Genetic architecture of 11 abdominal organ traits derived from abdominal MRI using deep learning

Yi Liu[1], Nicolas Basty[2], Brandon Whitcher[2], Jimmy D Bell[2], Elena Sorokin[1], Nick van Bruggen[1], E. Louise Thomas[2]*, Madeleine Cule[1]*

[1]Calico Life Sciences LLC, South San Francisco, California, USA

[2]Research Centre for Optimal Health, School of Life Sciences, University of Westminster, London, UK

* These authors contributed equally to this work

Abstract

Cardiometabolic diseases are an increasing global health burden. While well established socioeconomic, environmental, behavioural, and genetic risk factors have been identified, our understanding of the drivers and mechanisms underlying these complex diseases remains incomplete. A better understanding is required to develop more effective therapeutic interventions. Magnetic resonance imaging (MRI) has been used to assess organ health in a number of studies, but large-scale population-based studies are still in their infancy. Using 38,683 abdominal MRI scans in the UK Biobank, we used deep learning to systematically quantify parameters from individual organs (liver, pancreas, spleen, kidneys, lungs and adipose depots), and demonstrate that image derived phenotypes (volume, fat and iron content) reflect organ health and disease. We show that these traits have a substantial heritable component (8%-44%), and identify 93 independent genome-wide significant associations, including 3 associations with liver fat and one with liver iron that have not previously been reported, and 73 in traits that have not previously been studied. Overall our work demonstrates the utility of deep learning to systematically quantify health parameters from high-throughput MRI across a range of organs and tissues of the abdomen, and to generate new insights into the genetic architecture of complex traits.

Introduction

The UK Biobank (UKBB) project has gathered lifestyle information, biometric, and genetic data for 500,000 individuals aged 40 to 69 years in the UK population with the goal of advancing our understanding of health and disease (Sudlow et al., 2015). It was expanded to include imaging for a 100,000-person sub-cohort, creating the largest and most extensive collection of structural and functional medical imaging data in the world (Littlejohns et al., 2020).

MRI has become the gold standard for clinical research, including body composition, with measurements of visceral adipose tissue (VAT), liver and pancreatic fat content having an enormous impact on our understanding of conditions such as type-2 diabetes (T2D) and nonalcoholic fatty liver disease (NAFLD) (Thomas et al., 2013). The MRI protocol in the UKBB includes multiple tissues and organs with the potential for a wide variety of clinically-relevant variables. However, genetic studies utilising the UKBB MRI-derived features have focused mainly on brain and cardiac traits (Elliott et al., 2018; Miller et al., 2016; Pirruccello et al., 2020), with some limited studies focussed on liver iron (n=8,289) and MRI-based corrected T1 (n=14,440) (Parisinos et al., 2020; Wilman et al., 2019). Thus, the full potential of the UKBB abdominal MRI data has not been realised, in part due to the lack of suitable automated methods to extract the variety and depth of relevant features from multiple organs in very large cohorts.

To address this issue, we trained models using deep learning on expert manual annotations, following preprocessing and quality control, to automatically segment key organs from the UKBB MRI data from 38,683 subjects (Table 1 and Methods). Additionally, we quantified fat and iron content where suitable acquisitions were available (Figure 1, Supplementary Table 1, and Methods). In total, we defined 11 Image Derived Phenotypes (IDPs): volume of the liver, pancreas, kidneys, spleen, lungs, VAT, and abdominal subcutaneous adipose tissue (ASAT), and fat and iron content of the liver and pancreas. By linking these traits to measures of risk factors, genetic variation, and disease outcomes, we are able to better characterise their role in disease risk.

Table 1: Study population characteristics.

	UK Biobank cohort (at time of baseline visit)	Imaging cohort (at time of imaging visit)	GWAS cohort (White British Ancestry and passing QC)			
			Organ volume (DIXON)	Pancreas volume	Pancreas Fat and Iron	Liver Fat and Iron
Number of participants	502520	38881	32860	31758	25617	32858
% Female	54.4	51.8	51.5	51.40%	51.2	51.5
Age	56.5 (8.1)	63.7 (7.56)	63.9 (7.52)	63.8 (7.52)	64.2 (7.48)	63.9 (7.52)

BMI (kg/m^2)	27.4 (4.8)	26.5 (4.39)	26.5 (4.37)	26.5 (4.34)	26.5 (4.31)	26.5 (4.36)
Height (cm)	168 (9.28)	169 (9.3)	169 (9.26)	169 (9.25)	169 (9.26)	169 (9.26)
% White British Ancestry	81.5	81.5	100	100	100	100

Results

Characterisation of IDPs in the UK Biobank population

Previous studies have derived measures of VAT and ASAT, liver fat and iron in the UK Biobank from a subset of the scanned participants (McKay et al., 2018; West et al., 2016; Wilman et al., 2017). Our IDPs show a correlation of 0.87 (liver iron) to 1.0 (fat volume) (Methods; Supplementary Figure 1). The distribution of each organ-specific measure in the scanned population is summarised in Figure 1E, F, and G and Table 2.

Table 2: Mean and standard deviations for 11 IDPs in our study, and number of independent GWAS associations found at study-wide significance (p<4.54e-9; see Methods).

Trait	Organ	Combined	Female	Male	# study-wide significant GWAS hits
Volume (L)	VAT	3.92 (2.3)	2.78 (1.6)	5.14 (2.3)	3
	ASAT	8.16 (4.1)	9.57 (4.3)	6.64 (3.2)	1
	Lungs	2.67 (0.73)	2.32 (0.53)	3.03 (0.75)	5
	Spleen	0.17 (0.072)	0.14 (0.054)	0.2 (0.078)	29
	Kidney	0.14 (0.03)	0.12 (0.023)	0.16 (0.028)	9
	Pancreas	0.06 (0.018)	0.06 (0.016)	0.06 (0.019)	11
	Liver	1.38 (0.3)	1.28 (0.25)	1.49 (0.3)	11
Fat (%)	Pancreas	10.41 (7.9)	8.34 (6.7)	12.6 (8.5)	8
	Liver	5.06 (5)	4.43 (4.7)	5.73 (5.2)	11
Iron (mg/g)	Pancreas	0.77 (0.097)	0.8 (0.1)	0.75 (0.084)	0
	Liver	1.22 (0.26)	1.2 (0.24)	1.24 (0.28)	6*

*Due to complex LD structure in this region, we were not able to finemap the HFE locus. We count two signals at this locus (rs1800562 and rs1799945).

All IDPs, except liver fat, showed a statistically significant association with age after adjusting for imaging centre and date (Figure 1B), although the magnitudes of the changes are generally small (e.g. -8.8ml or -0.03s.d./year for liver volume, -27.7ml or -0.0067s.d./year for ASAT, and 24.3ml or 0.011s.d./year for VAT). Liver, pancreas, kidney, spleen, and ASAT volumes decreased, while VAT and lung volumes increased with age. Liver and pancreatic iron and pancreatic fat increase slightly with age. Several IDPs (volumes of liver, kidney, lung, and pancreas, as well as liver fat and iron) showed statistically significant evidence of heterogeneity in age-related changes between men and women. We found excess liver iron (>1.8mg/g) in 3.22% of men and 1.75% of women.

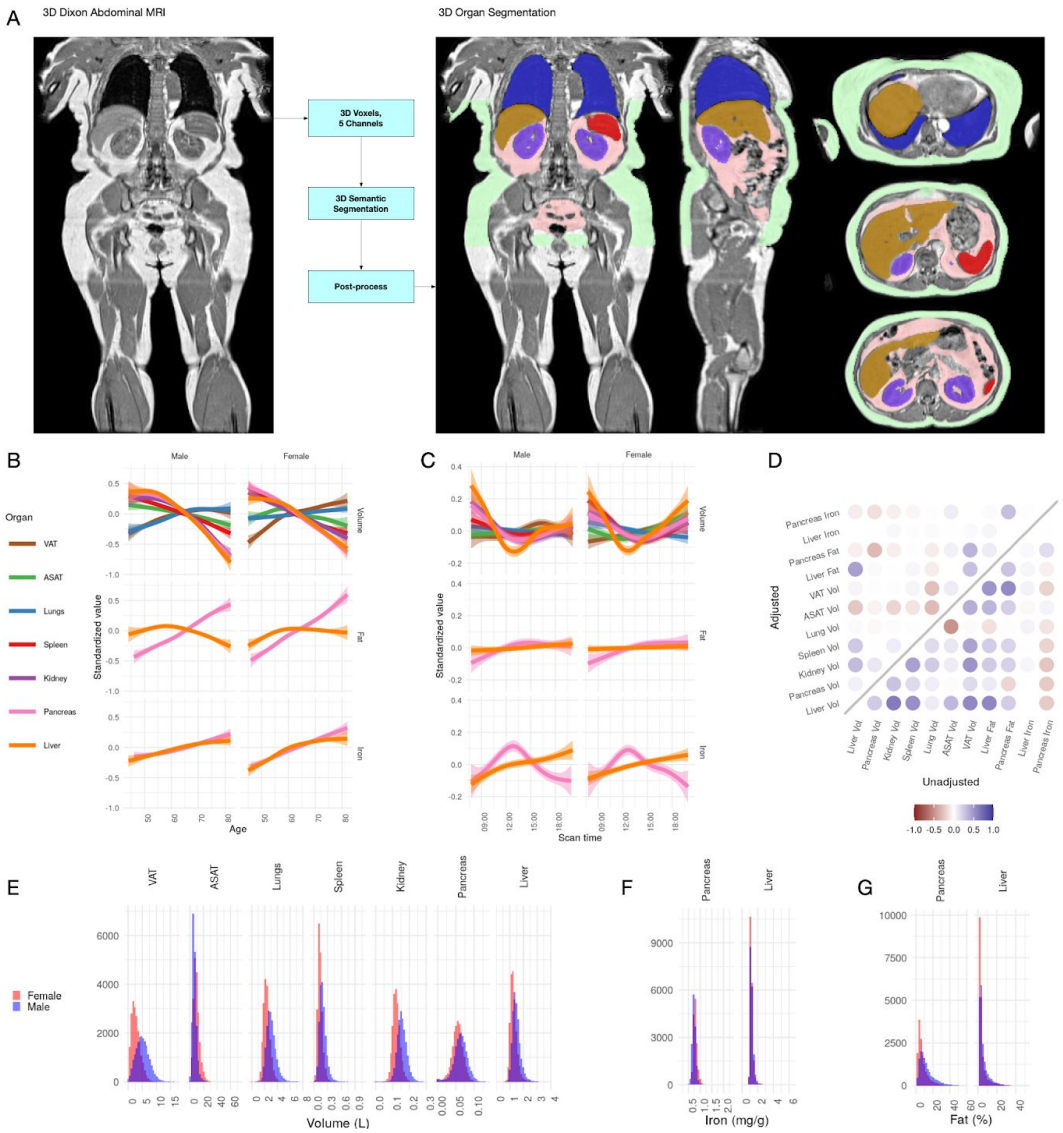

Figure 1. A: Example Dixon image before and after automated segmentation of ASAT, VAT, liver, lungs, left and right kidneys, and spleen. B: Relationship between IDPs and age and sex within the UKBB. Each trait is standardised within sex, so that the y axis represents standard deviations, after adjustment for imaging centre and date. The trend is smoothed using a generalised additive model with smoothing splines for visualisation purposes. C: Relationship between IDPs and scan time and sex within the UKBB. Each trait is standardised within sex, so that the y axis represents standard deviations, after adjustment for imaging centre and date. The trend is smoothed using a generalised additive model with smoothing splines for visualisation purposes. D: Correlation between IDPs. Lower right triangle: Unadjusted correlation (except for imaging centre and date). Upper left triangle: Correlation after adjustment for age, sex, height, and BMI. E: Histograms showing the distribution of the eleven IDPs in this study.

To explore diurnal variation, we investigated correlation between the imaging timestamp and IDPs. We find a decrease in liver volume during the day, with volume at 12noon being on average 112ml smaller than volume at 8am, and a return to almost the original volume by 8pm. This has previously been suggested in small ultrasound studies (n=8) which indicated that liver

volume is at its smallest between 12 and 2pm, attributed to changes in hydration and glycogen content (Leung et al., 1986). We also observe smaller, but still statistically significant, associations between time of day and liver and pancreas iron, as well as ASAT, VAT, kidney, and lung volume. Although these changes appear to be physiological in nature, we are currently unable to rule out other potential sources of confounding, however unlikely (for example, different groups of participants being more likely to attend the scanning appointment at different times of day).

IDPs are associated with organ-specific disease outcomes

To assess which IDPs are associated with health-related outcomes, we defined a set of diseases based on inpatient hospital episode statistics (Methods), and assessed the association between each IDP and disease diagnoses (Figure 2). Although we were not able to evaluate cause and effect, we found evidence that IDPs reflect organ function and health from the association with disease outcomes.

Liver volume was significantly associated with chronic liver disease and cirrhosis (p=4.5e-06, beta=0.389) as well as T2D (p=1.3e-92, beta=0.73) and hypertension (p=3.9e-17, beta=0.18). Kidney volume was associated with chronic kidney disease (CKD) (p=8.0e-23, beta=-1.0). Interestingly, pancreas volume was associated more strongly with Type 1 diabetes (T1D) (p=4.9e21, beta=-0.77), than T2D (p=1.1e-17, beta=-0.27), while pancreatic fat showed a small association with T2D (beta=0.181, p=1.16e-07) and not with T1D (p=0.241). Lung volume was most strongly associated with tobacco use (p=1.8e-46, beta=0.50) and disorders relating to chronic airway obstruction (COPD) (p=3.6e-35, beta=0.61), with larger lung volume corresponding to a greater likelihood of respiratory disease diagnosis. Spleen volume was associated with myeloproliferative disease (p=2.2e-33, beta=0.74), especially chronic lymphocytic leukaemia (p=9.9e-24, beta=0.78). Liver fat was associated with T2D (p=1.4e-34, beta=0.29). Liver iron was associated with T2D (p=3.1e-19, beta=-0.43) and iron deficiency anaemia (p=5.3e-12, beta=-0.44)

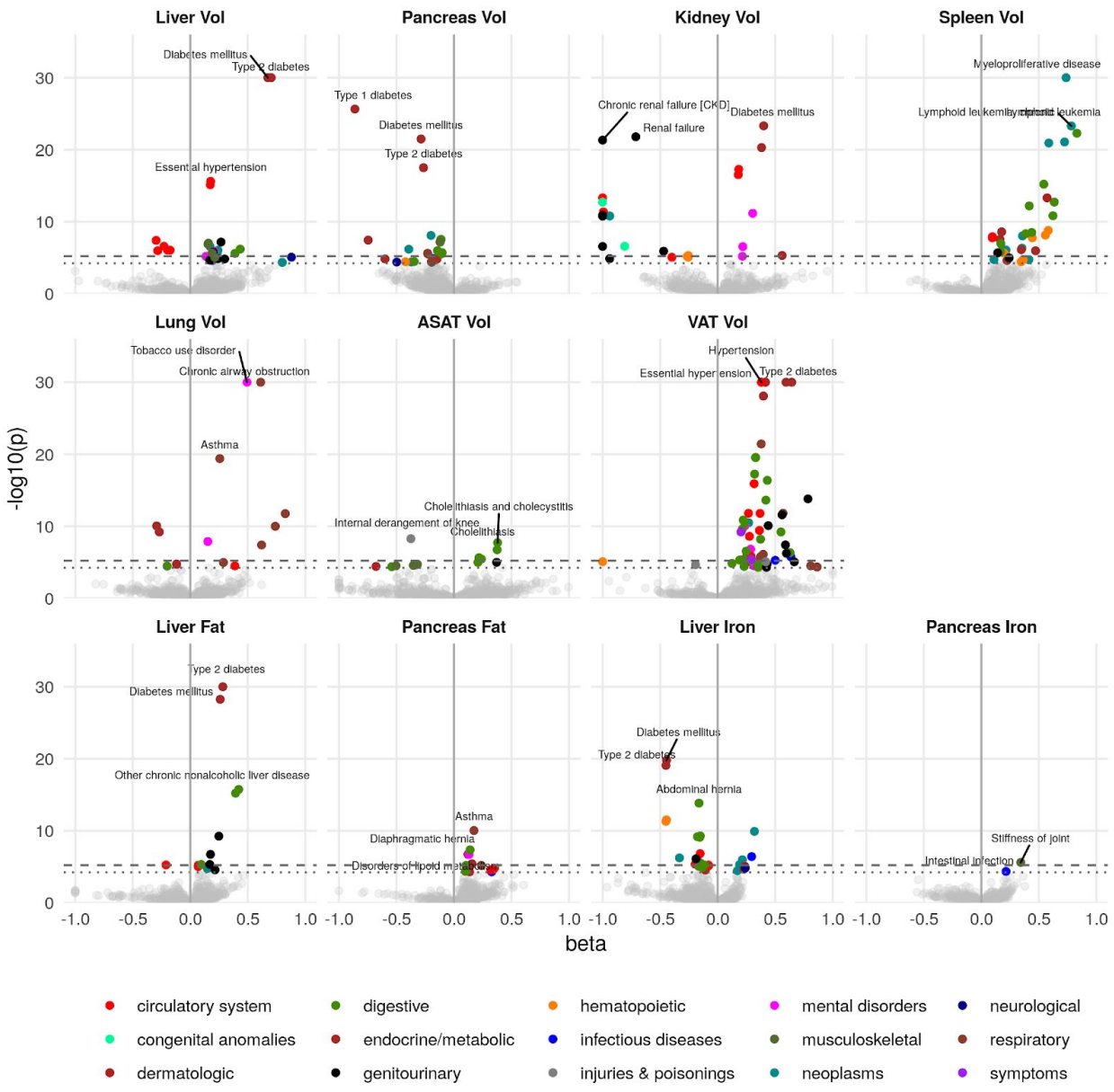

Figure 2: Disease phenome-wide association study across all IDPs and 754 disease codes (PheCodes). The x-axis gives the effect size per standard deviation, and the y-axis -log10(p-value). The top 3 associations for each phenotype are labelled. Horizontal lines at disease phenome-wide significance (dotted line, p=6.63e-05) and study-wide significance (dashed line, p=6.03e-06) after Bonferroni correction. Note that the PheCodes are not exclusive and have a hierarchical structure (for example, T1D and T2D are subtypes of Diabetes), so some diseases appear more than once in these plots.

VAT was associated with a wide range of cardiometabolic outcomes including hypertension (p=1e-49, beta=0.39), T2D (p=8.1e-44, beta=0.69), and lipid metabolism disorders (p=1.9e-33, beta=0.42), while ASAT was only associated with cholelithiasis and cholecystitis (p=1.3e-08, beta=0.38). This association remained statistically significant, after adjusting for VAT, counter to

reports that only VAT is predictive of gallstones (Radmard et al., 2015). Overall, this supports the key role of VAT and liver fat in the development of metabolic syndrome.

IDPs are associated with organ-specific biomarkers, physiological measures, and behavioural traits

To further explore the extent to which our IDPs reflect organ health, we assessed correlation between the IDPs and 87 biomarkers from blood, serum, and urine, chosen to reflect a range of health conditions (Methods, Supplementary Figure 3). We also investigated associations between IDPs and 352 lifestyle and exposure factors, 844 self-reported medical history factors, 500 physical and anthropometric measures, and 769 self-reported diet and exercise measures (Supplementary Figures s 4-7).

Across multiple abdominal organs, we observed strong correlations between IDPs and biomarkers reflective of organ function. For example, liver volume was associated with triglycerides (p=1.19e-242, beta=0.247) and sex hormone binding globulin (SHBG) (p=3.43e-210, beta=-0.216). Kidney volume was associated with serum cystatin C (p<1e-300, beta=-0.534), serum creatinine (p<1e-300, beta=-0.48), consistent with observations that smaller kidneys function less effectively (Jovanović et al., 2013). Pancreas volume was associated with glycated haemoglobin (HbA1c) (p=8.49e-28, beta=-0.0601), but the association with glucose was not statistically significant after Bonferroni correction (p=8.13e-05). Spleen volume was associated with multiple hematological measurements, including reticulocyte count (p<1e-300, beta=0.25), mean sphered cell volume (p<1e-300, beta=-0.323) and platelet distribution width (p<1e-300, beta=0.277).

Liver fat was associated with multiple liver function biomarkers including triglycerides (p=7.66e-219, beta-0.177), SHBG (p=4.75e-189, beta=-0.156) alanine aminotransferase (p<1e-300, beta=0.226, and gamma glutamyltransferase (p=1.63e-194, beta=0.162). Consistent with disease outcomes, which showed a correlation between hepatic iron, but not pancreatic iron, with iron deficiency anaemia, liver iron levels were correlated with measures of iron in the blood (e.g. mean corpuscular haemoglobin (MCH), p=1.71e-240, beta=0.174), while pancreatic iron did not show any such association (MCH p=0.218).

Consistent with previous reports (Harrison-Findik, 2007), we found that liver iron was associated with lower alcohol consumption (p=3e-116, beta=-0.247) and higher intake of red meat (beef intake p=1.61e-61, beta=0.168; lamb/mutton intake p=7.13e-56, beta=0.165). Liver iron was also associated with suppressed T2* derived from neuroimaging in the same UKBB cohort (Elliott et al., 2018), particularly in the putamen (left: p=1.53e-68, beta=-0.138; right: p=1.01e-69, beta=-0.14). There were no such associations for pancreatic iron (left p=0.223;

right p=0.194). Additionally, we found that liver fat was associated with lower birth weight (p=1.76e-30, beta=-0.0849) and comparative body size at age 10 (p=4.79e-76, beta=-0.22). Low birth weight has previously been associated with severity of pediatric non-alcoholic steatohepatitis (NASH) (Bugianesi et al., 2017), abnormal fat distribution (Parkinson et al., 2020) and liver fat levels in adults born prematurely (Thomas et al., 2011).

We found strong associations between increased lung volume and smoking status, tobacco smoking, COPD and lung disorders, wheeze, diagnosis of asthma and treatment for asthma, a decreased lung capacity as well as forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV1)/FVC ratio (Supplementary Figure 7). This is perhaps surprising in light of the age-related decreases in FEV1 and FVC, however it has been shown that lung volume increases with both age and as a consequence of obstructive pulmonary diseases(Lutfi, 2017). Although lung volume estimated via MRI is not a widely used clinical measure, our data suggests it may be a biomarker of ageing-related respiratory complications.

Genetic architecture of abdominal IDPs

To explore the genetic architecture of the IDPs, we performed a genome-wide association study (GWAS) for each IDP of 9 million single-nucleotide polymorphisms (SNPs) in the approximately 30,000 individuals of white British ancestry (Bycroft et al., 2018) (Methods). We verified that the test statistics showed no overall inflation compared to the expectation by examining the intercept of linkage disequilibrium (LD) score regression (LDSC) (Bulik-Sullivan, Loh, et al., 2015) (Supplementary Table 5). The number of individuals included in the analysis for each IDP is given in Table 1, together with the number of study-wide significant independent signals for each IDP. Utilizing a generalized linear mixed model framework and SKAT-O test implemented in SAIGE-GENE (Zhou et al., 2020), we performed gene-based exome-wide association studies in the 11,134 participants with IDP and exome sequencing data. Test statistics were well calibrated and we found no study-wide significant associations (Supplementary Figure 12).

Organ volume, fat, and iron are heritable

For each IDP, we estimated SNP-heritability using the BOLT-REML model (Loh, Bhatia, et al., 2015) (Methods). All IDPs showed a significant heritable component, indicating that genetic variation contributes substantially to the variation between individuals (Figure 3A). Heritability is largely unaffected by the inclusion of height and BMI as additional covariates, indicating that it is not a function of overall body size.

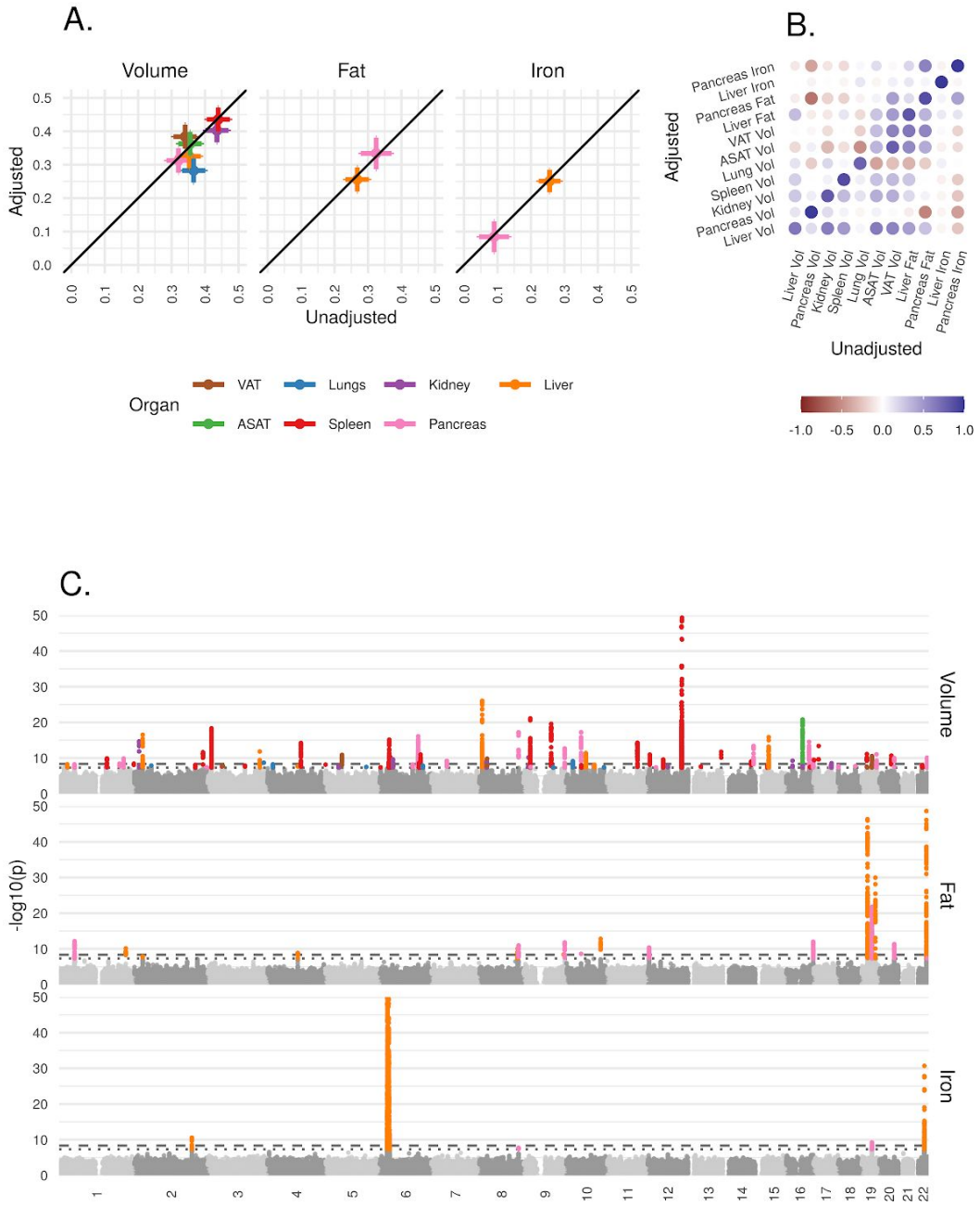

Figure 3. Genetic architecture of all IDPs. A: Heritability (point estimate and 95% confidence interval) for each IDP estimated using the BOLT-REML model. Y-axis: Adjusted for height and BMI. X-axis: Not adjusted for height and BMI. The three panels show volumes, fat, and iron respectively. B: Genetic correlation between IDPs estimated using bivariate LD score regression. The size of the points is given by -log10(p), where p is the p-value of the genetic correlation between the traits. Upper left triangle: Adjusted for height and BMI. Lower right triangle: Not adjusted for height and BMI. C: Manhattan plots showing genome-wide signals for all IDPs for volume (top panel), fat (middle panel), and iron concentration (lower panel). Horizontal lines at 5e-8 (blue dashed line, genome-wide significant association for a single trait) and 4.5e-9 (red dashed line, study-wide significant association). P-values are capped at 10e-50 for ease of display.

Genetic correlation between abdominal IDPs

To understand the extent to which genetic variation explains the correlation between traits, we used bivariate LD score regression (Bulik-Sullivan, Finucane, et al., 2015) to estimate the genetic correlation between all 55 IDP pairs, with and without including height and BMI as covariates (Methods). After Bonferroni correction, we found a statistically significant non-zero genetic correlation between 22 of the 55 unadjusted IDP-pairs traits (Figure 3B and Supplementary Table 6), the strongest (r_g=0.782, p=4.60e-137) between ASAT and VAT. There was substantial genetic correlation between VAT and liver fat (r_g==0.58, p=3.7e-38) and between VAT and pancreas fat (r_g=0.569, p=2.79E-16). We found a negative genetic correlation between pancreas volume and fat (r_g=-0.45, p=2.1e-06), and between pancreas volume and iron (r_g=-0.5, p=5.2e-05)

IDPs share a genetic basis with other physiological traits

To identify traits with a shared genetic basis, we estimated genetic correlation between IDPs and 282 complex traits with a heritable component (Methods). 650 IDP-trait pairs showed evidence of nonzero genetic correlation; 347 of these involved with measures of size or body composition (Supplementary Table 7 and Supplementary Figure 11). We found substantial genetic correlation between ASAT volume and other measures of body fat, such as whole-body fat mass (r_g=0.94, p=3.2e-143) and between VAT and conventional surrogate markers such as waist circumference (r_g=0.75, p=1.6e-109). The strongest genetic correlation with lung volume was with FVC (r_g=0.7, p=3.1e-71), with FEV and height also significant. We also found more modest genetic correlation between organ volumes and biochemical measures, such as liver fat and ALT (r_g=0.5, p=4.5e-23), kidney volume and serum creatinine (r_g=-0.4, p=3.9e-22), and liver iron and erythrocyte distribution width (r_g=-0.33, p=2.1e-14).

Heritability is enriched in organ-specific cell types

In order to identify tissues or cell types contributing to the heritability of each trait, we used stratified LD score regression (Finucane et al., 2015) (Methods). Liver fat showed evidence for enrichment in hepatocytes (p=4.20e-6) and liver tissue (p=2.2e-5), and pancreatic fat showed evidence for enrichment in pancreas tissue (smallest p=9.74e-5). Spleen volume showed enrichment in spleen cells (p=7.39e-10) and immune cell types including T cells, B cells, and natural killer cells, and neutrophils. VAT, ASAT, and lung volumes did not show evidence of significant heritability enrichment in any tissue or cell types (Supplementary Figures 8-10).

Genome-wide significant associations

For each locus containing at least one variant exceeding the study-wide significance threshold, we used GCTA COJO (Yang et al., 2012) to identify likely independent signals, and map likely causal variants (Methods, Supplementary Table 8). To better understand the biology of each signal, we explored traits likely to share the same underlying signal (colocalised signals) among 973 traits and 356 diseases measured in UKBB (Methods, Supplementary Table 9), and gene expression in 49 tissues (Methods, Supplementary Table 10).

Liver IDPs recapitulate known biology and point to new genes of interest

The strongest association with liver volume (lead SNP rs4240624, p=2.1e-34, beta=-0.15), lies on chromosome 8, 175kb from the nearest protein-coding gene, *PPP1R3B*. *PPP1R3B* is expressed in liver and skeletal muscle, and promotes hepatic glycogen biosynthesis (Mehta et al., 2017). Although this variant has been associated with attenuated signal on hepatic computed tomography (Stender et al., 2018); in our study it was not associated with liver fat (p=0.007) or iron (p=0.001).

We also detected an association between liver volume and a missense SNPs in *GCKR* (rs1260326, p=5.4e-19, beta=-0.061). This signal colocalised with T2D, hypercholesterolemia and hyperlipidemia, gout and gallstones, as well as other lipid and cardiovascular traits in the UKBB. This locus has previously been associated with NAFLD (Kawaguchi et al., 2018) as well as multiple metabolic traits including triglycerides, lipids, and C-reactive protein (Wojcik et al., 2019).

Of the eight study-wide independent signals associated with liver fat, three (rs58542926 in *TM6SF2* rs429358 in *APOE*; and rs738409 in *PNPLA3*) have previously been associated with NAFLD (Kozlitina et al., 2014; Romeo et al., 2008; Speliotes et al., 2011), and were also reported in a GWAS of liver fat in a subset of this cohort (Parisinos et al., 2020). The fourth SNP identified in that study, rs1260326 in *GCKR*, did not reach our stringent threshold of study-wide significance threshold (p=1.9e-8, beta=-0.044).

Two of the remaining five signals have previously been linked to liver disorders or lipid traits, although not specifically to liver fat. A signal near *TRIB1* (lead SNP rs112875651) colocalises with hyperlipidemia and atherosclerosis and has been linked to lipid levels in previous studies, and SNPs in this gene have an established role in the development of NAFLD (Liu et al., 2019). A missense SNP in *TM6SF2* (lead SNP rs188247550) is also associated with hyperlipidemia and has previously been linked to alcohol-induced cirrhosis (Buch et al., 2015).

Three further signals have not previously been associated with any liver traits, although some have been associated with other metabolic phenotypes. On chromosome 1, a SNP intronic to

MARC1 (lead SNP rs2642438) colocalises with cholesterol, LDL-cholesterol, and HDL-cholesterol levels, with the risk allele for higher fat associated with higher LDL-cholesterol. While this variant has not previously been associated with liver fat, missense and protein truncating variants in *MARC1* have been associated with protection from all-cause cirrhosis, and also associated with liver fat and circulating lipids (Emdin et al., 2020).

We found an association between intronic and *GPAM*, which encodes an enzyme responsible for catalysis in phospholipid biosynthesis (lead SNP rs11446981). This signal colocalises aspartate aminotransferase (AST), and HDL cholesterol levels in serum. *GPAM* knockout mice have reduced adiposity and its inhibition reduces food intake and increases insulin sensitivity in diet-induced obesity (Kuhajda et al., 2011). Our data suggests that this enzyme may play a role in the liver fat accumulation in humans.

A region overlapping to *MTTP* with 67 variants in the 95% credible set was associated with liver fat. Candidate gene studies have linked missense mutations in *MTTP* to NAFLD (Hsiao et al., 2015). Rare nonsense mutations in this gene cause abetalipoproteinemia, an inability to absorb and knockout studies in mice recapitulate this phenotype (Partin et al., 1974; Raabe et al., 1998). Inhibition of MTTP is a treatment for familial hypercholesterolemia and is associated with increased liver fat (Cuchel et al., 2007).

We replicate previously reported associations with liver iron at *HFE* (rs1800562 and rs1799945) and *TMPRSS6* (Wilman et al., 2019), although we were unable to accurately finemap at this locus. We found evidence for two independent additional signals on chromosome 2 between *ASND1* and *SLC40A1* (lead SNP rs7577758; conditional lead SNP rs115380467). *SLC40A1* encodes ferroportin, a protein essential for iron homeostasis (Donovan et al., 2005) that enables absorption of dietary iron into the bloodstream. Mutations in *SLC40A1* are associated with a form of hemochromatosis known as African Iron Overload (Mayr et al., 2011). This finding is consistent with a recent study which highlighted the role of hepcidin as a major regulator of hepatic iron storage (Wilman et al., 2019).

Novel associations with pancreas IDPs

We identified 11 study-wide significant associations with pancreatic volume. None were coding or colocalised with the expression of protein-coding genes. Two signals (rs72802342, nearest gene *CTRB2*; rs744103, nearest gene *ABO*) colocalised with diabetic-related traits. This is consistent with our findings that T1D was associated with smaller pancreatic volume.

We identified seven study-wide significant independent associations with pancreatic fat, with little overlap with liver-specific fat loci. Surprisingly, we found little evidence that loci associated with pancreatic fat were associated with other metabolic diseases or traits, suggesting that it

may have a more limited direct role in the development of T2D than previously suggested (R. Taylor, 2008).

The top association for pancreatic fat (lead SNP rs10422861) was intronic to *PEPD*, and colocalised with a signal for body and trunk fat percentage, leukocyte count, HDL-cholesterol, SHBG, total protein and triglycerides. *PEPD* codes for prolidase, an enzyme that degrades iminopeptides in which a proline or hydroxyproline lies at the C-terminus, with a special role in collagen metabolism (Kitchener & Grunden, 2012). There was an association at the *ABO* locus (lead SNP rs8176685) for pancreatic fat; rs507666, which tags the A1 allele, lies in the 95% credible set at this locus. This signal colocalises with lipid and cardiovascular traits and outcomes, and is consistent with previous reports that blood group A is associated with lipid levels, cardiovascular outcomes (H. Zhang et al., 2012) and increased risk of pancreatic cancer (B.-L. Zhang et al., 2014).

An association with pancreatic fat (lead SNP rs7405380) colocalises with the expression of *CBFA2T3* in the pancreas. rs7405380 lies in a promoter flanking region which is active in pancreatic tissue (ensemble regulatory region ENSR00000546057). *CBFA2T3* belongs to a family of ubiquitously expressed transcriptional repressors, highly expressed in the pancreas, about which little is known. A recent study identified Cbfa2t3 as a target of Hes1, which plays a critical role in regulating pancreatic development (de Lichtenberg et al., 2018). This SNP was not associated with any metabolic phenotypes.

We identified signals at a locus on chromosome 1 containing *FAF1* and *CDKN2C* (lead SNP rs775103516), and five other loci. In contrast to liver iron, where we identified strong signals at regions associated with ferroportin and hepcidin loci, we found no study-wide significant associations with pancreatic iron.

Novel associations with other organ volume IDPs

A locus on chromosome 2 was associated with average kidney volume. This signal colocalises with biomarkers of kidney function (cystatin C, creatinine, urate, and urea) and a SNP in the 95% credible set, rs807624, has previously been reported as associated with Wilms tumor (Turnbull et al., 2012), a pediatric kidney cancer rarely seen in patients over the age of five. However, this association raises the possibility that this locus plays a broader role in kidney structure and function in an adult population and warrants further study.

We also found a significant association at the *PDILT*/*UMOD* locus (lead SNP rs77924615), that colocalises with hypertension, cystatin C, creatine, and kidney and urinary calculus in the UKBB. This locus has previously been associated with hypertension as well as estimated

glomerular filtration rate (eGFR) and CKD (Wuttke et al., 2019) in other studies, supporting our finding that kidney volume reflects overall kidney function.

The trait with the most associations was the spleen, with 25 independent signals, of which 18 colocalised with at least one hematological measurement. We identified one association with ASAT volume (lead SNP rs1421085) at the well-known *FTO* locus which colocalised with many other body composition traits. The association with VAT volume at this SNP (p=3e-8, beta=0.037) was not study-wide significant. We identified three additional signals associated with VAT volume. rs559407214 (nearest gene *CEBPA*) is independent of the nearby pancreatic fat signal. rs73221948 lies 150kb from the nearest protein coding gene. This signal colocalizes with triglyceride levels and HDL levels. This has previously been reported (Richardson et al., 2020), in addition to an association with BMI-adjusted waist-hip circumference (Zhu et al., 2020) Finally, rs72276239 which is also associated with trunk fat percentage, diabetes-related traits, cardiovascular problems, and lipids, and and has previously been associated with waist-hip ratio (Kichaev et al., 2019).

Discussion

We have developed a pipeline to systematically measure IDPs in the UKBB MRI sub-cohort, and to explore epidemiological and genetic associations across seven organs and tissues. Our approach scaled to tens of thousands of subjects, and demonstrated the immense value of abdominal MRI data acquisition in large cohorts. We leveraged deep learning methods for semantic segmentation to address technical challenges, including visual heterogeneity arising from deformable tissues and joints. Systematic measurements of IDPs at this scale would have otherwise been insurmountable.

The observed age-related decrease in organ volume (liver, pancreas, kidney, spleen) appears to reflect the predicted organ atrophy associated with ageing, likely underpinned by mechanism(s) similar to those reported for brain and skeletal muscle (Mitchell et al., 2012; Svennerholm et al., 1997). However, individual organs exhibited distinct patterns of atrophy, with liver and pancreas exhibiting the largest reduction, probably reflecting genetic and environmental exposures. The continued increase in VAT (but not ASAT) and lung volume are interesting and may point at the overriding impact of environmental factors upon these tissues. Given that VAT and ASAT are exposed to similar exogenous factors, this suggests that the plasticity capacity of their adipocytes (hypertrophy and hyperplasia), and therefore tissue lipolysis and inflammation, ectopic fat deposition and insulin sensitivity are differentially affected by the ageing process (Mancuso & Bouchard, 2019). Clearly, access to these IDPs, in

combination with the rich clinical and phenotypic data, can be exploited to directly explore the impact of accelerated ageing on individual organs.

The liver plays a pivotal role in the regulation of iron homeostasis, with iron excess to requirements stored in hepatocytes (Terjung, 2013). Epidemiologic studies utilising indirect methods based on serum markers (ie. the ratio of serum transferrin receptor to serum ferritin) describe an age-related increase in total body iron, declining at a very late age (Cook et al., 2003). However, studies with direct measurements, although far more limited in scope and size, point towards a linear relationship with age (Kühn et al., 2017; McKay et al., 2018; Nomura et al., 1988; Schwenzer et al., 2008), similar to that observed in our study. The discrepancy between total and organ-specific changes with age may relate to the complex relationship between liver iron storage and circulating iron, which is known to be compromised by age related organ dysfunction and the inflammasome (Terjung, 2013). Similar patterns for pancreatic iron were observed (Schwenzer et al., 2008), again reflecting the overall iron homeostasis in the body.

Ectopic fat accumulation showed a more complex relationship with ageing. Although pancreatic fat increased with age for both men and women (Schwenzer et al., 2008), liver fat increased only up to approximately 60 years of age before plateauing in women and decreasing in men (Kühn et al., 2017; Nomura et al., 1988). Previous studies have suggested a linear relationship (Thomas et al., 2012; Wilman et al., 2017); but this may reflect the paucity of older participants (>60 years) in those cohorts, thus lacking the power to detect the true effects of age on liver fat. Both liver fat and iron were associated with T2D, consistent with previous studies (McKay et al., 2018). No association was observed between pancreatic fat or iron content with either T1D or T2D, despite the observed association between pancreas volume and T1D. This is surprising given its proposed causal role assigned to this fat depot in T2D (Roy Taylor, 2013). Interestingly, although both liver and pancreas volume decreased with age, pancreatic fat did not, in agreement with previous observations (Majumder et al., 2017). Additionally, there was considerably greater diurnal variation in liver volume compared with the pancreas. These observations add credence to the growing evidence of disparate mechanisms for the accumulation of fat in these organs (Hellerstein, 1999). Furthermore, given the observed diurnal variation in organ volume, fat and iron content, coupled to the known effects of feeding on the circadian clock on organ function (Kalhan & Ghosh, 2015), scheduling of MRI measurements of participants may be an important consideration in longitudinal studies.

Most organ volumes were associated with disease, i.e.: kidney volume with CKD (Grantham et al., 2006), and lung volumes with COPD, bronchitis, and respiratory disease. Liver volume was associated with chronic liver disease (Lin et al., 1998) and cirrhosis (Hagan et al., 2014) as well

as diabetes and hypertension. Although there is a strong correlation between liver volume and liver fat, liver volume is not generally measured in relation to metabolic disease. Whilst spleen volumes can be enlarged in response to a whole host of diseases such as infection, hematological, congestive, inflammatory and neoplastic (Pozo et al., 2009), we found spleen volume to be most strongly associated with leukaemia. Although organ volume is not a widely-used measure for disease diagnosis, spleen volume is a useful metric for predicting outcome and response to treatment (Shimomura et al., 2018), and a robust automated measure of this IDP could be a powerful auxiliary clinical tool.

The strong association between VAT and development of metabolic dysfunction is well established (Lee et al., 2018), and confirmed herein on a much larger cohort. No association between ASAT and disease, apart from incidence of gallstones, were observed. The overall role of subcutaneous fat in disease development is still debated. Viewed as benign or neutral in terms of risk of metabolic disease (Kuk et al., 2006), especially subcutaneous fat around the hips, ASAT does appear to be associated with components of the metabolic syndrome, though not after correcting for VAT or waist circumference (Elffers et al., 2017; Irlbeck et al., 2010). It has been suggested that subdivisions of ASAT may convey different risks, with superficial ASAT conferring little or no risk compared to deeper layers (Kelley et al., 2000). These conflicting results may reflect different approaches to ASAT and VAT measurement (MRI vs indirect assessment), size and make-up of study cohorts. Future studies within the UKBB and other biobanks will allow these relationships to be explored in more depth.

Through GWAS, we identify a substantial heritable component to organ volume, fat and iron content, after adjusting for body size. We demonstrate heritability enrichment in relevant tissues and cell types, suggesting that there may be specific mechanisms underpinning organ morphology and function that warrant further investigation. As well as replicating previous observations, we identify several novel associations that may suggest mechanisms for further study, including an association between *GPAM* and liver fat, *PPP1R3B* and liver volume (but not fat), *CB2FAT3* and pancreatic fat, and *SLC40A1* and liver iron. The substantial heritable component suggests that the planned studies involving up to 100,000 scanned individuals will yield further insights into the basis of organ form, and its relationship to function.

This study has some limitations. Although recruitment into the UK Biobank study finished in 2010, scanning began in 2014. The median follow-up period from scanning is 2.5 years, limiting our power to evaluate the prognostic value of IDPs, or to evaluate whether they are a cause or consequence of the disease state. Since medical records will continue to be collected prospectively, we will be able to assess this more systematically in future studies. Our genetic studies were limited to participants of white British ancestry. While this did not greatly affect

power due to the demographics of the imaging cohort, future imaging studies which incorporate greater diversity of ancestry and environmental exposure will facilitate fine-mapping as well as potentially elucidate new mechanisms (Wojcik et al., 2019). Finally, while this study focussed on tractable measures derived from segmentation, we expect that future studies will allow us to define more sophisticated traits derived from organ segmentations and will give deeper insight into the relationship between organ form and function.

In conclusion, by systematically quantifying eleven IDPs covering several abdominal organs in the largest imaging cohort to date, we have been able to unravel the hitherto unexplored relationships between organ form and function, genetic and environmental exposures, and disease outcomes. Exploration of the UKBB cohort through the application of models trained using deep learning is thus enhancing our understanding of health and disease.

Materials and Methods

Abdominal imaging data in UK Biobank

All abdominal scans were performed using a Siemens Aera 1.5T scanner (Syngo MR D13) (Siemens, Erlangen, Germany). We analyzed four distinct groups of acquisitions: (1) the Dixon protocol with six separate series covering 1.1 m of the participants (neck-to-knees), (2) a high-resolution T1-weighted 3D acquisition of the pancreas volume, (3a) a single-slice multi-echo acquisition sequence for liver fat and iron, and (3b) a single-slice multi-echo acquisition sequence for pancreas fat and iron. Additional details of the MRI protocol may be found elsewhere (Littlejohns et al., 2020). The protocol covers the neck-to-knee region, including organs such as the lungs outside the abdominal cavity. For consistency with the UK Biobank terminology, we used the term “abdominal” throughout the text.

The UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) to obtain and disseminate data and samples from the participants (http://www.ukbiobank.ac.uk/ethics/), and these ethical regulations cover the work in this study. Written informed consent was obtained from all participants.

Image Preprocessing

Analysis was performed on all available datasets as of December 2019, with 38,971 MRI datasets released by the UK Biobank, where a total of 100,000 datasets are the ultimate goal for the imaging sub-study. We focus here on four separate acquisitions, with one sequence being applied twice (once for the liver and once for the pancreas).

Dixon pipeline

The Dixon sequence involved six overlapping series that were acquired using a common set of parameters: TR = 6.67 ms, TE = 2.39/4.77 ms, FA = 10° and bandwidth = 440 Hz. The first series, over the neck, consisted of 64 slices, voxel size 2.232 x 2.232 x 3.0 mm and 224 x 168 matrix; series two to four (covering the chest, abdomen and pelvis) were acquired during 17 sec expiration breath holds with 44 slices, voxel size 2.232 x 2.232 x 4.5 mm and 224 x 174 matrix; series five, covering the upper thighs, consisted of 72 slices, voxel size 2.232 x 2.232 x 3.5 mm and 224 x 162 matrix; series six, covering the lower thighs and knees, consisted of 64 slices, voxel size 2.232 x 2.232 x 4 mm and 224 x 156 matrix.

The six separate series associated with the two-point Dixon acquisition were positioned automatically after the initial location was selected by the radiographer (Littlejohns et al., 2020). Reconstruction of the fat and water channels from the two-point Dixon acquisition was performed on the scanner console. Four sets of DICOM files were generated for each of the six series in the neck-to-knee Dixon protocol: in-phase, opposed-phase, fat and water.

Bias-field correction (Tustison et al., 2010) was performed on the in-phase volume and the resulting bias field applied to the other channels (opposed-phase, fat, water) for each series. The series were resampled to a single dimension and resolution to facilitate merging the six series into a single three-dimensional volume (size = [224, 174, 370], voxel = 2.232 x 2.232 x 3.0 mm). To reduce the effect of signal loss when blending the series, we identified the fixed set of slices that form an overlap (inferior-superior direction) between adjacent series and applied a nonlinear function to blend the signal intensities on these regions of overlap. Slices in the interior of the volume were heavily weighted and slices near the boundary were suppressed. We repeated the bias-field correction on the blended in-phase volume and applied the estimated bias field to the other channels.

Fat-water swaps are a common issue in the reconstruction of Dixon acquisitions, where the fat and water labels attributed to the reconstructed images are reversed for all voxels in the acquired data series or cluster of voxels associated with separate anatomical structures (e.g., legs or arms). Once corrected, the fat and water channels are consistent. We used a convolutional neural network (CNN) model to detect swaps, with six individual models trained for each of the six acquired series. Only fat-water swaps that involved the entire series or the left-right halves in the final two series were considered. Partial fat-water swaps (e.g., the top of the liver) will be considered in future work. Each model used a sequential architecture with six layers that assigned a label to each of the series when given a central 2D slice from the series. Each convolution block (C_n) was made up of *n* convolutions that were 3 x 3 spatial filters applied

with stride of length two, followed by a leaky rectified linear unit (ReLU) activation with slope 0.2 and batch normalization. The final layer had stride of length three and a sigmoid activation for binary classification of the input as either water or fat. The number of convolution filters was doubled in each layer down the network as follows: C_{64} - C_{128} - C_{256} - C_{512} - C_{1024} - C_1. The two models covering the bottom two series that include the legs checked the right and left half of the input image separately to accommodate for the legs being separate structures with increased likelihood of swaps. Each of the six series for 462 subjects were individually inspected to ensure no swaps occurred and used to train the models. The ten central coronal slices of each subject were selected by checking the image profile of the slice in each volume, where the largest profile was assumed to be the centre of the body. Thus, a total of 4620 images were available for training each of the networks. No additional data augmentation was performed. Each 2D slice was normalized. The model was trained with a binary cross entropy loss function using the Adam optimizer and a batch size of 100 until convergence, which was between 150 and 200 epochs depending on the series. The models were validated on a separate set of 615 subjects, resulting in 4,920 individual swap detection operations performed as every set of Dixon data is subject to eight classification tests. The validation, via visual inspection of all the series and the swap detection results, revealed only two instances of the second series (the chest) and one instance of the fifth series (one of the two legs) were mislabeled out of the total 4,920 checks performed. A single false positive, in the second series, was observed.

Anomaly detection of the final reconstructed volumes was performed to identify potential data issues such as image artifacts, positioning errors or missing series. This was achieved checking the dimensions of the final reconstructed volume and edge detection performed on the binary body mask. To generate the body mask, we applied multiscale adaptive thresholding to the flattened in-phase signal intensities, keeping only the largest connected component, then performed a binary closing operation. The presence of sharp edges in the body mask highlighted discontinuities in the data and was used as an indicator of data inconsistencies. We used Canny edge detection on a central coronal slice and a sagittal slice of the body mask containing both background and subject labels. In a normal subject, edge detection should not highlight anything other than the vertical contour of the body from neck to knee. Presence of discontinuities or horizontal features in the body mask were indicators of anomalies. Clusters of voxels in the edge image corresponding to horizontal edges exceeding a threshold 10 voxels in the sagittal and coronal slice, or 25 in either slice, triggered the anomaly detection. Those values were selected based on results of 1,000 subjects. Field of view errors in positioning the subject were identified if the head or chin were partly or fully visible, or if the total volume did not match the standard 224 x 174 x 370 dimension of the correctly assembled Dixon acquisition. Signal dropout artifacts were caused by metal objects such as knee or tooth implants and identified when discontinuities appeared inside the body mask.

3D pancreas pipeline

A high-resolution T1w acquisition sequence for determining pancreas volume was acquired under a single expiration breath hold with TR = 3.11 ms, TE = 1.15 ms, FA = 10°, bandwidth = 650 Hz, voxel size 1.1875 x 1.1875 x 1.6 mm and 320 x 260 matrix. Two versions were provided, with and without normalization, from the scanner. Bias-field correction was performed to reduce signal inhomogeneities in the normalized volume. No additional preprocessing was applied to the high-resolution 3D T1w pancreas volumes.

Multiecho pipeline (Gradient Echo and IDEAL)

Two types of acquisitions were performed to quantify fat in the liver and pancreas:

1. A single-slice gradient echo acquisition sequence, for both the liver and pancreas, was acquired using the common set of parameters: TR = 27 ms, TE = 2.38/4.76/7.15/9.53/11.91/14.29/16.67/19.06/21.44/23.82 ms, FA = 20°, bandwidth = 710 Hz, voxel size 2.5 x 2.5 x 6.0 mm and 160 x 160 matrix. This acquisition was stopped for the liver after the first 10,000 subjects (approximately) and replaced by the IDEAL sequence, but was continued for the pancreas for all subjects.
2. A single-slice IDEAL sequence (Reeder et al., 2005) for the liver used the following parameters: TR = 14 ms, TE = 1.2/3.2/5.2/7.2/9.2/11.2 ms, FA = 5°, bandwidth = 1565 Hz, voxel size 1.719 x 1.719 x 10.0 mm and 256 x 232 matrix.

We applied bias-field correction to each echo time separately to facilitate 2D segmentation. Software (https://github.com/marcsous/pdff) available from Dr Mark Bydder, specifically the PRESCO (Phase Regularized Estimation using Smoothing and Constrained Optimization) algorithm (Bydder et al., 2020), was used to simultaneously estimate the proton density fat fraction (PDFF, referred to as fat in results) and transverse relaxivity (R2*) values voxelwise from the single-slice gradient echo (GRE) and IDEAL acquisitions. Essentially, a multi-peak spectrum was constructed from the echo times in the acquisition protocol and used to perform nonlinear least squares under multiple regularization constraints that extends the IDEAL (Iterative Decomposition of Water and Fat with Echo Asymmetry and Least-Squares Estimation) algorithm (Reeder et al., 2005; Yu et al., 2008).

For consistency with previous studies (McKay et al., 2018; Wood et al., 2005), we convert R2* into iron concentration (mg/g) using the formula:

$$\text{iron concentration} = 0.202 + 0.0254 \times R2^*.$$

Liver iron concentrations were not adjusted for the potential effects of hepatic cellular pathologies (Li et al., 2018) but we would expect it to be minimal given the relatively low level of hepatocellular clinical diagnosis in the UKBB cohort.

To minimise error and confounding effects, we applied one voxel erosion to the 2D mask prior to summarising fat and iron content. If the final size was < 1% of the organ's 3D volume, or < 20 voxels, we excluded the mask from analysis.

To account for systematic differences between the IDEAL and GRE acquisitions, we used the acquisitions of 1,487 subjects that both had GRE and IDEAL acquisitions to fit a linear model relating these two measurements. If both acquisitions were available, we used the IDEAL measurement. For those with only GRE, we used the following formulae:

$$PDFF_{IDEAL} = 1.09 + 0.763 * PDFF_{GRE}$$

$$Iron_{IDEAL} = 0.196 + 0.855 * Iron_{GRE}$$

Manual annotation of abdominal structures for model training data

For each organ, we defined a standard operating procedure and provided training to a team of radiographers, utilising MITK, a free open-source software system for development of interactive medical image processing software (mitk.org). All annotations were visually inspected at multiple stages by experienced analysts before use in modelling.

Segmentation of organs, for volume assessment, from Dixon data

We re-purposed an updated 3D iteration of the U-net architecture (Ronneberger et al., 2015) based on label-free segmentation from 3D microscopy (Ounkomol et al., 2018). In order to produce sensible segmentations for QC purposes on minimal data, we made the following choices. Training data is intrinsically scarce, and performance can always be improved with additional data. We pursued a multi-task approach (Y. Zhang & Yang, 2017) so as to improve data efficiency. The supervision loss consists of binary heads as opposed to multi-class classification because compartments can overlap spatially. We annotated multiple compartments and organs on the same individuals. Although not intrinsically novel, we are the first to scale this application to a very large UKBB imaging cohort. All weights and pipelines and data augmentation details are available to download (https://github.com/calico/ukbb-mri-sseg). This is the first time that segmentations for multiple major organs and compartments have been published on the UKBB dataset. Comparisons across datasets are also difficult because

evaluation would be confounded by the specifics of how individuals are chosen, the conventions of annotation, and specifics of data acquisition or processing.

Our implementation of U-net had 72 channels on the outside, and we capped the maximum number of channels in deeper layers of the network to 1152. We used concatenation on skip connections, and convolution-transposes when upsampling. A heavily-engineered system was used to stream large datasets efficiently and perform data augmentation on demand. To address computational bottlenecks, we encoded the 3D multichannel images as urolled PNGs inside TFrecords. We relied on TensorFlow best practices to parallelise and streamline random batching during training. Data augmentation was performed on the fly on the GPU, and not pre-computed. We used a batch size of six, and some customized engineering was needed to accommodate very large tensors and total GPU memory use.

Input voxels were encoded into five channels: fat, water, in-phase, out-of-phase, and body mask. The body mask indicated whether a given voxel was inside the body The neural network branched into a different logit head for supervision on each organ. Supervision included the sum of Dice coefficient (Milletari et al., 2016) and binary cross-entropy across all organs.

Inspection of validation loss curves indicated that use of batch normalization and data augmentation provided sufficient regularization. During training, the model utilised 80,000 96x96x96 patches as subsequently described, and the Adam optimizer learning rate was reduced from 1e-5 to 1e-7 following a quadratic decay. During inference, we used Otsu thresholding (Otsu, 1979) to decode a binary decision for each voxel as to whether it was part of each given organ or not.

Data augmentation

Data augmentation included a 3D deformation to locally transform 3D data smoothly as a whole, rather than by slice. We iteratively batched a small number of individual voxels, assigned random Gaussian values and convolved noise with random width Gaussian filters. The summed result was treated as a noise vector and added to the raw image dynamically. We also used a smooth elastic warp to augment the data. This augmentation assigned a different smooth 3D optical flow offset to each voxel in any spatial direction, which was effective since it could locally subsume a heterogeneous combination of commonly used spatial distortions. The same warping function was applied to training masks to ensure that supervision was consistent with input data.

Each final voxel obtained its value from a location offset by an optical flow vector sampled from a Gaussian process. To preserve visual details, voxels that were close together were sampled with strongly correlated optical flow offsets, while pairs further away were less correlated. To reduce the computational load in the optical flow sampling process, we cropped the image to a 174x174x174 window and placed a 4x4x4 lattice of equispaced points centered inside it. These 64 lattice points had fixed relative spatial positions. Based on pairwise distances, we created a (4x4x4)-by-(4x4x4) covariance matrix to describe how correlated distortions should be in the warping. We applied a Gaussian kernel with a width of 24 voxels. These 3x64 values were multiplied by a random scaling chosen uniformly in [0, 4], treated as optical flow values and applied to the image in the distortion along three spatial directions for each of the 64 lattice points. Next, we extrapolated optical flow values to each underlying voxel position with a polyharmonic spline, and applied the warp by resampling the image at each voxel with its own floating point offsets in 3D. From the center of the warped and resampled image, we cropped a 96x96x96 patch and used this as training data. When interpolating supervision segmentation masks, we converted the masks to floating-point probabilities and applied clipping heuristics after the warp and resampling to ensure that probabilities were valid. Finally, we obtained volume measurements by thresholding the model output, removing disconnected structures, and multiplying the number of mask voxels by the image resolution.

Quality control consisted of iterations of visual inspection of extreme volumes for each distinct organ/structure, as well as spot checks of hundreds of random subjects. The training data was regularly enriched to include problematic cases. We repeated this procedure and retrained the model until results did not display outliers for extreme subjects nor any of the random spot checks. Performance metrics are available in Supplementary Table 1.

Abdominal Subcutaneous Adipose Tissue (ASAT) and Visceral Adipose Tissue (VAT)

Two structures, the ‘body cavity’ and ‘abdominal cavity’, were segmented using neural-network based methods from the Dixon segmentation to estimate ASAT and VAT. For estimation of VAT, the abdominal cavity was used to isolate only tissue in the abdomen and pelvis. The fat channel was thresholded, small holes filled, and segmentations of abdominal organs (e.g., liver, spleen, kidneys) were removed to produce the final mask of VAT. For ASAT estimation, the body cavity was used to exclude all tissue internal to the body. A bounding box was computed based on the abdominal cavity, where the upper and lower bounds in the superior-inferior (z) direction were used to define the limits of the ASAT compartment.

Segmentation of the liver, for fat and iron content assessment, from single-slice data

We applied a standard 2D U-net to segment the IDEAL and GRE liver data, training one model for each of the two liver acquisitions. We split 507 annotations of the IDEAL acquisition into a training set of 456 training images and 51 validation images. Similarly, we split 373 annotations of the GRE acquisition into 335 training images and 38 validation images. The raw data consisted of complex numbers in six channels in IDEAL and 10 in GRE, resulting in input shapes of (256, 232, 18) for IDEAL and (160, 160, 30) for GRE. We encoded the complex number as a triplet: magnitude, sine and cosine of the angle. We applied mild data augmentation in the form of small rotations, translations, zoom, shears, and flips. We used the Adam optimizer on 100 steps with batch size 32 for each of the following learning rates in the schedule: [1e-4, 1e-5, 1e-5, 1e-6, 1e-7]. To ensure high specificity at the cost of recall during inference (and thus ensure that our derived values do not include non-liver tissue), we used Otsu to propose a threshold based on the voxelwise prediction probabilities and adjusted the threshold to further ablate the 25% of the foreground 25%.

Pancreas segmentation from T1w MRI (volume) and extraction (fat and iron content assessment), from single-slice data

We performed pancreas 3D segmentation on the high-resolution T1w 3D acquisition based on a recent iteration of the U-net architecture used in label-free microscopy (Ounkomol et al., 2018), using 123 manual annotations. Segmentation was not performed using the Dixon data since the pancreas has a complex morphology and benefited from improved contrast and resolution. The network trunk ranged from 16 channels in the outer layers and grew to 256 in the deepest layer. Skip layers were added rather than concatenated. The learning rate was reduced by a factor of 10 and 100 after 12 and 25 epochs, respectively, starting with an initial learning rate 1e-3. We optimized with Adam, using an L2-regularization coefficient of 1e-4, on batch size six, and supervised segmentation with the mean of the Dice similarity coefficient and binary cross-entropy. For data augmentation, we translated the pancreas randomly. The segmented volume was resampled to extract an equivalent 2D mask for the single-slice data (Basty et al., 2020).

Statistical analysis of IDPs

All statistical analysis was performed using R version 3.6.0.

Comparison with previous studies

We compared the values extracted in our study with those from previous studies, available from the following UK Biobank fields:

- VAT (Field 22407) and ASAT (Field 22408) (West et al., 2016)
- Liver fat (22400) and liver iron (22402) (Wilman et al., 2017)

Relationship between age, scan time, and IDPs

For fitting linear models, we used the R function `lm`. For fitting smoothing splines, we used the `splines` package. To determine whether a coefficient was statistically significant in a set of models, we adjusted the p-values for each coefficient using Bonferroni correction. We compared models with and without scan time using ANOVA.

We looked for systematic differences between scanning centre, and trends by scan date (Supplementary Figure 2). Because there were some minor differences unlikely to be of biological interest, we included scanning centre and scan date as covariates in all subsequent analyses.

Disease phenome defined from hospital records

We used the R package PheWAS (Carroll et al., 2014) to combine ICD10 codes (Field 41270) into distinct diseases or traits (PheCodes). The raw ICD10 codes were grouped into 1283 PheCodes; of these, 754 PheCodes had at least 20 cases for all IDPs dataset allowing for a meaningful regression model. For each IDP-PheCode pair, we performed a logistic regression adjusted for age, sex, height, and BMI, and imaging center and imaging date, scan time, and ethnicity.

We defined two Bonferroni-adjusted p-values: a single-trait value of 6.63e-5, and a study-wide value of 6.03e-6. As many of the diagnoses are correlated, we expect this threshold to be conservative.

Other traits

We used the R package PHESANT (Millard et al., 2018) to generate an initial list of variables derived from raw data. We manually curated this list to remove variables related to procedural metrics (e.g., measurement date, time and duration; sample volume and quality), duplicates (e.g., data collected separately on a small number of participants during the pilot phase), and raw measures (e.g., individual components of the fluid intelligence score). This resulted in a total of 1824 traits. For each trait, we performed a regression (linear regression for quantitative traits, and logistic regression for binary traits) on the abdominal IDP, including imaging center, imaging date, scan time, age, sex, BMI, and height, and ethnicity as covariates.

We defined two Bonferroni-adjusted p-values: a single-trait value of 2.75e-5, and a study-wide value of 2.49e-6. As many traits are correlated, we expect this threshold to be conservative.

Genetics

We follow the methods described in a previous study (Sethi et al., 2020)

Genome-wide association study

We used the UKBB imputed genotypes version 3(Bycroft et al., 2018), excluding single nucleotide polymorphisms (SNPs) with minor allele frequency < 1% and imputation quality < 0.9. We excluded participants not recorded as Caucasian, exhibiting sex chromosome aneuploidy, with a discrepancy between genetic and self-reported sex, heterozygosity and missingness outliers, and genotype call rate outliers (Bycroft et al., 2018). We used BOLT-LMM version 2.3.2 (Loh, Tucker, et al., 2015) to conduct the genetic association study. We included age at imaging visit, age squared, sex, imaging centre, scan date, scan time, and genotyping batch as fixed-effect covariates, and genetic relatedness derived from genotyped SNPs as a random effect to control for population structure and relatedness. The genomic control parameter ranged from 1.05-1.15 across eleven IDPs (Supplementary Table 4). We verified that the test statistics showed no overall inflation compared to the expectation by examining the intercept of linkage disequilibrium (LD) score regression (LDSC) (Bulik-Sullivan, Loh, et al., 2015) (Supplementary Table 5). In addition to the commonly-used genome-wide significance threshold of p=5e-8, we defined an additional study-wide significance threshold using Bonferroni correction for the number of traits, p=5e-8/11=4.5e-9. For this analysis and all other analyses using LDSC, we followed the recommendation of the developers and (i) removed variants with imputation quality (info) <0.9 because the info value is correlated with the LD score and could introduce bias, (ii) excluded the major histocompatibility complex (MHC) region due to the complexity of LD structure at this locus (GRCh37::6:28,477,797-33,448,354; see https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC), and (ii) restricted to HapMap3 SNPs(International HapMap 3 Consortium et al., 2010).

For each IDP, we performed a secondary analysis with height and BMI as additional covariates.

Exome-wide association study

Exome sequencing variant calls from the raw FE variant calling pipeline (Regier et al., 2018) were downloaded from the UK Biobank website (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=23160). QC was performed in PLINK v.1.90 using the following criteria: removal of samples with discordant sex (no self-reported sex provided, ambiguous genetic sex, or discordance between genetic and self-reported sex), sample-level missingness <0.02, European genetic ancestry as defined by the UK Biobank (Bycroft et al., 2018). Variant annotation was performed using VEP v100, filtered for rare

(MAF<0.01) putative loss-of-function variants including predicted high-confidence loss-of-function variants, predicted using the LOFTEE plugin (Karczewski et al., 2020). 11,134 samples and 11,939 genes were analyzed in a generalized linear mixed model as implemented in SAIGE-GENE (Zhou et al., 2020). A filtering step of at least five loss-of-function carriers per gene was applied, resulting in 6,745 genes. Outcome variables were inverse normal transformed and regressed on gene carrier status, adjusted for genetic sex, age, age^2, the first ten principal components of genetic ancestry, scaled scan date, scaled scan time, and study center as fixed effects and genetic relatedness as a random effects term. A kinship matrix was built in SAIGE off of a filtered set of array-genotyped variants (r^2<0.2, MAF>=0.05, HWE p>1e-10 in European population).

Heritability estimation and enrichment

We estimated the heritability of each trait using restricted maximum likelihood as implemented in BOLT version 2.3.2 (Loh, 2018).

To identify relevant tissues and cell types contributing to the heritability of IDPs, we used stratified LD score regression (Finucane et al., 2018) to examine enrichment in regions of the genome containing genes specific to particular tissues or cell types. We used three types of annotations to define: (i) regions near genes specifically expressed in a particular tissue/cell type, (i) regions near chromatin marks from cell lines and tissue biopsies of specific cell types, and (iii) genomic regions near genes specific to cells from immune genes. For functional categories, we used the baseline v2.2 annotations provided by the developers (https://data.broadinstitute.org/alkesgroup/LDSCORE). Following the original developers of this method(Finucane et al., 2018), we calculated tissue-specific enrichments using a model that includes the full baseline annotations as well as annotations derived from (i) chromatin information from the NIH Roadmap Epigenomic (Roadmap Epigenomics Consortium et al., 2015) and ENCODE (ENCODE Project Consortium, 2012) projects (including the EN-TEx data subset of ENCODE which matches many of the GTEx tissues, but from different donors), (ii) tissue/cell type specific expression markers from GTEx v6p (GTEx Consortium et al., 2017) and other datasets (Fehrmann et al., 2015); (Pers et al., 2015), and (iii) immune cell type expression markers from the ImmGen Consortium (Heng et al., 2008). For each annotation set, we controlled for the number of tests using the Storey and Tibshirani procedure (Storey & Tibshirani, 2003). Although heritability is non-negative, the unbiased LDSC heritability estimate is unbounded; thus, it is possible for the estimated heritability, and therefore enrichment, to be negative (e.g., if the true heritability is near zero and/or the sampling error is large due to small sample sizes).

To enable visualization, we grouped tissue/cell types into systems (e.g., "blood or immune", "central nervous system") as used in Finucane et al. (Finucane et al., 2018).

Genetic correlation

We computed genetic correlation between traits using bivariate LDSC (Bulik-Sullivan, Finucane, et al., 2015).

Statistical fine-mapping

We performed approximate conditional analysis using genome-wide complex trait analysis (GCTA) (Yang et al., 2012), considering all variants that passed quality control measures and were within 500kb of a locus index variant. As a reference panel for LD calculations, we used genotypes from 5,000 UKBB participants (Bycroft et al., 2018) that were randomly selected after filtering for unrelated, participants of white British ancestry. We excluded the major histocompatibility complex (MHC) region due to the complexity of LD structure at this locus (GRCh37::6:28,477,797-33,448,354; see https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC). For each locus, we considered variants with genome-wide evidence of association (Pjoint<10−8) to be conditionally independent. We annotated each independent signal with the nearest known protein-coding gene using the OpenTargets genetics resource (May 2019 version).

Construction of genetic credible sets

For each distinct signal, we calculated credible sets (Wellcome Trust Case Control Consortium et al., 2012) with 95% probability of containing at least one variant with a true effect size not equal to zero. We first computed the natural log approximate Bayes factor (Wakefield, 2007) Λj , for the j-th variant within the fine-mapping region:

$$\Lambda_j = ln\left(\sqrt{\frac{V_j}{V_{j+\omega}}}\right)\frac{\omega\beta^2}{2V_j(V_j+\omega)}$$

where βj and Vj denote the estimated allelic effect (log odds ratio for case control studies) and corresponding variance. The parameter ω denotes the prior variance in allelic effects and is set to $(0.2)^2$ for case control studies (Wakefield, 2007) and $(0.15\sigma)^2$ for quantitative traits (Giambartolomei et al., 2014), where σ is the standard deviation of the phenotype estimated using the variance of coefficients (Var(βj)), minor allele frequency (fj), and sample size (nj; see the sdY.est function from the coloc R package):

$$2n_j f_j(1-f_j) \sim \sigma^2\frac{1}{Var(\beta_j)} - 1$$

Here, σ2 is the coefficient of the regression, estimating σ such that $\sigma = \sqrt{\sigma^2}$.

We calculated the posterior probability, πj, that the j th variant is driving the association, given l variants in the region, by:

$$\pi_j = \frac{(1-\gamma)\Lambda_j}{l \sum_{k=0}^{l} \Lambda_k}$$

where γ denotes the prior probability for no association at this locus and k indexes the variants in the region (with k=0 allowing for the possibility of no association in the region). We set γ=0.05 to control for the expected false discovery rate of 5%, since we used a threshold of P marginal <5x10−8 to identify loci for fine-mapping. To construct the credible set, we (i) sorted variants by increasing Bayes factors (natural log scale), (ii) included variants until the cumulative sum of the posterior probabilities was >=1−c, where c corresponds to the credible set cutoff of 0.95.

Colocalization of independent signals

To identify other traits potentially sharing the same underlying causal variant, we downloaded a catalog of summary statistics using the UK Biobank cohort from http://www.nealelab.is/uk-biobank (Version 2). For disease phenotypes, we additionally downloaded summary statistics computed using SAIGE (Zhou et al., 2018) from https://www.leelabsg.org/resources. After de-duplication, removal of biologically uninformative traits, and removal of traits with no genome-wide significant associations, we considered a total of 974 complex traits and, and 356 disease phenotypes. To identify potentially causal genes at each locus, additionally explored expression QTL data from GTEx (version 7, dbGaP accession number dbGaP accession number phs000424.v7.p2) to seek evidence for colocalization with expression in one of 49 tissues.

We performed colocalization analysis using the coloc R package (Giambartolomei et al., 2014) using default priors and all variants within 500kb of the index variant of each signal. Following previous studies (Guo et al., 2015), we considered two genetic signals to have strong evidence of colocalization if PP3+PP4≥0.99 and PP4/PP3≥5.

Identifying other associations with our lead signals

In addition to the colocalization analysis with UK Biobank traits, order to identify GWAS signals tagged by any of our associations from previous studies (not including the UK Biobank traits described above), we queried the Open Targets Genetics Resource (Carvalho-Silva et al., 2019), version 190505. We identified for studies where our lead variant was in LD (r>0.7) with the lead SNP of a published study. We also searched for our lead SNPs in the NHGRI-EBI GWAS catalog (Buniello et al., 2019) in October 2020.

References

Basty, N., Liu, Y., Cule, M., Thomas, E. L., Bell, J. D., & Whitcher, B. (2020). Automated Measurement of Pancreatic Fat and Iron Concentration Using Multi-Echo and T1-Weighted MRI Data. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 345–348. https://doi.org/10.1109/ISBI45749.2020.9098650

Buch, S., Stickel, F., Trépo, E., Way, M., Herrmann, A., Nischalke, H. D., Brosch, M., Rosendahl, J., Berg, T., Ridinger, M., Rietschel, M., McQuillin, A., Frank, J., Kiefer, F., Schreiber, S., Lieb, W., Soyka, M., Semmo, N., Aigner, E., … Hampe, J. (2015). A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nature Genetics*, *47*(12), 1443–1448. https://doi.org/10.1038/ng.3417

Bugianesi, E., Bizzarri, C., Rosso, C., Mosca, A., Panera, N., Veraldi, S., Dotta, A., Giannone, G., Raponi, M., Cappa, M., Alisi, A., & Nobili, V. (2017). Low Birthweight Increases the Likelihood of Severe Steatosis in Pediatric Non-Alcoholic Fatty Liver Disease. *The American Journal of Gastroenterology*, *112*(8), 1277–1286. https://doi.org/10.1038/ajg.2017.140

Bulik-Sullivan, B. K., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J.,

Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., … Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. In *Nucleic Acids Research* (Vol. 47, Issue D1, pp. D1005–D1012). https://doi.org/10.1093/nar/gky1120

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., 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. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Bydder, M., Ghodrati, V., Gao, Y., Robson, M. D., Yang, Y., & Hu, P. (2020). Constraints in estimating the proton density fat fraction. *Magnetic Resonance Imaging*, *66*, 1–8. https://doi.org/10.1016/j.mri.2019.11.009

Carroll, R. J., Bastarache, L., & Denny, J. C. (2014). R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* , *30*(16), 2375–2376. https://doi.org/10.1093/bioinformatics/btu197

Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., Carmona, M., Faulconbridge, A., Hercules, A., McAuley, E., Miranda, A., Peat, G., Spitzer, M., Barrett, J., Hulcoop, D. G., Papa, E., Koscielny, G., & Dunham, I. (2019). Open Targets Platform: new developments and updates two years on. *Nucleic Acids Research*, *47*(D1), D1056–D1065.

https://doi.org/10.1093/nar/gky1133

Cook, J. D., Flowers, C. H., & Skikne, B. S. (2003). The quantitative assessment of body iron. *Blood*, *101*(9), 3359–3364. https://doi.org/10.1182/blood-2002-10-3071

Cuchel, M., Bloedon, L. T., Szapary, P. O., Kolansky, D. M., Wolfe, M. L., Sarkis, A., Millar, J. S., Ikewaki, K., Siegelman, E. S., Gregg, R. E., & Rader, D. J. (2007). Inhibition of microsomal triglyceride transfer protein in familial hypercholesterolemia. *The New England Journal of Medicine*, *356*(2), 148–156. https://doi.org/10.1056/NEJMoa061189

de Lichtenberg, K. H., Seymour, P. A., Jørgensen, M. C., Kim, Y.-H., Grapin-Botton, A., Magnuson, M. A., Nakic, N., Ferrer, J., & Serup, P. (2018). Notch Controls Multiple Pancreatic Cell Fate Regulators Through Direct Hes1-mediated Repression. In *bioRxiv* (p. 336305). https://doi.org/10.1101/336305

Donovan, A., Lima, C. A., Pinkus, J. L., Pinkus, G. S., Zon, L. I., Robine, S., & Andrews, N. C. (2005). The iron exporter ferroportin/Slc40a1 is essential for iron homeostasis. *Cell Metabolism*, *1*(3), 191–200. https://doi.org/10.1016/j.cmet.2005.01.003

Elffers, T. W., de Mutsert, R., Lamb, H. J., de Roos, A., van Dijk, K. W., Rosendaal, F. R., Wouter Jukema, J., & Trompet, S. (2017). Body fat distribution, in particular visceral fat, is associated with cardiometabolic risk factors in obese women. In *PLOS ONE* (Vol. 12, Issue 9, p. e0185403). https://doi.org/10.1371/journal.pone.0185403

Elliott, L. T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K. L., Douaud, G., Marchini, J., & Smith, S. M. (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*, *562*(7726), 210–216. https://doi.org/10.1038/s41586-018-0571-7

Emdin, C. A., Haas, M. E., Khera, A. V., Aragam, K., Chaffin, M., Klarin, D., Hindy, G., Jiang, L., Wei, W.-Q., Feng, Q., Karjalainen, J., Havulinna, A., Kiiskinen, T., Bick, A., Ardissino, D., Wilson, J. G., Schunkert, H., McPherson, R., Watkins, H., … Kathiresan, S. (2020). A

missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. *PLoS Genetics*, *16*(4), e1008629. https://doi.org/10.1371/journal.pgen.1008629

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Fehrmann, R. S. N., Karjalainen, J. M., Krajewska, M., Westra, H.-J., Maloney, D., Simeonov, A., Pers, T. H., Hirschhorn, J. N., Jansen, R. C., Schultes, E. A., van Haagen, H. H. H. B. M., de Vries, E. G. E., te Meerman, G. J., Wijmenga, C., van Vugt, M. A. T. M., & Franke, L. (2015). Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nature Genetics*, *47*(2), 115–125. https://doi.org/10.1038/ng.3173

Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., … Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, *47*(11), 1228–1235. https://doi.org/10.1038/ng.3404

Finucane, H. K., Reshef, Y. A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shoresh, N., Genovese, G., Saunders, A., Macosko, E., Pollack, S., Brainstorm Consortium, Perry, J. R. B., Buenrostro, J. D., Bernstein, B. E., Raychaudhuri, S., … Price, A. L. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics*, *50*(4), 621–629. https://doi.org/10.1038/s41588-018-0081-4

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., & Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association

studies using summary statistics. *PLoS Genetics*, *10*(5), e1004383. https://doi.org/10.1371/journal.pgen.1004383

Grantham, J. J., Torres, V. E., Chapman, A. B., Guay-Woodford, L. M., Bae, K. T., King, B. F., Wetzel, L. H., Baumgarten, D. A., Kenney, P. J., Harris, P. C., Klahr, S., Bennett, W. M., Hirschman, G. N., Meyers, C. M., Zhang, X., Zhu, F., & Miller, J. P. (2006). Volume Progression in Polycystic Kidney Disease. In *New England Journal of Medicine* (Vol. 354, Issue 20, pp. 2122–2130). https://doi.org/10.1056/nejmoa054341

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory, Data Analysis &Coordinating Center (LDACC):, … Montgomery. (2017). Genetic effects on gene expression across human tissues. *Nature*, *550*(7675), 204–213. https://doi.org/10.1038/nature24277

Guo, H., Fortune, M. D., Burren, O. S., Schofield, E., Todd, J. A., & Wallace, C. (2015). Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Human Molecular Genetics*, *24*(12), 3305–3313. https://doi.org/10.1093/hmg/ddv077

Hagan, M. T., Sayuk, G. S., Lisker-Melman, M., Korenblat, K. M., Kerr, T. A., Chapman, W. C., & Crippin, J. S. (2014). Liver volume in the cirrhotic patient: does size matter? *Digestive*

Diseases and Sciences, *59*(4), 886–891. https://doi.org/10.1007/s10620-014-3038-1

Harrison-Findik, D. D. (2007). Role of alcohol in the regulation of iron metabolism. *World Journal of Gastroenterology: WJG*, *13*(37), 4925–4930. https://doi.org/10.3748/wjg.v13.i37.4925

Hellerstein, M. K. (1999). De novo lipogenesis in humans: metabolic and regulatory aspects. *European Journal of Clinical Nutrition*, *53 Suppl 1*, S53–S65. https://doi.org/10.1038/sj.ejcn.1600744

Heng, T. S. P., Painter, M. W., & Immunological Genome Project Consortium. (2008). The Immunological Genome Project: networks of gene expression in immune cells. *Nature Immunology*, *9*(10), 1091–1094. https://doi.org/10.1038/ni1008-1091

Hsiao, P.-J., Lee, M.-Y., Wang, Y.-T., Jiang, H.-J., Lin, P.-C., Yang, Y.-H. C., & Kuo, K.-K. (2015). MTTP-297H polymorphism reduced serum cholesterol but increased risk of non-alcoholic fatty liver disease-a cross-sectional study. *BMC Medical Genetics*, *16*, 93. https://doi.org/10.1186/s12881-015-0242-6

International HapMap 3 Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., Peltonen, L., Dermitzakis, E., Bonnen, P. E., Altshuler, D. M., Gibbs, R. A., de Bakker, P. I. W., Deloukas, P., Gabriel, S. B., Gwilliam, R., … McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52–58. https://doi.org/10.1038/nature09298

Irlbeck, T., Massaro, J. M., Bamberg, F., O'Donnell, C. J., Hoffmann, U., & Fox, C. S. (2010). Association between single-slice measurements of visceral and abdominal subcutaneous adipose tissue with volumetric measurements: the Framingham Heart Study. In *International Journal of Obesity* (Vol. 34, Issue 4, pp. 781–787). https://doi.org/10.1038/ijo.2009.279

Jovanović, D., Gasic, B., Pavlovic, S., & Naumovic, R. (2013). Correlation of kidney size with kidney function and anthropometric parameters in healthy subjects and patients with chronic kidney diseases. *Renal Failure*, *35*(6), 896–900. https://doi.org/10.3109/0886022X.2013.794683

Kalhan, S. C., & Ghosh, A. (2015). Dietary iron, circadian clock, and hepatic gluconeogenesis [Review of *Dietary iron, circadian clock, and hepatic gluconeogenesis*]. *Diabetes*, *64*(4), 1091–1093. https://doi.org/10.2337/db14-1697

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kawaguchi, T., Shima, T., Mizuno, M., Mitsumoto, Y., Umemura, A., Kanbara, Y., Tanaka, S., Sumida, Y., Yasui, K., Takahashi, M., Matsuo, K., Itoh, Y., Tokushige, K., Hashimoto, E., Kiyosawa, K., Kawaguchi, M., Itoh, H., Uto, H., Komorizono, Y., … Okanoue, T. (2018). Risk estimation model for nonalcoholic fatty liver disease in the Japanese using multiple genetic markers. *PloS One*, *13*(1), e0185490. https://doi.org/10.1371/journal.pone.0185490

Kelley, D. E., Leland Thaete, F., Troost, F., Huwe, T., & Goodpaster, B. H. (2000). Subdivisions of subcutaneous abdominal adipose tissue and insulin resistance. In *American Journal of Physiology-Endocrinology and Metabolism* (Vol. 278, Issue 5, pp. E941–E948). https://doi.org/10.1152/ajpendo.2000.278.5.e941

Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M. K., Schoech, A., Pasaniuc, B., & Price, A. L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS

Power. *American Journal of Human Genetics*, *104*(1), 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008

Kitchener, R. L., & Grunden, A. M. (2012). Prolidase function in proline metabolism and its medical and biotechnological applications. *Journal of Applied Microbiology*, *113*(2), 233–247. https://doi.org/10.1111/j.1365-2672.2012.05310.x

Kozlitina, J., Smagris, E., Stender, S., Nordestgaard, B. G., Zhou, H. H., Tybjærg-Hansen, A., Vogt, T. F., Hobbs, H. H., & Cohen, J. C. (2014). Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nature Genetics*, *46*(4), 352–356. https://doi.org/10.1038/ng.2901

Kuhajda, F. P., Aja, S., Tu, Y., Han, W. F., Medghalchi, S. M., El Meskini, R., Landree, L. E., Peterson, J. M., Daniels, K., Wong, K., Wydysh, E. A., Townsend, C. A., & Ronnett, G. V. (2011). Pharmacological glycerol-3-phosphate acyltransferase inhibition decreases food intake and adiposity and increases insulin sensitivity in diet-induced obesity. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, *301*(1), R116–R130. https://doi.org/10.1152/ajpregu.00147.2011

Kühn, J.-P., Meffert, P., Heske, C., Kromrey, M.-L., Schmidt, C. O., Mensel, B., Völzke, H., Lerch, M. M., Hernando, D., Mayerle, J., & Reeder, S. B. (2017). Prevalence of Fatty Liver Disease and Hepatic Iron Overload in a Northeastern German Population by Using Quantitative MR Imaging. *Radiology*, *284*(3), 706–716. https://doi.org/10.1148/radiol.2017161228

Kuk, J. L., Katzmarzyk, P. T., Nichaman, M. Z., Church, T. S., Blair, S. N., & Ross, R. (2006). Visceral Fat Is an Independent Predictor of All-cause Mortality in Men*. In *Obesity* (Vol. 14, Issue 2, pp. 336–341). https://doi.org/10.1038/oby.2006.43

Lee, J. J., Pedley, A., Hoffmann, U., Massaro, J. M., Levy, D., & Long, M. T. (2018). Visceral

and Intrahepatic Fat Are Associated with Cardiometabolic Risk Factors Above Other Ectopic Fat Depots: The Framingham Heart Study. *The American Journal of Medicine*, *131*(6), 684–692.e12. https://doi.org/10.1016/j.amjmed.2018.02.002

Leung, N. W., Farrant, P., & Peters, T. J. (1986). Liver volume measurement by ultrasonography in normal subjects and alcoholic patients. *Journal of Hepatology*, *2*(2), 157–164. https://doi.org/10.1016/s0168-8278(86)80074-5

Li, J., Lin, H., Liu, T., Zhang, Z., Prince, M. R., Gillen, K., Yan, X., Song, Q., Hua, T., Zhao, X., & Others. (2018). Quantitative susceptibility mapping (QSM) minimizes interference from cellular pathology in R2* estimation of liver iron concentration. *Journal of Magnetic Resonance Imaging: JMRI*, *48*(4), 1069–1079. https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.26019

Lin, X. Z., Sun, Y. N., Liu, Y. H., Sheu, B. S., Cheng, B. N., Chen, C. Y., Tsai, H. M., & Shen, C. L. (1998). Liver volume in patients with or without chronic liver diseases. *Hepato-Gastroenterology*, *45*(22), 1069–1074. https://www.ncbi.nlm.nih.gov/pubmed/9756008

Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boultwood, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., … Allen, N. E. (2020). The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications*, *11*(1), 2624. https://doi.org/10.1038/s41467-020-15948-9

Liu, Q., Liu, S.-S., Zhao, Z.-Z., Zhao, B.-T., Du, S.-X., Jin, W.-W., & Xin, Y.-N. (2019). TRIB1 rs17321515 gene polymorphism increases the risk of coronary heart disease in general population and non-alcoholic fatty liver disease patients in Chinese Han population. *Lipids*

in Health and Disease, *18*(1), 165. https://doi.org/10.1186/s12944-019-1108-2

Loh, P.-R. (2018). BOLT-LMM v2. 3.2 User Manual. *Available Oniline at: Https://data. Broadinstitute. org/alkesgroup/BOLT-LMM/(accessed May 2, 2019)*. https://data.broadinstitute.org/alkesgroup/BOLT-LMM/downloads/old/BOLT-LMM_v2.3.3_manual.pdf

Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J., Schizophrenia Working Group of Psychiatric Genomics Consortium, de Candia, T. R., Lee, S. H., Wray, N. R., Kendler, K. S., O'Donovan, M. C., Neale, B. M., Patterson, N., & Price, A. L. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics*, *47*(12), 1385–1392. https://doi.org/10.1038/ng.3431

Loh, P.-R., Tucker, G., Bulik-Sullivan, B. K., Vilhjálmsson, B. J., Finucane, H. K., Salem, R. M., Chasman, D. I., Ridker, P. M., Neale, B. M., Berger, B., Patterson, N., & Price, A. L. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*, *47*(3), 284–290. https://doi.org/10.1038/ng.3190

Lutfi, M. F. (2017). The physiological basis and clinical significance of lung volume measurements. *Multidisciplinary Respiratory Medicine*, *12*, 3. https://doi.org/10.1186/s40248-017-0084-5

Majumder, S., Philip, N. A., Takahashi, N., Levy, M. J., Singh, V. P., & Chari, S. T. (2017). Fatty Pancreas: Should We Be Concerned? *Pancreas*, *46*(10), 1251–1258. https://doi.org/10.1097/MPA.0000000000000941

Mancuso, P., & Bouchard, B. (2019). The Impact of Aging on Adipose Function and Adipokine Synthesis. *Frontiers in Endocrinology*, *10*, 137. https://doi.org/10.3389/fendo.2019.00137

Mayr, R., Griffiths, W. J. H., Hermann, M., McFarlane, I., Halsall, D. J., Finkenstedt, A., Douds,

A., Davies, S. E., Janecke, A. R., Vogel, W., Cox, T. M., & Zoller, H. (2011). Identification of mutations in SLC40A1 that affect ferroportin function and phenotype of human ferroportin iron overload. *Gastroenterology*, *140*(7), 2056–2063, 2063.e1. https://doi.org/10.1053/j.gastro.2011.02.064

McKay, A., Wilman, H. R., Dennis, A., Kelly, M., Gyngell, M. L., Neubauer, S., Bell, J. D., Banerjee, R., & Thomas, E. L. (2018). Measurement of liver iron by magnetic resonance imaging in the UK Biobank population. *PloS One*, *13*(12), e0209340. https://doi.org/10.1371/journal.pone.0209340

Mehta, M. B., Shewale, S. V., Sequeira, R. N., Millar, J. S., Hand, N. J., & Rader, D. J. (2017). Hepatic protein phosphatase 1 regulatory subunit 3B (Ppp1r3b) promotes hepatic glycogen synthesis and thereby regulates fasting energy homeostasis. *The Journal of Biological Chemistry*, *292*(25), 10444–10454. https://doi.org/10.1074/jbc.M116.766329

Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G., & Tilling, K. (2018). Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *International Journal of Epidemiology*, *47*(1), 29–35. https://doi.org/10.1093/ije/dyx204

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., Griffanti, L., Douaud, G., Okell, T. W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., … Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), 1523–1536. https://doi.org/10.1038/nn.4393

Milletari, F., Navab, N., & Ahmadi, S. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *2016 Fourth International Conference on 3D*

Vision (3DV), 565–571. https://doi.org/10.1109/3DV.2016.79

Mitchell, W. K., Williams, J., Atherton, P., Larvin, M., Lund, J., & Narici, M. (2012). Sarcopenia, dynapenia, and the impact of advancing age on human skeletal muscle size and strength; a quantitative review. *Frontiers in Physiology*, *3*, 260. https://doi.org/10.3389/fphys.2012.00260

Nomura, H., Kashiwagi, S., Hayashi, J., Kajiyama, W., Tani, S., & Goto, M. (1988). Prevalence of fatty liver in a general population of Okinawa, Japan. *Japanese Journal of Medicine*, *27*(2), 142–149. https://doi.org/10.2169/internalmedicine1962.27.142

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. http://webserver2.tecgraf.puc-rio.br/~mgattass/cg/trbImg/Otsu.pdf

Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F., & Johnson, G. R. (2018). Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nature Methods*, *15*(11), 917–920. https://doi.org/10.1038/s41592-018-0111-2

Parisinos, C. A., Wilman, H. R., Thomas, E. L., Kelly, M., Nicholls, R. C., McGonigle, J., Neubauer, S., Hingorani, A. D., Patel, R. S., Hemingway, H., Bell, J. D., Banerjee, R., & Yaghootkar, H. (2020). Genome-wide and Mendelian randomisation studies of liver MRI yield insights into the pathogenesis of steatohepatitis. *Journal of Hepatology*. https://doi.org/10.1016/j.jhep.2020.03.032

Parkinson, J. R. C., Emsley, R., Adkins, J. L. T., Longford, N., Ozanne, S. E., Holmes, E., & Modi, N. (2020). Clinical and molecular evidence of accelerated ageing following very preterm birth. *Pediatric Research*, *87*(6), 1005–1010. https://doi.org/10.1038/s41390-019-0709-9

Partin, J. S., Partin, J. C., Schubert, W. K., & McAdams, A. J. (1974). Liver ultrastructure in

abetalipoproteinemia: Evolution of micronodular cirrhosis. *Gastroenterology*, *67*(1), 107–118. https://www.ncbi.nlm.nih.gov/pubmed/4135110

Pers, T. H., Karjalainen, J. M., Chan, Y., Westra, H.-J., Wood, A. R., Yang, J., Lui, J. C., Vedantam, S., Gustafsson, S., Esko, T., Frayling, T., Speliotes, E. K., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Boehnke, M., Raychaudhuri, S., Fehrmann, R. S. N., Hirschhorn, J. N., & Franke, L. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications*, *6*, 5890. https://doi.org/10.1038/ncomms6890

Pirruccello, J. P., Bick, A., Wang, M., Chaffin, M., Friedman, S., Yao, J., Guo, X., Venkatesh, B. A., Taylor, K. D., Post, W. S., Rich, S., Lima, J. A. C., Rotter, J. I., Philippakis, A., Lubitz, S. A., Ellinor, P. T., Khera, A. V., Kathiresan, S., & Aragam, K. G. (2020). Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nature Communications*, *11*(1), 2254. https://doi.org/10.1038/s41467-020-15823-7

Pozo, A. L., Godfrey, E. M., & Bowles, K. M. (2009). Splenomegaly: investigation, diagnosis and management. *Blood Reviews*, *23*(3), 105–111. https://doi.org/10.1016/j.blre.2008.10.001

Raabe, M., Flynn, L. M., Zlot, C. H., Wong, J. S., Véniant, M. M., Hamilton, R. L., & Young, S. G. (1998). Knockout of the abetalipoproteinemia gene in mice: reduced lipoprotein secretion in heterozygotes and embryonic lethality in homozygotes. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(15), 8686–8691. https://doi.org/10.1073/pnas.95.15.8686

Radmard, A. R., Merat, S., Kooraki, S., Ashraf, M., Keshtkar, A., Sharafkhah, M., Jafari, E., Malekzadeh, R., & Poustchi, H. (2015). Gallstone disease and obesity: a population-based study on abdominal fat distribution and gender differences. *Annals of Hepatology*, *14*(5),

702–709. https://www.ncbi.nlm.nih.gov/pubmed/26256899

Reeder, S. B., Pineda, A. R., Wen, Z., Shimakawa, A., Yu, H., Brittain, J. H., Gold, G. E., Beaulieu, C. H., & Pelc, N. J. (2005). Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): application with fast spin-echo imaging. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *54*(3), 636–644. https://doi.org/10.1002/mrm.20624

Regier, A. A., Farjoun, Y., Larson, D. E., Krasheninina, O., Kang, H. M., Howrigan, D. P., Chen, B.-J., Kher, M., Banks, E., Ames, D. C., English, A. C., Li, H., Xing, J., Zhang, Y., Matise, T., Abecasis, G. R., Salerno, W., Zody, M. C., Neale, B. M., & Hall, I. M. (2018). Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nature Communications*, *9*(1), 4038. https://doi.org/10.1038/s41467-018-06159-4

Richardson, T. G., Sanderson, E., Palmer, T. M., Ala-Korpela, M., Ference, B. A., Davey Smith, G., & Holmes, M. V. (2020). Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Medicine*, *17*(3), e1003062. https://doi.org/10.1371/journal.pmed.1003062

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–330. https://doi.org/10.1038/nature14248

Romeo, S., Kozlitina, J., Xing, C., Pertsemlidis, A., Cox, D., Pennacchio, L. A., Boerwinkle, E.,

Cohen, J. C., & Hobbs, H. H. (2008). Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nature Genetics*, *40*(12), 1461–1465. https://doi.org/10.1038/ng.257

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Schwenzer, N. F., Machann, J., Haap, M. M., Martirosian, P., Schraml, C., Liebig, G., Stefan, N., Häring, H.-U., Claussen, C. D., Fritsche, A., & Schick, F. (2008). T2* Relaxometry in Liver, Pancreas, and Spleen in a Healthy Cohort of One Hundred Twenty-Nine Subjects–Correlation With Age, Gender, and Serum Ferritin. *Investigative Radiology*, *43*(12), 854. https://doi.org/10.1097/RLI.0b013e3181862413

Sethi, A., Taylor, L., Ruby, G., Venkataraman, J., Cule, M., & Melamud, E. (2020). Calcification of abdominal aorta is an underappreciated cardiovascular disease risk factor. In *Epidemiology* (No. medrxiv;2020.05.07.20094706v1). medRxiv. https://doi.org/10.1101/2020.05.07.20094706

Shimomura, Y., Hara, M., Katoh, D., Hashimoto, H., & Ishikawa, T. (2018). Enlarged spleen is associated with low neutrophil and platelet engraftment rates and poor survival after allogeneic stem cell transplantation in patients with acute myeloid leukemia and myelodysplastic syndrome. In *Annals of Hematology* (Vol. 97, Issue 6, pp. 1049–1056). https://doi.org/10.1007/s00277-018-3278-9

Speliotes, E. K., Yerges-Armstrong, L. M., Wu, J., Hernaez, R., Kim, L. J., Palmer, C. D., Gudnason, V., Eiriksdottir, G., Garcia, M. E., Launer, L. J., Nalls, M. A., Clark, J. M., Mitchell, B. D., Shuldiner, A. R., Butler, J. L., Tomas, M., Hoffmann, U., Hwang, S.-J., Massaro, J. M., … GOLD Consortium. (2011). Genome-wide association analysis identifies

variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genetics*, *7*(3), e1001324. https://doi.org/10.1371/journal.pgen.1001324

Stender, S., Smagris, E., Lauridsen, B. K., Kofoed, K. F., Nordestgaard, B. G., Tybjaerg-Hansen, A., Pennacchio, L. A., Dickel, D. E., Cohen, J. C., & Hobbs, H. H. (2018). Relationship between genetic variation at PPP1R3B and levels of liver glycogen and triglyceride. *Hepatology* , *67*(6), 2182–2195. https://doi.org/10.1002/hep.29751

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9440–9445. https://doi.org/10.1073/pnas.1530509100

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Svennerholm, L., Boström, K., & Jungbjer, B. (1997). Changes in weight and compositions of major membrane components of human brain during the span of adult human life of Swedes. *Acta Neuropathologica*, *94*(4), 345–352. https://doi.org/10.1007/s004010050717

Taylor, R. (2008). Pathogenesis of type 2 diabetes: tracing the reverse route from cure to cause. *Diabetologia*, *51*(10), 1781–1789. https://doi.org/10.1007/s00125-008-1116-7

Taylor, R. (2013). Type 2 diabetes: etiology and reversibility. *Diabetes Care*, *36*(4), 1047–1055. https://doi.org/10.2337/dc12-1805

Terjung, R. (Ed.). (2013). Iron Homeostasis in the Liver. In *Comprehensive Physiology* (Vol. 275, p. 19906). John Wiley & Sons, Inc. https://doi.org/10.1002/cphy.c120016

Thomas, E. L., Fitzpatrick, J., Frost, G. S., & Bell, J. D. (2013). Metabolic syndrome, overweight and fatty liver. In C. D. Berdanier, J. T. Dwyer, & D. Heber (Eds.), *Handbook of nutrition and food, 3rd edition* (pp. 763–768). CRC Press. https://westminsterresearch.westminster.ac.uk/item/8z1w4/metabolic-syndrome-overweight-and-fatty-liver

Thomas, E. L., Parkinson, J. R., Frost, G. S., Goldstone, A. P., Doré, C. J., McCarthy, J. P., Collins, A. L., Fitzpatrick, J. A., Durighel, G., Taylor-Robinson, S. D., & Bell, J. D. (2012). The missing risk: MRI and MRS phenotyping of abdominal adiposity and ectopic fat. *Obesity* , *20*(1), 76–87. https://doi.org/10.1038/oby.2011.142

Thomas, E. L., Parkinson, J. R., Hyde, M. J., Yap, I. K. S., Holmes, E., Doré, C. J., Bell, J. D., & Modi, N. (2011). Aberrant adiposity and ectopic lipid deposition characterize the adult phenotype of the preterm infant. *Pediatric Research*, *70*(5), 507–512. https://doi.org/10.1203/PDR.0b013e31822d7860

Turnbull, C., Perdeaux, E. R., Pernet, D., Naranjo, A., Renwick, A., Seal, S., Munoz-Xicola, R. M., Hanks, S., Slade, I., Zachariou, A., Warren-Perry, M., Ruark, E., Gerrard, M., Hale, J., Hewitt, M., Kohler, J., Lane, S., Levitt, G., Madi, M., … Rahman, N. (2012). A genome-wide association study identifies susceptibility loci for Wilms tumor. *Nature Genetics*, *44*(6), 681–684. https://doi.org/10.1038/ng.2251

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Wakefield, J. (2007). A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *American Journal of Human Genetics*, *81*(2), 208–227. https://doi.org/10.1086/519024

Wellcome Trust Case Control Consortium, Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J. M. M., Auton, A., Myers, S., Morris, A., Pirinen, M., Brown, M. A., Burton, P. R., Caulfield, M. J., Compston, A., Farrall, M., Hall, A. S., Hattersley, A. T., … Donnelly, P. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genetics*, *44*(12), 1294–1301. https://doi.org/10.1038/ng.2435

West, J., Dahlqvist Leinhard, O., Romu, T., Collins, R., Garratt, S., Bell, J. D., Borga, M., & Thomas, E. L. (2016). Feasibility of MR-Based Body Composition Analysis in Large Scale Population Studies. *PloS One*, *11*(9), e0163332. https://doi.org/10.1371/journal.pone.0163332

Wilman, H. R., Kelly, M., Garratt, S., Matthews, P. M., Milanesi, M., Herlihy, A., Gyngell, M., Neubauer, S., Bell, J. D., Banerjee, R., & Thomas, E. L. (2017). Characterisation of liver fat in the UK Biobank cohort. In *PLOS ONE* (Vol. 12, Issue 2, p. e0172921). https://doi.org/10.1371/journal.pone.0172921

Wilman, H. R., Parisinos, C. A., Atabaki-Pasdar, N., Kelly, M., Thomas, E. L., Neubauer, S., IMI DIRECT Consortium, Mahajan, A., Hingorani, A. D., Patel, R. S., Hemingway, H., Franks, P. W., Bell, J. D., Banerjee, R., & Yaghootkar, H. (2019). Genetic studies of abdominal MRI data identify genes regulating hepcidin as major determinants of liver iron concentration. *Journal of Hepatology*, *71*(3), 594–602. https://doi.org/10.1016/j.jhep.2019.05.032

Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., Highland, H. M., Patel, Y. M., Sorokin, E. P., Avery, C. L., Belbin, G. M., Bien, S. A., Cheng, I., Cullina, S., Hodonsky, C. J., Hu, Y., Huckins, L. M., Jeff, J., Justice, A. E., … Carlson, C. S. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature*, *570*(7762), 514–518. https://doi.org/10.1038/s41586-019-1310-4

Wood, J. C., Enriquez, C., Ghugre, N., Tyzka, J. M., Carson, S., Nelson, M. D., & Coates, T. D.

(2005). MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood*, *106*(4), 1460–1465. https://doi.org/10.1182/blood-2004-10-3982

Wuttke, M., Li, Y., Li, M., Sieber, K. B., Feitosa, M. F., Gorski, M., Tin, A., Wang, L., Chu, A. Y., Hoppmann, A., Kirsten, H., Giri, A., Chai, J.-F., Sveinbjornsson, G., Tayo, B. O., Nutile, T., Fuchsberger, C., Marten, J., Cocca, M., … Pattaro, C. (2019). A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nature Genetics*, *51*(6), 957–972. https://doi.org/10.1038/s41588-019-0407-x

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden, P. A. F., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., Frayling, T. M., McCarthy, M. I., Hirschhorn, J. N., Goddard, M. E., & Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*, *44*(4), 369–375, S1–S3. https://doi.org/10.1038/ng.2213

Yu, H., McKenzie, C., Shimakawa, A., Reeder, S., & Brittain, J. (2008). Bipolar Multi-Echo Water-Fat Separation: Phase Correction Using Parallel Imaging. *Proceedings of the 16th Annual Meeting of ISMRM, Toronto, Canada*, 648. https://cds.ismrm.org/ismrm-2008/files/00648.pdf

Zhang, B.-L., He, N., Huang, Y.-B., Song, F.-J., & Chen, K.-X. (2014). ABO blood groups and risk of cancer: a systematic review and meta-analysis. *Asian Pacific Journal of Cancer Prevention: APJCP*, *15*(11), 4643–4650. https://doi.org/10.7314/apjcp.2014.15.11.4643

Zhang, H., Mooney, C. J., & Reilly, M. P. (2012). ABO Blood Groups and Cardiovascular Diseases. *International Journal of Vascular Medicine*, *2012*, 641917.

https://doi.org/10.1155/2012/641917

Zhang, Y., & Yang, Q. (2017). A Survey on Multi-Task Learning. In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1707.08114

Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., LeFaive, J., VandeHaar, P., Gagliano, S. A., Gifford, A., Bastarache, L. A., Wei, W.-Q., Denny, J. C., Lin, M., Hveem, K., Kang, H. M., Abecasis, G. R., Willer, C. J., & Lee, S. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*, *50*(9), 1335–1341. https://doi.org/10.1038/s41588-018-0184-y

Zhou, W., Zhao, Z., Nielsen, J. B., Fritsche, L. G., LeFaive, J., Gagliano Taliun, S. A., Bi, W., Gabrielsen, M. E., Daly, M. J., Neale, B. M., Hveem, K., Abecasis, G. R., Willer, C. J., & Lee, S. (2020). Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nature Genetics*, *52*(6), 634–639. https://doi.org/10.1038/s41588-020-0621-6

Zhu, Z., Guo, Y., Shi, H., Liu, C.-L., Panganiban, R. A., Chung, W., O'Connor, L. J., Himes, B. E., Gazal, S., Hasegawa, K., Camargo, C. A., Jr, Qi, L., Moffatt, M. F., Hu, F. B., Lu, Q., Cookson, W. O. C., & Liang, L. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *The Journal of Allergy and Clinical Immunology*, *145*(2), 537–549. https://doi.org/10.1016/j.jaci.2019.09.035

Acknowledgements

We thank Adam Baker, Garret Fitzgerald, Frank Li, Anil Raj, and Amoolya Singh for input on the manuscript, and Leland Taylor for writing the genetic analysis pipeline used in this manuscript. We thank Stefan Stender for feedback on a draft. This study was carried out using UK Biobank

Application number 44584, and we thank the participants in the UK Biobank imaging study. This study was funded by Calico Life Sciences LLC.

Author contributions

JB, MC, YL, ELT, and NvB conceived the study. BW designed the annotation protocol and supervised annotation. NB and BW designed and implemented the image processing pipelines, and MC executed the pipelines. NB and YL conceived, implemented, and executed segmentation models. MC and ES performed statistical analysis. JB, NB, MC, YL, ES, ELT, and BW interpreted results. JB, NB, MC, YL, ELT, and BW wrote the first draft of the manuscript. All authors edited the final manuscript.

Data Availability

Summary statistics from all genome-wide association studies will be made available from NHGRI-EBI GWAS Catalog. IDPs will be available via application to the UKBB at www.ukbiobank.ac.uk.

Code Availability

MATLAB code to estimate the PDFF is available from Dr Mark Bydder at github.com/marcsous/pdff.

Code to preprocess the imaging data will be made available upon publication from github.com/recoh/pipeline. Fitted models and code to apply the models will be available upon publication from https://github.com/calico/ukbb-mri-sseg/.

Competing interests

MC, YL, ES and NvB are employees of Calico Life Sciences LLC. This work was funded by Calico Life Sciences LLC.

Materials & Correspondence

Correspondence should be addressed to MC (cule@calicolabs.com)

Supplementary Tables

Supplementary Table 1: Segmentation performance metrics.

Supplementary Table 2: Significant PheWAS associations. Only associations which are statistically significant after correction for multiple testing are shown.

Supplementary Table 3: Significant PHESANT associations. Only associations which are statistically significant after correction for multiple testing are shown.

Supplementary Table 4: Genomic control parameter for each trait.

Supplementary Table 5: LDSC intercept.

Supplementary Table 6: Genetic correlations between abdominal IDPs.

Supplementary Table 7: Genetic correlation between abdominal IDPs and other heritable complex traits. Only associations which are statistically significant after correction for multiple testing are shown.

Supplementary Table 8: Genome-wide significant lead SNPs. Columns are as follows

- trait: One of: volume, fat or iron
- organ: Organ
- var_index Index variant (in the format chr:pos:ref:alt:build (All index variants are listed in GRCh37 coordinates)
- rs_id: dbSNP ID
- var_conditional: If a conditional signal, variants conditioned on, in the same format as var_index
- pv P-value
- pp: Probability that the lead SNP is the causal variant
- beta: Effect size (in standard deviations)
- closest_gene: Closest protein-coding gene
- closest_gene_dist: Distance to TSS of closest gene

Supplementary Table 9: Significant colocalization with complex trait GWAS signals.

Supplementary Table 10: Significant colocalization with gene expression

Supplementary Figures

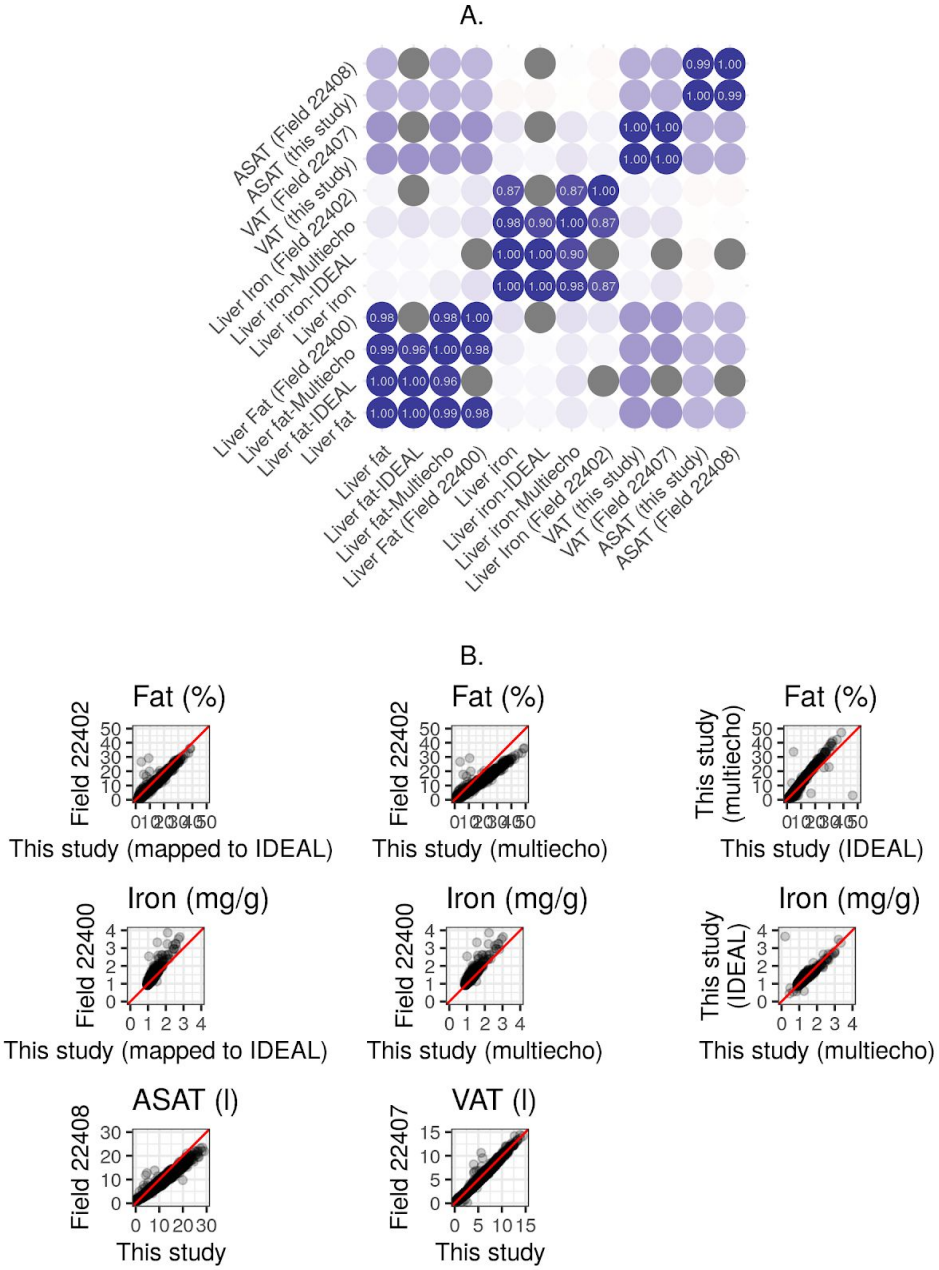

Figure S1. A. Correlation between multiple measurements of liver fat, liver iron, ASAT volume, and VAT volume in the UK Biobank. B. Scatter plots showing the relationship between multiple measurements of liver fat, liver iron, ASAT volume, and VAT volume in the UK Biobank.

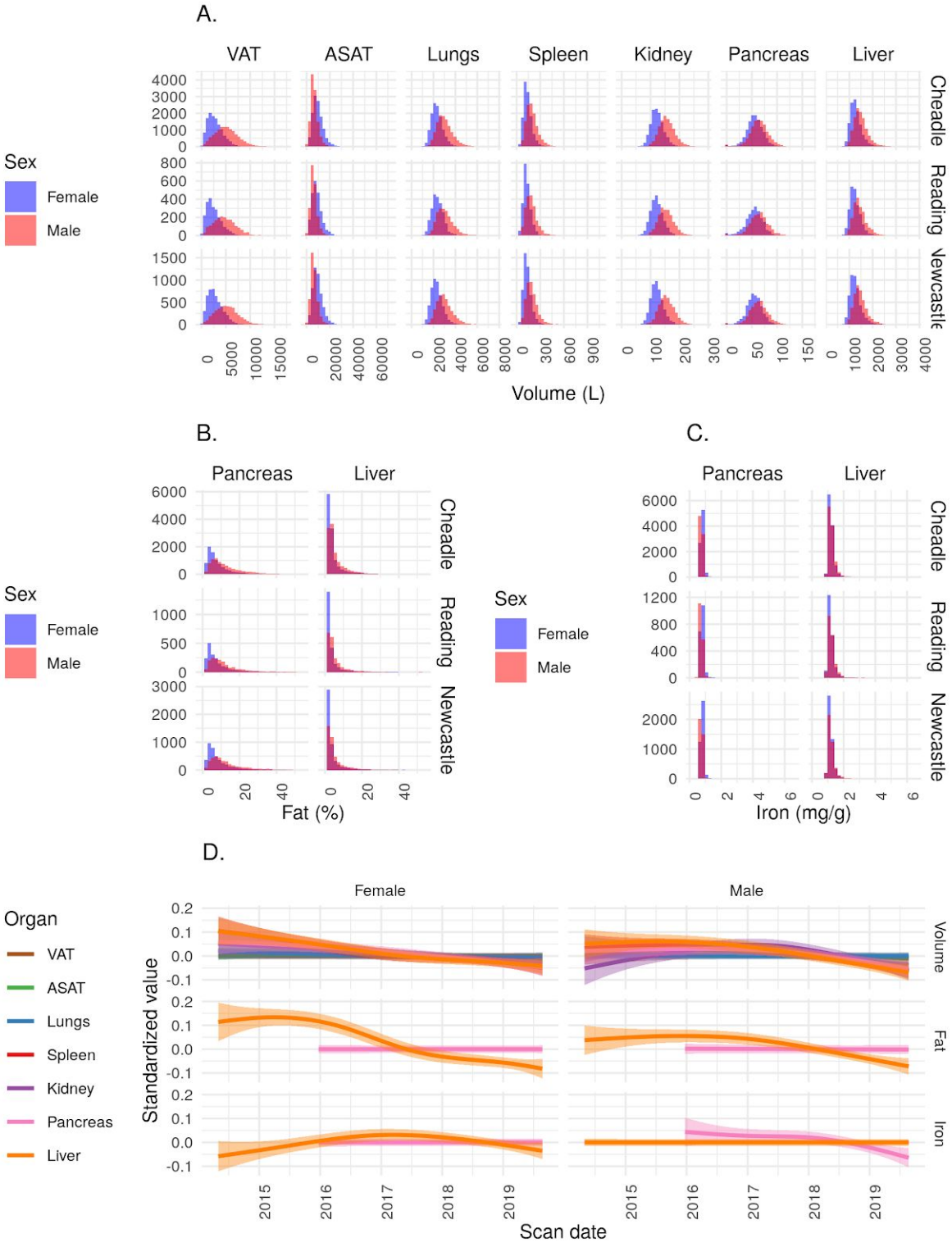

Figure S2. A. Organ volume IDPs, split by imaging centre. B. Fat IDPs, split by imaging centre. C. Iron IDPs, split by imaging centre. D. Relationship between scan date and IDPs.

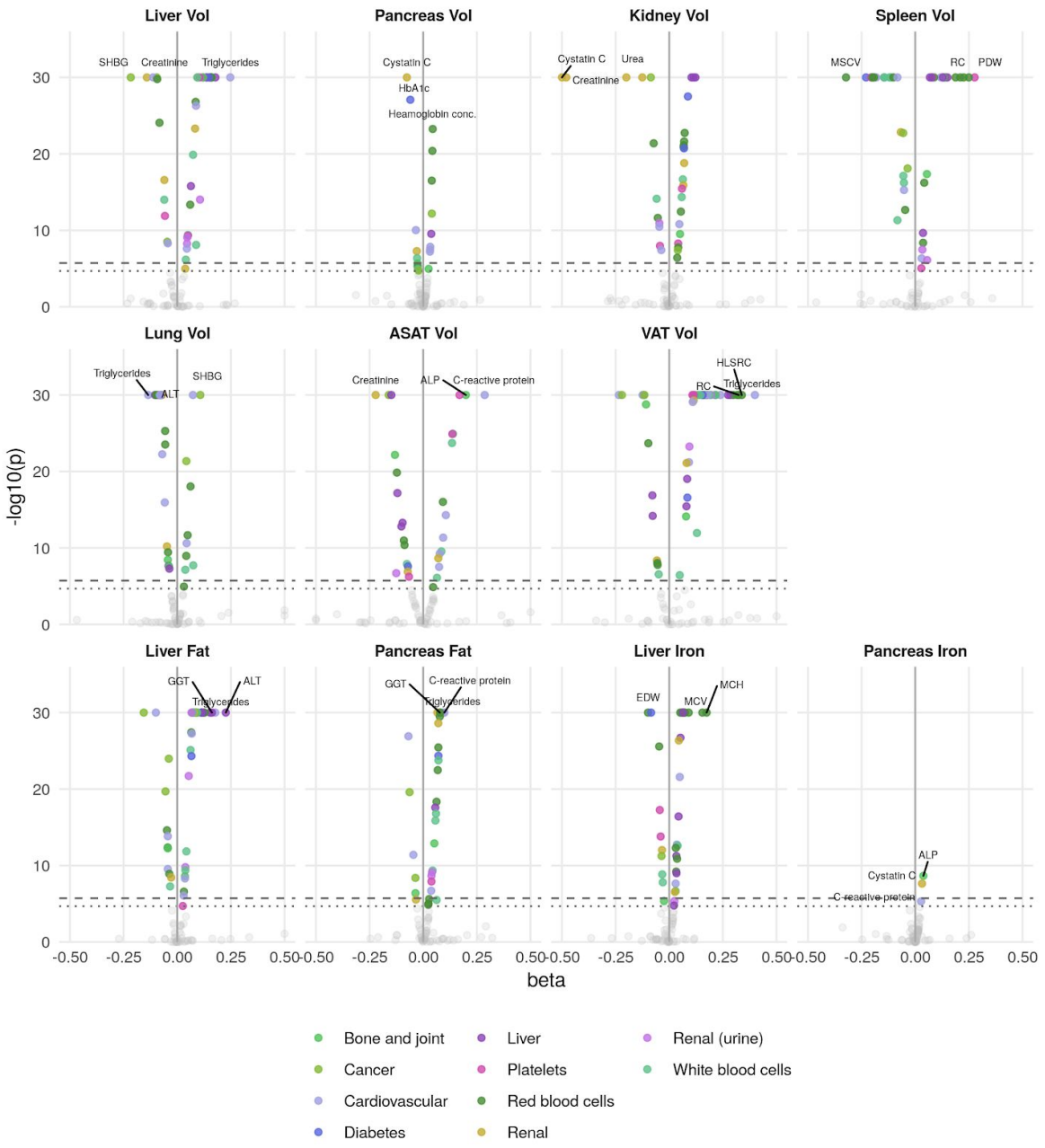

Figure S3. Phenome-wide associations across all IDPs and 83 biomarkers. The x-axis gives the effect size per standard deviation, and the y-axis -log10(p-value). The top 3 associations for each phenotype are labelled. Horizontal lines at phenome-wide significance (dotted line, p=2.7e-05) and study-wide significance (dashed line, p=2.48e-06) after Bonferroni correction for the total number of measures. Abbreviations: SHBG: Sex hormone binding globulin. MSCV: Mean sphered cell volume. MCH: Mean corpuscular hemoglobin. RC: Reticulocyte count. PDW: Platelet distribution width. ALT: alanine transaminase. ALP: Alkaline phosphatase. HLSRC: High light scatter reticulocyte count. GGT: Gamma glutamyl transferase.

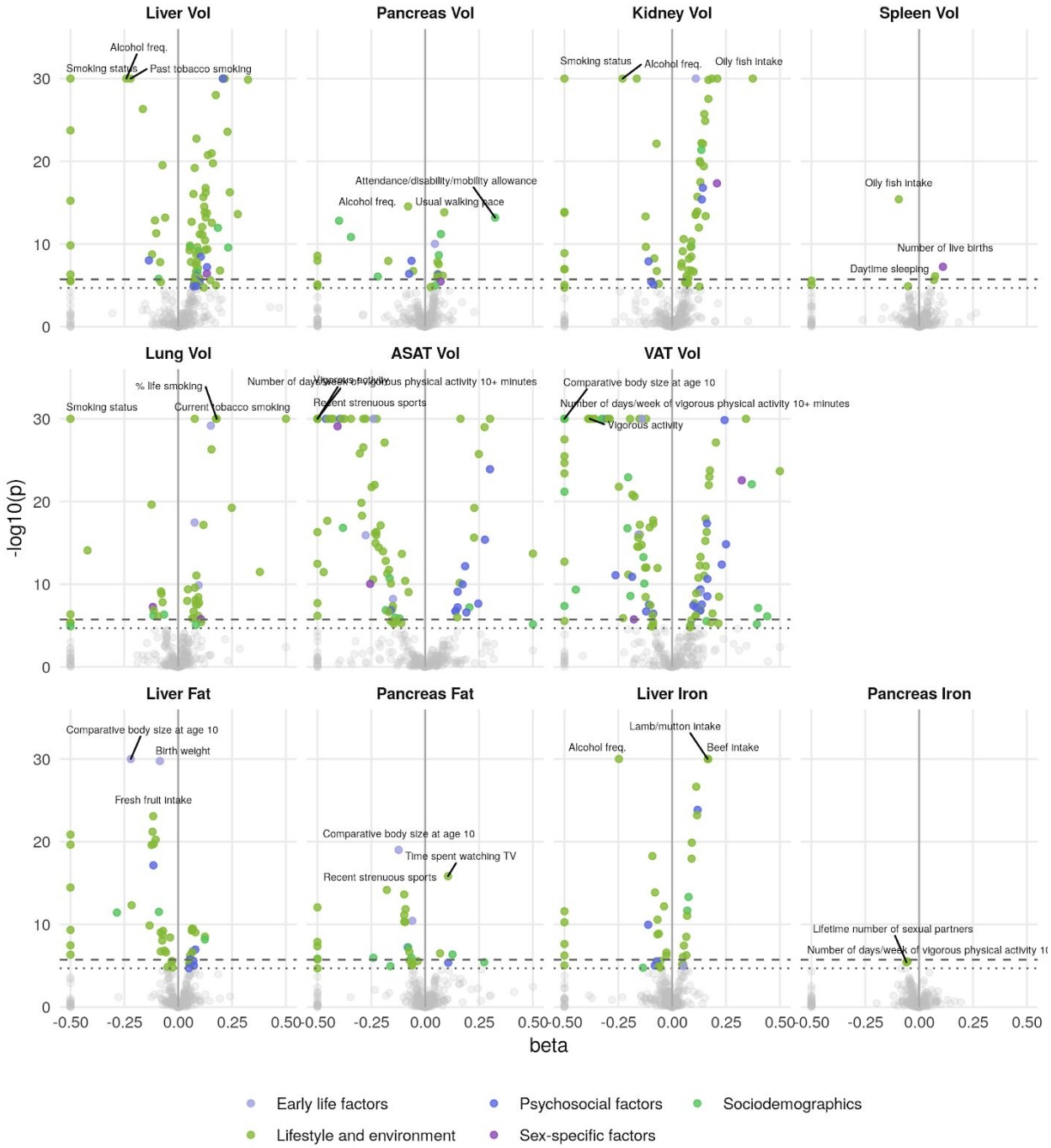

Figure S4. Phenome-wide associations across all IDPs and 199 lifestyle and history traits. The x-axis gives the effect size per standard deviation, and the y-axis -log10(p-value). The top 3 associations for each phenotype are labelled. Horizontal lines at phenome-wide significance (dotted line, p=2.7e-05) and study-wide significance (dashed line, p=2.48e-06) after Bonferroni correction for the total number of measures.

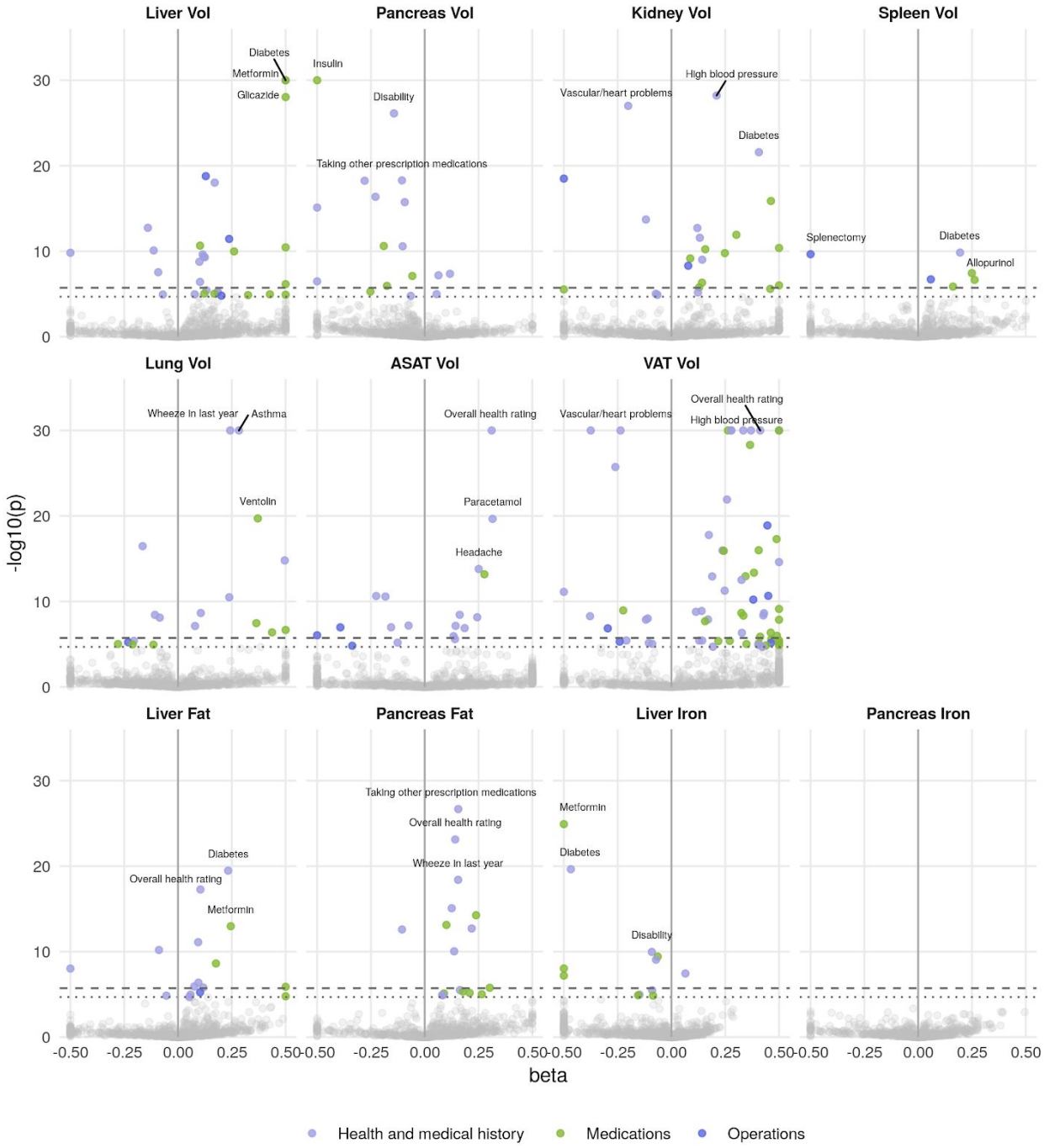

Figure S5. Phenome-wide associations across all IDPs and 770 medical history traits. The x-axis gives the effect size per standard deviation, and the y-axis -log10(p-value). The top 3 associations for each phenotype are labelled. Horizontal lines at phenome-wide significance (dotted line, p=2.7e-05) and study-wide significance (dashed line, p=2.48e-06) after Bonferroni correction for the total number of measures.

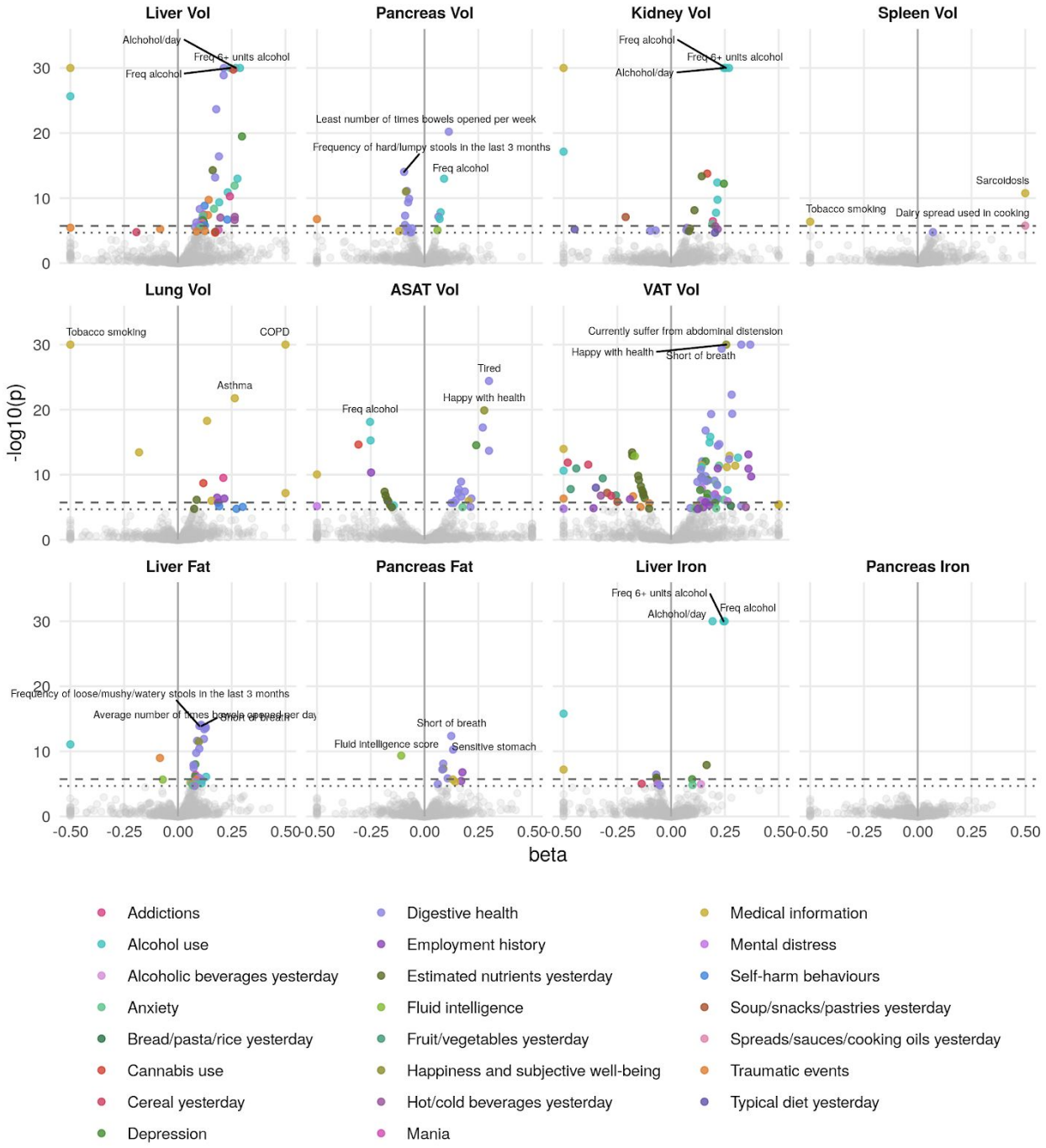

Figure S6. Phenome-wide associations across all IDPs and 444 traits measured in online follow-up. The x-axis gives the effect size per standard deviation, and the y-axis -log10(p-value). The top 3 associations for each phenotype are labelled. Horizontal lines at phenome-wide significance (dotted line, p=2.7e-05) and study-wide significance (dashed line, p=2.48e-06) after Bonferroni correction for the total number of measures.

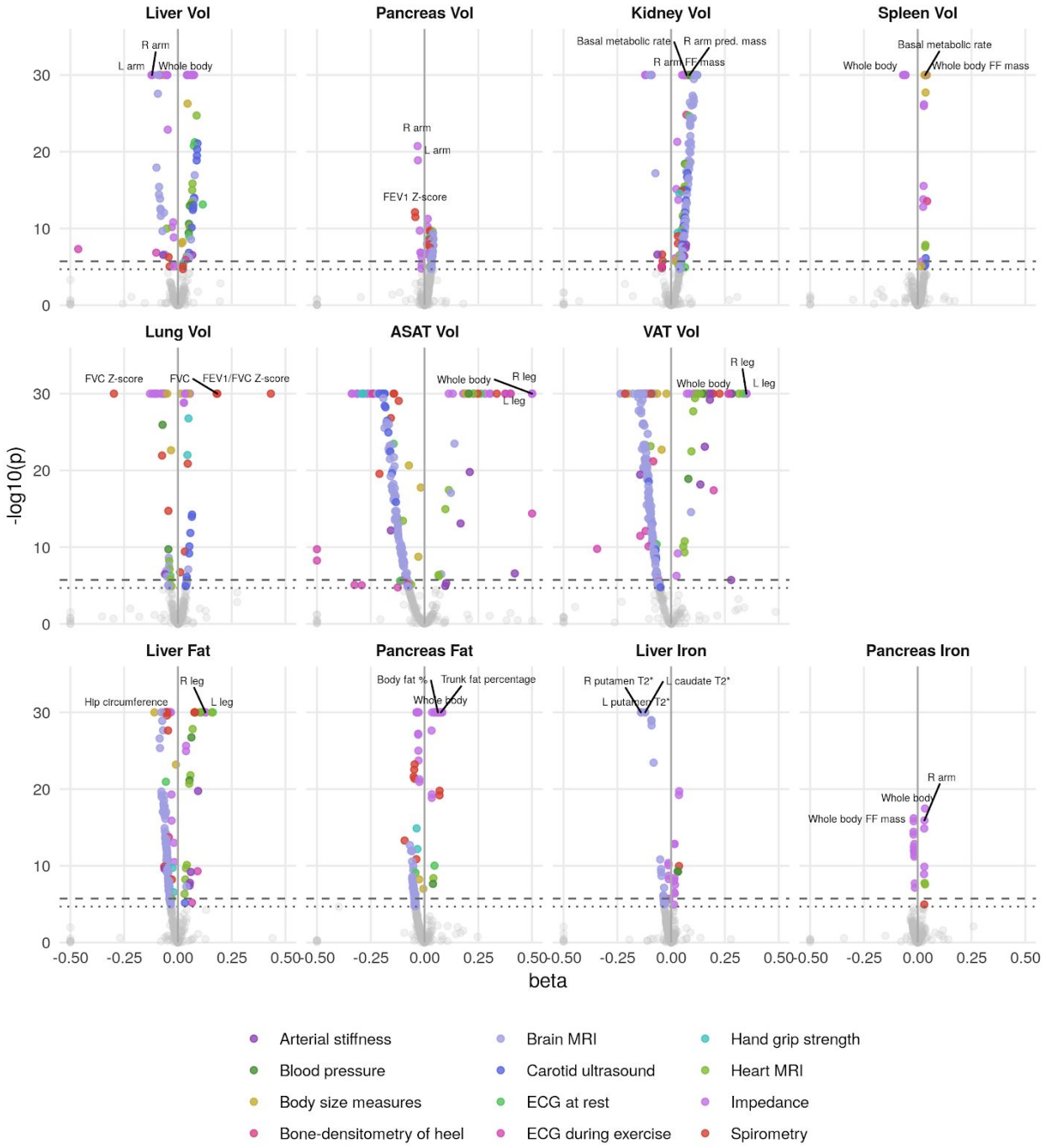

Figure S7. Phenome-wide associations across all IDPs and 335 physical measures. The x-axis gives the effect size per standard deviation, and the y-axis -log10(p-value). The top 3 associations for each phenotype are labelled. Horizontal lines at phenome-wide significance (dotted line, p=2.7e-05) and study-wide significance (dashed line, p=2.48e-06) after Bonferroni correction for the total number of measures. Abbreviations: FVC forced vital capacity. FEV1 Forced expiratory volume in 1 second. FF fat-free.

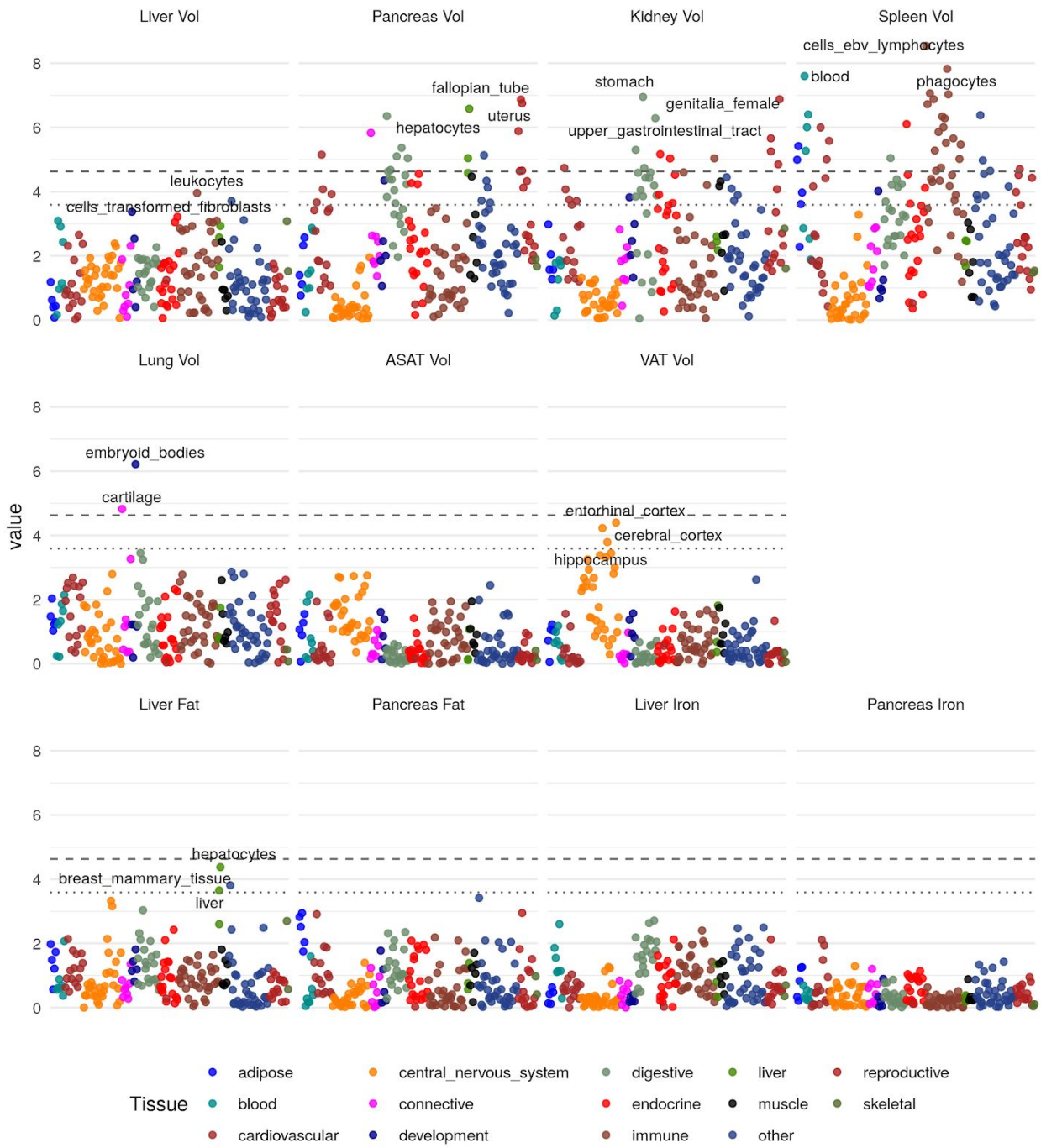

Figure S8. Heritability enrichment in tissues and cell types for annotations based on gene expression (see Methods). The top 3 enrichments for each phenotype passing a trait-wide significance threshold are labelled. Horizontal lines and trait-wide (dotted line) and study-wide (dashed line) significance after Bonferroni correction.

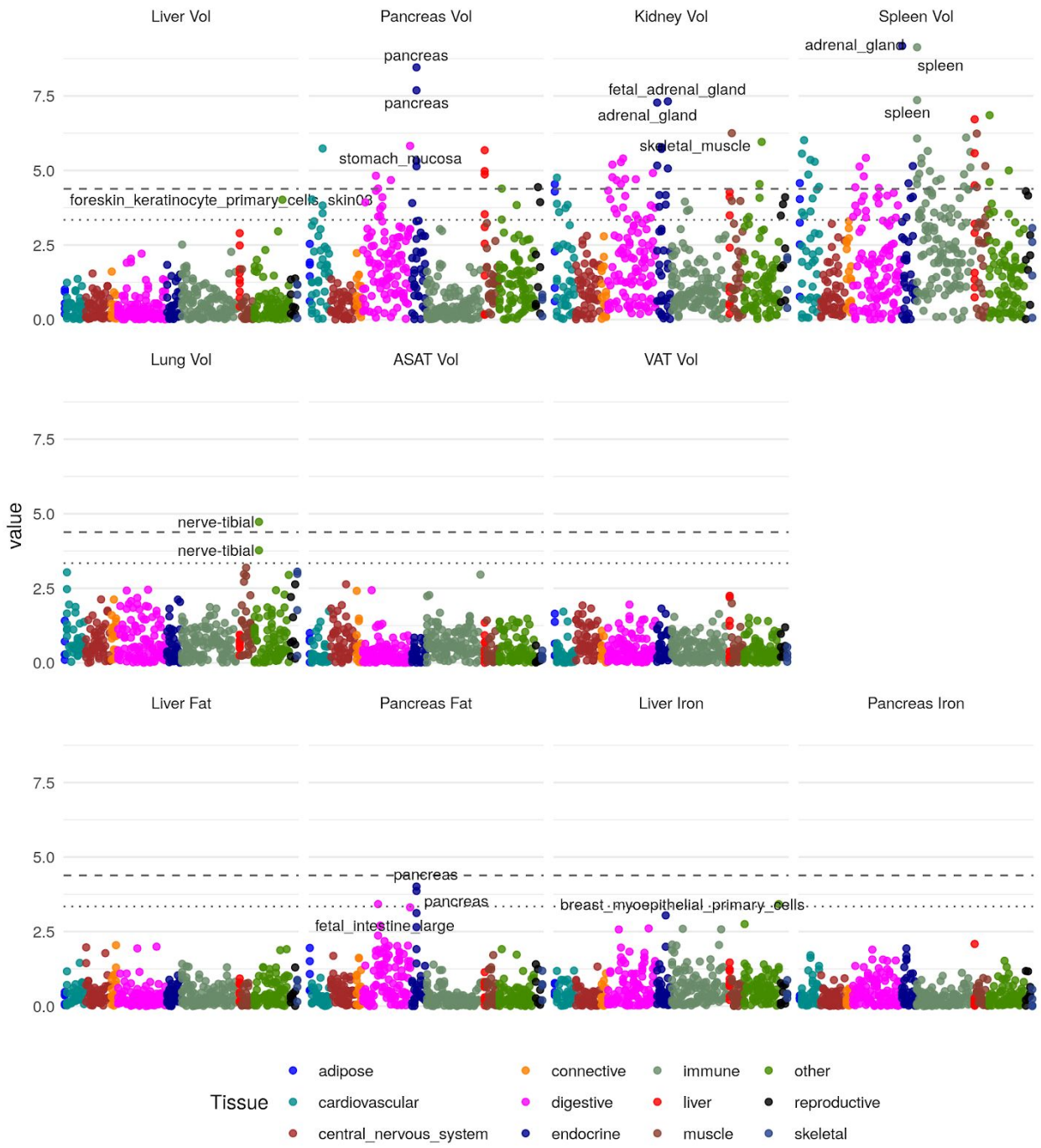

Figure S9. Heritability enrichment in tissues and cell types for annotations based on chromatin accessibility (see Methods). The top 3 enrichments for each phenotype passing a trait-wide significance threshold are labelled. Horizontal lines and trait-wide (dotted line) and study-wide (dashed line) significance after Bonferroni correction.

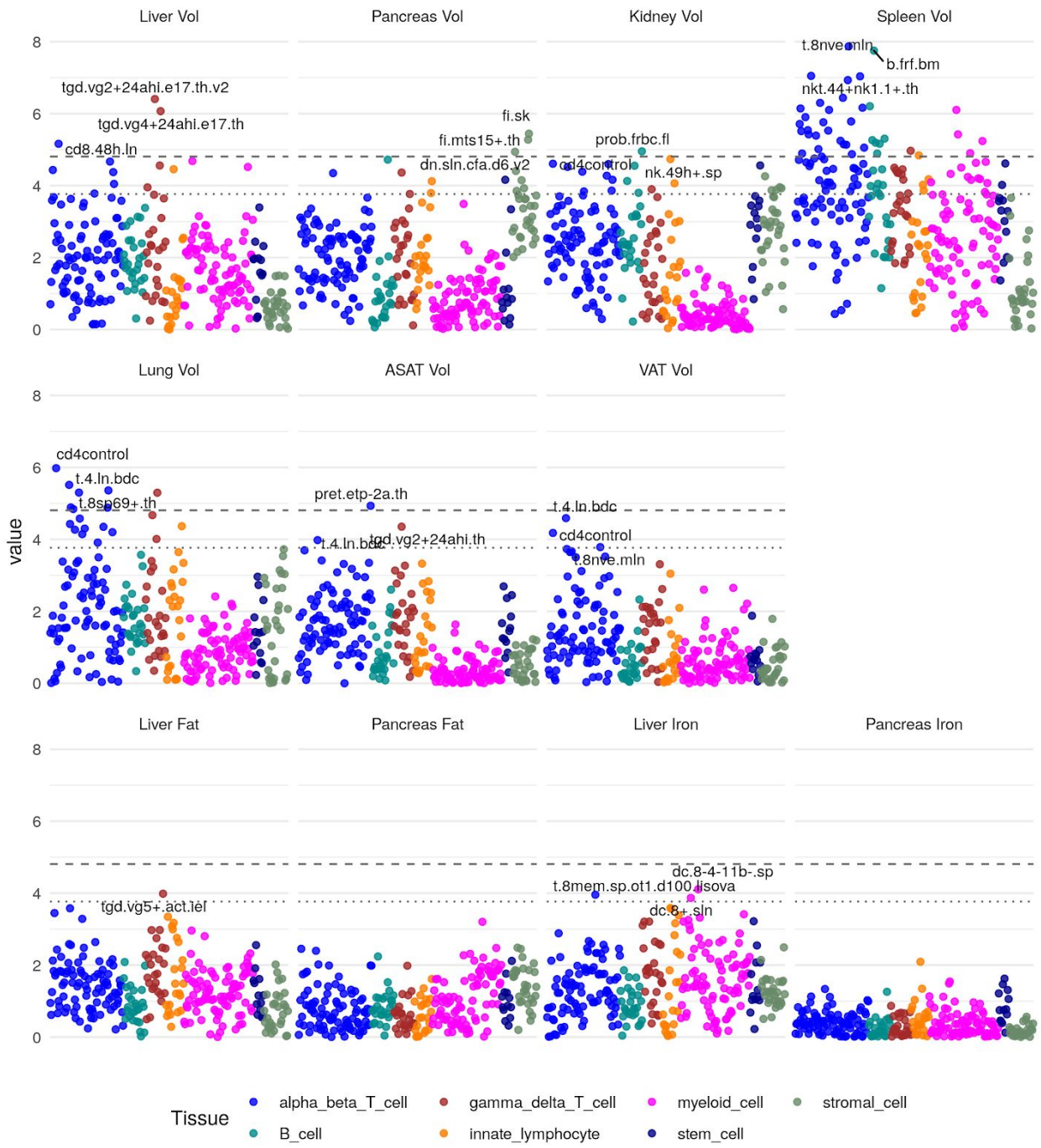

Figure S10. Heritability enrichment in tissues and cell types in immune cell types (see Methods). The top 3 enrichments for each phenotype passing a trait-wide significance threshold are labelled. Horizontal lines and trait-wide (dotted line) and study-wide (dashed line) significance after Bonferroni correction.

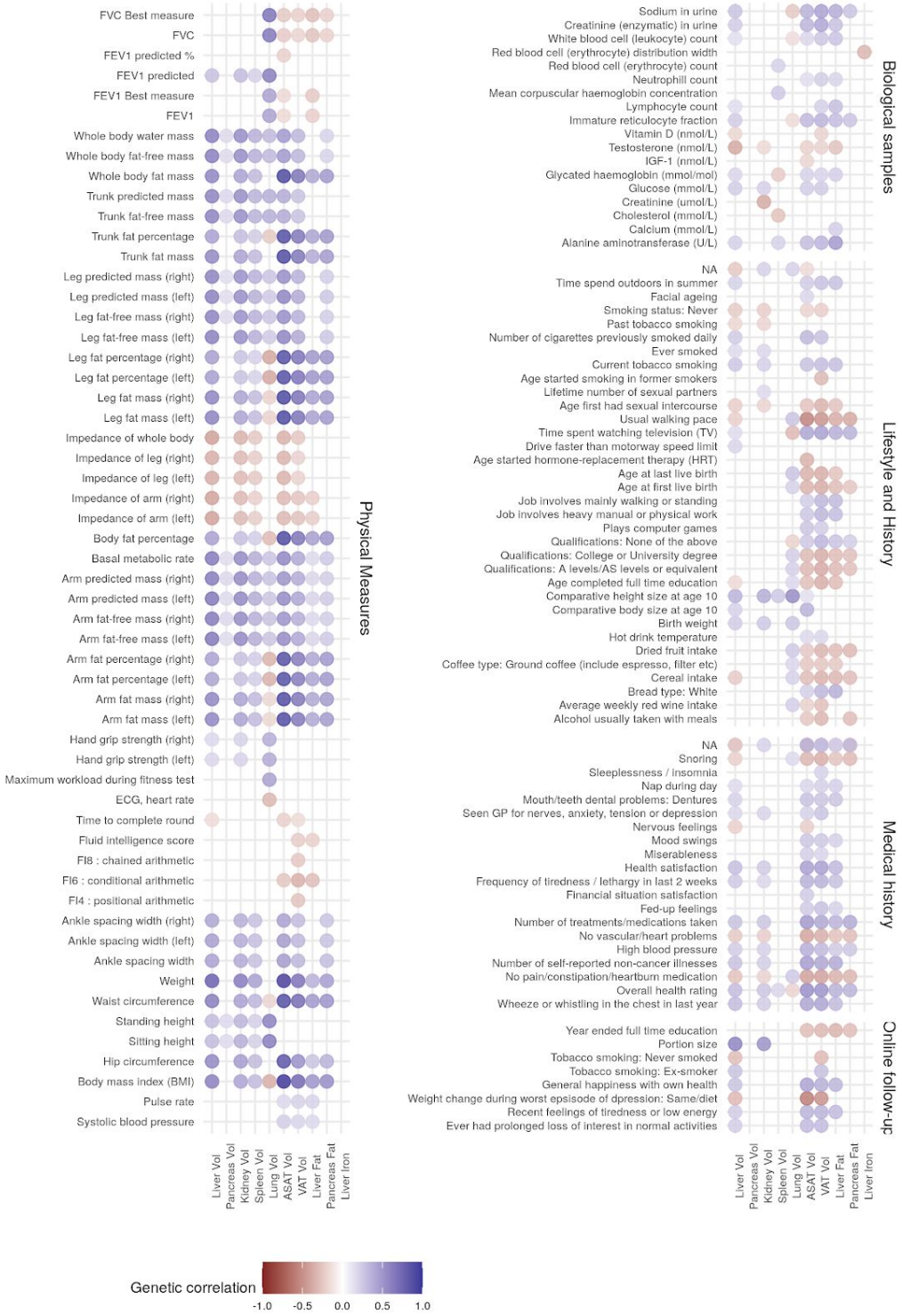

Figure S11. Genetic correlation between IDPs and complex traits. Only IDPs and traits with statistically significant genetic correlation ($p < 1.61e-05$ after Bonferroni correction for multiple testing) are shown.

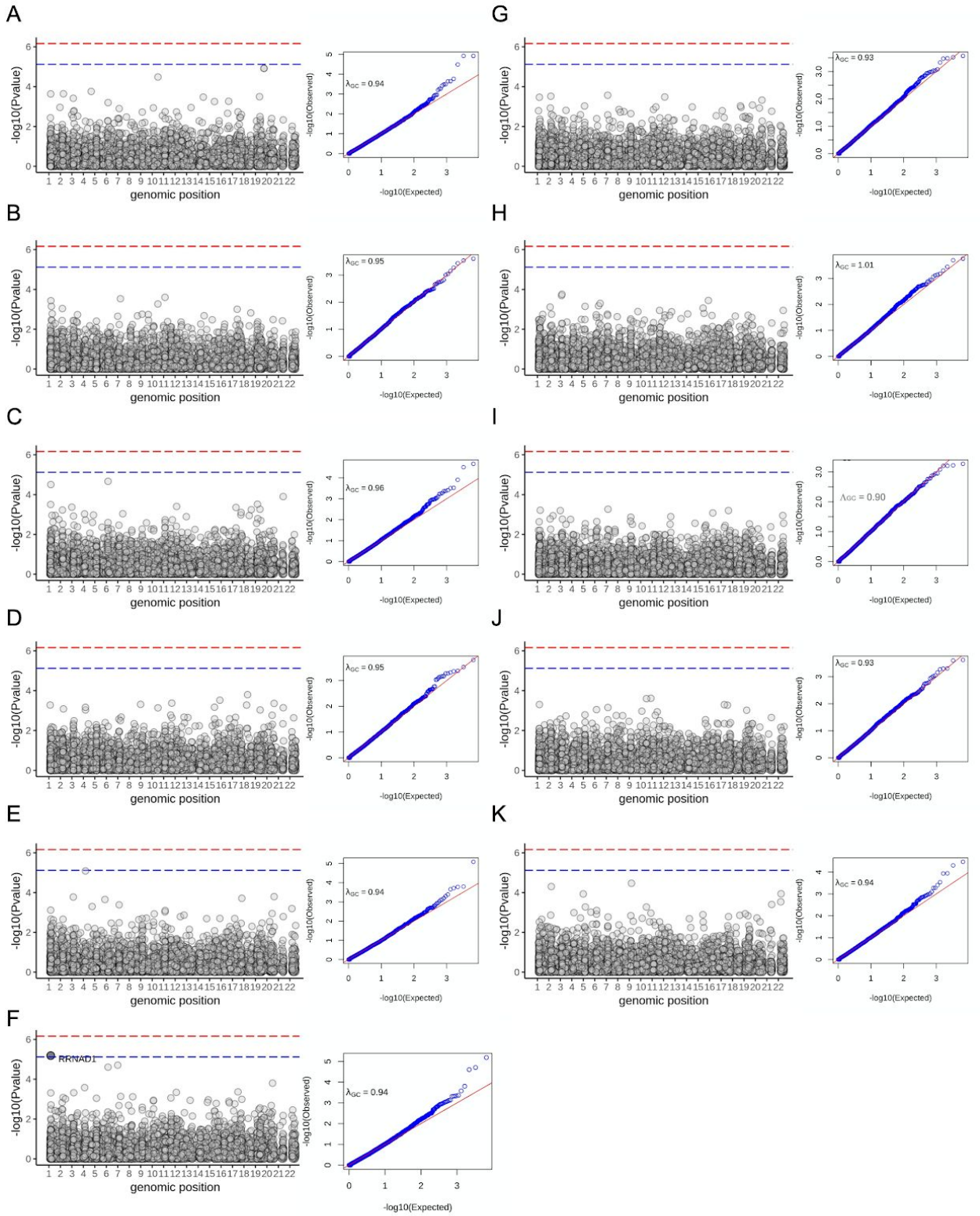

Figure S12. Rare association studies in the subcohort with both exome sequence data and

imaging-derived quantitative phenotypes. Left: Manhattan plot shows the association between each gene organised by genomic coordinates. Right: QQ-plot showing calibration of SKAT-O test statistics. λ_{GC}: Genomic control parameter for each trait. Blue dashed line indicates Bonferroni significance threshold genome-wide (p = 7.4e-06). Red dashed line indicates overall study significance threshold (p = 6.7e-07). (A) volume of visceral fat (n=11,069 samples) (B) volume of spleen (n=11,134) (C) volume of the lungs (n=11,134) (D) liver volume (n=11,134) (E) kidney volume (n=11,134) (F) abdominal subcutaneous fat (n=11,134). One gene achieved genome-wide significance but not study wide significance (*RRNAD1*: $p_{SKAT\text{-}O}$ = 6.5e-06; $beta_{burden}$ = -0.08).(G) pancreas volume (n=11,093) (H) pancreas iron level (n=5,525) (I) liver iron (n=11,069) (J) pancreatic fat (n=5525) (K) liver fat (n=11,069).