Systematic quantification of health parameters from UK Biobank abdominal MRI using deep learning

Yi Liu[1], Nicolas Basty[2], Brandon Whitcher[2], Jimmy D Bell[2], 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 population health burden and while several well established socioeconomic, environmental, behavioural, and genetic risk factors have been identified, our understanding of their drivers and mechanisms remains incomplete. Thus, a better understanding of these factors is required for the development of more effective 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 deep learning to segment individual organs from up to 38,683 abdominal MRI scans in the UK Biobank, we demonstrate that image derived phenotypes such as volume, fat and iron content reflect overall organ health. We further show that these traits have a substantial heritable component which is enriched in organ-specific cell types. We also identify several novel genome-wide significant associations. Overall our work demonstrates the feasibility and power of high-throughput MRI for the multi-organ study of cardiometabolic disease, health, and ageing.

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[1]. 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[2].

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)[3]. The abdominal 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[4–6], with some limited studies focussed on liver iron (n=8,289) and MRI-based corrected T1 (n=14,440)[7,8]. 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[9], to automatically segment key organs from the UKBB MRI data from 38,683 subjects (Table 1A) and quantified fat and iron content where suitable sequences were available (Figure 1, Supplementary Table 1, and Methods). In total, we defined 11 Image Derived Phenotypes (IDPs), including volume of the liver, pancreas, kidneys, spleen, lungs, VAT, and abdominal subcutaneous adipose tissue (ASAT), as well as 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.

Results

Characterisation of IDPs in the UK Biobank population

Our measures show correlations with previously published by others utilizing the same, albeit much smaller, UKBB data: liver fat (r=0.98), liver iron (r=0.87), VAT (r=1.00) and ASAT (r=0.99) (Methods; Extended Data Figure 1). The distribution of each organ-specific measure is summarised in Figure 1E, F, and G and Table 1B. 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, -24.4ml or -0.0067s.d./year for ASAT, and +24.3ml or 0.011s.d/year for ASAT). 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 find excess liver iron (>1.8m/g/g) in 3.22% of men and 1.75% of women.

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[10]. 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). We found evidence that IDPs reflect organ function and health. Liver volume was significantly associated with chronic liver disease (p=6.36e-07, beta=0.434) and cirrhosis (p=2.47e-06, beta=0.389) as well as T2D (p=1.81e-87, beta=0.704) and hypertension (p=2.48e-16, beta=0.177). Kidney volume was associated with chronic kidney disease (CKD) (p=4.66e-22, beta=-1.02). Interestingly, pancreas volume was associated most strongly with Type 1 diabetes (T1D) (p=2.27e-26, beta=-0.861), while the effect size on T2D was smaller (p=3.16e-18, beta=-0.265). Lung volume was most strongly associated with tobacco use (p=5.83e-46, beta=0.493), chronic obstructive pulmonary disease (COPD) (p=9.09e-36, beta=0.612), and bronchitis (p=1.01e-10, beta=0.74) , with larger lung volume corresponding to a greater likelihood of respiratory disease diagnosis. Spleen volume was associated with myeloproliferative disease (p=2.86e-33, beta=0.737), especially chronic lymphocytic leukaemia (p=5.03e-24, beta=0.782). Liver fat was associated with T2D (p=2.32e-33, beta=0.283). Liver iron was associated with T2D (p=8.2e-20, beta=-0.449) and disorders of iron metabolism (p=3e-12, beta=-0.445).

VAT was associated with a wide range of cardiometabolic outcomes including hypertension (p=1.36e-47, beta=0.38), T2d (p=1.36e-39, beta=0.644), and lipid metabolism disorders (p=2.36e-33, beta=0.416), , while ASAT was only associated with gallstones (p=1.9e-08, beta=0.38) . This association remained statistically significant, after adjusting for VAT, counter to reports that only VAT is predictive of gallstones[11]. 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, Extended Data Figure 3).

Again, we observed strong correlations between IDPs and biomarkers reflective of organ function. Liver volume was associated with triglycerides (p=1.33e-162, beta=0.193) and sex hormone binding globulin (SHBG) (p=1.55e-135, beta=-0.167). Kidney volume was associated with circulating cystatin C (p<1e-300, beta=-0.466), creatinine (p<1e-300, beta=-0.428) and urea in urine (p=1.67e-123, beta=-0.168, confirming observations that smaller kidneys function less effectively[12]. Pancreas volume was associated with glycated haemoglobin (HbA1c) (p=2.32e-27, beta=-0.060), but the association with glucose was not statistically significant (p=2.51e-05, beta=-0.0251). The strongest association with kidney volume was cystatin C (p=5.53e-53, beta=-0.0772). Spleen volume was associated with multiple hematological measurements, including reticulocyte count (p<1e-300, beta=0.233), mean sphered cell volume (p<1e-300, beta=-0.310) and platelet distribution width (p<1e-300, beta=0.267).

Liver fat was associated with multiple biomarkers including triglycerides (p=5.95e-225, beta-0.180), SHBG (p=2.87e-195, beta=-0.158) alanine aminotransferase (p<1e-300, beta=0.226, and gamma glutamyltransferase (p=9.51e-193, beta=0.160). Consistent with disease outcomes,which showed a correlation between hepatic iron, but not pancreatic iron, and disorders of iron metabolism, liver iron levels were correlated with measures of iron in the blood (e.g. mean corpuscular haemoglobin (MCH), p=1.07e-247 beta=0.178), while pancreatic iron did not show any such association (MCH p=0.511, beta=-0.00421).

Correlation with lifestyle factors and exposures

We 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 (Extended Data Figures 4-7).

Consistent with previous reports[13], we found that liver iron was associated with lower alcohol consumption (p=1.2e-122, beta=-0.254) and higher intake of red meat (p=2.7e-65, beta=0.172). Liver iron was also associated with suppressed T2* derived from neuroimaging in the same UKBB cohort[6], particularly in the putamen (left: p=2.1e-71, beta=-0.141; right: p=8.1e-74, beta=-0.145). There were no such associations for pancreatic iron (left: p=0.29, beta=0.00997, right: p=0.23, beta=0.0114). Additionally, we found that liver fat was associated with lower birth weight (p=1.2e-30, beta=-0.0849) and body size at age 10 (p=6.9e-77, beta=-0.221). Low birth weight has previously been associated with severity of pediatric non-alcoholic steatohepatitis (NASH)[14], abnormal fat distribution[15] and liver fat levels in adults born prematurely[16].

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 (Extended Data 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[17]. 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 IPDs, 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[18] (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)[19] (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.

Organ volume, fat, and iron are heritable

For each IDP, we estimated SNP-heritability using the BOLT-REML model[20] (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.

Genetic correlation between abdominal IDPs

To understand the extent that genetic variation underlying variability in IDPs is shared, we used bivariate LD score regression[21] to estimate the genetic correlation between all 55 IDP pairs, with and without including height and BMI as covariates (Methods). We found a statistically significant non-zero genetic correlation between 24 of the 55 unadjusted IDP-pairs traits (Figure 3B and Supplementary Table 6), the strongest (r_g=0.792, p=5.93e-151) between ASAT and VAT. There was substantial genetic correlation between VAT and liver fat (r_g=0.66, p=2.51e-2) and between VAT and pancreas fat (r_g=0.584, p=2.87e- 18). We found a negative genetic correlation between pancreas volume and fat (r_g=-0.554, p=1.21e-9) between pancreas volume and iron (r_g=-0.436, p=7.16e-4), and between lung volume and VAT (r_g=-0.274, p=2.52e- 5),

ASAT (r_g=-0.409, p=9.35e-11), liver fat (r_g=-0.333, p=1.41e-6) and pancreas fat (r_g=-0.221, p=6.30e-3).

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). 656 IDP-trait pairs showed evidence of nonzero genetic correlation; 268 of these involved with measures of size or body composition (Supplementary Table 7 and Extended Data 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=7.21e-133) and between VAT and conventional surrogate markers such as waist circumference (r_g=0.747, p=7.45e-107). The strongest genetic correlation with lung volume was with FVC (r_g=0.707, p=2.48e-73), 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=0.498, p=7.67e-24), kidney volume and serum creatinine (r_g=-0.38, p=1.65e-18), and liver iron and erythrocyte distribution width (r_g=-0.335, p=1.84e-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[22] (Methods). Liver fat showed evidence for enrichment in hepatocytes (p=4.50e-6) and liver (p=1.80e=5), and pancreatic fat showed evidence for enrichment in pancreas (smallest p=6.17e-5). Spleen volume showed enrichment in spleen cells (p=1.24e-9) 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 (Extended Data Figure 8-10).

We found some evidence for enrichment of liver fat heritability (though not liver volume or iron) in T-cell specific genomic regions (smallest p=9.47e-5). This is consistent with studies suggesting a CD4 T-cell mediated component in development of NASH[23]. Interestingly, we found no corresponding enrichment of pancreatic fat heritability, leaving open the possibility of a different mechanism of fat accumulation.

Genome-wide significant associations

For each locus containing at least one variant exceeding the study-wide significance threshold, we used GCTA COJO[24] to identify likely independent signals, and map likely causal variants. 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, and gene expression in 49 tissues (Methods, Supplementary Table 8).

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

The strongest association with liver volume (lead SNP rs4240624, p=9.10e-27, beta=-0.13), 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[25]. Although this variant has been associated with attenuated signal on hepatic computed tomography[26]; 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=3.10e-17, 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[27] as well as multiple metabolic traits including triglycerides, lipids, and C-reactive protein[28].

Of the nine study-wide independent signals associated with liver fat, three (rs58542926 in *TM6SF2*, p=9.6e-117, beta=0.333; rs429358 in *APOE*, p=1.1e-30, beta=-0.123; and rs738409 in *PNPLA3*, p=8e-134, beta=0.229) have previously been associated with NAFLD[29–31], and were reported in a GWAS of liver fat in a subset of this cohort[8]. The fourth SNP identified in that study, rs1260326 in *GCKR*, did not reach our stringent threshold of study-wide significance threshold (p=2.2e-8, beta=-0.044).

Two of the remaining six signals have previously been linked to liver disorders or lipid traits, although not specifically to liver fat. A signal near *TRIB1* (lead SNP rs112875651 p=8.1e-11, beta=-0.05) 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[32]. A missense SNP in *TM6SF2* (lead SNP rs188247550, p=4.79e-26, beta=0.35) is also associated with hyperlipidemia and has previously been linked to alcohol-induced cirrhosis[33].

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, p=7.5e-11, beta=0.055) 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[34].

We found an association between intronic to *GPAM*, which encodes an enzyme responsible for catalysis in phospholipid biosynthesis) (lead SNP rs11446981, p=1.6e-13, beta=-0.064). 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[35]. Our data suggests that this enzyme may play a role in the liver fat accumulation in humans.

A region overlapping to *MTTP* tagging 67 variants (lead SNP rs11274750, p=1.5e-9, beta=-0.053l) was associated with liver fat. Candidate gene studies have linked missense mutations in *MTTP* to NAFLD[36]. Rare nonsense mutations in this gene cause abetalipoproteinemia, an inability to absorb and knockout studies in mice recapitulate this phenotype[37,38]. Inhibition of *MTTP* is a treatment for familial hypercholesterolemia and is associated with increased liver fat[39].

Our final association has not previously been associated with any traits, and lies in a non-coding exon of *TOR1B* (lead SNP rs7029757, p=4.0e-9, beta=-0.07). This signal colocalises with *TOR1B* expression in adrenal gland, fibroblasts, esophagus, lung, and pancreas.

We replicate previously reported associations at *HFE* (rs1800562, p<1e-300, beta=0.565 and rs1799945, p=1.10e-61, beta=0.177)) and *TMPRSS6* (rs855791, p=2.0e-31, beta=0.09)[7] (p=2.60e-39, beta=0.1), and found evidence for two independent additional signal on chromosome 2 between *ASND1* and *SLC40A1* (lead SNPs rs7577758, p=2.8e-11, beta=0.06 and rs115380467, p=1.52e-12, beta=0.18)). *SLC40A1* encodes ferroportin, a protein essential for iron homeostasis[40] that enables absorption of dietary iron into the bloodstream. Mutations in *SLC40A1* are associated with a form of hemochromatosis known as African Iron Overload[41]. This finding is consistent with a recent study which highlighted the role of hepcidin as a major regulator of hepatic iron storage[7].

Novel associations with pancreas IDPs

We identified 13 study-wide significant associations with pancreatic volume. None were coding or colocalised with the expression of protein-coding genes. Three signals (rs35323862, p=8e-13, beta=0.07, nearest gene *RSPO3*; rs72802342, p=2.9e-15, beta=0.11, nearest gene *CTRB2*; rs8176685, p=2.1e-13, beta=-0.07, nearest gene *ABO*) colocalised with diabetic-related traits. This is consistent with our findings that T1D was associated with smaller pancreatic volume.

We identified nine 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[42].

The top association for pancreatic fat (lead SNP rs10422861, p=1.80e-22, beta=0.087) 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[43]. We also found evidence of a secondary signal at this locus (lead SNP rs118005033, conditional p=5.71e-10, conditional beta=0.087).

There was an association at the *ABO* locus, lead SNP (rs8176685, p=1.6e-12, beta=0.077 for pancreatic fat and p=2.1e-13, beta=-0.07 volume) is in complete LD in a European population with rs507666, which tags the A1 allele. 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[44] and increased risk of pancreatic cancer[45].

An association with pancreatic fat (lead SNP rs7404039, p=1.1e-12, beta=0.06) colocalises with the expression of *CBFA2T3* in the pancreas. rs7404039 lies in a promoter flanking region which is active in pancreatic tissue (ensemble regulatory region ENSR00000546074). *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[46]. 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, p=7.2e-13, beta=-0.064), and five other loci. In contrast to liver iron, where we identified strong signals at regions associated with ferroportin and hepcidin loci, we found one study-wide significant association with pancreatic iron that has not been associated with any other traits (lead SNP rs1005731, p=5.9e-10, beta=0.06).

Novel associations with other organ volume IDPs

rs807624 was associated with average kidney volume (p=2.20e-15; beta 0.0534). This signal colocalises with biomarkers of kidney function (cystatin C, creatinine, urate, and urea) and has previously been reported as associated with Wilms tumor[47], 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, p=4e-11, beta=0.0558) 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[48] in other studies,supporting our finding that kidney volume reflects overall kidney function.

The trait with the most associations was the spleen, with 24 independent signals, of which 21 colocalised with at least one hematological measurement. We identified one association with ASAT volume (rs1421085, p=1.6e-21, beta=0.07) at the well-known *FTO* locus which colocalised with many other body composition traits. The association with VAT volume at this SNP (p=1.7e-8, beta=0.038) was not study-wide significant. We identified three additional signals associated with VAT volume: rs153701 (p=2.9e-11, beta=-0.04) is intronic to *PEPD* and is independent from the pancreatic fat signal, rs73221948 (p=1.6e-10, beta=-0.04) which lies 150kb from the nearest protein coding gene and is not associated with any other traits, and rs72276239 (p=1.3e-11, beta=-0.057) which is also associated with trunk fat percentage and has previously been associated with waist-hip ratio[49].

Discussion

We developed a pipeline to systematically measure IDPs in UKBB MRI, and to explore epidemiological and genetic associations across multiple organs. 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[50,51]. 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[52]. 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[53]. 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[54]. However, studies with direct measurements although far more limited in scope and size, point towards a linear relationship with age[55–58], 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[53]. Similar patterns for pancreatic iron were observed[56], 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, fat[56]; liver fat increased only up to approximately 60 years of age before plateauing in women and decreasing in men[57,58]. Previous studies have suggested a linear relationship[59,60] but this may reflect the paucity of older participants (>60 years) in those cohorts that lack 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 [55]. 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[61]. Interestingly, although both liver and pancreas volume decreased with age, pancreatic fat did not, in agreement with previous observations[62]. Additionally, there was considerably greater diurnal variation in liver volume and compared with the pancreas. These observations add credence to the growing evidence of disparate mechanisms for the accumulation of fat in these organs[63]. 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[64], 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[65], and lung volumes with COPD, bronchitis, and respiratory disease. Liver volume was associated with chronic liver disease[66] and cirrhosis[67] 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[68], we found spleen volume to be most strongly associated with leukaemia. Although organ volume per se may have its limitation in disease diagnosis, spleen volume is a useful metric for predicting

outcome and response to treatment[69], 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[70], and confirmed herein on a much larger and diverse 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[71], 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[72,73]. It has been suggested that subdivisions of ASAT may convey different risks, with superficial ASAT conferring little or no risk compared to deeper layers of adipose tissue[74]. 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 UK 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.

In conclusion, by systematically quantifying 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. As the bulk of our study focused mainly on subjects of European ancestry, inclusion of greater diversity of genetic backgrounds as well as environmental exposures should significantly enrich our understanding of health, disease, and ageing. Exploration of the UKBB cohort through the application of models trained using deep learning is thus enhancing our understanding of health and disease.

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: (i) the Dixon protocol with six separate series covering 1.1 m of the participants (neck-to-knees), (ii) a high-resolution T1-weighted 3D acquisition of the pancreas, (iii) a single-slice multi-echo acquisition sequence for the liver, and (iv) a single-slice multi-echo acquisition sequence for the pancreas. Additional details of the MRI protocol may be found elsewhere[2]. 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 here.

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 pipeline

We describe the preprocessing pipeline for Dixon, high-resolution T1-weighted (T1w) and quantitative acquisitions from the UKBB Abdominal MRI protocol in detail elsewhere [9]. In brief, the Dixon data, used in 3D segmentations, is assembled from six individual series covering subjects from the neck to the knee.

The T1w data, used for 3D pancreas segmentation, is centred on the pancreas and only requires bias field correction reducing signal inhomogeneities.

The single-slice data (Gradient Multi-Echo (GRE) and IDEAL) used for liver and pancreas quantification had bias-field correction applied to each echo time separately for 2D segmentation. We used the PRESCO (Phase Regularized Estimation using Smoothing and Constrained Optimization) algorithm[75] to simultaneously estimate PDFF (referred to as fat in results) and R2* values voxelwise from the (uncorrected) GRE and IDEAL acquisitions. For consistency with previous studies[55,76], 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[77] 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 the2D mask prior to summarising fat and iron content. If the final size was <1% of its 3D organ 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 structure, we defined a standard operating procedure and provided training to a team of radiographers. All annotations were visually inspected at multiple stages by experienced analysts before use in modelling.

Segmentation of organs from Dixon data

We re-purposed an updated 3D iteration of the U-net architecture[78] based on label-free segmentation from 3D microscopy[79]. In order to produce sensible segmentations for QC purposes on minimal data, we made the following choices.

We used a batch size of four. 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[80]. Data augmentation was performed on the fly on the GPU, and not pre-computed.

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[2]. The neural network branched into a different logit head for supervision on each organ. Supervision included the sum of Dice coefficient[81] 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[82] to decode a binary decision for each voxel as to whether it was part of a 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 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 and extraction from single-slice data

We performed pancreas 3D segmentation on the T1w acquisition based on a recent iteration of the U-net architecture used in label-free microscopy[79]. We chose not to use the Dixon 3D U-net implementation as the pancreas data were sufficiently different to require a dedicated model,

trained with 99 manual annotations. 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 Ll2-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[83].

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)[84]
- Liver fat (22400) and liver iron (22402)[59]

Relationship between age, scan time, and IDPs

All statistical analysis was performed using R version 3.6.0.

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 (Extended Data Figure 2). Because there were some differences, 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[85] 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.

Other traits

We used the R package PHESANT[86] 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). 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, age, sex, BMI, and height as covariates.

Genetics

We follow the methods described in a previous study[87]

Genome-wide association study

We used the UKBB imputed genotypes version 3[18], 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[18]. We used BOLT-LMM version 2.3.2[88] to conduct the genetic association study. We included age at imaging visit, age squared, sex, imaging centre, 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 1.05-1.1 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)[19] (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, p=4e-9, which accounts for the multiple testing due to the number of IDPs in this study. 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[89].

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

Heritability estimation and enrichment

We estimated the heritability of each trait using restricted maximum likelihood as implemented in BOLT version 2.3.2[90].

To identify relevant tissues and cell types contributing to the heritability of IDPs, we used stratified LD score regression[91] 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[91], 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[92] and ENCODE[93] 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[94] and other datasets[95,96], and (iii) immune cell type expression markers from the ImmGen Consortium[97]. For each annotation set, we controlled for the number of tests using the Storey and Tibshirani procedure[98]. 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.[91].

Genetic correlation

We computed genetic correlation between traits using bivariate LDSC[21].

Statistical fine-mapping

We performed approximate conditional analysis using genome-wide complex trait analysis (GCTA)[24], 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[18] that were randomly selected after filtering for unrelated, Caucasian participants. 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 locus-wide evidence of association (Pjoint<10−6) to be conditionally independent. We annotated each independent signal with the nearest protein-coding gene using the OpenTargets genetics resource (May 2019 version).

Construction of genetic credible sets

For each distinct signal, we calculated credible sets [99] 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[100] 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[100] and $(0.15\sigma)^2$ for quantitative traits[101], 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 [102] 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[101] using default priors and all variants within 500kb of the index variant of each signal. Following previous studies[103], 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 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[104], version 190505. We identified for studies where our lead variant was in LD (r>0.7) with the lead SNP of a published study.

References

1. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
2. Littlejohns, T. J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat. Commun.* **11**, 2624 (2020).
3. Thomas, E. L., Fitzpatrick, J., Frost, G. S. & Bell, J. D. Metabolic syndrome, overweight and fatty liver. in *Handbook of nutrition and food, 3rd edition* (eds. Berdanier, C. D., Dwyer, J. T. & Heber, D.) 763–768 (CRC Press, 2013).
4. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
5. Pirruccello, J. P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat. Commun.* **11**, 2254

(2020).

6. Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
7. Wilman, H. R. *et al.* Genetic studies of abdominal MRI data identify genes regulating hepcidin as major determinants of liver iron concentration. *J. Hepatol.* **71**, 594–602 (2019).
8. Parisinos, C. A. *et al.* Genome-wide and Mendelian randomisation studies of liver MRI yield insights into the pathogenesis of steatohepatitis. *J. Hepatol.* (2020) doi:10.1016/j.jhep.2020.03.032.
9. Basty, N. *et al.* Image Processing and Quality Control for Abdominal Magnetic Resonance Imaging in the UK Biobank. *arXiv [eess.IV]* (2020).
10. Leung, N. W., Farrant, P. & Peters, T. J. Liver volume measurement by ultrasonography in normal subjects and alcoholic patients. *J. Hepatol.* **2**, 157–164 (1986).
11. Radmard, A. R. *et al.* Gallstone disease and obesity: a population-based study on abdominal fat distribution and gender differences. *Ann. Hepatol.* **14**, 702–709 (2015).
12. Jovanović, D., Gasic, B., Pavlovic, S. & Naumovic, R. Correlation of kidney size with kidney function and anthropometric parameters in healthy subjects and patients with chronic kidney diseases. *Ren. Fail.* **35**, 896–900 (2013).
13. Harrison-Findik, D. D. Role of alcohol in the regulation of iron metabolism. *World J. Gastroenterol.* **13**, 4925–4930 (2007).
14. Bugianesi, E. *et al.* Low Birthweight Increases the Likelihood of Severe Steatosis in Pediatric Non-Alcoholic Fatty Liver Disease. *Am. J. Gastroenterol.* **112**, 1277–1286 (2017).
15. Parkinson, J. R. C. *et al.* Clinical and molecular evidence of accelerated ageing following very preterm birth. *Pediatr. Res.* **87**, 1005–1010 (2020).
16. Thomas, E. L. *et al.* Aberrant adiposity and ectopic lipid deposition characterize the adult

phenotype of the preterm infant. *Pediatr. Res.* **70**, 507–512 (2011).

17. Lutfi, M. F. The physiological basis and clinical significance of lung volume measurements. *Multidiscip. Respir. Med.* **12**, 3 (2017).
18. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
19. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
20. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
21. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
22. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
23. Van Herck, M. A. *et al.* The Differential Roles of T Cells in Non-alcoholic Fatty Liver Disease and Obesity. *Front. Immunol.* **10**, 82 (2019).
24. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–75, S1–3 (2012).
25. Mehta, M. B. *et al.* Hepatic protein phosphatase 1 regulatory subunit 3B (Ppp1r3b) promotes hepatic glycogen synthesis and thereby regulates fasting energy homeostasis. *J. Biol. Chem.* **292**, 10444–10454 (2017).
26. Stender, S. *et al.* Relationship between genetic variation at PPP1R3B and levels of liver glycogen and triglyceride. *Hepatology* **67**, 2182–2195 (2018).
27. Kawaguchi, T. *et al.* Risk estimation model for nonalcoholic fatty liver disease in the

Japanese using multiple genetic markers. *PLoS One* **13**, e0185490 (2018).

28. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
29. Speliotes, E. K. *et al.* Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7**, e1001324 (2011).
30. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **40**, 1461–1465 (2008).
31. Kozlitina, J. *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **46**, 352–356 (2014).
32. Liu, Q. *et al.* 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 Health Dis.* **18**, 165 (2019).
33. Buch, S. *et al.* A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat. Genet.* **47**, 1443–1448 (2015).
34. Emdin, C. A. *et al.* A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. *PLoS Genet.* **16**, e1008629 (2020).
35. Kuhajda, F. P. *et al.* 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**, R116–30 (2011).
36. Hsiao, P.-J. *et al.* MTTP-297H polymorphism reduced serum cholesterol but increased risk of non-alcoholic fatty liver disease-a cross-sectional study. *BMC Med. Genet.* **16**, 93 (2015).
37. Partin, J. S., Partin, J. C., Schubert, W. K. & McAdams, A. J. Liver ultrastructure in

abetalipoproteinemia: Evolution of micronodular cirrhosis. *Gastroenterology* **67**, 107–118 (1974).

38. Raabe, M. *et al.* Knockout of the abetalipoproteinemia gene in mice: reduced lipoprotein secretion in heterozygotes and embryonic lethality in homozygotes. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8686–8691 (1998).
39. Cuchel, M. *et al.* Inhibition of microsomal triglyceride transfer protein in familial hypercholesterolemia. *N. Engl. J. Med.* **356**, 148–156 (2007).
40. Donovan, A. *et al.* The iron exporter ferroportin/Slc40a1 is essential for iron homeostasis. *Cell Metab.* **1**, 191–200 (2005).
41. Mayr, R. *et al.* Identification of mutations in SLC40A1 that affect ferroportin function and phenotype of human ferroportin iron overload. *Gastroenterology* **140**, 2056–63, 2063.e1 (2011).
42. Taylor, R. Pathogenesis of type 2 diabetes: tracing the reverse route from cure to cause. *Diabetologia* **51**, 1781–1789 (2008).
43. Kitchener, R. L. & Grunden, A. M. Prolidase function in proline metabolism and its medical and biotechnological applications. *J. Appl. Microbiol.* **113**, 233–247 (2012).
44. Zhang, H., Mooney, C. J. & Reilly, M. P. ABO Blood Groups and Cardiovascular Diseases. *Int. J. Vasc. Med.* **2012**, 641917 (2012).
45. Zhang, B.-L., He, N., Huang, Y.-B., Song, F.-J. & Chen, K.-X. ABO blood groups and risk of cancer: a systematic review and meta-analysis. *Asian Pac. J. Cancer Prev.* **15**, 4643–4650 (2014).
46. de Lichtenberg, K. H. *et al.* Notch Controls Multiple Pancreatic Cell Fate Regulators Through Direct Hes1-mediated Repression. *bioRxiv* 336305 (2018) doi:10.1101/336305.
47. Turnbull, C. *et al.* A genome-wide association study identifies susceptibility loci for Wilms

tumor. *Nat. Genet.* **44**, 681–684 (2012).

48. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).
49. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
50. Svennerholm, L., Boström, K. & Jungbjer, B. Changes in weight and compositions of major membrane components of human brain during the span of adult human life of Swedes. *Acta Neuropathol.* **94**, 345–352 (1997).
51. Mitchell, W. K. *et al.* Sarcopenia, dynapenia, and the impact of advancing age on human skeletal muscle size and strength; a quantitative review. *Front. Physiol.* **3**, 260 (2012).
52. Mancuso, P. & Bouchard, B. The Impact of Aging on Adipose Function and Adipokine Synthesis. *Front. Endocrinol.* **10**, 137 (2019).
53. Iron Homeostasis in the Liver. in *Comprehensive Physiology* (ed. Terjung, R.) vol. 275 19906 (John Wiley & Sons, Inc., 2013).
54. Cook, J. D., Flowers, C. H. & Skikne, B. S. The quantitative assessment of body iron. *Blood* **101**, 3359–3364 (2003).
55. McKay, A. *et al.* Measurement of liver iron by magnetic resonance imaging in the UK Biobank population. *PLoS One* **13**, e0209340 (2018).
56. Schwenzer, N. F. *et al.* T2* Relaxometry in Liver, Pancreas, and Spleen in a Healthy Cohort of One Hundred Twenty-Nine Subjects–Correlation With Age, Gender, and Serum Ferritin. *Invest. Radiol.* **43**, 854 (2008).
57. Kühn, J.-P. *et al.* Prevalence of Fatty Liver Disease and Hepatic Iron Overload in a Northeastern German Population by Using Quantitative MR Imaging. *Radiology* **284**, 706–716 (2017).

58. Nomura, H. *et al.* Prevalence of fatty liver in a general population of Okinawa, Japan. *Jpn. J. Med.* **27**, 142–149 (1988).

59. Wilman, H. R. *et al.* Characterisation of liver fat in the UK Biobank cohort. *PLOS ONE* vol. 12 e0172921 (2017).

60. Thomas, E. L. *et al.* The missing risk: MRI and MRS phenotyping of abdominal adiposity and ectopic fat. *Obesity* **20**, 76–87 (2012).

61. Taylor, R. Type 2 diabetes: etiology and reversibility. *Diabetes Care* **36**, 1047–1055 (2013).

62. Majumder, S. *et al.* Fatty Pancreas: Should We Be Concerned? *Pancreas* **46**, 1251–1258 (2017).

63. Hellerstein, M. K. De novo lipogenesis in humans: metabolic and regulatory aspects. *Eur. J. Clin. Nutr.* **53 Suppl 1**, S53–65 (1999).

64. Kalhan, S. C. & Ghosh, A. Dietary iron, circadian clock, and hepatic gluconeogenesis. *Diabetes* vol. 64 1091–1093 (2015).

65. Grantham, J. J. *et al.* Volume Progression in Polycystic Kidney Disease. *New England Journal of Medicine* vol. 354 2122–2130 (2006).

66. Lin, X. Z. *et al.* Liver volume in patients with or without chronic liver diseases. *Hepatogastroenterology* **45**, 1069–1074 (1998).

67. Hagan, M. T. *et al.* Liver volume in the cirrhotic patient: does size matter? *Dig. Dis. Sci.* **59**, 886–891 (2014).

68. Pozo, A. L., Godfrey, E. M. & Bowles, K. M. Splenomegaly: investigation, diagnosis and management. *Blood Rev.* **23**, 105–111 (2009).

69. Shimomura, Y., Hara, M., Katoh, D., Hashimoto, H. & Ishikawa, T. 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. *Annals of Hematology* vol. 97 1049–1056 (2018).

70. Lee, J. J. *et al.* Visceral and Intrahepatic Fat Are Associated with Cardiometabolic Risk Factors Above Other Ectopic Fat Depots: The Framingham Heart Study. *Am. J. Med.* **131**, 684–692.e12 (2018).
71. Kuk, J. L. *et al.* Visceral Fat Is an Independent Predictor of All-cause Mortality in Men*. *Obesity* vol. 14 336–341 (2006).
72. Irlbeck, T. *et al.* Association between single-slice measurements of visceral and abdominal subcutaneous adipose tissue with volumetric measurements: the Framingham Heart Study. *International Journal of Obesity* vol. 34 781–787 (2010).
73. Elffers, T. W. *et al.* Body fat distribution, in particular visceral fat, is associated with cardiometabolic risk factors in obese women. *PLOS ONE* vol. 12 e0185403 (2017).
74. Kelley, D. E., Leland Thaete, F., Troost, F., Huwe, T. & Goodpaster, B. H. Subdivisions of subcutaneous abdominal adipose tissue and insulin resistance. *American Journal of Physiology-Endocrinology and Metabolism* vol. 278 E941–E948 (2000).
75. Bydder, M. *et al.* Constraints in estimating the proton density fat fraction. *Magn. Reson. Imaging* **66**, 1–8 (2020).
76. Wood, J. C. *et al.* MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood* **106**, 1460–1465 (2005).
77. Li, J. *et al.* Quantitative susceptibility mapping (QSM) minimizes interference from cellular pathology in R2* estimation of liver iron concentration. *J. Magn. Reson. Imaging* **48**, 1069–1079 (2018).
78. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention –*

MICCAI 2015 234–241 (Springer International Publishing, 2015).

79. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
80. TFRecord and tf.Example | TensorFlow Core. *TensorFlow* https://www.tensorflow.org/tutorials/load_data/tfrecord.
81. Milletari, F., Navab, N. & Ahmadi, S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. in *2016 Fourth International Conference on 3D Vision (3DV)* 565–571 (2016).
82. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
83. Basty, N. *et al.* Automated Measurement of Pancreatic Fat and Iron Concentration Using Multi-Echo and T1-Weighted MRI Data. in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* 345–348 (2020).
84. West, J. *et al.* Feasibility of MR-Based Body Composition Analysis in Large Scale Population Studies. *PLoS One* **11**, e0163332 (2016).
85. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).
86. Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G. & Tilling, K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* **47**, 29–35 (2018).
87. Sethi, A. *et al.* Calcification of abdominal aorta is an underappreciated cardiovascular disease risk factor. *Epidemiology* (2020) doi:10.1101/2020.05.07.20094706.

88. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

89. International HapMap 3 Consortium *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).

90. Loh, P.-R. BOLT-LMM v2. 3.2 User Manual. *Available oniline at: https://data. broadinstitute. org/alkesgroup/BOLT-LMM/(accessed May 2, 2019)* (2018).

91. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

92. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

93. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

94. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

95. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).

96. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

97. Heng, T. S. P., Painter, M. W. & Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).

98. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).

99. Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals

for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).

100. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).

101. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).

102. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

103. Guo, H. *et al.* Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum. Mol. Genet.* **24**, 3305–3313 (2015).

104. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2019).

Tables

Table 1A: Study population characteristics. 1B: 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)

	UK Biobank cohort (at time of baseline visit)	Imaging cohort (at time of imaging visit)	GWAS cohort			
			Volumes	Pancreas 3D	Pancreas 2D	Liver 2D
n	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)
% Caucasian	81.5	81.5	100	100	100	100

			Mean (standard deviation)	

Trait	Organ	# Quantified	Combined	Female	Male	# indep. GWAS
Volume (l)	Liver	38683	1.38 (0.298)	1.28 (0.255)	1.49 (0.304)	7
	Pancreas	37362	0.0587 (0.0177)	0.0548 (0.0157)	0.063 (0.0186)	15
	Kidney	38683	0.139 (0.0304)	0.123 (0.0226)	0.156 (0.0283)	8
	Spleen	38683	0.167 (0.0723)	0.139 (0.0536)	0.196 (0.0779)	24
	Lungs	38683	2.67 (0.734)	2.32 (0.526)	3.03 (0.747)	2
	ASAT	38683	8.16 (4.09)	9.57 (4.35)	6.64 (3.15)	1
	VAT	38683	3.92 (2.3)	2.78 (1.55)	5.14 (2.35)	3
Fat (%)	Liver	38639	5.06 (5)	4.43 (4.73)	5.73 (5.2)	12
	Pancreas	29091	10.4 (8.02)	8.34 (6.71)	12.7 (8.67)	10
Iron (%)	Liver	38639	1.22 (0.26)	1.2 (0.243)	1.24 (0.276)	7
	Pancreas	29091	0.769 (0.0998)	0.796 (0.104)	0.741 (0.0863)	1

Figure legends

Figure 1. A: Example Dixon image before and after 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.

Figure 2: Disease phenome-wide association study across all IDPs and 754 disease codes (PheCodes). Top panel: Organ volumes. Middle panel: Fat. Bottom panel: iron. The top 3 associations for each phenotype are labelled. Associations reaching disease phenome-wide significance are coloured by organ, with the color scheme as in Figure 1. Horizontal lines at disease phenome-wide significance (dotted line, p=6.6313e-05) and study-wide significance (dashed line, p=6.03e-06) after correction for multiple testing. Arrowheads indicate direction of the association between the organ volume and the PheCode. Abbreviations: T1D: Type 1

diabetes. T2D: Type 2 diabetes. CKD: Chronic kidney disease. LL: Lymphoid leukaemia. CLL: Chronic lymphoid leukaemia.

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.

Acknowledgements

We thank Adam Baker, Garret Fitzgerald, Frank Li, Anil Raj, Amoolya Singh, and Elena Sorokin 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 performed statistical analysis. JB, NB, MC, YL, 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 Statement

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 Statement

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, 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) or ELT (L.Thomas3@westminster.ac.uk).

Additional information

Additional information may be found in the “Supplementary Information” file.

Extended data figure legends

1. 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.
2. A. Organ volume IDPs, faceted by imaging centre. B. Fat IDPs, faceted by imaging centre. C. Iron IDPs, faceted by imaging centre. D. Relationship between scan date and imaging IDP (not adjusted for imaging centre mix).
3. Phenome-wide association plots for biomarkers. Top panel: Organ volumes. Middle panel: Fat. Bottom panel: iron. The top 3 associations for each phenotype are labelled. Associations reaching disease phenome-wide significance are coloured by organ, with the color scheme as in Figure 1. Horizontal lines at phenome-wide significance (dotted

line, 2.03e-05) and study-wide significance (dashed line, p=1.85e-06) after correction for multiple testing.

4. Phenome-wide association plots for online follow-up traits.
5. Phenome-wide association plots for lifestyle and history traits.
6. Phenome-wide association plots for medical history traits.
7. Phenome-wide association plots for physical measures.
8. Heritability enrichment in tissues and cell types for volume IDPs. The top 3 enrichments for each phenotype are labeled. Associations significant after Bonferroni correction for multiple testing are colored by organ, with the color scheme as in Figure 1. Horizontal lines and trait-wide (dotted line) and study-wide (dashed line) significance.
9. Heritability enrichment in tissues and cell types for fat IDPs.
10. Heritability enrichment in tissues and cell types for iron IDPs.
11. 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.

Figures

Figure 1. A: Example Dixon image before and after 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.

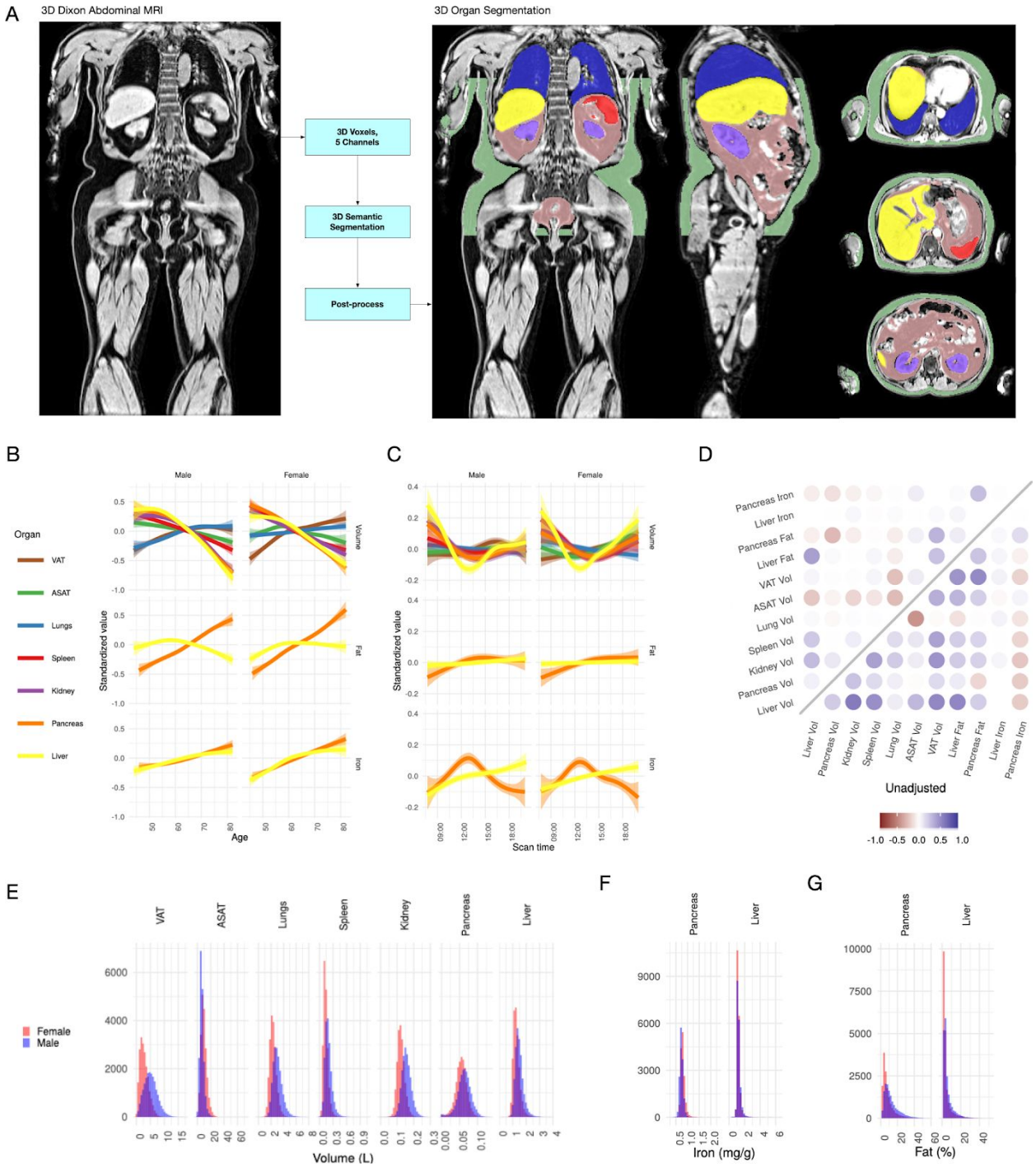

A
3D Dixon Abdominal MRI
3D Voxels, 5 Channels
3D Semantic Segmentation
Post-process
3D Organ Segmentation
B
Male
Female
Organ
VAT
ASAT
Lungs
Spleen
Kidney
Pancreas
Liver
Standardized value
Volume
Fat
Iron
Age
C
Scan time
D
Pancreas Iron
Liver Iron
Pancreas Fat
Liver Fat
VAT Vol
ASAT Vol
Lung Vol
Spleen Vol
Kidney Vol
Pancreas Vol
Liver Vol
Unadjusted
E
Female
Male
Volume (L)
F
Iron (mg/g)
G
Fat (%)

Figure 2: Disease phenome-wide association study across all IDPs and 754 disease codes (PheCodes). Top panel: Organ volumes. Middle panel: Fat. Bottom panel: iron. The top 3 associations for each phenotype are labelled. Associations reaching disease phenome-wide significance are coloured by organ, with the color scheme as in Figure 1. Horizontal lines at disease phenome-wide significance (dotted line, p=6.6313e-05) and study-wide significance (dashed line, p=6.03e-06) after correction for multiple testing. Arrowheads indicate direction of the association between the organ volume and the PheCode. Abbreviations: T1D: Type 1 diabetes. T2D: Type 2 diabetes. CKD: Chronic kidney disease. LL: Lymphoid leukaemia. CLL: Chronic lymphoid leukaemia.

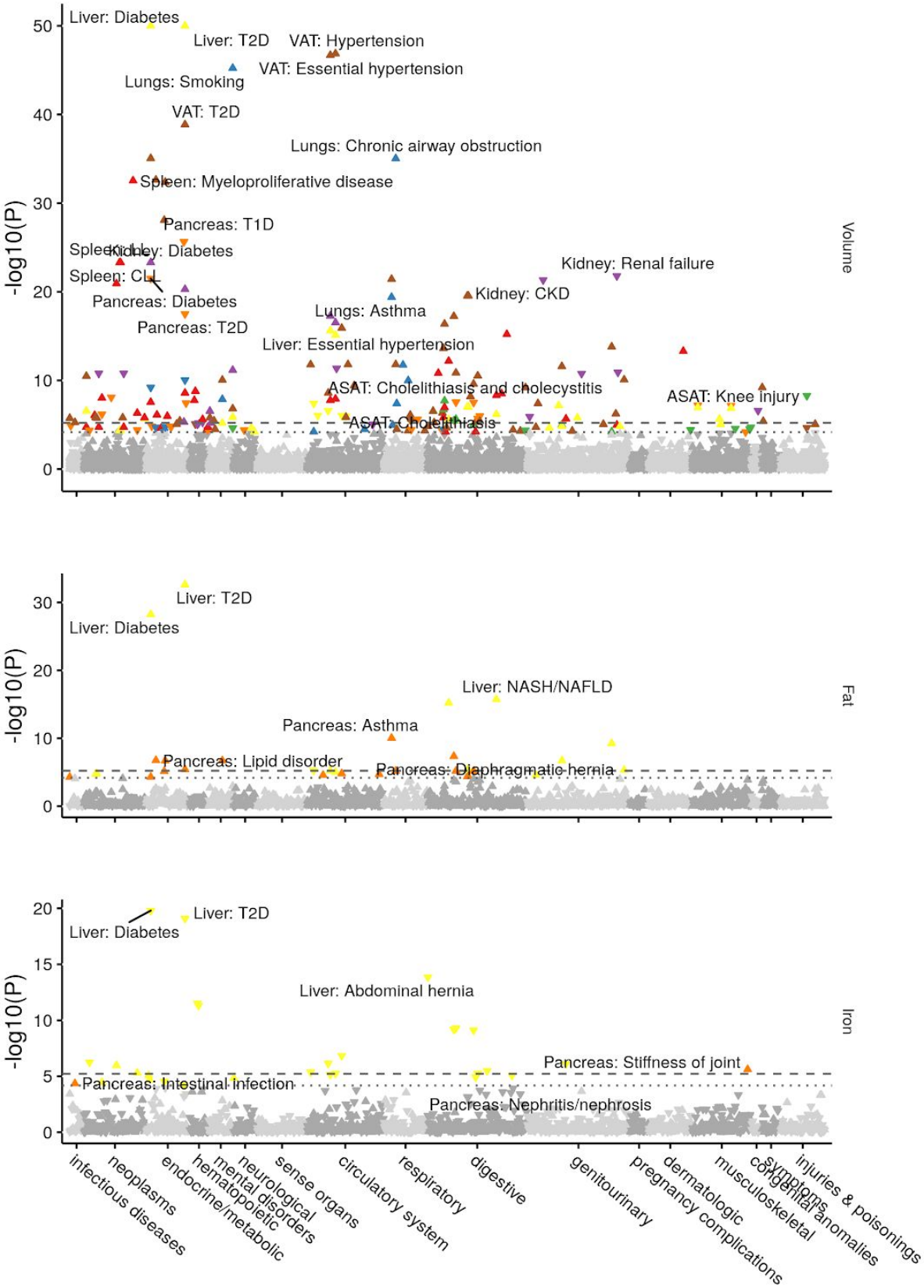
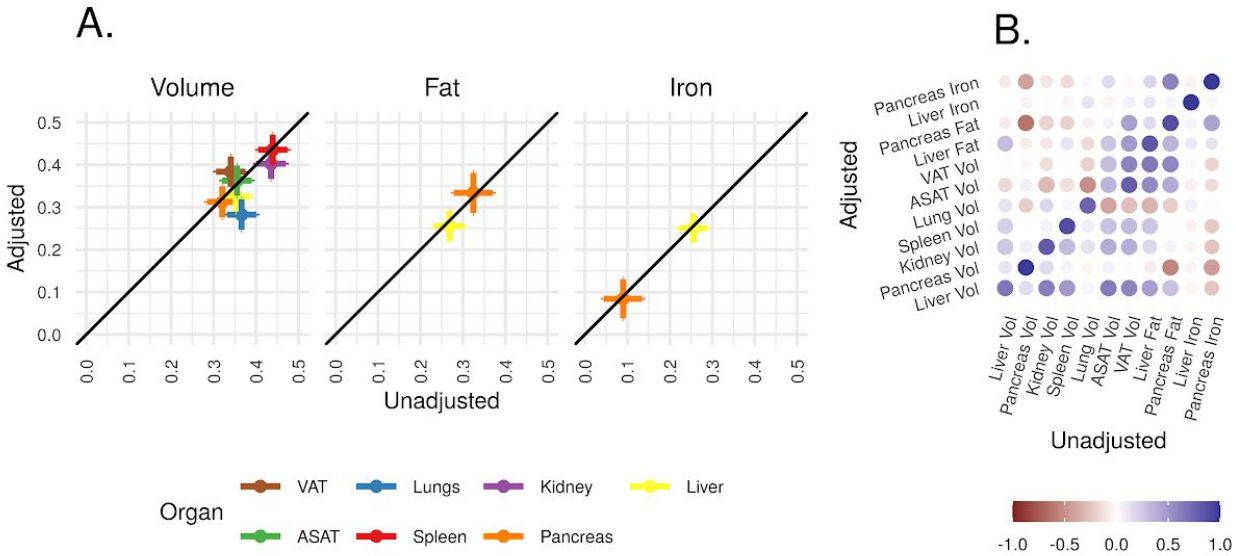
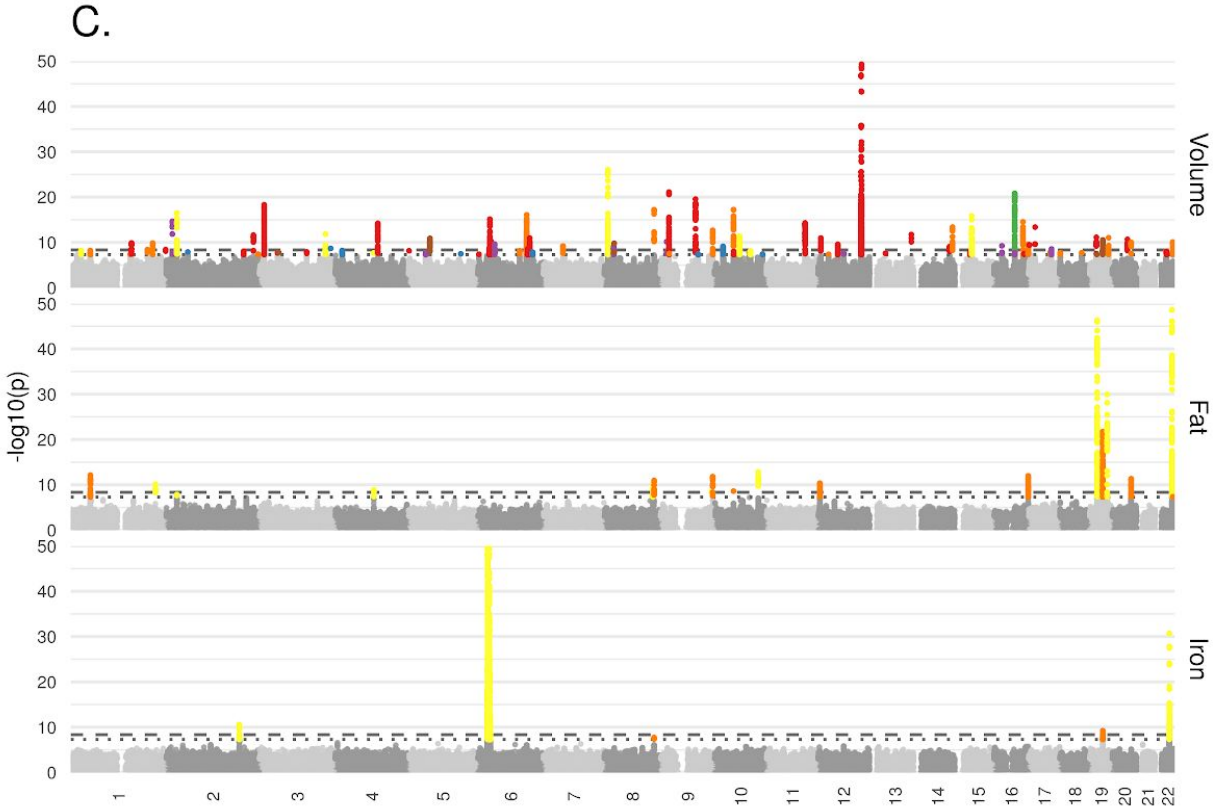

Liver: Diabetes
Liver: T2D
VAT: Hypertension
VAT: Essential hypertension
Lungs: Smoking
VAT: T2D
Lungs: Chronic airway obstruction
Spleen: Myeloproliferative disease
Pancreas: T1D
Spleen: Kidney: Diabetes
Spleen: CLL
Pancreas: Diabetes
Pancreas: T2D
Lungs: Asthma
Kidney: Renal failure
Kidney: CKD
Liver: Essential hypertension
ASAT: Cholelithiasis and cholecystitis
ASAT: Cholelithiasis
ASAT: Knee injury
Volume
Liver: T2D
Liver: Diabetes
Liver: NASH/NAFLD
Pancreas: Asthma
Pancreas: Lipid disorder
Pancreas: Diaphragmatic hernia
Fat
Liver: T2D
Liver: Diabetes
Liver: Abdominal hernia
Pancreas: Stiffness of joint
Pancreas: Intestinal infection
Pancreas: Nephritis/nephrosis
Iron
-log10(P)
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
pregnancy complications
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries & poisonings

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.


A.
Volume
Fat
Iron
Adjusted
Unadjusted
0.0
0.1
0.2
0.3
0.4
0.5
Organ
VAT
ASAT
Lungs
Spleen
Kidney
Pancreas
Liver
B.
Adjusted
Unadjusted
Pancreas Iron
Liver Iron
Pancreas Fat
Liver Fat
VAT Vol
ASAT Vol
Lung Vol
Spleen Vol
Kidney Vol
Pancreas Vol
Liver Vol
-1.0
-0.5
0.0
0.5
1.0


C.
-log10(p)
Volume
Fat
Iron
0
10
20
30
40
50
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22