

1 Mapping rare protein-coding variants on multi-organ imaging traits

2

3 Yijun Fan¹, Jie Chen², Zirui Fan³, Julio Chirinos^{4,5}, Jason L. Stein^{6,7}, Patrick F. Sullivan⁶, Rujin
4 Wang⁸, Ajay Nadig^{9,10,11,12}, David Y. Zhang^{13,14}, Shuai Huang², Zhiwen Jiang², Peter Yi Guan²,
5 Xinjie Qian², Ting Li², Haoyue Li², Zehui Sun², Marylyn D. Ritchie^{13,15}, Joan O'Brien^{16,17},
6 Walter Witschey¹⁸, Daniel J. Rader^{13,14}, Tengfei Li^{19,20}, Hongtu Zhu^{2,6,20-22*}, and Bingxin
7 Zhao^{1,3,15,23-26*#}

8

9 ¹Graduate Group in Applied Mathematics and Computational Science, University of Pennsylvania,
10 Philadelphia, PA 19104, USA.

11 ²Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599,
12 USA.

13 ³Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104,
14 USA.

15 ⁴Division of Cardiovascular Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA
16 19104, USA.

17 ⁵University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.

18 ⁶Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

19 ⁷UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

20 ⁸Regeneron Genetics Center, 777 Old Saw Mill River Rd., Tarrytown, NY, 10591, USA.

21 ⁹Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

22 ¹⁰Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
23 02142, USA

24 ¹¹Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge,
25 MA 02142, USA

26 ¹²Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114,
27 USA

28 ¹³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia,
29 PA 19104, USA.

30 ¹⁴Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia,
31 PA 19104, USA.

1 ¹⁵Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania
2 Philadelphia, PA 19104, USA.

3 ¹⁶Scheie Eye Institute, University of Pennsylvania, Philadelphia, PA 19104, USA

4 ¹⁷Penn Medicine Center for Ophthalmic Genetics in Complex Diseases, Philadelphia, PA 19104,
5 USA

6 ¹⁸Department of Radiology, Perelman School of Medicine, University of Pennsylvania,
7 Philadelphia, PA 19104, USA

8 ¹⁹Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

9 ²⁰Biomedical Research Imaging Center, School of Medicine, University of North Carolina at Chapel
10 Hill, Chapel Hill, NC 27599, USA.

11 ²¹Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC
12 27599, USA.

13 ²²Department of Statistics and Operations Research, University of North Carolina at Chapel Hill,
14 Chapel Hill, NC 27599, USA.

15 ²³Center for AI and Data Science for Integrated Diagnostics, Perelman School of Medicine,
16 University of Pennsylvania, Philadelphia, PA 19104, USA.

17 ²⁴Population Aging Research Center, University of Pennsylvania, Philadelphia, PA 19104, USA.

18 ²⁵Institute for Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia,
19 PA 19104, USA.

20 ²⁶Penn Center for Eye-Brain Health, Perelman School of Medicine, University of Pennsylvania,
21 Philadelphia, PA 19104, USA.

22

23 **Corresponding authors:*

24 Hongtu Zhu (htzhu@email.unc.edu) and Bingxin Zhao (bxzhao@upenn.edu)

25

26 *#Lead contact:*

27 Bingxin Zhao

28 413 Academic Research Building

29 265 South 37th Street, Philadelphia, PA 19104

30 E-mail: bxzhao@upenn.edu Phone: (215) 898-8222

1 **Abstract**

2 Human organ structure and function are important endophenotypes for clinical outcomes.
3 Genome-wide association studies (GWAS) have identified numerous common variants
4 associated with phenotypes derived from magnetic resonance imaging (MRI) of the brain
5 and body. However, the role of rare protein-coding variations affecting organ size and
6 function is largely unknown. Here we present an exome-wide association study that
7 evaluates 596 multi-organ MRI traits across over 50,000 individuals from the UK Biobank.
8 We identified 107 variant-level associations and 224 gene-based burden associations (67
9 unique gene-trait pairs) across all MRI modalities, including *PTEN* with total brain volume,
10 *TTN* with regional peak circumferential strain in the heart left ventricle, and *TNFRSF13B*
11 with spleen volume. The singleton burden model and AlphaMissense annotations
12 contributed 8 unique gene-trait pairs including the association between an approved drug
13 target gene of *KCNA5* and brain functional activity. The identified rare coding signals
14 elucidate some shared genetic regulation across organs, prioritize previously identified
15 GWAS loci, and are enriched for drug targets. Overall, we demonstrate how rare variants
16 enhance our understanding of genetic effects on human organ morphology and function
17 and their connections to complex diseases.

1 Magnetic resonance imaging (MRI)-derived traits enable us to study the structure,
2 function, and abnormalities of human organs *in vivo*. Many of these traits serve as
3 established endophenotypes implicated in complex diseases and related traits. Therefore,
4 it is of great interest to uncover genetic effects using imaging data to better understand
5 the biology of human organs in health and disease. Recent genome-wide association
6 studies (GWAS) have successfully identified common variants associated with multi-organ
7 imaging traits, including brain structural, diffusion, and functional MRI¹⁻⁸, as well as
8 cardiovascular magnetic resonance imaging (CMR)⁹⁻¹¹ and abdominal MRI¹²⁻¹⁴. However,
9 a limitation of common variant signals identified by GWAS is that they often reside in non-
10 coding regions and exhibit small effect sizes, which complicates the direct derivation of
11 biological insights or the identification of causal genes¹⁵⁻¹⁹. By focusing on rare variants in
12 protein-coding regions of the genome, whole exome sequencing (WES) studies aim to
13 directly identify genes of interest.

14

15 Although existing exome-wide association studies^{16,17,20,21} (ExWAS) have identified
16 associations for some imaging traits, our knowledge of the rare variant genetic
17 architectures of human organs and their roles in various diseases remains substantially
18 limited. Specifically, most previous ExWAS have focused on a single organ and/or a small
19 set of imaging traits, lacking a multi-organ perspective that simultaneously explores the
20 genetic effects of the human brain and body. For example, Park et al.²⁰ analyzed CT
21 imaging-derived hepatic fat, Haas et al.¹⁴ studied liver fat based on abdominal MRI using
22 machine learning, Jurgen et al.²¹ studied several CMR traits, and Backman, et al.¹⁶ and
23 Karczewski et al.¹⁷ included brain MRI traits in their UK Biobank (UKB) WES studies of a
24 wide range of phenotypes. Therefore, analyzing more and refined imaging traits, spanning
25 multiple organs, in larger sample sizes will yield deeper insights into the genetic effects of
26 rare variants across the whole body and shared genetic regulation across organs¹⁰.

27

28 Here we conducted ExWAS for 596 MRI traits derived from the brain, heart, liver, kidney,
29 and lung (**Table S1**) of over 50,000 participants from the UKB study. We used an internal
30 discovery-replication design to make the best use of available data resources to identify
31 novel rare variants and genes associated with human organ structure and function (**Fig.**
32 **1**). We evaluated various functional annotation approaches for missense variants in gene-

1 level set-based association testing, including conventional methods such as SIFT²²,
2 PolyPhen2 HDIV²³, PolyPhen2 HVAR²³, LRT²⁴, and MutationTaster²⁵, as well as the deep
3 learning-based method AlphaMissense²⁶. We compared our results with previous GWAS
4 on the same set of MRI traits in order to provide shared common and rare variant
5 evidence for a gene's involvement in a trait. We also used burden heritability regression
6 (BHR)²⁷ tests to characterize the genetic architecture for ultra-rare coding variants (minor
7 allele frequency [MAF] < 1×10^{-4}) across functional classes. In summary, our study
8 identified novel exome-wide associations for multi-organ structure and function,
9 providing a valuable source of evidence that could be useful in drug discovery and clinical
10 therapeutics. These findings may also enhance our understanding of the complex
11 interrelations between the human organs, health, and disease.

12

13 RESULTS

14 Overview of variant-level associations with multi-organ MRI traits

15 Based on European individuals from UKB phases 1 to 5 MRI data (released up to late 2023,
16 average $n = 40,038$; **Fig. 1**), we identified 107 rare (MAF < 0.01) variant-level associations
17 (**Table S2**) using a conservative P -value threshold of 2.8×10^{-10} (Bonferroni adjusted for
18 all variant-trait tests with minor allele count [MAC] > 5 as $0.05/178,280,016$, Methods)
19 (**Fig. 2A**). Only predicted loss-of-function variants (pLoF) and missense variants were
20 included in our study. There were 75 gene-trait pairs between 24 unique genes and 62
21 MRI traits, including 3 abdominal MRI traits, 7 CMR traits of the heart and the aorta, 10
22 brain structural MRI traits (regional brain volumes), as well as 42 brain diffusion MRI traits
23 (diffusion tensor imaging [DTI] parameters) (**Fig. 2B**). Additionally, 138 associations (92
24 unique genes and 107 MRI traits) showed suggestive evidence, surviving a more liberal P -
25 value threshold of 1×10^{-8} (**Table S3**, Methods). These associations spanned across all
26 categories of MRI traits (**Table S2**). As expected, larger effect sizes were linked with lower
27 MAF, which revealed the process of negative selection^{17,28,29} (**Fig. 2C**).

28

29 Due to the lack of independent data sources and the inherent rarity of the variants,
30 identifying rare variant associations typically requires large sample sizes, and findings are
31 difficult to fully replicate. To assess the reliability of our findings and maximize the use of
32 available data, we adopted a procedure that includes internal replication and joint

1 analysis³⁰ (**Fig. 1** and Methods). First, we limited our discovery dataset to include only
2 individuals from phases 1 to 3 of MRI release (average $n = 30,739$). This subset yielded 59
3 associations that passed the Bonferroni threshold of $P < 3.5 \times 10^{-10}$ (Bonferroni adjusted
4 for all variant-level tests with MAC > 5 in this subset, $0.05/142,935,657$, Methods) (**Table**
5 **S4**). Among these 59 associations, 56 had mutations in the independent sample from
6 phases 4 and 5 of the MRI release. Of these, 47 (84% = 47/56) passed the P -value
7 threshold of 2.9×10^{-2} (Benjamini-Hochberg false discovery rate [FDR] at the 0.05 level),
8 and all had consistent effect size directions as those in the phases 1 to 3 data.
9 Furthermore, 84.7% (= 50/59) of these associations had smaller P -values in the combined
10 phases 1 to 5 sample, and all of them had consistent effect directions. Variants with
11 decreased P -values in the joint analysis implied similar effects of the two sub-cohorts, and
12 we found that the 50 signals that had smaller P -values in the combined sample indeed
13 included all the 47 signals that were replicated in the independent phases 4 and 5 sub-
14 cohort (**Table S4**). Overall, this internal replication analysis showed the robustness and
15 validity of our data results. At least 84% of the 107 variant-level associations identified in
16 the UKB phases 1 to 5 sample could potentially be replicated at the 5% FDR level, should
17 a replication dataset become available. In the following two sections, we highlighted
18 some interesting findings across different organs. The complete list of all these 107
19 variant-level associations is presented in **Table S2**.

20

21 **Variant-level tests identified novel associations for hepatic and spleen MRI traits**

22 To our knowledge, associations between rare variants and abdominal MRI-derived traits
23 were only studied in Hass et al.¹⁴, where the focus was restricted to liver fat and only
24 18,103 subjects were included. Another study²⁰, with a similar focus on hepatic fat trait,
25 used CT imaging data from 9,594 individuals to study the exome-wide associations. Thus,
26 the roles of rare variants in a wider range of abdominal MRI traits were largely unexplored.
27 In our variant-level analysis, we identified associations between spleen volume and two
28 missense variants in *TNFRSF13B* (rs72553883, effect_{org} = 0.025 L, effect = 0.37 s.d. units,
29 95% CI = [0.29, 0.45], $P = 5.6 \times 10^{-19}$, and rs34557412, effect_{org} = 0.028 L, effect = 0.42 s.d.
30 units, 95% CI = [0.34, 0.49], $P = 2.0 \times 10^{-28}$), while the latter one was the top hit among all
31 variant-level associations. *TNFRSF13B* is a well-known risk gene of common variable
32 immunodeficiency and mutations in *TNFRSF13B* typically occurred in patients who

1 developed splenomegaly³¹⁻³³. In addition, the missense variant rs72553883 was
2 associated with a series of blood-related traits including platelet, myeloid white cell, and
3 lymphoid white cell indices in a previous GWAS³⁴ using genotyping array data while
4 similar associations between rare variants in *TNFRSF13B* and multiple blood biomarkers
5 were also observed in Karczewski et al.¹⁷ using exome data (**Fig. 2D**). Consistent with these
6 results on spleen abnormalities derived from immunodeficiency and the central role of
7 the spleen in blood filter and blood cell turnover, our finding points out a direct
8 association between spleen volume and missense mutations in *TNFRSF13B*. Another
9 missense variant in the manganese transporter *SLC30A10* (rs188273166, effect_{org} = 63.72
10 milliseconds^{35,36} [ms], effect = 1.18 s.d. units, 95% CI = [0.94, 1.42], $P = 2.1 \times 10^{-22}$) was
11 associated with higher corrected T1 liver iron. *SLC30A10* is known to be involved in
12 maintaining the manganese level³⁷. This association aligned with recent GWAS that
13 reported the role of rs188273166 in hypermagnesemia symptoms and a *SLC30A10*-
14 targeted study found its positive effect on corrected T1 liver iron^{38,39}. Interestingly, we
15 also identified a missense variant in *DDX51* associated with the fat-free muscle volume of
16 right posterior thigh (rs200735214, effect_{org} = -0.87 L, effect = -1.09 s.d. units, 95% CI = [-
17 1.42, -0.76], $P = 1.5 \times 10^{-10}$), which constitutes the only signal that passed our stringent P
18 value threshold for human muscle/fat composition traits. Notably, at a more permissive
19 P -value threshold at 1×10^{-8} as our suggestive evidence, we observed that rs200735214
20 in *DDX51* was also associated with the fat-free muscle volume for total thigh (effect_{org} = -
21 2.26 L, effect = -0.90 s.d. units, 95% CI = [-1.19, -0.60], $P = 1.95 \times 10^{-9}$) among other signals
22 for human muscle/fat composition traits such as *PLIN4* and posterior thigh muscle fat
23 infiltration (effect_{org} = -0.95%, effect = -0.40 s.d. units, 95% CI = [-0.53, -0.27], $P = 1.2 \times 10^{-9}$
24 ⁹ for the left; effect_{org} = -0.91%, effect = -0.38 s.d. units, 95% CI = [-0.51, -0.26], $P = 4.2 \times$
25 10^{-9} for the right). *DDX51*, a member of the DEAD-box helicase family, plays a crucial role
26 in ribosomal RNA processing and is essential for ribosome biogenesis^{40,41}. Although no
27 direct link between *DDX51* and muscle function has been reported previously, the strong
28 association identified in our exome data highlights its potential involvement in muscle
29 composition traits. Meanwhile, *PLIN4* encodes a protein known as Perilipin 4, which
30 belongs to the perilipin family of proteins that coat lipid droplets and involve in regulating
31 lipid metabolism⁴². Perilipins are essential for the proper storage and release of lipids

1 within cells. In particular, *PLIN4* has been shown to be highly expressed in skeletal muscle
2 and is found at periphery of skeletal muscle fibers⁴², consistent with our findings.

3 4 **Variant-level tests identified novel associations for brain and heart MRI traits**

5 We highlight some rare variants associations with regional brain volumes, brain DTI
6 parameters, and CMR traits of the heart and aorta. A missense variant in *ADRA1A* was
7 associated with the volume of the left ventral diencephalon (rs771722367, effect_{org} =
8 683.56 mm³, effect = 1.28 s.d. units, 95% CI = [0.90, 1.67], $P = 6.5 \times 10^{-11}$), though the P
9 value for this variant and right ventral diencephalon was $P = 3.6 \times 10^{-8}$, which did not pass
10 our liberal threshold at 1×10^{-8} . *ADRA1A* was consistently reported to be associated with
11 schizophrenia^{43,44} while enlarged brain regional volume of the ventral diencephalon area
12 in patients with schizophrenia was previously observed⁴⁵. The ventral diencephalon area
13 was also associated with a missense variant in *PKD1* (rs1181041827, effect_{org} = 923.88
14 mm³, effect = 1.73 s.d. units, 95% CI = [1.21, 2.24], $P = 5.2 \times 10^{-11}$ for right ventral
15 diencephalon; effect_{org} = 945.24 mm³, effect = 1.77 s.d. units, 95% CI = [1.26, 2.29], $P =$
16 1.4×10^{-11} for left ventral diencephalon). For DTI parameters, two missense variants
17 (rs2652098 and rs143368552) in *Versican (VCAN)* contributed the largest number of
18 associations (63 in total, effect range = [-0.48, 0.42] s.d. units, $P < 2.8 \times 10^{-10}$). **Figure 2E**
19 visualizes the associations between these two missense variants and 12 mean diffusivity
20 traits. Common variants in *VCAN* have been extensively associated with white matter
21 traits in previous GWAS^{2,5,8,46-48}. *VCAN* plays a pivotal role in various neural processes,
22 which may influence the pathophysiology of neurological disorders such as multiple
23 sclerosis⁴⁹⁻⁵¹. Another missense variant rs201680145 in *NOTCH3* was also associated with
24 multiple DTI parameters (effect range = [-1.50, 1.70] s.d. units, $P < 2.5 \times 10^{-10}$). Previous
25 studies have revealed the role of *NOTCH3* in white matter hyperintensities and several
26 neurodegenerative diseases⁵²⁻⁵⁴.

27
28 We found associations between a missense variant rs189569984 in *RBM20* and three
29 CMR traits, including left ventricular end-systolic volume (LVESV), right ventricular end-
30 systolic volume (RVESV) (effect_{org} = -3.77 mL, effect = -0.20 s.d. units, 95% CI = [-0.25, -
31 0.15], $P = 9.1 \times 10^{-16}$ for LVESV; effect_{org} = -3.37 mL, effect = -0.16 s.d. units, 95% CI = [-
32 0.21, -0.11], $P = 2.1 \times 10^{-11}$ for RVESV), and left ventricular ejection fraction (LVEF)

1 (effect_{org} = 1.49 %, effect = 0.24 s.d. units, 95% CI = [0.18, 0.30], $P = 6.0 \times 10^{-14}$). Mutations
2 in *RBM20* were previously known to be related to cardiovascular diseases including heart
3 failure and dilated cardiomyopathy⁵⁵⁻⁵⁷, and were associated with LVEF in a previous
4 GWAS on CMR traits⁹. Our rare variant analysis prioritized *RBM20* and provided additional
5 evidence for its role in regulating heart structure and function. Furthermore, we found
6 that ascending aorta maximum/minimum areas were associated with a missense variant
7 in *ANO1* (rs201870990, effect_{org} = 41.95 mm², effect = 0.22 s.d. units, 95% CI = [0.15, 0.29],
8 $P = 7.6 \times 10^{-11}$ for maximum; effect_{org} = 42.05 mm², effect = 0.23 s.d. units, 95% CI = [0.16,
9 0.29], $P = 1.4 \times 10^{-11}$ for minimum). *ANO1* was among the loci identified by a recent
10 GWAS⁵⁸ on ascending aorta diameter (see Figure 1 in their study⁵⁸) but was not pointed
11 out or discussed explicitly. Indeed, *ANO1* was also reported to be effective in preventing
12 cardiac fibrosis and may be a potential target for therapy^{59,60}. However, future research
13 might elucidate its role in aortic development and/or geometric remodeling. In summary,
14 our analysis of rare variants directly prioritized and implied a small set of genes related to
15 human organs. Due to the rarity of these variants and their lack of linkage disequilibrium
16 (LD) with common variants, they could not be effectively studied in previous GWAS that
17 focused on common variants.

18

19 **Gene-based burden tests identified complementary signals**

20 Gene-based burden tests enable us to study the collective effects of rare variants within
21 a gene, thus boosting the power. However, the involved burden models pose challenges
22 to the adjustment for multiple testing since they have unknown correlated structure. Thus,
23 we prioritized a P -value threshold at 1×10^{-9} based on two empirical null distribution^{61,62}
24 (Methods). Using the same dataset as in variant-level tests, we identified 224 significant
25 associations in gene-based tests ($P < 1 \times 10^{-9}$, Methods; **Table S5** for all the significant
26 results). As different burden models or MAF cutoffs may implicate the same gene-trait
27 associations (Methods), we further summarize the nonredundant results of 67 unique
28 gene-trait pairs⁶¹ (involving 26 genes and 57 MRI traits) in **Table 1** (a more comprehensive
29 version is provided in **Table S6**). For suggestive evidence, we additionally put all the
30 associations that passed a more relaxed P -value threshold at 1×10^{-8} in **Table S7**.
31 Manhattan plot for all the gene-level associations were presented in **Figure 3A**, across all
32 MRI phenotype categories. Consistent with observations in previous ExWAS^{17,62}, we

1 identified 21 additional genes, highlighting the power of gene-based burden tests in rare
2 variant association studies. In particular, we found 3 genes and 18 gene-trait associations
3 for brain functional MRI (fMRI) traits, which did not have any signals in variant-level tests.
4 To evaluate the reportability of our findings, we wanted to follow a similar procedure in
5 variant-level tests to perform an internal replication. However, many rare mutations,
6 particularly those with a MAF of less than 0.01%, accounted for a large proportion of our
7 discoveries but were not observed in the smaller independent dataset from phases 4 to
8 5 (average $n = 8,989$). Therefore, we only examined whether the associations in the
9 combined phases 1 to 5 sample had smaller P -values than those in the phases 1 to 3 data
10 (Methods). When we restricted the discovery sample to individuals in phases 1 to 3, there
11 were 96 significant associations for 25 unique gene-trait pairs ($P < 1 \times 10^{-9}$, **Table S8**).
12 Within these 96 associations, 79.2% (76/96, 19 gene-trait pairs) had smaller P -values in
13 the combined phases 1 to 5 sample, all of which had concordant directions of effects
14 (**Table S8**). We highlighted these 19 gene-trait pairs in **Table 1** and discussed some of the
15 67 gene-trait pairs in the following two sections.

16

17 **Genes associated with abdominal, brain, and heart MRI traits in burden tests**

18 As shown in **Table 1**, we identified 5 genes for abdominal MRI traits, 7 for CMR traits, 7
19 for regional brain volumes, 8 for brain DTI parameters, and 3 for brain fMRI traits. **Figure**
20 **3B** highlights the genes with large effect sizes. Below we highlight several interesting
21 findings for each trait category.

22

23 We found that abdominal MRI signals were sparse and robust, with all 5 signals from the
24 combined phases 1 to 5 sample also present in the phases 1 to 3 data. These signals were
25 the top-ranking signals among all imaging categories. The highest hit among all the gene-
26 level signals was the association between spleen volume and *TNFRSF13B* when
27 aggregating the effects of pLoF and Alpha damaging (Methods) missense variants
28 (effect_{org} = 0.027 L, effect = 0.38 s.d. units, 95% CI = [0.29, 0.48], $P = 1.7 \times 10^{-55}$). In addition
29 to *TNFRSF13B*, we observed that *SH2B3* was also associated with spleen volume (effect_{org}
30 = 0.025 L, effect = 0.40 s.d. units, 95% CI = [0.35, 0.45], $P = 1.7 \times 10^{-15}$), where spleen
31 abnormalities (e.g. splenomegaly) and associated blood traits have been previously
32 reported⁶³⁻⁶⁶. For muscle measurements, we found a novel association between fat-free

1 muscle volume of the posterior thigh and *TANC1* (effect_{org} = -1.06 L, effect = -1.32 s.d.
2 units, 95% CI = [-1.74, -0.90], $P = 8.5 \times 10^{-10}$). *TANC1* had genetic overlaps with the
3 identified loci in previous GWAS on heel bone mineral density⁶⁷ and its role in muscle
4 development and rhabdomyosarcoma has been discussed in previous studies^{68,69}.

5

6 As shown in **Figure 4A**, *TTN* was associated with 8 CMR traits, including LVESV (effect_{org} =
7 7.34 mL, effect = 0.40 s.d. units, 95% CI = [0.32, 0.47], $P = 1.2 \times 10^{-25}$), LVEF (effect_{org} = -
8 3.48 %, effect = -0.56 s.d. units, 95% CI = [-0.66, -0.47], $P = 2.1 \times 10^{-32}$), as well as global
9 and regional peak circumferential strain measurements (effect range = [0.32, 0.60] s.d.
10 units, P range = [1.2×10^{-21} , 2.4×10^{-11}]). These results make sense as *TTN* is a well-known
11 gene associated with cardiac structure⁹ and cardiovascular diseases such as heart failure⁷⁰,
12 dilated cardiomyopathy^{71,72}, atrial fibrillation⁷³, supraventricular tachycardia, and mitral
13 valve disease²¹. In addition to the previously known exome-wide association between
14 *TTN* and LVESV²¹, our findings suggest a broader influence of this gene on cardiac
15 structure and function, using a larger sample size and an expanded set of CMR traits.

16

17 We would like to highlight our exome-wide associations with CMR traits of aorta. Notably,
18 three genes (*ANO1*, *COL21A1*, *GEM*) were associated with both maximum and minimum
19 area of ascending aorta, while *PLCE1* was associated with maximum and minimum area
20 of descending aorta (**Fig. 4B**). In particular, both *COL21A1* and *GEM* lower the size of the
21 ascending aorta (effect_{org} range = [-61.16, -54.95] mm², P range = [1.7×10^{-13} , 6.4×10^{-13}]
22 for *COL21A1* and effect_{org} range = [-87.68, -87.06] mm², P range = [1.3×10^{-12} , 1.06×10^{-11}]
23 for *GEM*, respectively). Rare variants in *COL21A1* and *GEM* were reported to be
24 associated with higher pulse blood pressure in a previous study for blood pressure⁷⁴. A
25 smaller aorta leads to a wider pulse pressure because it increases aortic characteristic
26 impedance, which is highly dependent on aortic diameter. Our results on these aorta
27 area-lowering genes further provide the evidence for the underlying mechanisms of
28 previous observations. Moreover, *GEM* has a role in regulating cardiac activities⁷⁵ and is
29 related to heart failure^{75,76}. Another study⁷⁷ pointed out that such effects of *GEM* could
30 potentially be used as a gene therapy target in treating ventricular arrhythmias and heart
31 failure. Our novel discovery indicated that *GEM*'s regulatory effects may also impact the
32 structural integrity of the ascending aorta, thus improving our understanding of the role

1 of *GEM* in cardiovascular health. In contrast, missense variants in *ANO1* have positive
2 effect. The role of *ANO1* has been discussed in our variant-level analysis and we
3 recaptured these associations again by aggregating the missense variants in *ANO1* using
4 burden test (effect_{org} range = [32.43, 32.79] mm², *P* range = [1.7×10^{-13} , 6.4×10^{-13}]),
5 suggesting the similar effects and aligned directions across a group of missense variants
6 within *ANO1*. In addition, *TAGLN* was also associated with descending aorta minimum
7 area (effect_{org} = 62.28 mm², effect = 0.67 s.d. units, 95% CI = [0.46, 0.88], *P* = 4.91×10^{-11})
8 while its association with descending aorta maximum area was observed at a more liberal
9 *P* value threshold at 1×10^{-8} (**Table S7**). *GOLM1* was found to be associated with
10 descending aorta distensibility (effect_{org} = 2.85mmHg, effect = 2.5 s.d. units, 95% CI = [1.77,
11 3.24], *P* = 2.93×10^{-11}). A recent study demonstrated that *GOLM1* may present a potential
12 therapeutic target for treat sepsis-induced cardiac dysfunction in animal models⁷⁸.
13 *GOLM1* was also related to Alzheimer's disease and the implied cognitive deficits⁷⁹.
14 However, its role in aortic wall remodeling requires further study. Using GTEx data
15 resources⁸⁰, we found that four (*ANO1*, *COL21A1*, *GEM*, *TAGLN*) out of five genes (*ANO1*,
16 *COL21A1*, *GEM*, *TAGLN*, *PLCE1*) associated with ascending/descending aorta areas had
17 moderate to high expression and overexpressed in aorta while *PLCE1* had high expression
18 in artery tissues in general (Methods), which further supports their roles in regulating the
19 structure of aorta.

20

21 For brain MRI traits, we highlight in **Figures 3C** and **3D** the strongest association between
22 *PTEN* and total brain volume with extraordinary effect size when aggregating pLoF and
23 missense variants (effect_{org} = 386703.40 mm³, effect = 2.77 s.d. units, 95% CI = [2.22, 3.32],
24 *P* = 5.8×10^{-23}). The distribution of total brain volume in cm³ for *PTEN* mutation carriers
25 versus non-carriers is provided in **Figure 3C**. Notably, as illustrated in **Figure 3D**, different
26 burden masks would yield associations with various strengths, and we observed that the
27 signal became stronger if we aggregate ultrarare (MAF < 1×10^{-4}) pLoF variants and
28 damaging missense variants in *PTEN*, which may not be observed and testable in the
29 previous ExWAS^{16,17} with a sample size smaller than ours. In a recent ExWAS⁸¹, *PTEN* was
30 associated with autism spectrum disorder. Indeed, *PTEN* is a well-known gene that has
31 been consistently linked to autism spectrum disorder⁸²⁻⁸⁵, brain development⁸⁶⁻⁸⁸, and
32 brain clinical phenotypes including brain overgrowth⁸⁹⁻⁹¹, with an enlarged brain volume

1 representing an associated phenotype in these conditions. To our knowledge, no previous
2 GWAS or ExWAS has linked *PTEN* to any brain MRI measurements, thus our analysis
3 provided direct evidence connecting rare variants in *PTEN* with total brain volume.
4 Another gene *OMA1* was observed to be associated with 7 regional brain volume traits
5 including superior frontal gyrus, lateral orbitofrontal cortex, and precentral gyrus for both
6 sides of the brain and right superior parietal lobule (effect_{org} range = [-524.83, -171.02]
7 mm³ units, *P* range = [1.77×10^{-14} , 6.35×10^{-10}]). Notably, *PTEN* and *OMA1* regulate the
8 *PTEN*-induced kinase 1 (*PINK1*) and thus may be useful in preventing epileptogenesis⁹²
9 while a protective role of *OMA1* in neurodegeneration⁹³⁻⁹⁵ was previously reported. The
10 remarkably strong association *PTEN* demonstrate that even within a general healthy
11 cohort (as opposed to disease-focused or case-control studies), we can identify biological
12 links between genes associated with neurodevelopmental disorders using MRI data.

13

14 For DTI parameters, in contrast to variant-level analysis where missense variants in *VCAN*
15 accounted for the most associations, our gene-based burden tests revealed a more
16 diverse range of effects on white matter, with 8 distinct genes accounting for 10 gene-
17 trait associations. For brain fMRI traits, *PLCE1*, *GLUL*, and *KCNA5* were identified to be
18 associated with 18 phenotypes, where *PLCE1* contributed to 16 of them. We found that
19 many gene-trait pairs would not have been identified if singleton burden models or the
20 recent deep learning-based AlphaMissense²⁶ burden models were not included. We will
21 provide more detailed discussions about these interesting associations in the following
22 sections.

23

24 Overall, gene-based burden tests examined the aggregated effects of rare variations
25 within genes and advanced our understanding of the underlying genetic dispositions
26 across the abdomen, heart, and brain organs. Our novel findings prioritized a set of rare
27 genes previously unknown to be associated with human organs. We further discussed
28 examples of genetic overlaps between these MRI signals and other health-related
29 phenotypes, which may help understand complex diseases by providing exome-wide
30 insights into the genetic mechanisms underlying multiple human organs including
31 abdominal, cardiac, and neurological pathologies.

32

1 **Pleiotropic effects of *PLCE1* and *COL21A1* on brain and heart MRI traits**

2 In our gene-based burden tests, *PLCE1* was associated with the largest number of
3 phenotypes including 16 brain fMRI traits (including 5 traits from parcellation-based
4 approach⁹⁶ and 11 traits from whole brain spatial independent component analysis
5 [ICA]⁹⁷⁻⁹⁹) and 2 CMR traits (descending aorta maximum/minimum areas) (**Fig. 4B**). In
6 addition, *COL21A1* was associated with both brain cerebrospinal fluid (CSF) volume and
7 ascending aorta maximum/minimum areas (**Fig. 4B**). We found that a recent GWAS¹⁰⁰
8 discovered the associations of descending and ascending aorta areas separately with
9 *PLCE1* and *COL21A1* while *PLCE1* was also found in previous GWAS^{3,5,101} for brain fMRI.
10 *PLCE1* encodes a phospholipase (PLC ϵ) that catalyzes the hydrolysis of
11 phosphatidylinositol-4,5-bisphosphate to generate inositol 1,4,5-triphosphate and
12 diacylglycerol. They are common second messengers regulating multiple cellular
13 processes, including cell activation, growth, differentiation, and gene
14 expression¹⁰². Interestingly, prior mechanistic study identified a causal role for PLC ϵ on
15 the development of thoracic dilation¹⁰³ and dissection¹⁰⁴. Similarly, *COL21A1* encodes the
16 alpha-1 chain of collagen 21, a known extracellular matrix component in the arterial wall,
17 secreted by vascular smooth muscle cells. While the previous GWAS¹⁰⁰ suggested distinct
18 genetic bases between the ascending and descending aorta and linked aortic traits to
19 brain small vessel disease, it did not establish direct associations of these genes with other
20 brain MRI traits. Our results provide exome-wide evidence highlighting the associations
21 of *PLCE1* and *COL21A1* with both brain and heart MRI traits. Additionally, a previous study
22 suggested that the observed GWAS effects of *PLCE1* on fMRI traits may be blood-
23 derived¹⁰¹. Our exome-wide evidence supporting *PLCE1*'s role in both aortic and fMRI
24 traits aligns with this hypothesis.

25

26 The connection between the heart and brain has increasingly garnered attention^{10,105}.
27 Moreover, the role of hemodynamic role of the aorta involved the cushioning of pressure
28 and flow pulpability caused by the intermittent ventricular ejection, and an impaired
29 cushioning function for changes in aortic stiffness and/or diameter has been identified as
30 a key mechanism for target organ damage, including the brain and the heart¹⁰⁶. The
31 associations we identified, involving rare coding mutations in the same genes (*PLCE1* and
32 *COL21A1*) linked to heart, aortic and brain structure and function, suggest that similar

1 biological mechanisms may underlie the shared pathways of these organs that influence
2 both cardiac and neurological health.

3

4 **Novel associations uniquely identified by singletons and AlphaMissense**

5 Singleton variants are those observed only once in the study cohort. AlphaMissense²⁶ is a
6 recent deep learning-based method that offers a novel approach to annotating damaging
7 missense variants. Singleton burden model captured the rarest category in our
8 association tests and may be of special interest^{16,107}, and the incorporation of
9 AlphaMissense offered novel genetic findings in our study. Therefore, we would like to
10 highlight the contributions of these two groups of burden models in this section. In our
11 gene-based burden tests, 10 gene-trait pairs would not be detected without the singleton
12 burden models or including AlphaMissense as part of our annotation resources. Among
13 these signals, 8 of them even did not have any counterparts that passed the relaxed $P < 1$
14 $\times 10^{-8}$ threshold (**Table S9**).

15

16 Singleton damaging missense variants in glutamate-ammonia ligase *GLUL*, which encodes
17 the glutamine synthetase protein, were associated with functional connectivity of the
18 somatomotor and secondary visual networks (effect = 2.48 s.d. units, 95% CI = [1.68, 3.27],
19 $P = 8.9 \times 10^{-10}$). *GLUL* knockout mice were reported to be neuronally affected in multiple
20 regions including somatosensory and visual cortices¹⁰⁸. Our singleton analysis provides
21 consistent evidence in human genetics. In addition, a recent study suggested that visual-
22 somatosensory integration may be a new biomarker for preclinical Alzheimer's disease¹⁰⁹.
23 Together, rare mutations in *GLUL* may play a potential role in human neurodegenerative
24 diseases¹¹⁰⁻¹¹³. Another gene, *KCNA5* related to potassium voltage-gated channel, was
25 associated with functional activity in the subcortical-cerebellum network when we
26 combined the rare pLoF variants and Alpha damaging missense variants (effect = 0.59 s.d.
27 units, 95% CI = [0.41, 0.76], $P = 2.9 \times 10^{-11}$). Notably, voltage-gated potassium channels
28 are essential for neurons and cardiac activities and previous GWAS have discovered the
29 common variants mapped to *KCNA5* associated with cortical surface⁷ and thickness¹¹⁴. In
30 addition, a previous study on the rat cerebellum model revealed strong *KCNA5*
31 immunoreactivities in the cerebellar nuclei¹¹⁵. Consistent with these findings, we
32 provided additional evidence for *KCNA5*'s role in brain function. Moreover, some studies

1 also found close relationships between *KCNA5* and heart diseases including atrial
2 fibrillation¹¹⁶ and pulmonary arterial hypertension¹¹⁷. Indeed, *KCNA5* is an approved drug
3 target for the treatment of cardiac arrhythmias. Inspired by this additional context, we
4 investigated whether *KCNA5* was associated with any CMR traits. Interestingly, we found
5 that when aggregating the singleton pLoF and damaging missense variants, *KCNA5* was
6 associated with a regional peak circumferential strain measurement (effect_{org} = 5.11%,
7 effect = 1.08 s.d. units, 95% CI = [0.63, 1.54], $P = 3.2 \times 10^{-6}$, **Table S1**) at a P -value threshold
8 of 1.74×10^{-5} (0.05/2,870, Bonferroni adjusted for 82 CMR phenotypes across all the 7
9 variant function classes and 5 MAF classes). These discussions further suggested that
10 multi-organ imaging genetic studies would bring about insights to the complex interplay
11 across the brain-heart system. In summary, our results underscore the efficacy of our
12 gene-based burden test approach and highlight the importance of using multiple MAF
13 cutoffs and innovative annotation tools.

14

15 **Comparison with previous ExWAS for brain MRI and CMR traits**

16 In this section, we discuss the links between our gene-based rare variant signals and
17 existing ExWAS results on MRI traits. To our knowledge, there were no rare variant
18 associations reported for brain fMRI or abdominal MRI traits. The studies most similar to
19 ours are those by Backman et al.¹⁶, which included regional brain volumes and DTI
20 parameters among a broad range of phenotypes, and Jurgen et al.²¹, which focused on
21 CMR traits, while Pirruccello et al.¹¹⁸ declared no findings of rare variant associations for
22 aorta traits. Both studies used data from the UKB cohort but with smaller sample sizes
23 compared to our current study. Importantly, we analyzed different brain MRI and DTI
24 parameters, extracted from raw images using our own pipeline, compared to those
25 investigated by Backman et al.¹⁶. Additionally, we incorporated a broader set of CMR
26 traits, including strain and thickness metrics, expanding upon the traits analyzed by
27 Jurgen et al.²¹.

28

29 For regional brain volumes and DTI parameters (corresponding to STR and dMRI traits in
30 Backman et al.¹⁶), 5 genes (*AMPD3*, *HTRA1*, *MYCBP2*, *RBL1*, *SCUBE2*) whose associations
31 passed our stringent P -value threshold ($P < 1 \times 10^{-9}$) were also significantly associated
32 with their brain MRI traits. For example, associations between *AMPD3* and mean MD of

1 splenium of corpus callosum passed the stringent P value threshold both in our study and
2 in Backman et al.¹⁶. *SCUBE2* was significantly associated with the regional volume of the
3 left cerebellum exterior in our analysis, while its association with the volume of the
4 cerebellum cortex in the left hemisphere was reported in their result as suggestive
5 evidence. *MYCBP2* was associated with the external capsule in both our analysis and
6 Backman et al.¹⁶. Additionally, several associations between *PLEKHG3* and multiple dMRI
7 traits were identified in Backman et al.¹⁶ and two gene-trait pairs within these
8 associations passed our relaxed threshold ($P < 1 \times 10^{-8}$). For CMR traits, we replicated the
9 only exome-wide significant association between *TTN* and LVEF in Jurgen et al.²¹, while
10 their suggestive association between *TTN* and LVESV also passed our stringent threshold.
11 In addition, we found 7 novel associations between *TTN* and 7 peak circumferential strain
12 metrics. We summarized the overlapping associations in these previous ExWAS studies in
13 **Table S10**.

14

15 **Concordant evidence with GWAS signals**

16 Associations with rare coding variants could prioritize genes among the numerous loci
17 identified in GWAS for polygenic complex traits by providing concordant evidence^{16,21}. In
18 this section, we leveraged the previous GWAS summary statistics^{1-3,10,119} of all the same
19 596 imaging traits to compare the identified signals between common variants and rare
20 coding variants (Methods). As expected, we observed convergent evidence for all
21 categories of MRI traits. Briefly, more than half of the 174 rare coding signals (53.4% =
22 93/174, that is, 64/107 variant-trait associations and 29/67 gene-trait pairs) were within
23 the 1Mb range of the independent GWAS signals, which is consistent with the observation
24 in a previous large-scale UKB phenotype screening¹⁶. Based on our results, we found that
25 variant-level associations with regional brain volumes and CMR traits were not within the
26 neighborhood of GWAS signals, while all categories of MRI traits had at least one rare
27 coding association within the 1Mb of GWAS signal for signals from gene-level tests.

28

29 We zoomed into these shared signals between GWAS and ExWAS to provide more
30 detailed insights. The association between *SH2B3* and spleen volume identified in our
31 gene-level burden test was further supported by an independent GWAS signal rs2239194,
32 which is an expression quantitative trait locus (eQTL) in spleen tissue⁸⁰. For DTI

1 parameters, two missense variants (rs2652098 and rs143368552) in *VCAN* contributed
2 the most to the variant-level associations and all these associations were within 1Mb of
3 the GWAS signals. However, similar to findings from a previous GWAS study⁴⁸, we did not
4 identify any brain tissue-related eQTLs. Associations between *AMPD3* and DTI parameters
5 of splenium of corpus callosum were also accompanied by signals from GWAS. These
6 common variants are eQTLs of multiple brain tissues including cerebellum and spinal cord
7 where brain white matter also presents. For fMRI traits, all the associations between
8 *PLCE1* and 18 MRI traits (16 brain fMRI traits and 2 CMR traits) were consistent with
9 previous GWAS results, though there were not eQTLs in brain cortex tissues. On the other
10 hand, we found that the only signal for brain volumes that fell into the GWAS loci was the
11 association between *COL21A1* and CSF volume. The corresponding common variants
12 rs3857615 and rs9475654 were eQTLs in cerebellum and cerebellum hemisphere.

13

14 In addition, we identified gene-trait associations that were not reported in our previous
15 GWAS of the same MRI traits but appeared in other studies (e.g. *RBM20* and LVEF in
16 Pirruccello et al.⁹). We also observed that some rare coding associations in our study were
17 linked to GWAS loci of related traits within the same phenotype cluster. For example,
18 though both ExWAS and GWAS revealed the associations between *TTN* and multiple
19 regional peak circumferential strain measurements, the phenotype for the signals did not
20 exactly match. In summary, rare variant associations with MRI traits not only identified
21 novel signals but also prioritized genes among thousands of GWAS loci. These associations
22 provide new insights and context for understanding the genetic basis of structure and
23 function in the human brain and body. We provided these exome-wide signals that
24 showed convergent evidence with previous GWAS signals in **Tables S11-S12**.

25

26 **Multi-organ imaging genetics and drug targets**

27 We used the Therapeutic Target Database (TTD)¹²⁰ to query potential drug target genes
28 (Methods). Among the 26 unique genes identified by gene-based burden tests, 6 genes
29 were approved drug targets (*KCNA5* and *ANO1*) or in clinical trials (*TTK*, *GEM*, *LPAR3*, and
30 *TNFRSF13B*). The identified genes in gene-based burden tests were significantly enriched
31 for potential drug targets (6 out of 26, compared with 1,735 out of 17,448 genes, OR =
32 2.7, $P = 0.0391$, **Table S13**). Moreover, when including a broader set of potential drug

1 genes that were documented in TTD to be reported in the literature, the enrichment
2 became stronger (13 out of 26, compared with 3,468 out of 17,448 genes, OR = 4.0, $P =$
3 5.737×10^{-4} , **Table S13**). Consistent with the procedure in a previous study⁶² and
4 observations in other ExWAS^{16,62,121}, our results suggest that rare coding variants
5 identified in imaging genetics may facilitate the discovery of promising drug target
6 genes^{122,123}. Given the complicated nature of human organ interplay and the
7 corresponding biological pathways involved, we would emphasize that such enrichment
8 should be interpreted carefully as they might not directly point to the corresponding
9 diseases and related traits for existing drugs. However, these associations with human
10 organs could provide novel drug targets if they are supported by further evidence, and in
11 turn these MRI traits themselves might be useful to validate the new drugs if the
12 underlying mechanisms are well established¹²⁴. Moreover, As endophenotypes for
13 complex diseases, MRI traits can offer additional information on known drug targets, such
14 as potential off-targets and/or side effects^{122,125}, and also aid in drug repurposing (e.g.
15 *ANO1* and cardiac disease as discussed in the previous section^{59,60}).

16

17 **Burden heritability and genetic correlation**

18 To investigate the rare coding genetic architecture for multi-organ MRI traits, we applied
19 burden heritability regression (BHR)²⁷ to estimate the heritability for pLoF and damaging
20 missense variants (i.e. “int1 missense variants”, Methods), while the heritability for
21 synonymous variants is also calculated and it serves as negative control. Designed for rare
22 variants with $MAF < 1 \times 10^{-3}$, BHR stratifies variants based on functional annotation and
23 MAF to allow for different variant classes to have different mean effect sizes. Given our
24 relatively small sample size (average $n = 40,038$), we restricted our analysis to the ultra-
25 rare MAF bin ($MAF < 1 \times 10^{-4}$) which contained most of the coding variants in our analysis
26 (**Table S14**). As expected, ultra-rare pLoF variants consistently demonstrated higher
27 heritability than damaging missense variants across all MRI categories (**Fig. 5A**). **Tables**
28 **S15-S16** show the complete list of burden heritability for ultra-rare pLoF variants and
29 damaging missense variants. In contrast, there is no evidence that synonymous variants
30 have significant heritability, showing that the estimates from our results for rare variants
31 are well calibrated (**Table S17**). Furthermore, the highest average heritability of ultra-rare
32 pLoF variants were observed in the DTI parameters category. For ultra-rare pLoF variants,

1 the trait had the highest heritability was the caudal anterior cingulate ($h^2 = 0.012$, SE =
2 0.0035) among regional brain volumes, the mean axial diffusivity of the body of the corpus
3 callosum ($h^2 = 0.020$, SE = 0.0048) among DTI parameters, functional connectivity
4 between the second visual and auditory networks ($h^2 = 0.020$, SE = 0.0047) among brain
5 fMRI traits, regional peak circumferential strain ($h^2 = 0.016$, SE = 0.0053) among CMR
6 traits, and pancreas iron ($h^2 = 0.019$, SE = 0.0053) among abdominal MRI traits. We note
7 that these burden h^2 estimates are relatively large, which is similar to the high common
8 variant heritability for many imaging traits. This in general supports an endophenotype
9 model where genes have more direct connections with imaging phenotypes than
10 downstream health outcomes.

11

12 We further explored the burden genetic correlation pattern for ultra-rare pLoF variants
13 between MRI traits ($MAF < 1 \times 10^{-4}$) and 186 complex traits and diseases ($MAF < 1 \times 10^{-5}$)
14 downloaded from Genebass¹⁷ (**Table S18**). We focused on 67 heritable MRI traits that
15 passed the Bonferroni adjusted threshold at $P < 8.4 \times 10^{-5}$ ($= 0.05/596$). Consequently, we
16 estimated 12,462 (67×186) pairs of burden genetic correlations. We found 3 pairs that
17 passed the MRI-multiple testing threshold at $P < 7.5 \times 10^{-4}$ ($0.05/67$) and 591 pairs passed
18 a nominal P -value threshold at $P < 0.05$. Although no pairs survived the most stringent P -
19 value threshold adjusted for both the imaging traits and Genebass traits at $P < 4.0 \times 10^{-6}$
20 ($0.05/12,462$), we observed that there were interesting pairs in the top-ranking subset for
21 continuous traits from Genebass (**Table S19**). The results for binary traits can be found in
22 **Table S20**. **Figure 5B** shows the selected top hits of burden genetic correlation pairs,
23 including insular volume and reaction time¹²⁶, regional radial strain and systolic blood
24 pressure, left atrium maximum volume and triglycerides to high-density lipoprotein
25 cholesterol ratio, and abdominal fat ratio and birth weight. These burden genetic
26 correlations between MRI traits and health-related traits may reveal a shared genetic
27 basis in ultra-rare pLoF variants. While the current MRI sample size was slightly
28 underpowered, future larger samples may enable a more robust quantification and
29 confirmation of these genetic links.

30

31 **DISCUSSION**

1 We conducted a large-scale ExWAS for 596 multi-organ imaging traits including brain,
2 heart, liver, kidney, and lung. By using both WES and MRI data from over 50,000
3 participants from the UKB study, we uncovered how rare coding variants contribute to
4 human organ structure and function. The role of rare variants in regulating human organs
5 was largely understudied. For example, in abdominal organs, a prior ExWAS focused solely
6 on hepatic fat measurement from CT imaging²⁰, while another GWAS included a broader
7 set of traits¹³. However, with a relatively limited sample size of over 10,000, it failed to
8 identify rare variant associations. For CMR traits, our study incorporated refined
9 measurements of the left ventricle, right ventricle, ascending/descending aorta, and
10 left/right atrium. These were not examined in previous ExWAS²¹ but resulted in novel
11 discoveries in our analysis. Additionally, our results helped prioritize genes among
12 previously identified GWAS loci^{9,10}. For brain MRI traits, we analyzed a different set of
13 regional brain volumes and DTI parameters compared to previous ExWAS^{16,17}.
14 Additionally, we discovered associations between rare variants and fMRI traits, which had
15 not been previously reported.

16
17 Variant-level tests and gene-based burden tests provided different sets of associated
18 genes. Although it has been widely accepted that gene-based tests enhance signal
19 detection^{16,17}, annotating the deleterious variants is still an active research area. In group-
20 based tests, variants are collapsed into functional-frequency groups, but different
21 approaches may lead to inconsistent and/or incomplete results¹²⁷. Therefore, it is helpful
22 to leverage the annotation information from diverse resources to better empower the
23 set-based association test. We incorporated AlphaMissense²⁶ as part of our annotations,
24 resulting in the discovery of an additional set of genes (*GOLM1*, *KCNA5*, *LPAR3*, *PIGX*,
25 *TAGLN*, and *WIPF3*). AlphaMissense is a novel deep learning model for pathogenic
26 missense variant prediction and achieved good performance in predicting the unknown
27 clinical significance of many missense variants.

28
29 Over half of the identified rare variant signals were within 1Mb of GWAS common variants.
30 These rare variant associations could prioritize the causal genes among many identified
31 GWAS loci and uncover novel signals. Human organ MRI traits are widely used as
32 endophenotypes for diseases and health-related traits; our rare variant results deepen

1 our understanding from two aspects. First, we showed that many of the identified gene-
2 trait pairs were consistent with existing literature on the associated genes and their links
3 to organ-related diseases or complex traits. This alignment includes findings from
4 previous GWAS and ExWAS, observed clinical outcomes related to these genes, and
5 animal models that elucidate the biological mechanisms connecting the gene to the
6 phenotype. Second, we observed a significant enrichment of potential drug targets
7 among our rare variant signals. This suggests that rare variant imaging genetics can play
8 a role in identifying and repurposing drug targets, as well as understanding potential side
9 effects during drug development. These findings have practical implications for both real-
10 world clinical applications and future scientific research.

11

12 Along with these significant new insights, the present study has a few limitations. A
13 primary limitation is the relatively modest sample size for identifying rare variant
14 associations with MRI traits. For example, when the sample size increased from an
15 average of $n = 30,739$ in the phases 1 to 3 analysis to an average of $n = 40,038$ in the
16 phases 1 to 5 analysis, we observed a substantial increase in identified signals (84 vs. 174,
17 that is 59 variant-level associations and 25 unique gene-trait pairs in the phases 1 to 3
18 analysis, and 107 variant-level associations and 67 unique gene-trait pairs in the phases 1
19 to 5 analysis). It is reasonable to hypothesize that a large number of rare variant
20 associations have not been discovered for many MRI traits. Compared with recent ExWAS
21 studies for other phenotypes encompassing nearly or more than 500,000
22 participants^{16,17,21,62,128,129}, there remains substantial room for improvement in imaging
23 studies. This is especially important for MRI traits that have a highly polygenic
24 architecture, such as regional brain volumes, and is also critical for simultaneously
25 comparing multiple traits. As the UKB imaging project completes data collection from
26 100,000 subjects¹³⁰, we anticipate detecting additional signals and obtaining more robust
27 results. This will not only help identify associated rare coding variants and genes but also
28 elucidate the genetic architecture, such as burden heritability and burden genetic
29 correlation, facilitating downstream analyses for rare coding variants.

30

31 Another limitation is the lack of diversity in imaging ExWAS data resources. First, no
32 independent non-UKB database currently exists that combines both MRI traits and WES

1 data in a sample size comparable to the UKB study. As discussed above, rare variants
2 typically require a large sample size for detection and replication. Although our study
3 leveraged the largest available dataset and had enhanced statistical power, we were
4 unable to perform independent non-UKB replications because many rare variants may
5 not be observed in smaller samples. Consequently, we used an internal validation
6 procedure to assess the robustness of our findings. Nonetheless, future independent
7 studies with more diverse data resources are essential to replicate the signals we
8 identified. Second, our analysis primarily focused on individuals of European ancestry. It
9 is critical to extend these studies to underrepresented ancestry groups in genetic research
10 as data become available. Different ethnicities may exhibit heterogeneous genetic
11 architectures, and including a diverse range of ancestries could provide a more
12 comprehensive understanding of genetic influences on organ structure and function.

13

14 In conclusion, we have identified associations between rare coding variants and imaging
15 traits across human organs. These findings enhance our understanding of the genomic
16 mechanisms that regulate organ structure and function, potentially contributing to the
17 identification and prioritization of novel targets for pre-clinical and clinical drug
18 development. Looking ahead, it is important to increase sample sizes, integrate data from
19 diverse populations, and expand ExWAS to include a broader spectrum of imaging
20 phenotypes. We anticipate that future collaborative efforts will further elucidate the
21 genetic landscape of human organs, advancing our knowledge of human biology and
22 health.

23

24 **METHODS**

25 Methods are available in the *Methods* section.

26

27 **ACKNOWLEDGEMENTS**

28 Research reported in this publication was supported by the National Institute On Aging of
29 the National Institutes of Health under Award Number RF1AG082938 (B.Z. and H.Z.). The
30 content is solely the responsibility of the authors and does not necessarily represent the
31 official views of the National Institutes of Health. The study has also been partially
32 supported by funding from the Purdue University Statistics Department, Department of

1 Statistics and Data Science at the University of Pennsylvania, Wharton Dean’s Research
2 Fund, Analytics at Wharton, and Perelman School of Medicine CCEB Innovation Center
3 Grant (B.Z.). The study has also been partially supported by the National Institute On
4 Aging of the National Institutes of Health under Award Numbers U01AG079847 and
5 R01AR082684 (H.Z.). This research has been conducted using the UK Biobank resource
6 (application number 76139), subject to a data transfer agreement. We would like to thank
7 the individuals who represented themselves in the UK Biobank for their participation and
8 the research teams for their efforts in collecting, processing, and disseminating these
9 datasets. We would like to thank the research computing groups at the University of
10 North Carolina at Chapel Hill, Purdue University, and the Wharton School of the University
11 of Pennsylvania for providing computational resources and support that have contributed
12 to these research results.

13

14 **AUTHOR CONTRIBUTIONS**

15 Y.F. and B.Z. designed the study. Y.F., Jie C., Z.F., D.Y.Z., and Z.S analyzed the data and
16 generated the results. Tengfei L., S.H., Z. J, P.Y.G., X.Q., Ting L., and H.L processed the MRI
17 data. Julio C., J.L.S., P.F.S., R.W., A.N., M.D.R., J.O., W.W., D.J.R., and H.Z. provided
18 feedback on study design and helped results interpretations. Y.F. and B.Z. wrote the
19 manuscript with feedback from all authors.

20

21 **COMPETING INTERESTS**

22 R.W. is a current employee and/or stockholder of Regeneron Genetics Center or
23 Regeneron Pharmaceuticals. Other authors declare no competing interests.

24

25 **REFERENCES**

- 26 1. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies
27 variants influencing regional brain volumes and refines their genetic co-
28 architecture with cognitive and mental health traits. *Nat Genet* **51**, 1637-1644
29 (2019).
- 30 2. Zhao, B. *et al.* Common genetic variation influencing human white matter
31 microstructure. *Science* **372**(2021).

- 1 3. Zhao, B. *et al.* Common variants contribute to intrinsic human brain functional
2 networks. *Nat Genet* **54**, 508-517 (2022).
- 3 4. Zhao, B. *et al.* Genetic influences on the intrinsic and extrinsic functional
4 organizations of the cerebral cortex. *medRxiv*, 2021.07.27.21261187 (2023).
- 5 5. Elliott, L.T. *et al.* Genome-wide association studies of brain imaging phenotypes
6 in UK Biobank. *Nature* **562**, 210-216 (2018).
- 7 6. Smith, S.M. *et al.* An expanded set of genome-wide association studies of brain
8 imaging phenotypes in UK Biobank. *Nat Neurosci* **24**, 737-745 (2021).
- 9 7. Grasby, K.L. *et al.* The genetic architecture of the human cerebral cortex. *Science*
10 **367**(2020).
- 11 8. Zhao, B. *et al.* Large-scale GWAS reveals genetic architecture of brain white
12 matter microstructure and genetic overlap with cognitive and mental health
13 traits (n = 17,706). *Mol Psychiatry* **26**, 3943-3955 (2021).
- 14 9. Pirruccello, J.P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000
15 individuals yields genetic insights into dilated cardiomyopathy. *Nat Commun* **11**,
16 2254 (2020).
- 17 10. Zhao, B. *et al.* Heart-brain connections: Phenotypic and genetic insights from
18 magnetic resonance images. *Science* **380**, abn6598 (2023).
- 19 11. Khurshid, S. *et al.* Clinical and genetic associations of deep learning-derived
20 cardiac magnetic resonance-based left ventricular mass. *Nat Commun* **14**, 1558
21 (2023).
- 22 12. Streicher, S.A. *et al.* Genome-wide association study of abdominal MRI-measured
23 visceral fat: The multiethnic cohort adiposity phenotype study. *PLoS One* **18**,
24 e0279932 (2023).
- 25 13. Liu, Y. *et al.* Genetic architecture of 11 organ traits derived from abdominal MRI
26 using deep learning. *Elife* **10**(2021).
- 27 14. Haas, M.E. *et al.* Machine learning enables new insights into genetic
28 contributions to liver fat accumulation. *Cell Genom* **1**(2021).
- 29 15. Edwards, S.L., Beesley, J., French, J.D. & Dunning, A.M. Beyond GWASs:
30 illuminating the dark road from association to function. *Am J Hum Genet* **93**, 779-
31 97 (2013).

- 1 16. Backman, J.D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank
2 participants. *Nature* **599**, 628-634 (2021).
- 3 17. Karczewski, K.J. *et al.* Systematic single-variant and gene-based association
4 testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom*
5 **2**, 100168 (2022).
- 6 18. Van Hout, C.V. *et al.* Exome sequencing and characterization of 49,960
7 individuals in the UK Biobank. *Nature* **586**, 749-756 (2020).
- 8 19. Uffelmann, E. *et al.* Genome-wide association studies. *Nature Reviews Methods*
9 *Primers* **1**, 59 (2021).
- 10 20. Park, J. *et al.* Exome-wide association analysis of CT imaging-derived hepatic fat
11 in a medical biobank. *Cell Rep Med* **3**, 100855 (2022).
- 12 21. Jurgens, S.J. *et al.* Analysis of rare genetic variation underlying cardiometabolic
13 diseases and traits among 200,000 individuals in the UK Biobank. *Nat Genet* **54**,
14 240-250 (2022).
- 15 22. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. & Ng, P.C. SIFT missense predictions
16 for genomes. *Nat Protoc* **11**, 1-9 (2016).
- 17 23. Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. Predicting functional effect of human
18 missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7**, Unit7
19 20 (2013).
- 20 24. Chun, S. & Fay, J.C. Identification of deleterious mutations within three human
21 genomes. *Genome Res* **19**, 1553-61 (2009).
- 22 25. Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster
23 evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-
24 6 (2010).
- 25 26. Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with
26 AlphaMissense. *Science* **381**, eadg7492 (2023).
- 27 27. Weiner, D.J. *et al.* Polygenic architecture of rare coding variation across 394,783
28 exomes. *Nature* **614**, 492-499 (2023).
- 29 28. Park, J.H. *et al.* Distribution of allele frequencies and effect sizes and their
30 interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci*
31 *U S A* **108**, 18026-31 (2011).

- 1 29. Zeng, J. *et al.* Widespread signatures of natural selection across human complex
2 traits and functional genomic categories. *Nat Commun* **12**, 1164 (2021).
- 3 30. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient
4 than replication-based analysis for two-stage genome-wide association studies.
5 *Nat Genet* **38**, 209-13 (2006).
- 6 31. Salzer, U. *et al.* Mutations in TNFRSF13B encoding TACI are associated with
7 common variable immunodeficiency in humans. *Nat Genet* **37**, 820-8 (2005).
- 8 32. Artac, H., Bozkurt, B., Talim, B. & Reisli, I. Sarcoid-like granulomas in common
9 variable immunodeficiency. *Rheumatol Int* **30**, 109-12 (2009).
- 10 33. Pulvirenti, F. *et al.* Clinical Associations of Biallelic and Monoallelic TNFRSF13B
11 Variants in Italian Primary Antibody Deficiency Syndromes. *J Immunol Res* **2016**,
12 8390356 (2016).
- 13 34. Astle, W.J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and
14 Links to Common Complex Disease. *Cell* **167**, 1415-1429 e19 (2016).
- 15 35. Parisinos, C.A. *et al.* Genome-wide and Mendelian randomisation studies of liver
16 MRI yield insights into the pathogenesis of steatohepatitis. *J Hepatol* **73**, 241-251
17 (2020).
- 18 36. Mojtahed, A. *et al.* Reference range of liver corrected T1 values in a population
19 at low risk for fatty liver disease-a UK Biobank sub-study, with an appendix of
20 interesting cases. *Abdom Radiol (NY)* **44**, 72-84 (2019).
- 21 37. Mercadante, C.J. *et al.* Manganese transporter Slc30a10 controls physiological
22 manganese excretion and toxicity. *J Clin Invest* **129**, 5442-5461 (2019).
- 23 38. Ward, L.D. *et al.* GWAS of serum ALT and AST reveals an association of SLC30A10
24 Thr95Ile with hypermanganesemia symptoms. *Nat Commun* **12**, 4571 (2021).
- 25 39. Seidelin, A.S., Nordestgaard, B.G., Tybjaerg-Hansen, A., Yaghootkar, H. &
26 Stender, S. A rare genetic variant in the manganese transporter SLC30A10 and
27 elevated liver enzymes in the general population. *Hepatol Int* **16**, 702-711 (2022).
- 28 40. Martin, R., Straub, A.U., Doebele, C. & Bohnsack, M.T. DExD/H-box RNA helicases
29 in ribosome biogenesis. *RNA Biol* **10**, 4-18 (2013).
- 30 41. Hussain, A. DEAD Box RNA Helicases: Biochemical Properties, Role in RNA
31 Processing and Ribosome Biogenesis. *Cell Biochem Biophys* **82**, 427-434 (2024).

- 1 42. Pourteymour, S. *et al.* Perilipin 4 in human skeletal muscle: localization and
2 effect of physical activity. *Physiol Rep* **3**(2015).
- 3 43. Cheng, C. *et al.* Association of the ADRA1A gene and the severity of metabolic
4 abnormalities in patients with schizophrenia. *Prog Neuropsychopharmacol Biol*
5 *Psychiatry* **36**, 205-10 (2012).
- 6 44. Clark, D.A. *et al.* Polymorphisms in the promoter region of the alpha1A-
7 adrenoceptor gene are associated with schizophrenia/schizoaffective disorder in
8 a Spanish isolate population. *Biol Psychiatry* **58**, 435-9 (2005).
- 9 45. Goldman, A.L. *et al.* Heritability of brain morphology related to schizophrenia: a
10 large-scale automated magnetic resonance imaging segmentation study. *Biol*
11 *Psychiatry* **63**, 475-83 (2008).
- 12 46. Persyn, E. *et al.* Genome-wide association study of MRI markers of cerebral small
13 vessel disease in 42,310 participants. *Nat Commun* **11**, 2175 (2020).
- 14 47. Sargurupremraj, M. *et al.* Cerebral small vessel disease genomics and its
15 implications across the lifespan. *Nat Commun* **11**, 6285 (2020).
- 16 48. Rutten-Jacobs, L.C.A. *et al.* Genetic Study of White Matter Integrity in UK
17 Biobank (N=8448) and the Overlap With Stroke, Depression, and Dementia.
18 *Stroke* **49**, 1340-1347 (2018).
- 19 49. Ghorbani, S. *et al.* Versican promotes T helper 17 cytotoxic inflammation and
20 impedes oligodendrocyte precursor cell remyelination. *Nat Commun* **13**, 2445
21 (2022).
- 22 50. Gu, W.L. *et al.* Expression and regulation of versican in neural precursor cells and
23 their lineages. *Acta Pharmacol Sin* **28**, 1519-30 (2007).
- 24 51. Dours-Zimmermann, M.T. *et al.* Versican V2 assembles the extracellular matrix
25 surrounding the nodes of ranvier in the CNS. *J Neurosci* **29**, 7731-42 (2009).
- 26 52. Ramirez, J. *et al.* Parkinson's Disease, NOTCH3 Genetic Variants, and White
27 Matter Hyperintensities. *Mov Disord* **35**, 2090-2095 (2020).
- 28 53. Ceroni, M. *et al.* Migraine with aura and white matter abnormalities: Notch3
29 mutation. *Neurology* **54**, 1869-71 (2000).
- 30 54. Masoli, J.A.H., Pilling, L.C., Kuchel, G.A. & Melzer, D. Clinical Outcomes of
31 CADASIL-Associated NOTCH3 Mutations in 451,424 European Ancestry
32 Community Volunteers. *Transl Stroke Res* **10**, 339-341 (2019).

- 1 55. Larson, E.J. *et al.* Rbm20 ablation is associated with changes in the expression of
2 titin-interacting and metabolic proteins. *Mol Omics* **18**, 627-634 (2022).
- 3 56. Li, N., Hang, W., Shu, H. & Zhou, N. RBM20, a Therapeutic Target to Alleviate
4 Myocardial Stiffness via Titin Isoforms Switching in HFpEF. *Front Cardiovasc Med*
5 **9**, 928244 (2022).
- 6 57. Radke, M.H. *et al.* Therapeutic inhibition of RBM20 improves diastolic function in
7 a murine heart failure model and human engineered heart tissue. *Sci Transl Med*
8 **13**, eabe8952 (2021).
- 9 58. Tcheandjieu, C. *et al.* High heritability of ascending aortic diameter and trans-
10 ancestry prediction of thoracic aortic disease. *Nat Genet* **54**, 772-782 (2022).
- 11 59. Gao, Y. *et al.* ANO1 inhibits cardiac fibrosis after myocardial infraction via TGF-
12 beta/smad3 pathway. *Sci Rep* **7**, 2355 (2017).
- 13 60. Tian, X. *et al.* ANO1 regulates cardiac fibrosis via ATI-mediated MAPK pathway.
14 *Cell Calcium* **92**, 102306 (2020).
- 15 61. Dhindsa, R.S. *et al.* Rare variant associations with plasma protein levels in the UK
16 Biobank. *Nature* **622**, 339-347 (2023).
- 17 62. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK
18 Biobank exomes. *Nature* **597**, 527-532 (2021).
- 19 63. Blombery, P. *et al.* Biallelic deleterious germline SH2B3 variants cause a novel
20 syndrome of myeloproliferation and multi-organ autoimmunity. *EJHaem* **4**, 463-
21 469 (2023).
- 22 64. Morales, C.E., Stieglitz, E., Kogan, S.C., Loh, M.L. & Braun, B.S. Nf1 and Sh2b3
23 mutations cooperate in vivo in a mouse model of juvenile myelomonocytic
24 leukemia. *Blood Adv* **5**, 3587-3591 (2021).
- 25 65. Perez-Garcia, A. *et al.* Genetic loss of SH2B3 in acute lymphoblastic leukemia.
26 *Blood* **122**, 2425-32 (2013).
- 27 66. Arfeuille, C. *et al.* Germline bi-allelic SH2B3/LNK alteration predisposes to a
28 neonatal juvenile myelomonocytic leukemia-like disorder. *Haematologica*
29 (2023).
- 30 67. Kim, S.K. Identification of 613 new loci associated with heel bone mineral density
31 and a polygenic risk score for bone mineral density, osteoporosis and fracture.
32 *PLoS One* **13**, e0200785 (2018).

- 1 68. Avirneni-Vadlamudi, U. *et al.* Drosophila and mammalian models uncover a role
2 for the myoblast fusion gene TANC1 in rhabdomyosarcoma. *J Clin Invest* **122**,
3 403-7 (2012).
- 4 69. Granados, V.A. *et al.* Selective Targeting of Myoblast Fusogenic Signaling and
5 Differentiation-Arrest Antagonizes Rhabdomyosarcoma Cells. *Cancer Res* **79**,
6 4585-4591 (2019).
- 7 70. Povysil, G. *et al.* Assessing the Role of Rare Genetic Variation in Patients With
8 Heart Failure. *JAMA Cardiol* **6**, 379-386 (2021).
- 9 71. Micaglio, E. *et al.* Clinical Considerations for a Family with Dilated
10 Cardiomyopathy, Sudden Cardiac Death, and a Novel TTN Frameshift Mutation.
11 *Int J Mol Sci* **22**(2021).
- 12 72. Liu, J.S. *et al.* Whole-Exome Sequencing Identifies Two Novel TTN Mutations in
13 Chinese Families with Dilated Cardiomyopathy. *Cardiology* **136**, 10-14 (2017).
- 14 73. Choi, S.H. *et al.* Monogenic and Polygenic Contributions to Atrial Fibrillation Risk:
15 Results From a National Biobank. *Circ Res* **126**, 200-209 (2020).
- 16 74. Surendran, P. *et al.* Discovery of rare variants associated with blood pressure
17 regulation through meta-analysis of 1.3 million individuals. *Nat Genet* **52**, 1314-
18 1332 (2020).
- 19 75. Murata, M., Cingolani, E., McDonald, A.D., Donahue, J.K. & Marban, E. Creation
20 of a genetic calcium channel blocker by targeted gene transfer in the heart.
21 *Circ Res* **95**, 398-405 (2004).
- 22 76. Tan, F.L. *et al.* The gene expression fingerprint of human heart failure. *Proc Natl*
23 *Acad Sci U S A* **99**, 11387-92 (2002).
- 24 77. Paar, V. *et al.* Pathophysiology of Calcium Mediated Ventricular Arrhythmias and
25 Novel Therapeutic Options with Focus on Gene Therapy. *Int J Mol Sci* **20**(2019).
- 26 78. Xing, Y. *et al.* Golgi Protein 73 Promotes LPS-Induced Cardiac Dysfunction via
27 Mediating Myocardial Apoptosis and Autophagy. *J Cardiovasc Pharmacol* **83**,
28 116-125 (2024).
- 29 79. Inkster, B. *et al.* Genetic variation in GOLM1 and prefrontal cortical volume in
30 Alzheimer's disease. *Neurobiol Aging* **33**, 457-65 (2012).
- 31 80. Consortium, G.T. The GTEx Consortium atlas of genetic regulatory effects across
32 human tissues. *Science* **369**, 1318-1330 (2020).

- 1 81. Satterstrom, F.K. *et al.* Large-Scale Exome Sequencing Study Implicates Both
2 Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**,
3 568-584 e23 (2020).
- 4 82. Frazier, T.W. *et al.* Molecular and phenotypic abnormalities in individuals with
5 germline heterozygous PTEN mutations and autism. *Mol Psychiatry* **20**, 1132-8
6 (2015).
- 7 83. Rademacher, S. & Eickholt, B.J. PTEN in Autism and Neurodevelopmental
8 Disorders. *Cold Spring Harb Perspect Med* **9**(2019).
- 9 84. Busch, R.M. *et al.* Neurobehavioral phenotype of autism spectrum disorder
10 associated with germline heterozygous mutations in PTEN. *Transl Psychiatry* **9**,
11 253 (2019).
- 12 85. Cummings, K., Watkins, A., Jones, C., Dias, R. & Welham, A. Behavioural and
13 psychological features of PTEN mutations: a systematic review of the literature
14 and meta-analysis of the prevalence of autism spectrum disorder characteristics.
15 *J Neurodev Disord* **14**, 1 (2022).
- 16 86. Clipperton-Allen, A.E. *et al.* Pten haploinsufficiency disrupts scaling across brain
17 areas during development in mice. *Transl Psychiatry* **9**, 329 (2019).
- 18 87. Clipperton-Allen, A.E. *et al.* Pten haploinsufficiency causes desynchronized
19 growth of brain areas involved in sensory processing. *iScience* **25**, 103796 (2022).
- 20 88. Chen, Y., Huang, W.C., Sejourne, J., Clipperton-Allen, A.E. & Page, D.T. Pten
21 Mutations Alter Brain Growth Trajectory and Allocation of Cell Types through
22 Elevated beta-Catenin Signaling. *J Neurosci* **35**, 10252-67 (2015).
- 23 89. Koboldt, D.C. *et al.* PTEN somatic mutations contribute to spectrum of cerebral
24 overgrowth. *Brain* **144**, 2971-2978 (2021).
- 25 90. Vanderver, A. *et al.* Characteristic brain magnetic resonance imaging pattern in
26 patients with macrocephaly and PTEN mutations. *Am J Med Genet A* **164A**, 627-
27 33 (2014).
- 28 91. Dhamija, R. & Hoxworth, J.M. Imaging of PTEN-related abnormalities in the
29 central nervous system. *Clin Imaging* **60**, 180-185 (2020).
- 30 92. Cui, Y. *et al.* Inhibition of PTEN-induced kinase 1 autophosphorylation may assist
31 in preventing epileptogenesis induced by pentylentetrazol. *Neurochem Int* **172**,
32 105644 (2024).

- 1 93. Korwitz, A. *et al.* Loss of OMA1 delays neurodegeneration by preventing stress-
2 induced OPA1 processing in mitochondria. *J Cell Biol* **212**, 157-66 (2016).
- 3 94. Franchino, C.A. *et al.* Sustained OMA1-mediated integrated stress response is
4 beneficial for spastic ataxia type 5. *Brain* **147**, 1043-1056 (2024).
- 5 95. Anderson, C.J. *et al.* Prohibitin levels regulate OMA1 activity and turnover in
6 neurons. *Cell Death Differ* **27**, 1896-1906 (2020).
- 7 96. Glasser, M.F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature*
8 **536**, 171-178 (2016).
- 9 97. Beckmann, C.F. & Smith, S.M. Probabilistic independent component analysis for
10 functional magnetic resonance imaging. *IEEE transactions on medical imaging*
11 **23**, 137-152 (2004).
- 12 98. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component
13 analysis. *IEEE transactions on Neural Networks* **10**, 626-634 (1999).
- 14 99. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000
15 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400-424 (2018).
- 16 100. Francis, C.M. *et al.* Genome-wide associations of aortic distensibility suggest
17 causality for aortic aneurysms and brain white matter hyperintensities. *Nat*
18 *Commun* **13**, 4505 (2022).
- 19 101. J, P.O.F.T.G., Sprooten, E., Beckmann, C.F., Franke, B. & Bralten, J. Shared genetic
20 influences on resting-state functional networks of the brain. *Hum Brain Mapp*
21 **43**, 1787-1803 (2022).
- 22 102. Smrcka, A.V., Brown, J.H. & Holz, G.G. Role of phospholipase Cepsilon in
23 physiological phosphoinositide signaling networks. *Cell Signal* **24**, 1333-43
24 (2012).
- 25 103. Hodgkinson, C.P. & Gomez, J.A. Loss of PLCepsilon activity as a culprit of
26 ascending aortic dilation and aortic aneurysm. *Am J Physiol Heart Circ Physiol*
27 **324**, H659-H661 (2023).
- 28 104. Atchison, D.K. *et al.* Phospholipase Cepsilon insufficiency causes ascending aortic
29 aneurysm and dissection. *Am J Physiol Heart Circ Physiol* **323**, H1376-H1387
30 (2022).
- 31 105. Testai, F.D. *et al.* Cardiac Contributions to Brain Health: A Scientific Statement
32 From the American Heart Association. *Stroke* (2024).

- 1 106. Chirinos, J.A., Segers, P., Hughes, T. & Townsend, R. Large-Artery Stiffness in
2 Health and Disease: JACC State-of-the-Art Review. *J Am Coll Cardiol* **74**, 1237-
3 1263 (2019).
- 4 107. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association
5 studies. *Proc Natl Acad Sci U S A* **111**, E455-64 (2014).
- 6 108. Zhou, Y. *et al.* Selective deletion of glutamine synthetase in the mouse cerebral
7 cortex induces glial dysfunction and vascular impairment that precede epilepsy
8 and neurodegeneration. *Neurochem Int* **123**, 22-33 (2019).
- 9 109. Mahoney, J.R. & Verghese, J. Visual-Somatosensory Integration and Quantitative
10 Gait Performance in Aging. *Front Aging Neurosci* **10**, 377 (2018).
- 11 110. Chen, J. & Herrup, K. Chapter 70 - Glutamine as a Potential Neuroprotectant in
12 Alzheimer's Disease. in *Diet and Nutrition in Dementia and Cognitive Decline*
13 (eds. Martin, C.R. & Preedy, V.R.) 761-771 (Academic Press, San Diego, 2015).
- 14 111. Andersen, J.V. *et al.* Deficient astrocyte metabolism impairs glutamine synthesis
15 and neurotransmitter homeostasis in a mouse model of Alzheimer's disease.
16 *Neurobiol Dis* **148**, 105198 (2021).
- 17 112. Robinson, S.R. Neuronal expression of glutamine synthetase in Alzheimer's
18 disease indicates a profound impairment of metabolic interactions with
19 astrocytes. *Neurochem Int* **36**, 471-82 (2000).
- 20 113. Jayakumar, A.R. & Norenberg, M.D. Glutamine Synthetase: Role in Neurological
21 Disorders. *Adv Neurobiol* **13**, 327-350 (2016).
- 22 114. Shadrin, A.A. *et al.* Vertex-wise multivariate genome-wide association study
23 identifies 780 unique genetic loci associated with cortical morphology.
24 *Neuroimage* **244**, 118603 (2021).
- 25 115. Chung, Y.H., Shin, C., Kim, M.J., Lee, B.K. & Cha, C.I. Immunohistochemical study
26 on the distribution of six members of the Kv1 channel subunits in the rat
27 cerebellum. *Brain Res* **895**, 173-7 (2001).
- 28 116. Yang, Y. *et al.* Novel KCNA5 loss-of-function mutations responsible for atrial
29 fibrillation. *J Hum Genet* **54**, 277-83 (2009).
- 30 117. Vera-Zambrano, A. *et al.* Novel Loss-of-Function KCNA5 Variants in Pulmonary
31 Arterial Hypertension. *Am J Respir Cell Mol Biol* **69**, 147-158 (2023).

- 1 118. Pirruccello, J.P. *et al.* Deep learning enables genetic analysis of the human
2 thoracic aorta. *Nat Genet* **54**, 40-51 (2022).
- 3 119. Zhao, B. *et al.* An atlas of trait associations with resting-state and task-evoked
4 human brain functional organizations in the UK Biobank. *Imaging Neuroscience*
5 **1**, 1-23 (2023).
- 6 120. Zhou, Y. *et al.* TTD: Therapeutic Target Database describing target druggability
7 information. *Nucleic Acids Res* **52**, D1465-D1477 (2024).
- 8 121. Sun, B.B. *et al.* Genetic associations of protein-coding variants in human disease.
9 *Nature* **603**, 95-102 (2022).
- 10 122. Szustakowski, J.D. *et al.* Advancing human genetics research and drug discovery
11 through exome sequencing of the UK Biobank. *Nat Genet* **53**, 942-948 (2021).
- 12 123. Trajanoska, K. *et al.* From target discovery to clinical drug development with
13 human genetics. *Nature* **620**, 737-745 (2023).
- 14 124. Steckler, T. & Salvatore, G. Chapter 7 - Neuroimaging as a Translational Tool in
15 Animal and Human Models of Schizophrenia. in *Translational Neuroimaging* (ed.
16 McArthur, R.A.) 195-220 (Academic Press, 2013).
- 17 125. Nguyen, P.A., Born, D.A., Deaton, A.M., Nioi, P. & Ward, L.D. Phenotypes
18 associated with genes encoding drug targets are predictive of clinical trial side
19 effects. *Nat Commun* **10**, 1579 (2019).
- 20 126. Vicario, C.M., Nitsche, M.A., Salehinejad, M.A., Avanzino, L. & Martino, G. Time
21 Processing, Interoception, and Insula Activation: A Mini-Review on Clinical
22 Disorders. *Front Psychol* **11**, 1893 (2020).
- 23 127. McCarthy, D.J. *et al.* Choice of transcripts and software has a large effect on
24 variant annotation. *Genome Med* **6**, 26 (2014).
- 25 128. Akbari, P. *et al.* Sequencing of 640,000 exomes identifies GPR75 variants
26 associated with protection from obesity. *Science* **373**(2021).
- 27 129. Tian, R. *et al.* Whole-exome sequencing in UK Biobank reveals rare genetic
28 architecture for depression. *Nat Commun* **15**, 1755 (2024).
- 29 130. Littlejohns, T.J. *et al.* The UK Biobank imaging enhancement of 100,000
30 participants: rationale, data collection, management and future directions. *Nat*
31 *Commun* **11**, 2624 (2020).

- 1 131. Glasser, M.F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature*
2 **536**, 171-178 (2016).
- 3 132. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and
4 richer datasets. *GigaScience* **4**, s13742-015-0047-8 (2015).
- 5 133. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122
6 (2016).
- 7 134. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from
8 variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
- 9 135. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of
10 transcript-specific functional predictions and annotations for human
11 nonsynonymous and splice-site SNVs. *Genome Med* **12**, 103 (2020).
- 12 136. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human
13 nonsynonymous SNPs and their functional predictions. *Hum Mutat* **32**, 894-9
14 (2011).
- 15 137. Agrawal, S. *et al.* Inherited basis of visceral, abdominal subcutaneous and
16 gluteofemoral fat depots. *Nat Commun* **13**, 3771 (2022).
- 17 138. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for
18 quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021).
- 19 139. Fadista, J., Manning, A.K., Florez, J.C. & Groop, L. The (in)famous GWAS P-value
20 threshold revisited and updated for low-frequency variants. *Eur J Hum Genet* **24**,
21 1202-5 (2016).
- 22 140. Gao, X.R., Chiariglione, M. & Arch, A.J. Whole-exome sequencing study identifies
23 rare variants and genes associated with intraocular pressure and glaucoma. *Nat*
24 *Commun* **13**, 7376 (2022).
- 25 141. Dhindsa, R.S. *et al.* A minimal role for synonymous variation in human disease.
26 *Am J Hum Genet* **109**, 2105-2109 (2022).
- 27 142. Cui, X. *et al.* Endothelial nitric oxide synthase regulates white matter changes via
28 the BDNF/TrkB pathway after stroke in mice. *PLoS One* **8**, e80358 (2013).
- 29 143. Elahi, F.M. *et al.* "Liquid Biopsy" of White Matter Hyperintensity in Functionally
30 Normal Elders. *Front Aging Neurosci* **10**, 343 (2018).
- 31 144. Winkler, E.A. *et al.* A single-cell atlas of the normal and malformed human brain
32 vasculature. *Science* **375**, eabi7377 (2022).

- 1 145. Crouch, E.E., Joseph, T., Marsan, E. & Huang, E.J. Disentangling brain vasculature
2 in neurogenesis and neurodegeneration using single-cell transcriptomics. *Trends*
3 *Neurosci* **46**, 551-565 (2023).
4 146. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional
5 mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**,
6 1826 (2017).
7 147. Raney, B.J. *et al.* The UCSC Genome Browser database: 2024 update. *Nucleic*
8 *Acids Res* **52**, D1082-D1088 (2024).

9

10 **METHODS**

11 **Ethics**

12 This study made use of the data from UK Biobank (UKB) involving approximately 500,000
13 participants aged from 40 to 69 when recruited between 2006 and 2010
14 (<https://www.ukbiobank.ac.uk/>). The UKB study received the ethics approval from the
15 North West Multi-centre Research Ethics Committee (reference number: 11/NW/0382)
16 with informed consent obtained by all the participants. The present study was under UKB
17 application number 76139.

18

19 **Multi-organ imaging phenotypes**

20 The information of imaging phenotypes can be found in previous studies^{1-4,10}. Briefly, we
21 studied 596 imaging traits encompassing brain MRI traits¹⁻⁴, heart CMR traits¹⁰, and
22 abdominal MRI traits. Three major modalities of brain MRI were included. First, we
23 included 101 regional brain volume traits¹ derived from structural MRI. Second, we used
24 110 tract-averaged parameters² from DTI capturing the microstructure of brain white
25 matter. Third, for fMRI traits, we involved 76 node amplitude traits and six global
26 functional connectivity traits based on ICA³ as part of resting fMRI traits; we also included
27 180 (90+90) parcellation-based^{4,131} resting fMRI traits and task fMRI traits. Regarding CMR
28 traits, we made use of 82 traits¹⁰ including regional and local measurements from cardiac
29 chambers and the aorta. For abdominal MRI traits, we contained 41 imaging traits based
30 on MRI data of the liver, kidneys, lungs, pancreas, spleen, and body muscle/fat
31 composition. A full list of these 596 traits can be found in **Tables S1**.

32

1 **Internal replication and joint analysis design**

2 Given the limited sample size (average $n = 40,038$ across imaging traits with non-missing
3 data) and lack of independent data sources, we prioritized our signals through joint
4 analysis³⁰ and adopted an internal validation procedure to evaluate the robustness of our
5 discoveries. Specifically, we first restricted our sample to include only participants from
6 phases 1 to 3 of MRI release (average $n = 30,739$) and then excluded any individuals
7 related to phases 1 to 3 participants in the rest of the sample from phases 4 to 5 of MRI
8 release (average $n = 8,989$, relatives of the phases 1 to 3 were removed). The joint sample
9 included all the individuals from phases 1 to 5 and was used to report significant results.
10 To investigate the robustness of variant-level associations, we tested the associations on
11 phases 1 to 3 individuals and phases 4 to 5 individuals separately. Then, focusing on the
12 significant level and effect sign, we investigated (i) whether signals identified from the
13 phases 1 to 3 sample could be replicated in the independent phases 4 to 5 sample and (ii)
14 whether the signals identified from phases 1 to 3 sample had stronger evidence (i.e.
15 smaller P values and concordant effect directions) in the joint sample. For gene-based
16 burden tests, we noted that many rare mutations, especially those variants with $MAF < 1$
17 $\times 10^{-4}$, accounted for a large proportion of our discoveries but were not observed in the
18 smaller independent dataset from phases 4 to 5 (average $n = 8,989$; even within a burden
19 model with MAF cutoff at 0.01, the lack of those rare variants in the sample from phase
20 4 to phase 5 would make the burden associations incomparable). Thus, we only
21 investigated whether the identified associations from phases 1 to 3 dataset had stronger
22 evidence in the joint sample. Consequently, based on this internal replication procedure,
23 we were able to expect the level to which our discoveries from the joint analysis could be
24 replicated.

25

26 **Quality control for UK Biobank exome data**

27 We used Plink v.2.0¹³² to conduct the quality control steps, restricting the sample to
28 individuals with imaging data ($n = 54,365$). For exome sequencing data, we included all
29 the variants that had a MAF below 0.01. Variant-level quality control excluded all the
30 variants that had missing rate larger than 10% or had Hardy-Weinberg equilibrium $P < 1$
31 $\times 10^{-15}$. Sample-level quality control excluded any sample with missing rate over 10%.

1 Consequently, no individual was excluded from the quality control process, while
2 8,127,841 variants remained for the downstream analyses before annotation.

3

4 **Functional annotation for protein-coding variants**

5 We used Variant Effect Predictor¹³³ (VEP v.108) for variant annotation. Each variant was
6 mapped to the most severe consequence across the canonical transcripts. We defined
7 the loss-of-function variants using the Loss-of-Function Transcript Effect Estimator¹³⁴
8 (LOFTEE) plugin, which further collapsed stop-gained, essential splice, and frameshift
9 variants into high-confidence predicted loss-of-function variants (hcpLoF or pLoF
10 hereafter) or low-confidence predicted loss-of-function variants (lcpLoF). Missense
11 variants were prioritized by dbNSFP^{135,136} (v.4.5a) plugin using five prediction algorithms¹⁶:
12 SIFT²², PolyPhen2 HDIV²³, PolyPhen2 HVAR²³, LRT²⁴, and MutationTaster²⁵. Missense
13 variants were defined as “int5 damaging missense variants” if predicted damaging or
14 possible damaging by the intersection of all the five algorithms and “int1 damaging
15 missense variants” if predicted damaging or possible damaging by any one of the five
16 algorithms. Parallely, we also used AlphaMissense²⁶ to assign the missense variants to be
17 “Alpha damaging” if predicted as “pathogenic”. Synonymous variants served as empirical
18 null control to support the study-wise *P*-value threshold for burden tests. Predicted loss-
19 of-function variants (including both hcpLoF and lcpLoF) and missense variants were
20 included in our downstream association tests. They were referred to as non-synonymous
21 variants or coding variants of interest ($n = 2,143,707$).

22

23 **Exome-wide association tests**

24 The covariate adjustment for brain MRI, DTI parameters, brain fMRI, and heart CMR traits
25 was consistent with our previous GWAS^{1-3,10,119}. In short, we adjusted a set of basic
26 covariates for all the imaging traits including age (at imaging), age-squared, sex, age-sex
27 interaction, age-squared-sex interaction, imaging site, and the top 40 genetic PCs (for the
28 phases 4 to 5 sample replication analysis, we only adjusted for top 10 genetic PCs). For
29 brain structural MRI traits, we additionally adjusted for total brain volume (for traits other
30 than itself). For brain fMRI traits, we additionally adjusted for the effects of volumetric
31 scaling, head motion, head motion-squared, brain position, and brain position-squared.
32 For heart CMR traits, we additionally adjusted for the effects of standing height and

1 weight. For abdominal MRI traits¹³⁷, we additionally adjusted for the effects of standing
2 height and body mass index.

3

4 We used REGENIE (v. 3.1.3)¹³⁸ to conduct the exome-wide association tests. Common
5 variants from genotyping array data (MAF > 0.01, genotype missing rate < 10%, Hardy-
6 Weinberg equilibrium $P > 1 \times 10^{-15}$, LD pruning with $r^2 < 0.9$) were included in the step 1
7 of REGENIE to capture genome-wide polygenic effects. Then, the predictors obtained
8 from step 1 were used in step 2 for both variant-level tests and gene-based burden tests.
9 For variant-level tests, we tested all the coding variants of interest as defined above. For
10 gene-based burden tests, we included three general types of burden masks: (i) hcpLoF
11 variants only, (ii) damaging missense variants only, and (iii) the combinations hcpLoF,
12 lcpLoF, and damaging missense variants. For damaging missense variants, we further
13 defined three categories, that is “int5 damaging”, “int1 damaging”, and “Alpha damaging”
14 missense variants as described above, which resulted in seven finer variant sets for each
15 gene: (i) hcpLoF variants only, (ii) “Alpha damaging” missense variants only, (iii) “int5
16 damaging” missense variants only, (iv) “int1 damaging” missense variants only, (v) hcpLoF
17 variants and “Alpha damaging” missense variants, (vi) hcpLoF variants and “int5 damaging”
18 missense variants, and (vii) hcpLoF variants, lcpLoF variants, and “int1 missense” variants.
19 Additionally, for each gene set, we considered four levels of variants based on the
20 alternative allele frequency (REGENIE used alternative allele frequency for separation, but
21 we would keep the notation MAF as they were mostly concordant in our analysis):
22 singletons only, $MAF \leq 1 \times 10^{-4}$, $MAF \leq 1 \times 10^{-3}$, and $MAF \leq 1 \times 10^{-2}$. We noted that some
23 sets may not be testable due to lack of qualified variants (for example, considering a gene
24 that only had variants with $MAF > 1 \times 10^{-3}$) while some sets would produce repetitive
25 results (for example, considering a gene that only had variants with $MAF \leq 1 \times 10^{-4}$). To
26 account for these issues, we adopted an empirical null based P -value threshold for
27 multiple testing adjustment as described below.

28

29 **Study-wise significance level for association tests**

30 For variant-level associations, we applied the Bonferroni correction at 0.05 level, which
31 yields a conservative P -value cutoff at 2.8×10^{-10} ($= 0.05/178,280,016$, that is 0.05
32 adjusted for all the variant-level tests for coding variants of interest, see **Table S21**).

1 Moreover, we also provided the results for the variant-level associations that passed the
2 rare variant genome-wide significance threshold at 1×10^{-8} as suggestive evidence^{139,140}.
3 The P -value cutoff for phases 1 to 3 sample was 3.5×10^{-10} ($= 0.05/142,935,657$, **Table**
4 **S21**).

5

6 For gene-based burden tests, Bonferroni correction for all the tests conducted may be
7 too strict and inappropriate here since distinct burden models and MAF cutoffs were
8 highly correlated and may produce repetitive results. Moreover, the dependence
9 structure made it unclear whether the Benjamini and Hochberg procedure could provide
10 valid false discovery rate control. Alternatively, we used resampling-based methods to
11 derive the empirical distributions, which suggested setting 1×10^{-9} as our conservative P -
12 value threshold for burden tests. Specifically, we investigated the empirical null
13 distributions of P -values from both permutation test^{61,62} and synonymous burden
14 test^{62,141} separately. We conducted the permutation for the imaging traits once for every
15 phenotype while preserving the genetic structure and the burden models^{61,62}. At the tail
16 of 243,691,511 P -values derived from the permutation-based null test, we only observed
17 5 results that had P -values smaller than 1×10^{-9} (**Table S22**). Therefore, based on this
18 permutation-derived threshold at 1×10^{-9} , the expected false discoveries would be 5 out
19 of 224 significant associations (3 out of 67 significant gene-trait pairs) across all burden
20 models and imaging phenotypes.

21

22 Furthermore, we found that under this permutation-based P -value cutoff, only
23 synonymous variants in *NOSTRIN* and *HIGD1B* were observed to be associated with the
24 mean diffusivity of the cingulum (cingulate gyrus) tract in brain white matter and the
25 volume of left rostral middle frontal region, respectively, at the tail of the distribution of
26 40,545,911 P -values from synonymous burden tests (**Table S23**). These two genes were
27 previously reported to be associated with white matter^{142,143} and other brain-related
28 traits^{144,145}, although the biological significance of the roles of synonymous variants
29 remains unclear. Nevertheless, the very few significant associations further supported the
30 validity of our choice of this P -value cutoff through a complementary perspective as
31 synonymous variants generally would not contribute to the gene-trait associations^{62,141}.
32 We observed similar patterns of the tail distributions of P -values from permutation test

1 and synonymous burden test from the phases 1 to 3 sample (**Tables S24-S25**). Thus, the
2 *P*-value cutoff for burden tests of this sample (used in our internal replication) was set to
3 be 1×10^{-9} . We also provided a more liberal *P*-value cutoff at 1×10^{-8} as suggestive
4 evidence (approximately empirical false discovery proportion less than 0.1, **Tables S22-**
5 **S23**) for gene-based burden test.

6

7 **Cross-reference with previous GWAS**

8 For the comparison with GWAS signals, we leveraged the summary statistics derived from
9 previous studies^{1-4,10} on the same study cohort. The *P*-value cutoff for GWAS signals was
10 set to be the Bonferroni adjustment for all the traits ($P < 5 \times 10^{-8}$ / trait numbers for each
11 category) as described previously. Moreover, we focused on reported independent
12 variants¹⁴⁶ for GWAS signals and examined whether our exome-wide signals fell within
13 the 1Mb range of such independent GWAS signals to prioritize the effector genes among
14 thousands of identified GWAS loci. The genetic build for previous GWAS was GRCH37, so
15 we did liftover¹⁴⁷ for our exome-wide results back from GRCH38 to GRCH37 and made
16 comparisons based on the same genetic coordinates.

17

18 **Imaging genetics and drug detection**

19 We downloaded Therapeutic Target Database (TTD, last updated by January 10th,
20 2024)¹²⁰ and evaluated all the entries documented in TTD that were claimed to be
21 approved drug targets, in clinical trials, or supported by literature. We performed
22 enrichment tests to see whether the identified genes in our gene-based burden tests
23 were enriched in TTD drug target genes. We conducted Fisher's exact tests for two classes
24 of drug target genes in TTD. Specifically, we first included all drug targets that are either
25 approved drug targets or in clinical trials (documented as 'Successful' or 'Clinical trial' in
26 TTD), and then additionally included drug targets that were supported by literature
27 (documented as 'Literature-reported' in TTD).

28

29 **Ultra-rare burden heritability and genetic correlation with health-related traits**

30 We used the recently proposed BHR²⁷ method to quantify the burden heritability and the
31 burden genetic correlation for rare variants. BHR required one to divide the rare variants
32 into multiple subgroups based on MAF cutoff and annotation outcome. Following the

1 practical guide and previous examples, we investigated the genetic architecture of ultra-
2 rare variants ($MAF < 1 \times 10^{-4}$) in our sample across 596 imaging traits. We used “univariate”
3 mode and ran BHR to estimate the burden heritability for ultra-rare pLoF variants and
4 “int1 damaging” missense variants. Moreover, we downloaded 186 health-related traits
5 or diseases from Genebass¹⁷ to estimate the burden genetic correlation with imaging
6 traits in terms of the ultra-rare pLoF variants. Among the 596 imaging traits examined,
7 those that showed significant heritability after multiple testing adjustment ($P = 8.4 \times 10^{-5}$
8 $= 0.05/596$, that is Bonferroni adjusted for 596 imaging traits) were included in burden
9 correlation analysis. Then, we ran “bivariate” mode of BHR to calculate the burden
10 genetic correlation for the selected imaging traits and downloaded phenotypes.

11

12 **Code availability**

13 We made use of publicly available software and tools. The code used in this study will be
14 deposited in Zenodo upon publication.

15

16 **Data availability**

17 The individual-level data used in this study can be obtained from UK Biobank
18 (<https://www.ukbiobank.ac.uk/>). Other datasets in this paper include: Genebass
19 (<https://app.genebass.org/>), GWAS summary statistics for imaging traits in BIG-KP
20 (<https://bigkp.org/>), dbNSFP v.4.5a (<https://sites.google.com/site/jpopgen/dbNSFP>), the
21 Therapeutic Target Database (<https://idrblab.net/ttd/>), and the GTEx dataset v8
22 (<https://gtexportal.org/home/>). The full ExWAS summary statistics generated in this
23 study will be deposited in Zenodo upon publication.

24

25 **Figure legends**

26 **Fig. 1 Data overview and study design.**

27 **(A)** An overview of the data included in our study. For imaging phenotypes, we made use
28 of a broad range of multi-organ imaging phenotype data including brain imaging traits
29 such as regional brain volume from structural MRI, diffusion tensor imaging (DTI)
30 parameters from diffusion MRI, and functional connectivity and activity traits from resting
31 and task fMRI; CMR traits such as measurements of the aorta (ascending/descending
32 aorta) and heart chambers (left ventricle, right ventricle, left atrium, and right atrium);

1 abdominal MRI traits such as abdominal organ traits (liver, kidney, lung, spleen, and
2 pancreas) and abdominal composition measurements of fat and muscle. For genetic data,
3 we included whole exome sequencing data and focused on rare coding variants (minor
4 allele frequency [MAF] < 0.01). Only European individuals with imaging traits were
5 included in the present study ($n = 54,365$). **(B)** An overview of our study design. We
6 adopted an internal discovery-replication procedure and finalized the summary statistics
7 based on the joint sample (phases 1 to 5 sample). Specifically, we conducted association
8 tests on phases 1 to 3 sample, phases 4 to 5 sample, and phases 1 to 5 sample respectively.
9 We examined (i) whether the detected signals from phases 1 to 3 sample had concordant
10 directions and remained significant in phases 4 to 5 sample; (ii) whether the detected
11 signals from phases 1 to 3 sample had concordant directions and obtained stronger
12 evidence (smaller P -values) in the joint sample. Based on this validated procedure, the
13 significant results and other downstream analyses used the summary statistics generated
14 from this joint sample.

15

16 **Fig. 2 Results of variant-level association tests.**

17 **(A)** Manhattan plot for variant-level association tests across all 596 imaging phenotypes.
18 Only coding variants of interest were included in the variant-level association tests and
19 associations with $P < 1 \times 10^{-3}$ were plotted. The dashed lines indicate the variant-level
20 significant P -value threshold at 2.8×10^{-10} as well as the suggestive P -value threshold at
21 1×10^{-8} . The x-axis shows all the categories of our imaging phenotypes, while we further
22 separated the abdominal MRI traits into abdominal organs and body/fat composition, and
23 separated CMR traits into heart chambers and aorta areas based on the specific traits.
24 Significant coding-variant-level associations ($P < 2.8 \times 10^{-10}$) were labeled to the
25 corresponding genes. For DTI parameters, the unlabeled signals all belong to *VCAN* gene.
26 **(B)** Summary count for the number of signals, unique variants (unique genes in the
27 parentheses), and unique traits within each imaging category that has significant variant-
28 level associations. **(C)** The scatter plot for all the variant-level significant results (colored
29 by imaging categories) and suggestive results (in grey color). The x-axis shows the minor
30 allele frequency of each association and the y-axis is the absolute value of the
31 corresponding genetic effect size in s.d. units. **(D)** Graphical illustration of associations

1 between two missense variants (rs72553883 and rs34557412) in *TNFRSF13B* and spleen
2 volume. The rug plot at the first row demonstrates the locations of these two missense
3 variants. The rug plot at the second row indicates the locations of rare variants in the
4 ‘plof_alpha’ gene burden model, since gene-based burden test also revealed the
5 association between *TNFRSF13B* and spleen volume. The summary statistics (*P*-values) of
6 exome-wide associations between *TNFRSF13B* and blood traits based on burden tests
7 were downloaded from Genebase¹⁷. **(E)** Visualization of the associations between two
8 missense variants (rs2652098 and rs143368552) in *VCAN* and 12 mean diffusivity traits
9 (β_1 is the effect size in s.d. of rs2652098 and β_2 is the effect size in s.d. of rs143368552).

10

11 **Fig. 3 Results of gene-level burden tests.**

12 **(A)** Manhattan plot for gene-level burden tests across all 596 imaging phenotypes. For
13 significant gene-trait pairs appearing in multiple burden models, we plotted the
14 association with the smallest *P*-value. The dashed lines indicate the gene-level significant
15 *P*-value threshold at 1×10^{-9} as well as the suggestive *P*-value threshold at 1×10^{-8} . The *y*-
16 axis is capped at 1×10^{-30} and only gene-trait pairs with $P < 1 \times 10^{-3}$ were included. The
17 color legend and *x*-axis are same as the Manhattan plot for variant-level associations in
18 Figure 2A. Significant genes were labeled. **(B)** Effect sizes for all non-redundant significant
19 gene-trait pairs ($n = 67$, corresponding to **Table 1**) within each imaging category. The
20 dashed line indicates an effect size of zero. We labelled all the genes that had the largest
21 positive and negative effects within each imaging category and genes that had an
22 absolute effect size greater than 1.5. **(C)** Distribution of total brain volume (in cm^3) of
23 mutation carriers versus non-carriers of *PTEN*. To illustrate, we select the burden model
24 with the smallest *P*-value, that is “plof_int5”, which means the aggregation of pLoF
25 variants and missense variants with a minor allele frequency (MAF) cutoff at 1×10^{-4} . **(D)**
26 The effect size estimates and the corresponding 95% confidence intervals across different
27 burden models of *PTEN*. The color indicates distinct MAF cutoffs. The *y*-axis includes all
28 the burden models (Methods). Only burden models with minor allele count > 5 (after
29 aggregation) were tested (Methods) and plotted.

30

31 **Fig. 4 Heart and aorta-related associations and heart-brain connections.**

1 **(A)** Graphical illustration of 8 CMR traits of left ventricle (LV) associated with *TTN*. LVESV,
2 left ventricular end-systolic volume; LVEF, left ventricular ejection fraction; and Ecc, peak
3 circumferential strain, including both global (“Global”) and regional traits (such as
4 “AHA_9”). **(B)** Five genes (*ANO1*, *COL21A1*, *GEM*, *PLCE1*, and *TAGLN*) associated with
5 ascending aorta and descending aorta areas. *PLCE1* was also associated with brain fMRI
6 traits. Three ICA-based functional activity traits (Net25_Node3, Net25_Node5, and
7 Net25_Node9) associated with *PLCE1* are illustrated, with their major brain regions and
8 networks labeled. The color represents the weight profile of the ICA node. In addition,
9 *COL21A1* was associated with cerebrospinal fluid volume.

10

11 **Fig. 5 Burden heritability and genetic correlation.**

12 **(A)** The distribution of burden heritability point estimates for ultra-rare (minor allele
13 frequency $< 1 \times 10^{-4}$) pLoF variants and damaging missense variants (i.e. the int1 damaging
14 missense variants) across 591 multi-organ imaging traits that had a sample size larger
15 than 10,000 (5 abdominal imaging traits were excluded due to insufficient sample size).

16 **(B)** Burden genetic correlations of ultra-rare (minor allele frequency $< 1 \times 10^{-4}$) pLoF
17 variants for selected imaging phenotypes and other complex traits. Only top hits are
18 presented here. Specifically, we first selected seven trait pairs from pairs that had ten
19 smallest *P*-values among all the trait pairs and additionally included three pairs that had
20 smallest or second smallest *P*-values for regional brain volumes and abdominal MRI traits,
21 which resulted in nine imaging traits across all MRI categories and seven Genebass traits.
22 Filled circles indicate the selected top hits ($P < 0.0032$) while filled triangles indicate other
23 pairs that passed a nominal significance threshold at $P < 0.05$. “Phenotype_ID” was used
24 to define the selected imaging traits. Specifically, abdominal fat ratio from abdominal MRI
25 traits, left atrium maximum volume (LAV_max), regional radial strain (Err_AHA_13) and
26 peak circumferential strain (Err_AHA_4) traits from CMR, left and right insular volumes,
27 functional activity trait from resting fMRI (ICA-based, Net100_Node55) traits, Visual2-
28 Auditory network functional activity trait from task fMRI (parcellation-based), and
29 Somatomotor-Dorsal-Attention network functional activity trait from resting fMRI
30 (parcellation-based) traits were shown here (bottom-to-top order). **Table S1** includes
31 more details of these phenotypes.

1

2 **Table 1. Significant gene-trait pairs for exome-wide gene-based burden test.**

3 Here we present non-redundant results for gene-trait pairs that passed the stringent
4 threshold at $P < 1 \times 10^{-9}$. As a gene-trait pair may appear in multiple burden models and
5 MAF cutoffs, only the association with the smallest P -value is included. Column
6 “Phenotype_ID” uniquely defines a phenotype while column “Phenotype_info” provides
7 additional information of the specific phenotype if applicable. Specifically, for abdominal
8 MRI traits, the UKB data category is included in “Phenotype_info”; for CMR traits and DTI
9 parameters, the full name is included in “Phenotype_info”; for all the fMRI traits
10 (including ICA-based and parcellation-based ones), the network name is included in
11 “Phenotype_info”. In the column “Organ/Category”, we further separate non-brain traits
12 (i.e. abdominal MRI and CMR traits) to the corresponding organs. More details about the
13 phenotype information can be found in **Table S1**, and a more comprehensive version of
14 this table can be found in **Table S6**.

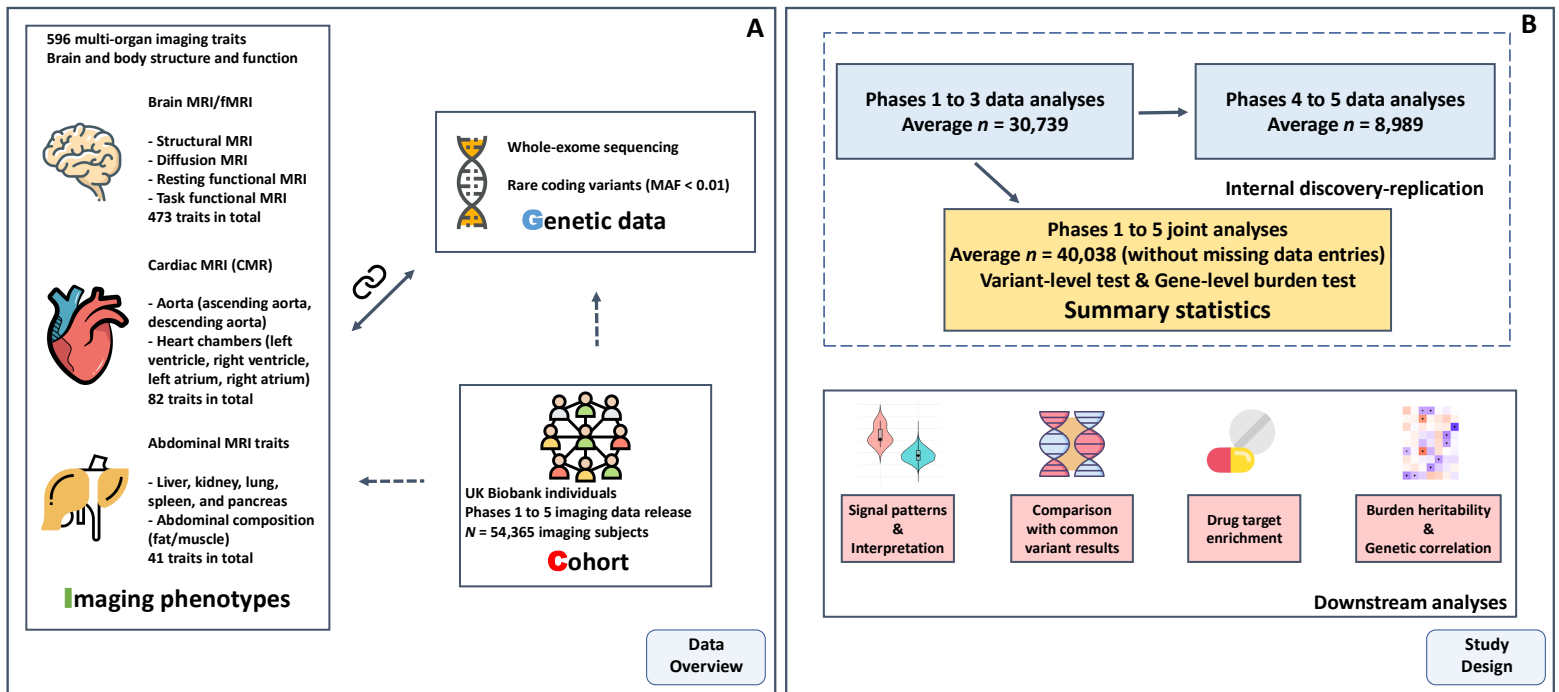


Figure 1

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

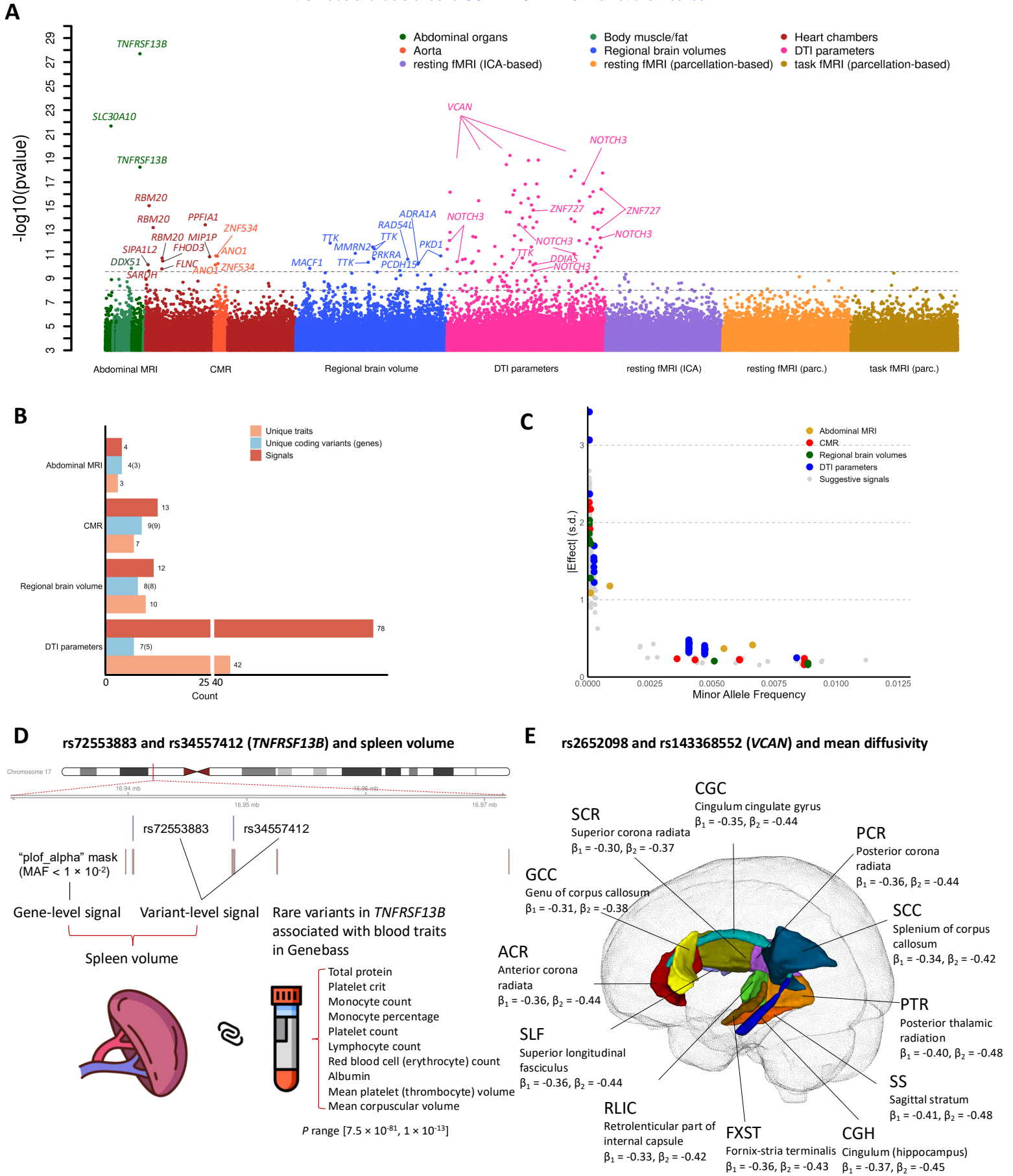


Figure 2

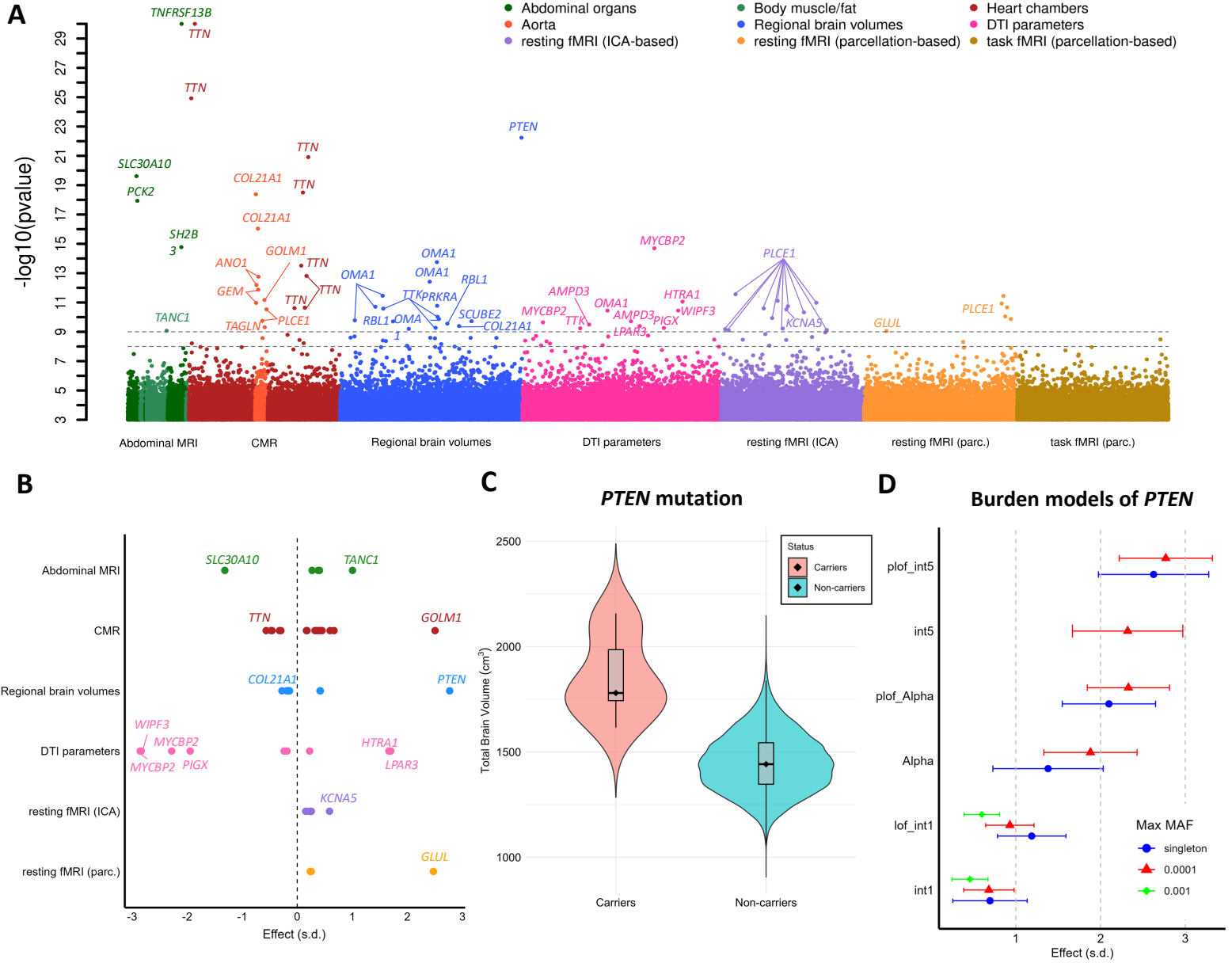


Figure 3

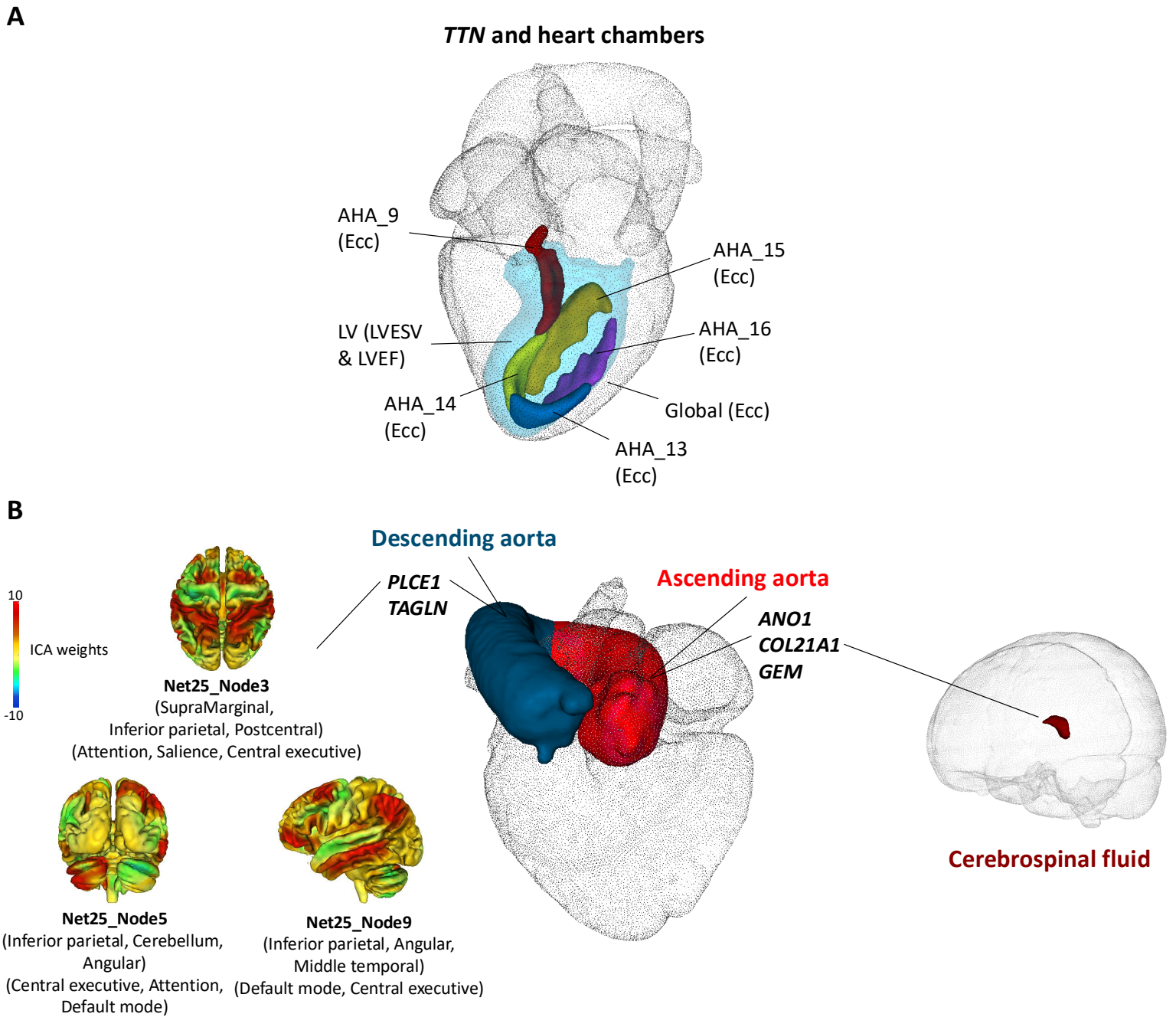
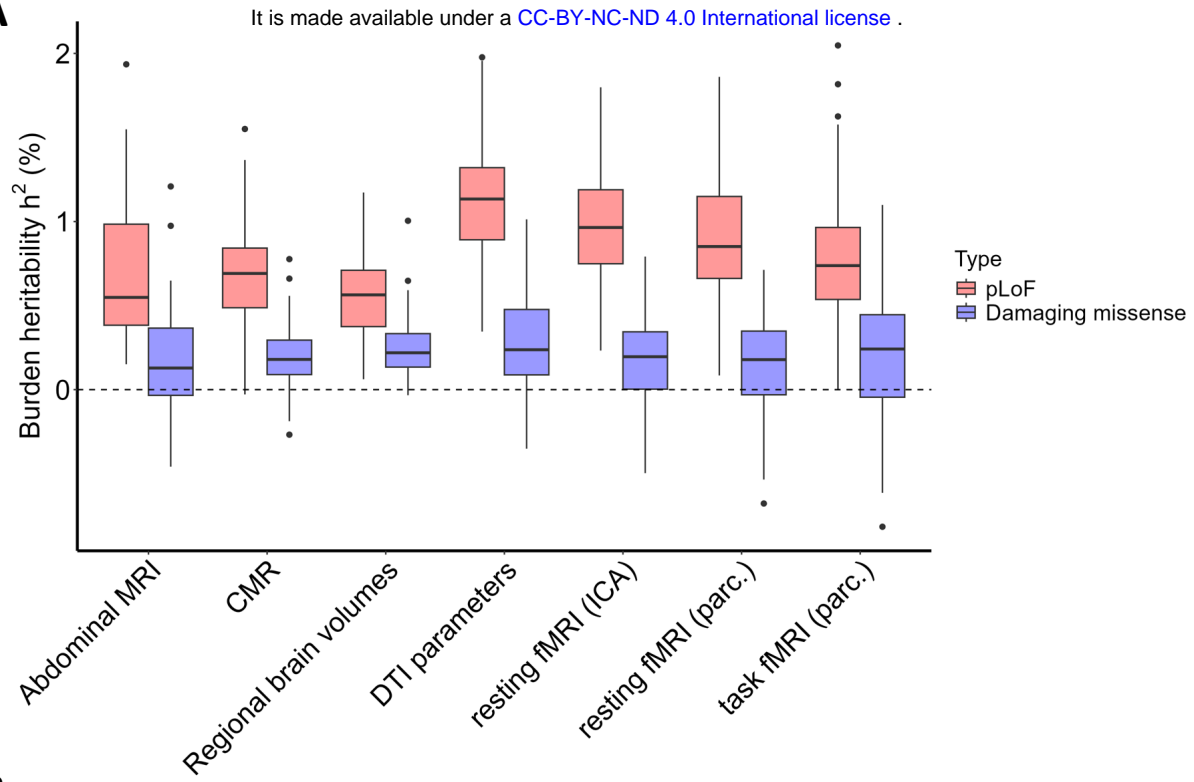


Figure 4

A



B

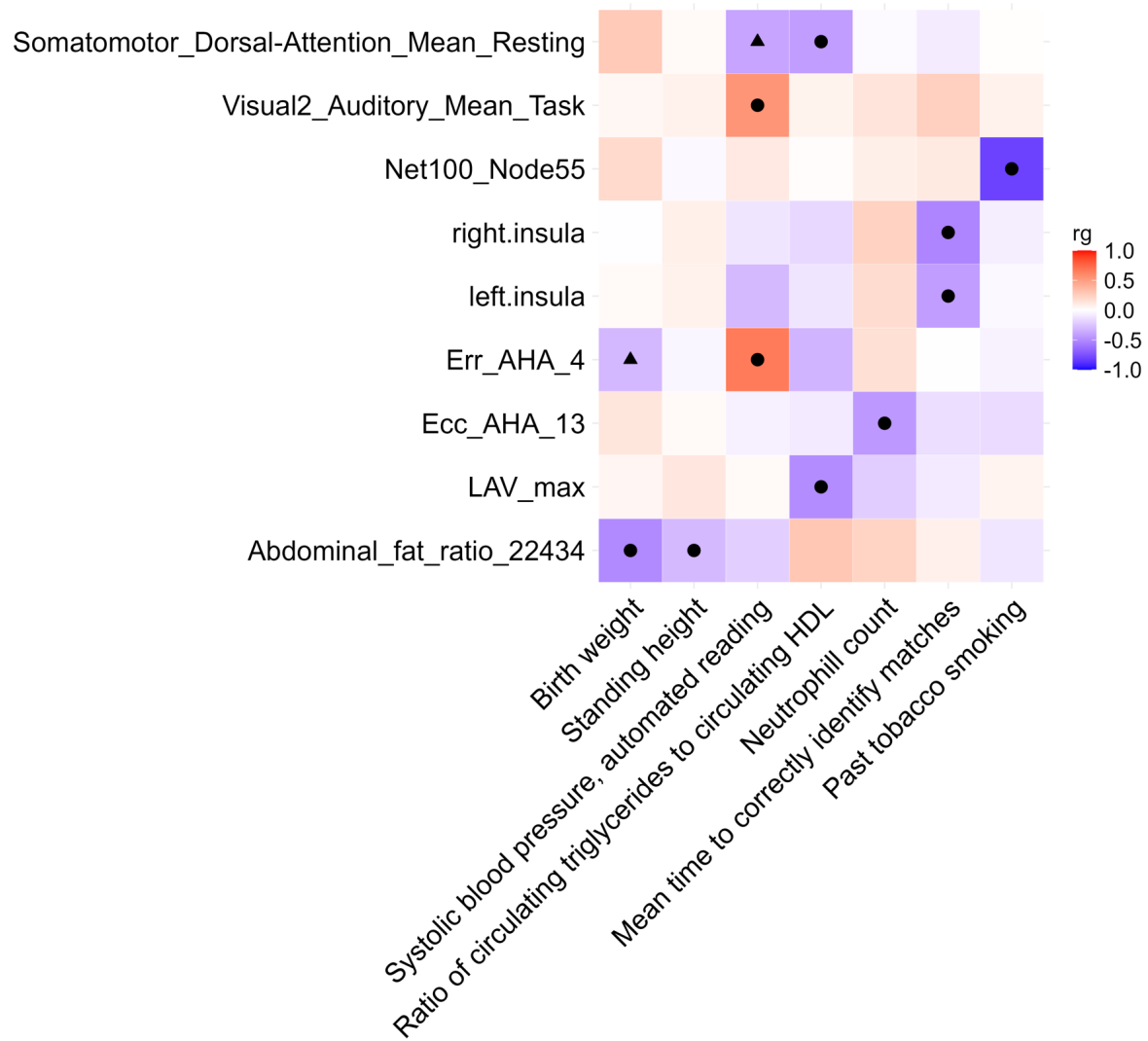


Figure 5