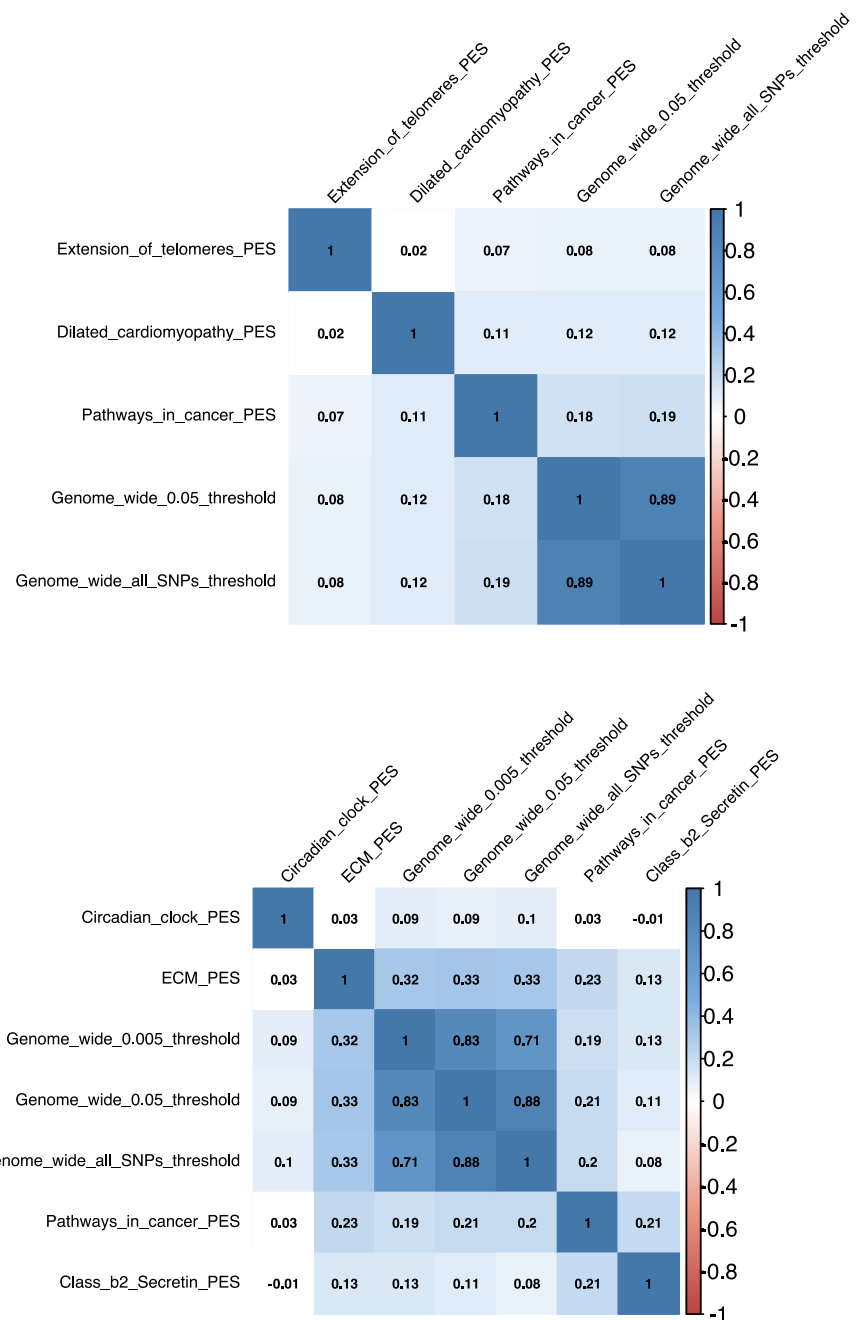


Supplementary Methods Figure 2: The distribution of spirometry measures in the HCS cohort. FEV₁ histograms with overlaid kernel density estimation plots are shaded red, whilst FVC are shaded blue. The red dotted line denotes the mean. The values plotted on the left represent the raw, untransformed spirometry indices, with the right plots of the normalised residuals of these variables.

Details regarding the genotyping of the HCS cohort were as follows. Participants provided blood samples from which DNA was extracted and genotyped using the Affymetrix Axiom Kaiser array. Quality control excluded SNPs with genotype call rate of < 0.95 , deviation from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$) or minor allele frequency (MAF) of < 0.01 . The input for relatedness testing and removal of population outliers were autosomal, common (MAF > 0.05), physically genotyped SNPs in relative linkage equilibrium ($r^2 < 0.02$), with

regions of long range LD removed, as is usual practice¹⁴. We used PLINK 1.9 to retain only unrelated individuals ($\hat{\pi} > 0.185$), with one participant from each related pair kept blinded to phenotype information. Population outliers were determined by performing principal component analysis (PCA) using PLINK 1.9. We clustered individuals in the HCS with the first two principal components from each 1000 genomes phase 3 superpopulation using k -means clustering. Thereafter, we conservatively excluded any HCS individual with a first or second principal component above or below the maximum or minimum 1000 genomes European values for these eigenvectors. PCA was repeated in the filtered European ancestry HCS subset, such that eigenvectors could be used as downstream covariates. Imputation to the Haplotype Reference Consortium panel involved a series of steps and additional data clean up, reference lift over to the hg19/GRCh37 and data submission to the Michigan imputation server, as specified in the submission guidelines^{15,16}. Post-imputation quality control was as follows: imputation $R^2 > 0.8$, MAF > 0.01 , and missingness < 0.02 .

Supplementary Figure 1



Supplementary Figure 1: Correlations between genome wide PGS and PES.

Each box denotes the correlation coefficient between one of the following, a) two PGS at different P -value thresholds, b) two PES, or c) a PGS and PES. Shaded boxes represent variables significantly correlated after the application of multiple testing correction. The top plot is for FEV₁, followed by FVC represented by the lower correlation plot.

REFERENCES

1. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology* **28**, 30–42 (2017).
2. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
3. Aiston, S., Trinh, K. Y., Lange, A. J., Newgard, C. B. & Agius, L. Glucose-6-phosphatase Overexpression Lowers Glucose 6-Phosphate and Inhibits Glycogen Synthesis and Glycolysis in Hepatocytes without Affecting Glucokinase Translocation: EVIDENCE AGAINST FEEDBACK INHIBITION OF GLUCOKINASE. *J. Biol. Chem.* **274**, 24559–24566 (1999).
4. Matschinsky, F. M. & Wilson, D. F. The Central Role of Glucokinase in Glucose Homeostasis: A Perspective 50 Years After Demonstrating the Presence of the Enzyme in Islets of Langerhans. *Front. Physiol.* **10**, 148 (2019).
5. Petrolonis, A. J. *et al.* Enzymatic characterization of the pancreatic islet-specific glucose-6-phosphatase-related protein (IGRP). *J. Biol. Chem.* **279**, 13976–13983 (2004).
6. Pound, L. D. *et al.* G6PC2: A Negative Regulator of Basal Glucose-Stimulated Insulin Secretion. *Diabetes* **62**, 1547–1556 (2013).
7. Kaneko, Y. K. & Ishikawa, T. Diacylglycerol Signaling Pathway in Pancreatic β -Cells: An Essential Role of Diacylglycerol Kinase in the Regulation of Insulin Secretion. *Biological & Pharmaceutical Bulletin* **38**, 669–673 (2015).
8. Kurohane Kaneko, Y. *et al.* Depression of Type I Diacylglycerol Kinases in Pancreatic β -Cells From Male Mice Results in Impaired Insulin Secretion. *Endocrinology* **154**, 4089–4098 (2013).

9. Briaud, I., Harmon, J. S., Kelpe, C. L., Segu, V. B. G. & Poitout, V. Lipotoxicity of the Pancreatic β -Cell Is Associated With Glucose-Dependent Esterification of Fatty Acids Into Neutral Lipids. *Diabetes* **50**, 315–321 (2001).
10. Santoro, N. *et al.* Hepatic De Novo Lipogenesis in Obese Youth Is Modulated by a Common Variant in the GCKR Gene. *J. Clin. Endocrinol. Metab.* **100**, E1125-1132 (2015).
11. Facchinello, N. *et al.* Tcf7l2 plays pleiotropic roles in the control of glucose homeostasis, pancreas morphology, vascularization and regeneration. *Sci Rep* **7**, 9605 (2017).
12. Zhou, Y. *et al.* TCF7L2 is a master regulator of insulin production and processing. *Hum. Mol. Genet.* **23**, 6419–6431 (2014).
13. McEvoy, M. *et al.* Cohort Profile: The Hunter Community Study. *International Journal of Epidemiology* **39**, 1452–1463 (2010).
14. Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135; author reply 135-139 (2008).
15. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443–1448 (2016).
16. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287 (2016).