nature aging

Article

https://doi.org/10.1038/s43587-025-00925-y

Large-scale genome-wide analyses with proteomics integration reveal novel loci and biological insights into frailty

Received: 5 July 2024

Accepted: 23 June 2025

Published online: 5 August 2025

Check for updates

Jonathan K. L. Mak[1,2,3], Chenxi Qin[1], Moritz Krüger[3], Anna Kuukka[3], FinnGen[5,*], Sara Hägg[1], Jake Lin[1,3,73] & Juulia Jylhävä[1,3,4,73]

Frailty is a clinically relevant phenotype with notable gaps in our understanding of its etiology. Using the Hospital Frailty Risk Score (HFRS) to define frailty, we performed a genome-wide association study in FinnGen (N = 500,737), replicated the results in the UK Biobank (N = 407,463) and performed a meta-analysis. We prioritized genes through colocalization with expression, splicing and protein quantitative trait loci and proteomics integration. We identified 53 independent lead variants associated with frailty ($P < 5 \times 10^{-8}$), of which 45 were novel and not previously reported in the GWAS Catalog. Replication at the individual variant and polygenic risk score of the HFRS ($P = 1.86 \times 10^{-522}$) levels and meta-analysis largely confirmed the findings. Colocalization analysis supported a causal role for several genes, including *CHST9*, *C6orf106* (*ILRUN*), *KHK*, *MET*, *APOE*, *CGREF1* and *PPP6C*. Additionally, plasma levels of MET, CGREF1 and APOE were associated with HFRS. Our results reveal new genetic contributions to frailty and shed light on its biological basis.

Aging is a highly complex process with substantial heterogeneity in health trajectories among individuals. Frailty represents a clinically relevant aging phenotype that gauges health in aging[1] and predicts various adverse outcomes independent of chronological age[2]. Frailty describes a syndrome of decreased physiological reserves across multiple homeostatic systems[1]. Currently, no gold standard exists to measure frailty; instead, several scales with different properties have been developed, each capturing partially different at-risk populations[3]. Created based on 109 weighted International Classification of Diseases, 10th Revision (ICD-10) codes characterizing older adults with high resource use and diagnoses associated with frailty, the HFRS presents a relatively new scale to measure frailty[4]. It has a fair overlap with existing frailty definitions based on the deficit accumulation (frailty index (FI)) and phenotypic (frailty phenotype (FP)) models of frailty and has a moderate agreement with the FI[4]. While the HFRS uses ICD-10 codes for administrative ease, enabling the measurement of frailty in real-world data, the FI[5] and FP[6] are rooted in clinical and functional data and are often assessed in cohort studies. The FI is a multidimensional measure of frailty, offering a comprehensive view of a person's overall health[5]. In contrast, the FP defines frailty through specific physical characteristics: weakness, slowness, exhaustion, low physical activity and weight loss[6]. While each measure captures distinct aspects of frailty, together they provide a more complete understanding of the condition.

The etiology of frailty remains incompletely understood. Twin studies by us and others suggest that frailty, measured using the FI, is up to 52% heritable[7,8], with relatively stable genetic influences across age[9]. To date, only two previous large-scale genome-wide association studies (GWASs) of frailty exist. Atkins et al. performed a meta-analysis GWAS of FI that

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [2]Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [3]Faculty of Medicine and Health Technology and Gerontology Research Center (GEREC), Tampere University, Tampere, Finland. [4]Tampere Institute for Advanced Study, Tampere, Finland. [5]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. [73]These authors contributed equally: Jake Lin, Juulia Jylhävä. *A list of authors and their affiliations appears at the end of the paper. e-mail: juulia.jylhava@ki.se

https://doi.org/10.1038/s43587-025-00925-y

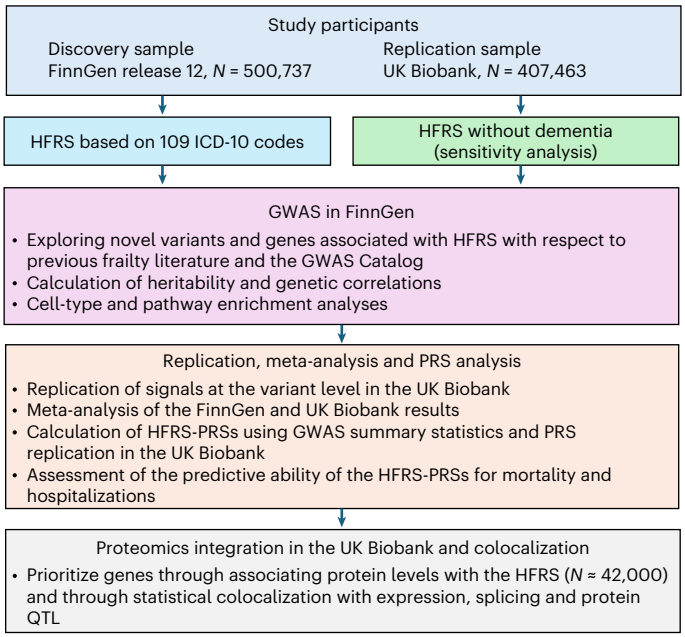


Fig. 1 | Outline of the study. Discovery GWASs of HFRS and HFRS without dementia were performed in FinnGen to identify genetic variants associated with frailty. The significant variants ($P < 5 \times 10^{-8}$) were then replicated in the UK Biobank, and a meta-analysis of the FinnGen and UK Biobank results was performed. The GWAS summary statistics of FinnGen were used to calculate HFRS-PRSs, which were then assessed for their association with mortality and hospitalizations in the UK Biobank. Finally, protein association and colocalization analyses were performed to prioritize genes and identify causal variants.

identified 34 loci and estimated the single nucleotide polymorphism (SNP) heritability of the FI at 11%[10]. Ye at al. identified 123 loci for FP and estimated the SNP heritability of the FP at 6%[11]. However, it is likely that additional genetic signals exist and analyses in other large populations can shed further light on the genetic underpinnings of frailty.

To date, no previous studies into the genetics of frailty using the HFRS exist. To this end, we performed a GWAS of the HFRS in FinnGen (N = 500,737), with replication of the results in the UK Biobank (N = 407,463), both at the individual variant level and through polygenic risk scores (PRSs). We also performed a meta-analysis on the results from both GWASs to capture the totality of the evidence. Given that dementia has the highest weight in the HFRS definition, we performed a sensitivity analysis by excluding dementia from the HFRS definition and similarly replicated the results in the UK Biobank and conducted a meta-analysis on the results. A functional follow-up to identify causal genetic loci was performed through colocalization analysis[12] with expression, splicing and protein quantitative trait loci (eQTL, sQTL and pQTL, respectively) and associating measured protein levels with the HFRS in the UK Biobank (N = up to 42,495).

Results

Sample characteristics

The workflow of the analyses is presented in Fig. 1. In the HFRS GWAS, we included 500,737 (282,202 females, 56.4%) FinnGen and 407,463 UK Biobank participants (220,208 females, 54.1%). Characteristics of the study populations are presented in Table 1.

Discovery GWAS of HFRS in FinnGen

We identified 1,588 variants associated ($P < 5 \times 10^{-8}$) with the HFRS in the main analysis and 492 variants in the sensitivity analysis, which removed the dementia weights from the HFRS (Fig. 2a,b and Supplementary Tables 1 and 2). Of these, 53 variants (at 50 loci) and 42 variants (at 42 loci) were identified as independent lead variants ($r^2 < 0.01$) for the HFRS and HFRS without dementia, respectively. As dementia diagnosis has the highest weight in the HFRS formula, the most influential peak expectedly resided in the *APOE* (rs7412) region on chromosome 19 (Fig. 2a). Sensitivity analysis confirmed the expected loss of the *APOE* peak (Fig. 2b). Of the independent lead variants associated with HFRS and HFRS without dementia, 45/53 and 36/42, respectively, were novel with respect to the GWAS Catalog and previously reported GWAS results of the FI[10], FP[11] and mvAge[13] (Fig. 3a and Supplementary Tables 1 and 2). The variants mapped to 41 (HFRS) and 30 (HFRS without dementia) genes of which 6 and 3, respectively, were novel, that is, previously unreported for any trait at $P < 5 \times 10^{-8}$. The results also demonstrated unique, non-shared associations in both analyses (Fig. 3b and Supplementary Tables 1 and 2). Supplementary Table 3 presents the shared and unique genes between the HFRS, FI and FP GWASs.

Table 1 | Characteristics of the study samples

Characteristic	FinnGen	UK Biobank
No. of individuals	519,200	407,463
Age at baseline assessment, mean (s.d.)	53.1 (17.9)	56.9 (8.0)
Age at end of follow-up/death, mean (s.d.)	60.8 (18.0)	70.9 (7.9)
Sex, *n* (%)		
Women	292,784 (56.4)	220,208 (54.1)
Men	226,416 (43.6)	187,255 (45.9)
BMI (kg/m^2), mean (s.d.)	27.35 (5.53)	27.41 (4.76)
Missing, *n* (%)	142,454 (27.4)	1273 (0.3)
Smoking, *n* (%)		
Nonsmoker	156,355 (50.9)	221,770 (54.6)
Former smoker	70,317 (22.9)	143,384 (35.3)
Current smoker	80,736 (26.2)	41,109 (10.1)
Missing	211,792	1,380
HFRS, median (IQR)	5.2 (1.6–10.4)	1.5 (0–5)
Women, median (IQR)	5.3 (1.6–10.5)	1.5 (0–4.7)
Men, median (IQR)	5.0 (1.5–10.3)	1.5 (0–5.4)
HFRS categories, *n* (%)		
Low risk (<5)	241,656 (48.4)	304,555 (74.7)
Intermediate risk (5–15)	188,147 (37.8)	74,386 (18.2)
High risk (>15)	65,925 (13.2)	28,702 (7.0)
HFRS>5, *n* (%)	254,874 (51.0)	101,326 (24.9)
HFRS>5 before age 65, *n* (%)	95,410 (18.4)	33,485 (8.2)
Died during follow-up, *n* (%)	62,764 (12.1)	36,795 (9.0)
Number of hospitalizations, median (IQR)	8 (4–17)	1 (0–3)

FinnGen participant characteristics are presented for the sample with non-missing data on the HFRS (N=519,200). IQR, interquartile range.

Replication in the UK Biobank and meta-analysis

For HFRS, 1,262/1,588 variants were available for replication and meta-analysis. In the UK Biobank, 73 variants (6%) replicated at $P < 5 \times 10^{-8}$ and 688 (55%) at $P < 0.05$, while in the meta-analysis, 357 variants (28%) replicated at $P < 5 \times 10^{-8}$ and 1,260 (100%) at $P < 0.05$ (Supplementary Table 1). Of the 53 lead variants, 36 were available; 2 lead variants (6%) replicated at $P < 5 \times 10^{-8}$ and 14 (39%) at $P < 0.05$ in the UK Biobank, while 6 (17%) replicated at $P < 5 \times 10^{-8}$ and 35 (97%) at $P < 0.05$ in the meta-analysis (Supplementary Table 1). For HFRS without dementia, 435/492 variants were available for replication and meta-analysis. In the UK Biobank, 21

https://doi.org/10.1038/s43587-025-00925-y

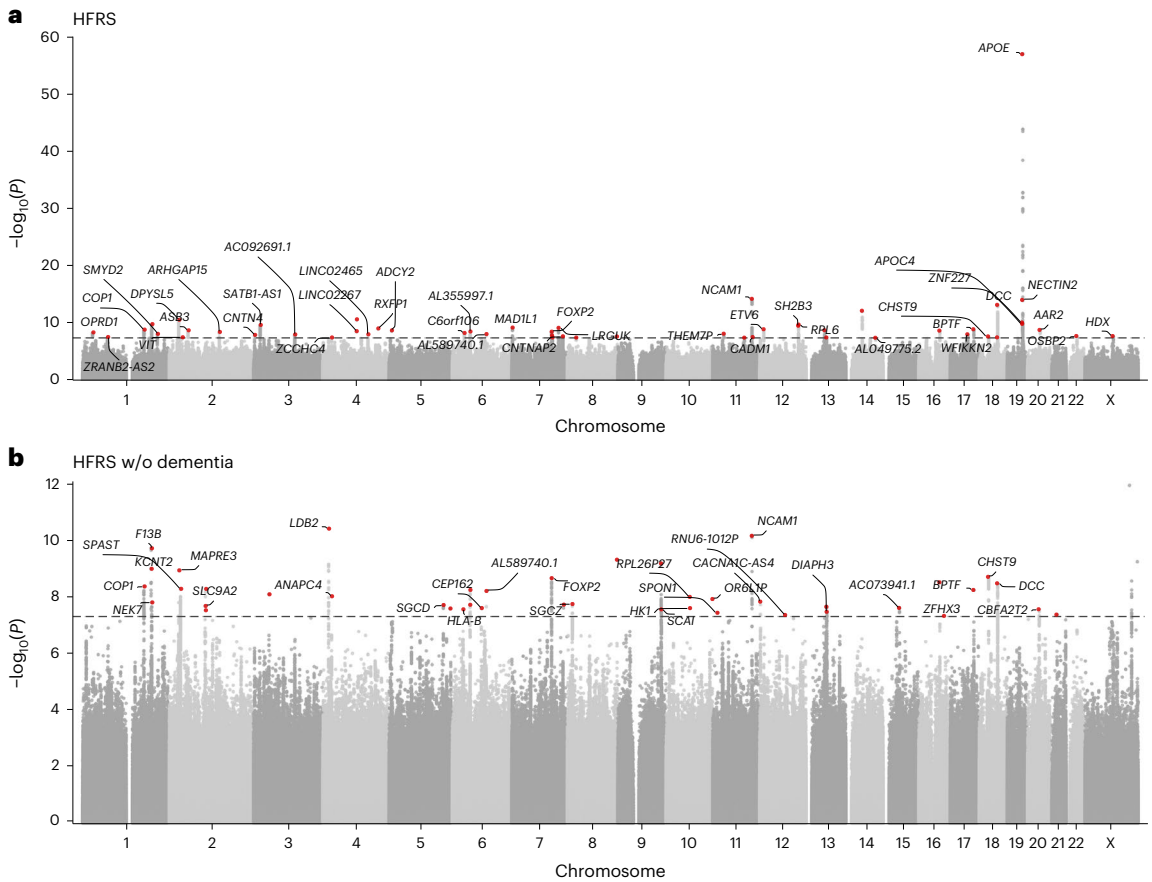


Fig. 2 | GWAS results in FinnGen. a,**b**, Manhattan plots for the associations with HFRS (**a**) and HFRS excluding dementia (**b**) in FinnGen using linear mixed-effects modeling adjusted for birth year, sex and the first ten PCs. The dashed lines indicate the genome-wide significance threshold ($P = 5 \times 10^{-8}$). The annotations represent the independent lead variants associated with frailty.

variants (5%) replicated at $P < 5 \times 10^{-8}$ and 118 (27%) at $P < 0.05$, while in the meta-analysis, 50 variants (11%) replicated at $P < 5 \times 10^{-8}$ and 435 (100%) at $P < 0.05$ (Supplementary Table 1). Of the 42 lead variants, 26 were available; 1 lead variant (4%) replicated at $P < 5 \times 10^{-8}$ and 10 (38%) at $P < 0.05$ in the UK Biobank, while 4 (17%) replicated at $P < 5 \times 10^{-8}$ and 26 (100%) at $P < 0.05$ in the meta-analysis (Supplementary Table 2). The effect direction was consistent for all variants that replicated at $P < 5 \times 10^{-8}$ in the meta-analysis (Supplementary Tables 1 and 2).

Genetic correlation and heritability

We observed a lambda genomic control value of 1.27 with an intercept of 1.19 (s.e. = 0.011) for HFRS and 1.11 with an intercept of 1.23 (s.e. = 0.010) for HFRS without dementia (QQ plots provided in Extended Data Fig. 1). Despite the relatively high lambda values, the intercepts suggest that the inflation in test statistics was mainly due to polygenicity, rather than bias due to population stratification. The SNP heritability was 0.06 (s.e. = 0.002) for HFRS and 0.04 (s.e. = 0.002) for HFRS without dementia. Statistically significant and positive genetic correlations ($P < 2.2 \times 10^{-16}$) were observed between HFRS and previous GWASs on frailty and mvAge (Fig. 3c).

Cell-type and pathway enrichment

For HFRS, the top ($P < 3.7 \times 10^{-5}$, corrected for multiple testing) cell types enriched for expression were limbic system neurons in cerebrum, excitatory neurons (Ex6) in visual cortex, oligodendrocyte precursor cells (OPCs) in cerebellar hemisphere and oligodendrocytes in cerebellum (Extended Data Fig. 2 and Supplementary Table 4). For HFRS without dementia, the top cell types were OPCs and astrocytes in cerebellar hemisphere, skeletal muscle satellite cells in muscle and endocrine cells in stromal cells in stomach (Extended Data Fig. 3 and Supplementary Table 5). Enrichr[14] pathway analysis (adjusted $P < 0.05$) showed that the top pathways for the HFRS signals were relevant to the nervous system functions (herpes simplex virus 1 infection, netrin-mediated repulsion signals), cell adhesion and lipid metabolism (Supplementary Table 6). Comparison of the pathways from the HFRS, FI and FP GWASs revealed overlap in herpes simplex virus 1 infection and cell adhesion molecules between HFRS and FI, and in multiple pathways related to lipid and lipoprotein metabolism, cellular interactions and adhesion between HFRS and FP (Supplementary Table 6). Each GWAS also had distinct pathways not shared with the others (Supplementary Table 6). For HFRS without dementia, several functions related to cell cycle were enriched at $P < 0.05$, although none of the pathways were statistically significant after correction for multiple testing (Supplementary Table 7).

Exploring causal variants through proteomics integration

To identify potentially causal and functional variants (that is, missense, splice region, loss of function and 5′ and 3′ untranslated region variants associated with the HFRS and HFRS without dementia at $P < 5 \times 10^{-7}$; Supplementary Tables 8 and 9), we associated the protein levels of the corresponding genes to HFRS (13 proteins available in UK Biobank Olink platform) and HFRS without dementia (8 proteins available in UK Biobank Olink platform). We adjusted the models for birth year, sex and the first ten principal components (PCs; model 1), as well as batch, baseline assessment center, body mass index (BMI) and smoking (model 2). Significantly associated proteins at a false discovery rate (FDR) < 0.05 in both models 1 and 2 were CGREF1, MET, ALDH2, NECTIN2, APOC1, APOE and FOSB for HFRS, and CDK and POF1B for HFRS without dementia (Fig. 4 and Supplementary Table 10).

https://doi.org/10.1038/s43587-025-00925-y

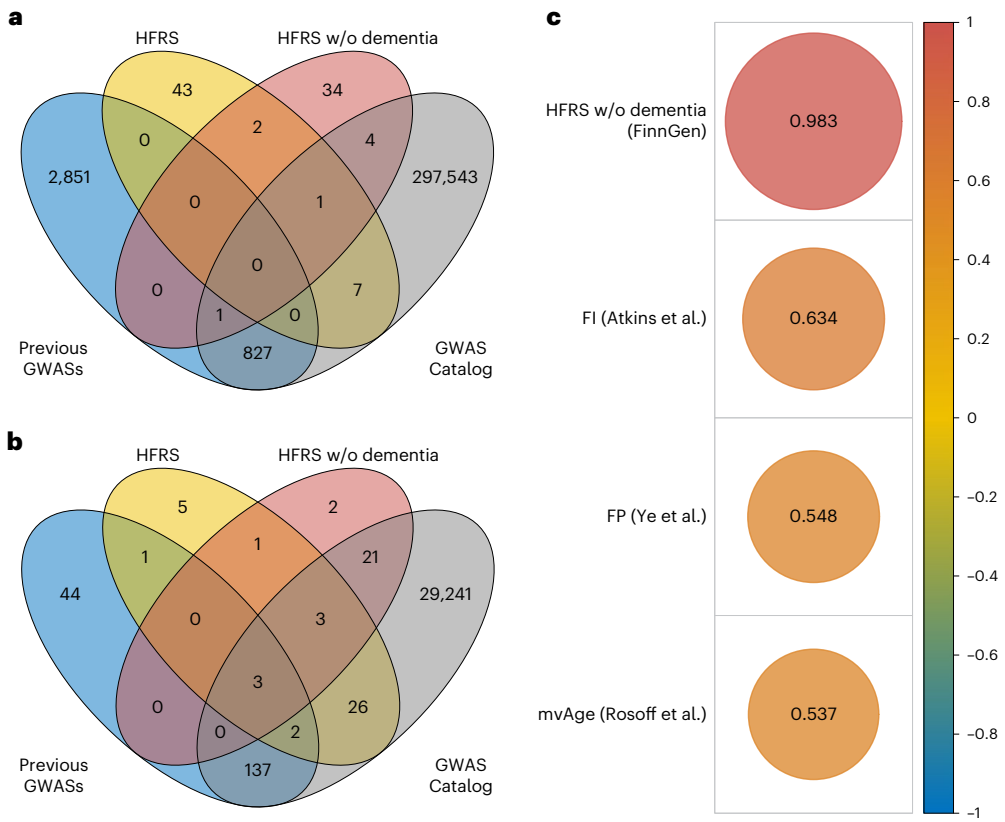


Fig. 3 | Lead variants and genes and genetic correlations of the HFRS. a, Venn diagram showing the overlap of the lead variants associated with the full HFRS and the HFRS without dementia in FinnGen and those reported in the literature. Previous GWASs refers to genes identified in for the FI[10], FP[11] and mvAge[13]. **b**, Venn diagram showing the overlap of the lead variant genes associated with the full HFRS and the HFRS without dementia in FinnGen and those reported in the literature. **c**, Genetic correlations between HFRS in FinnGen and other frailty-related traits estimated using linkage disequilibrium score regression. All the correlations were statistically significant at $P < 2.2 \times 10^{-16}$.

Colocalization analysis

Several gene loci, such as *CHST9*, *C6orf106* (*ILRUN*), *KHK*, *MET*, *CGREF1* and *PPP6C* had shared causal variants (posterior probability for H4 (PP.H4) > 80%) in eQTL and/or sQTL for HFRS. Several colocalized (PP.H4 > 80%) eQTL and/or sQTL loci were also identified for HFRS without dementia, including *CHST9*, *CGREF1*, *PPP6C*, *ADARB1* and *PSMB7*. The full eQTL and sQTL colocalization results for the HFRS and HFRS without dementia are presented in Supplementary Tables 11 and 12, and the colocalization results with a PP.H4 > 80% are summarized by tissue for each gene in Extended Data Fig. 4. In the pQTL analysis, of those genes that had a protein measurement available (that is, the protein was detectable in plasma), a total of 20 loci for HFRS and 9 loci for HFRS without dementia had enough power for the analyses (PP > 88%; Methods). Of them, a colocalized signal (that is, shared single causal variant, PP.H4 > 80%; Methods) was detected within *APOE* and *BRAP* genes for HFRS (Supplementary Table 13), whereas no colocalized signal with pQTL was detected within genes for HFRS without dementia. For most of the tested genes, the PP.H3 values were greater than or close to 90%, indicative of distinct causal variants for protein levels and HFRS (Supplementary Tables 13 and 14). Regional association plots of the *APOE* gene demonstrated that the strongest signal peak rs429358 and variants in high linkage disequilibrium with it fall in the vicinity (Extended Data Fig. 5).

HFRS-PRS analyses in the UK Biobank

The PRSs for HFRS (HFRS-PRSs) were statistically significantly associated with the HFRS in the UK Biobank ($\beta = 0.074$ per s.d. increase; $P = 1.86 \times 10^{-522}$) after adjusting for birth year, sex and the first ten PCs (Fig. 5a). Next, using similar adjustments, we analyzed whether the HFRS-PRSs could predict early-onset frailty in the UK Biobank (that is, HFRS > 5 before age 65) and observed an odds ratio of 1.25 ($P = 2.0 \times 10^{-322}$; Fig. 5b). We further examined whether the HFRS-PRSs could predict all-cause mortality and number of hospitalizations and found statistically significant associations with both outcomes (Fig. 5c,d). The estimates of the HFRS-PRSs were similar in men and women compared to the full sample, and also consistent for the HFRS-PRsS excluding dementia (Fig. 5a–d). Numeric estimates for all the HFRS-PRS analyses are presented in Supplementary Table 15. Lastly, we found that adding the HFRS-PRSs to a model with age, sex and the first ten PCs significantly improved model performance on mortality and hospitalizations, as assessed by likelihood-ratio and *F*-test statistics (Supplementary Table 16).

Prediction of mortality using HFRS

To assess the validity of HFRS in predicting mortality, we examined its association with all-cause mortality and found that higher HFRSs, both with and without dementia, were associated with mortality in FinnGen (hazard ratio 1.29 for both HFRS and without dementia) and UK Biobank (hazard ratio 1.48 for both HFRS without dementia), independent of age, birth year and sex (Supplementary Table 17).

Discussion

Our study represents a large GWAS of frailty using the HFRS. We identified 1,588 associated variants and 53 lead variants, of which 45 were novel, and not previously reported for any trait. The lead variants mapped to 41 genes, of which 6 were novel. Replication in the UK Biobank and subsequent meta-analysis showed that 28% of all variants and 17% of lead variants replicated at $P < 5 \times 10^{-8}$, while 100%

https://doi.org/10.1038/s43587-025-00925-y

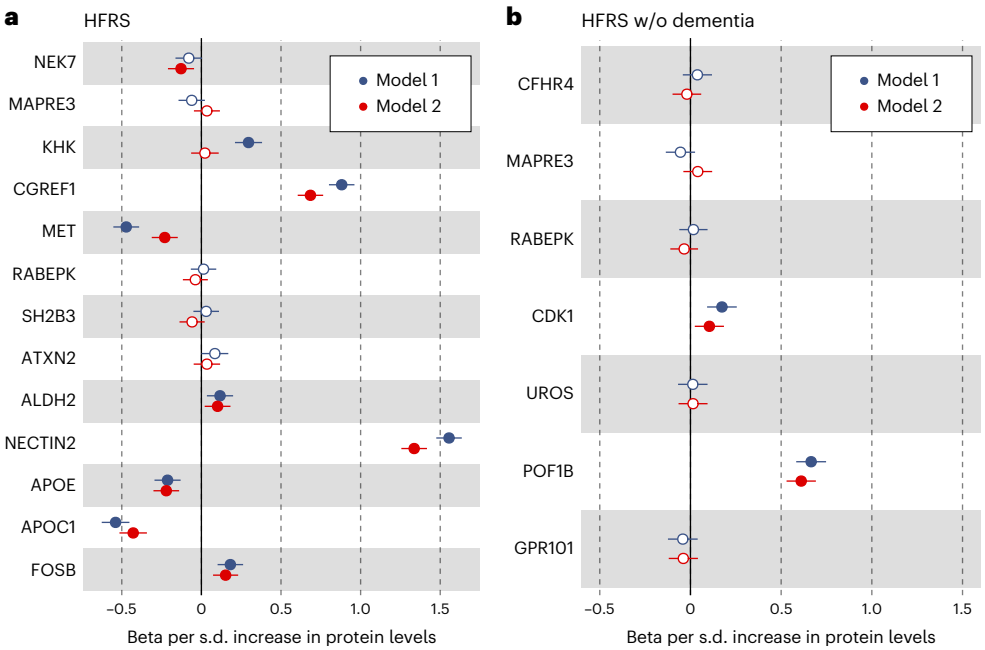


Fig. 4 | Proteomics integration in the UK Biobank. a,**b**, Protein associations (beta coefficients) with the full HFRS (**a**) and HFRS without dementia (**b**) the in the UK Biobank using linear regression models (N = 34,879–42,495; exact N for each model is given in Supplementary Table 10). All models were adjusted for birth year, sex and the first ten PCs (model 1), and additionally adjusted for batch, baseline assessment center, BMI and smoking (model 2). Solid dots indicate statistically significant associations at an FDR < 0.05. The bars indicate 95% confidence intervals.

of all variants and 97% of lead variants replicated at $P < 0.05$ in the meta-analysis. Colocalization analysis identified several causal candidate genes, including *CHST9*, *C6orf106* (*ILRUN*), *KHK*, *MET*, *APOE*, *CGREF1* and *PPP6C*. Additionally, plasma levels of MET, CGREF1 and APOE were associated with HFRS, further supporting their roles in frailty. We also derived PRSs for HFRS and showed that they predict frailty, early-onset frailty, mortality and hospitalizations in the UK Biobank.

The strongest GWAS signals were observed in the *TOMM40*/*APOE*/*APOC1*/*NECTIN2* locus on 19q13.3, a locus in strong linkage disequilibrium and known for its associations with cognitive[15] and cardiometabolic[16] traits. We observed the strongest signal for the missense variant rs429358 (388 T > C) that, together with rs7412, defines the *APOE* ε2, ε3 and ε4 haplotypes. The rs7412 was, however, not associated with frailty in our study. A similar finding has been observed for longitudinal weight loss—a feature that also characterizes frailty—where rs429358 increased the risk, while rs7412 did not[17]. Our sensitivity analysis, which removed dementia from the HFRS, truncated the chromosome 19 peak as expected and revealed additional loci. The HFRS lead variants mapped to 42 genes, 7 of which were shared with HFRS without dementia, while 31 genes were uniquely mapped in HFRS without dementia. The unique lead variant associations for both HFRS and HFRS without dementia are plausible, as dementia had the highest individual weight in the HFRS definition, and for highly polygenic traits like frailty, even small differences in phenotype definitions can influence which variants reach genome-wide significance. Genetic correlation between HFRS and HFRS without dementia was nevertheless almost perfect (0.98), indicating the same underlying genetic construct.

The genes to which the 45 novel lead variants for the HFRS mapped include *C6orf106* (*ILRUN*) and *CHST9*, both of which also displayed colocalized signals with eQTL across different tissues, supporting their potential causal roles. *C6orf106* (*ILRUN*) is a regulator of inflammation and lipid metabolism[18], while *CHST9* encodes an enzyme essential for cell–cell interactions and signal transduction[19]. Notably, several *CHST9* variants were also associated with HFRS without dementia and similarly exhibited colocalization with eQTL. *CGREF1*, a gene linked to cell cycle regulation and adhesion[20], and *PPP6C*, a gene involved in nuclear factor-κB pathway regulation[21], showed the same sQTL-colocalized gene–tissue pairs for HFRS and HFRS without dementia, supporting their functional roles in frailty, irrespective of the HFRS definition. While *C6orf106 (ILRUN)*, *CHST9*, *CGREF1* and *PPP6C* are functionally diverse, they collectively link immunoinflammatory modulation, cellular interactions and adhesion to frailty. Specific to HFRS, we additionally identified multiple colocalized signals in *KHK* and *MET*, while for HFRS without dementia, we identified additional colocalized signals in *ADARB1* and *PSMB7*. Aside from a few links to blood pressure, plasma lipids or BMI in the GWAS Catalog, *CHST9*, *CGREF1*, *PPP6C*, *KHK*, *MET*, *ADARB1* and *PSMB7* have no prior GWAS associations with the HFRS conditions, suggesting that HFRS, as a composite measure, can offer insights into frailty beyond its individual components.

Proteomics integration showed that CGREF1, NECTIN2, MET and APOC1 were associated with the HFRS with the largest effect sizes; elevated levels of the first two and lower levels of the latter two were associated with higher HFRS scores. Previous studies have linked elevated circulating NECTIN2 levels to Alzheimer's disease risk[22] and low APOC1 levels to cognitive decline and frailty, as defined using the FP[23], which likely explains their associations with the HFRS. In contrast, no prior studies have linked plasma CGREF1 or MET to frailty or HFRS conditions, highlighting a novel association. Additionally, as *CGREF1* and *MET* exhibited eQTL-colocalized and/or sQTL-colocalized signals across multiple tissues, their protein-level associations further support their biological relevance in frailty.

We estimated the SNP heritability of HFRS at 6%, an estimate in the same range as previously reported for the FI (11%)[10] and FP (6%)[11]. Genetic correlations between HFRS, FI and FP were moderate, ranging from 0.54 to 0.63, while gene-level overlap was limited: two shared genes between HFRS and FI and eight between HFRS and FP. The limited gene-level overlap is likely a result of frailty being a highly polygenic trait, where genome-wide significant variants represent only a fraction of the total genetic signal. Genetic correlation, in turn, reflects the combined influence of numerous variants, including those that do not reach the genome-wide significance threshold, but still make

https://doi.org/10.1038/s43587-025-00925-y

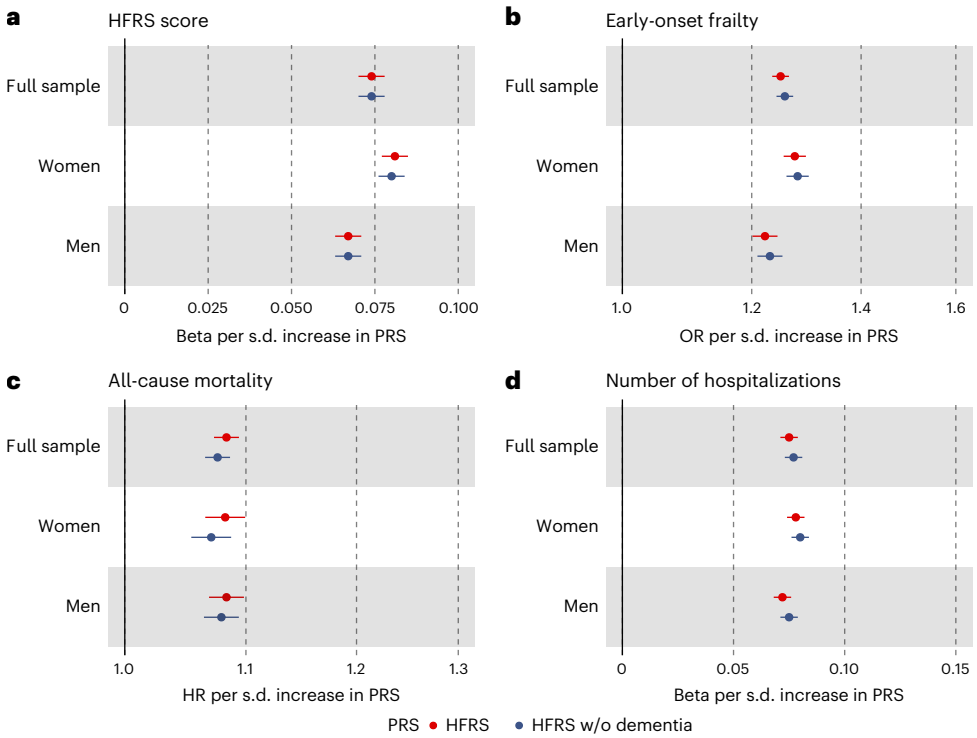


Fig. 5 | HFRS-PRSs, frailty, mortality, and hospitalizations. a–d, Associations of the HFRS-PRSs with the HFRS (**a**), early-onset frailty (**b**), all-cause mortality (**c**) and number of hospitalizations (**d**) in the UK Biobank (N = 407,463). All models included birth year, birth region, sex and the first ten PCs as covariates. The bars indicate 95% confidence intervals of the beta coefficients, odds ratios (ORs) and hazard ratios (HRs).

a substantial overall contribution to the trait. Moreover, it has been shown that different frailty scales identify only partially overlapping groups of individuals as frail[4,24], suggesting that these scales may capture somewhat distinct constructs. In our previous study[25], we assessed the phenotypic correlation between HFRS and FI at 0.21 and HFRS and FP at 0.31 in the UK Biobank participants, indicating somewhat lower phenotypic correlations compared to their genotypic counterparts. A possible explanation is that, because most UK Biobank participants are still relatively young, frailty may not yet be fully expressed, leading to many values being 0 and thereby diluting the phenotypic correlations. Additionally, environmental factors, such as physical activity, may directly influence phenotypic frailty, but might not affect the multidimensional FI or HFRS to the same extent, leading to reduced phenotypic correlations. The overall low prevalence of frailty in the UK Biobank participants may also have contributed to the low gene-level overlap between FI, FP and HFRS because both the FI[10] and FP[11] GWASs included UK Biobank participants. For the same reason, the overall lower HFRS scores in the UK Biobank and differences in the proportions of individuals with certain HFRS conditions between FinnGen and the UK Biobank may have also affected the replication results, potentially leading to underestimated effect sizes in the UK Biobank and the overlooking of some true associations.

Cell-type enrichment indicated enriched expression of the genes associated with the signals in various neuronal cells, such as limbic system neurons, excitatory neurons, OPCs and oligodendrocytes located in the cerebrum, visual cortex, cerebellar hemisphere and cerebellum, respectively. Enrichment of OPCs (cerebellar hemisphere) persisted even after removing the contribution of dementia diagnoses from the HFRS. Expression enrichment in brain tissues was likewise observed the GWAS of FI[10], which identified frontal cortex BA9, cerebellar hemisphere, spinal cord cervical C1 and hippocampus as statistically significant. The GWAS on FP[11] also identified the genetic signals enriched in brain tissues, such as cerebellar hemisphere, frontal cortex BA9 and cerebellum. It is noteworthy that neither FI nor FP in these GWASs included any items of cognition or dementia diagnosis in the frailty definition. Our findings thus reinforce the role of central nervous system functions in frailty, regardless of the frailty definition.

Our pathway analyses highlighted Herpes simplex virus 1 infection and various cell adhesion and lipid/lipoprotein metabolism pathways relevant to the signals. The first two pathways overlapped with the FI pathways, while lipid metabolism processes were shared with the FP pathways. However, several pathways were unique to each frailty measure: FI was enriched for immunoinflammatory functions, while FP included cardiac and membrane transport processes. These differences likely stem from the varying components of each frailty measure. The HFRS, which includes 109 conditions capturing both multisystem decline and core physiological senescence, showed enrichment in fundamental processes like cell adhesion and lipoprotein metabolism. The FI, also reflecting multisystem decline, appears particularly influenced by immunoinflammatory factors, as seen also in previous associations with GlycA[26], a marker of systemic inflammation, including studies supporting a causal link[27]. Many FI-related conditions, such as cardiovascular disease and diabetes, also have inflammatory components, potentially explaining the connection. The FP, which mostly focuses on physical frailty, was enriched for cardiac function and membrane transport pathways, both essential for muscle activity, ion flux regulation and nutrient uptake.

To assess the usefulness of the HFRS in our samples, we showed that it predicts mortality independent of sex and birth year and performs equally well even when dementia is excluded. Similarly, the HFRS-PRSs, also when dementia was removed, associated with the risk of frailty, early-onset frailty, mortality and hospitalizations. As frailty manifests relatively late in life for most individuals, risk assessment through PRSs may offer possibilities for early intervention to mitigate frailty before it escalates. Future studies are needed to ascertain the clinical utility of such approaches.

Our definition of frailty was based on clinical diagnoses in register data; such an approach has both advantages and disadvantages. A notable advantage is that in Finland and the United Kingdom, public

healthcare is primarily tax-funded, and each citizen has equal access. Issues pertinent to self-reported data, such as recall bias and missing information were also avoided. On the other hand, some conditions may be underreported in the registers, while others may have a lag from the onset of symptoms to assigning the diagnosis. We also note that the genetic associations were weaker in the UK Biobank compared to FinnGen, a finding likely explained by healthy selection due to volunteer-based participation in the UK Biobank[28] compared to FinnGen, which consists of national cohorts and biobank samples of hospitalized individuals. Also pertinent to all GWASs, the discovery samples tend to have stronger association statistics compared to replication, a phenomenon known as the winner's curse.

In conclusion, we provide a large GWAS on HFRS and reveal new genetic contributions and causal candidate genes. Overall, the results reinforce previous findings that immunoinflammatory and nervous system functions are relevant to the etiology of frailty, regardless of how frailty is defined. Future studies should thus explore the role of these functions in the development of frailty, including cognitive frailty, to better understand the etiology of frailty.

Methods

This work complies with all relevant ethical regulations. A full list of the ethics boards that approved the study protocols is provided at the end of this section.

Samples

FinnGen is a large national genetic resource (N = 520,210; release 12) established in 2017 and consists of Finnish individuals, aged 18 years and older at study baseline[29]. FinnGen includes prospective epidemiological and disease-based cohorts, as well as hospital biobank samples. Information on diagnoses since 1969 was linked by the unique national personal identification number to national healthcare, population and cause of death registries and recorded using the ICD Revisions 8–10. Information on dates and causes of death were obtained via linkages to the population and cause of death registers through 30 September 2023 (R12 v1). After excluding individuals with missing information on baseline age, birth year and sex, and samples not passing genotyping quality control (see below), we included 500,737 FinnGen participants in this study.

The UK Biobank includes 502,642 volunteer participants, aged 37 to 73 years old at baseline, recruited through 22 assessment centers across England, Scotland and Wales between 2006 and 2010 (ref. 30). The participants provided self-reported information on demographics, lifestyle and disease history via questionnaire and underwent physiological measurements, including providing a blood sample for genetics data. Hospital inpatient data were sourced from the Hospital Episode Statistics for England, Scottish Morbidity Record and Patient Episode Database for Wales, which contain electronic medical records (that is, ICD-10 codes) for all hospital admissions in England, Scotland and Wales, respectively. The hospital inpatient data were available through 31 October 2022 for England, 31 August 2022 for Scotland and 31 May 2022 for Wales. Death register data contained all deaths in the population through 30 November 2022, including primary and contributory causes of death. Participants in both UK Biobank and FinnGen have not received compensation for their participation.

Assessment of frailty

The HFRS was calculated according to a previously described protocol[4] based on 109 weighted ICD-10 codes. The codes were selected through a data-driven approach to include codes that were most prevalent in individuals with frailty and high healthcare resource use[4]. Each code was assigned with a weight ranging from 0.1 to 7.1, based on its association with frailty and predictive value for frailty-related outcomes[4]. The weights of all relevant ICD-10 codes present in an individual's records were then extracted and summed to calculate the HFRS score. The conditions, their respective weights and proportion of individuals with each condition in FinnGen and the UK Biobank are listed in Supplementary Table 18. The HFRS was used as a continuous variable in the GWAS. We also categorized the HFRS into low (<5), intermediate (5–15) and high (>15) risk of frailty as previously described[4] and used the cutoff points to describe frailty in our study populations. In the main analysis, we included all available ICD-10 codes for each person from age 30 years to the age at the end of follow-up to calculate the HFRS. As dementia diagnoses have the highest weight in the HFRS, we also calculated the HFRS by excluding dementia weights from the formula and performed sensitivity analyses on all analyses using the HFRS without dementia.

Genotyping and imputation

Genotyping in FinnGen was performed in Illumina and custom AxiomGT1 Affymetrix (Thermo Fisher Scientific) genome-wide arrays and imputed to 16,387,711 (imputation INFO score > 0.6) variants using a population-specific SISu v.3 imputation reference panel as previously described[31]. Individuals with ambiguous sex and non-Finnish ancestry were excluded. UK Biobank samples (v3 genotyping release) were genotyped on custom Affymetrix microarrays and imputed using the 1000 Genomes and the Haplotype Reference Consortium reference panels to ~93 million variants[32]. Participants were excluded if they were flagged as having unusually high heterozygosity or missing genotype calls (<5%). Our analysis was restricted to participants with European descent and white British ancestry (N = 407,463). Detailed procedures on genotype calling, quality controls and imputation have been previously described for FinnGen[29] and the UK Biobank[32].

Statistics and reproducibility

No statistical method was used to predetermine sample size, as the UK Biobank and FinnGen cohorts are sufficiently large and can be anticipated to provide adequate statistical power for the planned analyses. We have sought to include all samples after exclusion based only on incomplete data, such as sex, birth year and genotype quality control as called by the respective cohorts. In the case of the UK Biobank, non-European descent and non-white British ancestry participants were excluded to facilitate the comparison to the homogeneous FinnGen Finnish populations. Our study did not involve randomization/allocation into experimental groups, as it was an observational, hypothesis-free GWAS treating the HFRS as a continuous outcome. Therefore, no experimental manipulation or group assignment was performed. In a hypothesis-free GWAS, blinding is not possible/necessary as the analysis is fully automated and applies standardized statistical tests uniformly across the genome. Data distribution was assumed to be normal, but this was not formally tested.

Discovery GWAS, replication and meta-analysis

The analytical pipeline for GWAS and post-GWAS analyses is presented in Fig. 1. We first performed GWASs of HFRS and HFRS without dementia in FinnGen using the SAIGE[33] (v.0.35.8.8) software, which uses linear mixed-effects modeling to account for genetic relatedness and confounding by ancestry[34]. We included variants (N = 21,294,561) with minor allele frequency > 0.01%, Hardy–Weinberg P value > 1×10^{-9} and imputation INFO score ≥ 0.9. The models were adjusted for birth year, sex and the first ten PCs. The genome-wide significance level was set to 5×10^{-8}. The total number of genes to which the variants were mapped was determined by extracting variants with a $P < 5 \times 10^{-8}$, followed by variant mapping and annotation using the Variant Effect Predictor[35] in the standard FinnGen GWAS annotation pipeline[29]. Independent lead variants were identified using the R package gwasRtools[36]. We used a distance-based loci definition on the genome-wide significant variants (that is, 500 kb from index variant) to estimate the independent genomic loci. Independent lead variants were identified by linkage disequilibrium clumping and defined as those that were independent from each other at $r^2 < 0.01$.

To replicate the findings at the variant level, we performed both HFRS GWASs in the UK Biobank. To account for the related samples in the UK Biobank, we applied a mixed linear model-based GWAS analysis ('fastGWA')[37], which is an efficient method to control for relatedness by a sparse genetic relationship matrix, without the need of excluding related individuals. The models were adjusted for birth year, sex, genotyping array and the first ten PCs. Finally, to capture the totality of the evidence, we conducted a meta-analysis on the results from FinnGen and the UK Biobank using METAL[38]. A fixed-effect meta-analysis was performed using the default approach, with *P* value and direction of effect weighted according to sample size, and with adjustment for genomic control (lambda). Using the NHGRI-EBI GWAS Catalog[39] filtered for $P < 5 \times 10^{-8}$ and results of previous GWASs into frailty (using the FP[11] and FI[10] to measure frailty) and mvAge[13], a genomic structural equation modeling-derived composite construct of healthspan, parental lifespan, extreme longevity, frailty and epigenetic aging, we assessed the number of novel and previously unreported associations relative to the FinnGen results.

Genetic correlation and heritability

Using linkage disequilibrium score regression[40] (v1.0.1) and linkage disequilibrium merged with the HapMap3 reference panel of ~1.1 million variants, we estimated (1) the potential bias from, for example, population stratification and cryptic heritability in the GWAS results, (2) heritability of HFRS and (3) genetic correlations between HFRS and previous GWASs of FI[10], FP[11] and mvAge[13]. As the FI GWAS[10] used an opposite effect allele compared to the standard FinnGen workflow, we inverted the genetic correlation coefficient to prevent an artifactual negative correlation and facilitate interpretation.

Functional annotation: cell-type and pathway enrichment

To explore tissue and cell-type specificity of the annotated genes underlying HFRS, we applied WebCSEA, a web platform to derive context-specific expression patterns of genes underlying complex traits, encompassing the Human Cell Atlas and single-cell data resources[41,42]. Enrichr pathway analysis[14] based on KEGG[39] and Reactome[40] resources was applied to explore enriched pathways of the identified genes (GWAS $P < 5 \times 10^{-8}$). To effectively compare the enriched pathways of the HFRS with those of the FI and FP GWASs, we extracted all genome-wide significant variants from these GWASs and performed KEGG and Reactome pathway analyses using the same (default) settings.

Proteomics integration

To prioritize genes and identify potentially functional and causal variants, we narrowed down the association signals to a smaller number of missense, splice region, loss of function and 5′ and 3′ untranslated region variants (the two last mentioned potentially affecting transcript stability, localization and signal response), identified from the Variant Effect Predictor pipeline[35], that were associated with the HFRS at a slightly more relaxed threshold ($P < 5 \times 10^{-7}$). Using the Olink proteomics data, we then examined if the protein levels of the variants (at a gene-level resolution) were associated with HFRS in the UK Biobank. Details of the UK Biobank Olink proteomics assay, quality-control and data processing procedures have been described elsewhere[43]. Briefly, 54,239 UK Biobank participants were selected for the proteomics profiling using EDTA plasma samples collected at the baseline assessment. Of the 54,239 samples, 46,595 were randomly selected, while 6,376 were chosen by UKB-PPP consortium members and 1,268 were from participants in the coronavirus disease 2019 repeat imaging study, resulting in a sample that was predominantly, but not entirely, random. A total of 2,923 proteins were measured across 8 protein panels using the antibody-based Olink Explore 3072 platform. Protein levels were measured in Normalized Protein eXpression values, which represent the relative concentration of proteins on a $\log_2$ scale. All the protein levels were scaled to mean = 0 and s.d. = 1 before the association testing. Linear regression models were then performed to assess the associations between the proteins that were available in the Olink platform and HFRS, adjusting for (i) birth year, sex and the first ten PCs and (ii) batch, baseline assessment centers, BMI and smoking. We considered an FDR < 0.05 as statistically significant in the proteomics analysis.

Colocalization analyses

To further prioritize the genes and identify causal variants, we performed a Bayesian-based colocalization analysis with eQTL, sQTL and pQTL, using a flanking window of 1 Mb and default parameters for prior probabilities[12]. The analysis assumes that only one causal variant exists for each trait in a genomic locus and returns PPs indicating the likelihood that the following hypotheses (H) are true: there is no association at the locus with either expression/splicing/protein level or HFRS (H0); there is an association with expression/splicing/protein level but not HFRS (H1); there is no association with expression/splicing/protein level, but there is an association with HFRS (H2); there is an association with both expression/splicing/protein level and HFRS, but with distinct causal variants (H3); there is an association with both expression/splicing/protein level and HFRS with a shared causal variant (H4). We considered the analysis having enough power if the sum PPs had a distinct or shared causal variant exceeded 88%. A colocalized signal was detected if the PP of a shared causal variant (H4) existence was greater than 80%. The GTEx database[44] (v8) was interrogated for eQTL and sQTL, while the UK Biobank Pharma Proteomics Project[43] was used for pQTL.

PRS analyses

Using the GWAS summary statistics from FinnGen, we calculated the PRSs for HFRS in the UK Biobank by applying PRSs with continuous shrinkage[45,46] and using the European panel from the 1000 Genomes[46] linkage disequilibrium reference, where ~1.1 million variants were selected. Using linear regression, we fitted a linear model to assess how the HFRS-PRSs associate with the HFRS score. HFRS was considered as a standardized *z*-score in the linear regressions. We also performed logistic regressions to assess the associations of the HFRS-PRSs with early-onset frailty, defined as HFRS > 5 before age 65. Age 65 was chosen as the cutoff as it is commonly used to distinguish 'young' from 'old' in statistical and policy contexts. Our previous work also identified age 65 as the optimal threshold for distinguishing between early-life and late-life frailty[47]. The PRS was modeled per standard deviation change, and all the models included birth year, sex and the first ten PCs as covariates.

Lastly, as frailty manifests in late life for most individuals, we asked whether the HFRS-PRSs could be used in early risk stratification to identify individuals at risk of adverse outcomes. To this end, Cox models with attained age as the timescale and linear regression models were fitted to assess whether the HFRS-PRSs predict all-cause mortality and number of hospitalizations, respectively. The added value of the HFRS-PRSs beyond age and sex in the prediction was assessed using the *F*-test for linear regressions and likelihood-ratio test for Cox models. The number of hospitalizations was scaled to a mean = 0 and s.d. = 1 before modeling.

Prediction of mortality using HFRS

Cox models with attained age as the timescale, which inherently adjusts for age, were fitted to assess the association between HFRS, HFRS without dementia and all-cause mortality in FinnGen and the UK Biobank. Two models were fitted for each HFRS definition: one adjusting for sex and birth year, and one without adjustments.

Ethics statements of FinnGen and UK Biobank

FinnGen. Patients and control participants in FinnGen provided informed consent for biobank research, based on the Finnish Biobank

Act. Alternatively, separate research cohorts, collected before the Finnish Biobank Act came into effect (in September 2013) and the start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by the Finnish Medicines Agency, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by the Finnish Medicines Agency. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is HUS/990/2017. The FinnGen study is approved by Finnish Institute for Health and Welfare (permit nos. THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit nos. VRK43431/2017-3, VRK/6909/2018-3 and VRK/4415/2019-3), the Social Insurance Institution (permit nos. KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020 and KELA 16/522/2020), Findata (permit nos. THL/2364/14.02/2020, THL/4055/14.06.00/2020, THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021 and THL/4235/14.06.00/202), Statistics Finland (permit nos. TK-53-1041-17, TK/143/07.03.00/2020 (earlier TK-53-90-20), TK/1735/07.03.00/2021 and TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4 July 2019.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 9 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, Auria Biobank AB17-5154 and amendment no. 1 (17 August 2020), AB20-5926 and amendment no. 1 (23 April 2020) and it's modification (22 September 2021), Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21 February 2020 and 06 October 2020), Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25 August 2020.

UK Biobank. The UK Biobank study was approved by the North West Multi-centre Research Ethics Committee (approval no. 11/NW/03820). All participants provided written informed consent for data collection, analysis and record linkage. We have also obtained ethical approval for the use of UK Biobank data in Sweden (2016/1888-31/1).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Individual-level data cannot be stored in public repositories or otherwise made publicly available due to ethical and data protection restrictions. However, data are available upon request for researchers who meet the criteria for access to confidential data. Data from the UK Biobank are available to bona fide researchers upon application at https://www.ukbiobank.ac.uk/enable-your-research/. The following UK Biobank-associated data were accessed through, and as part of, our UK Biobank accession: Hospital Episode Statistics for England, Scottish Morbidity Record and Patient Episode Database for Wales. FinnGen results, according to the FinnGen consortium agreement, are subjected to a one-year embargo, and summary statistics are then made available to the scientific community and released two times a year. Information on accessing FinnGen data can be found at https://www.finngen.fi/en/access_results/. The national healthcare, population and cause of death registers were accessed through, and as part of, our FinnGen accession, implemented in the FinnGen pipelines.

Code availability

All the data processing, visualization and statistical analyses were performed using Python 3.8 (2.7 for LDSC) and R v.4.3.2 (R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org/). Meta-analyses were performed using METAL v.2011-03-25. Independent Genomic loci were identified using the R package gwasRtools[36] (version 0.1.7; https://lcpilling.github.io/gwasRtools/). Venn diagrams were created using the R package ggvenn (version 0.1.10; https://cran.r-project.org/web/packages/ggvenn/index.html). Correlation plots were created using the R package corrplot (v.0.92; https://cran.r-project.org/web/packages/corrplot/index.html). Forest plots were created using the R package ggforestplot (v.0.1.0; https://nightingalehealth.github.io/ggforestplot/). The analysis codes are available as follows: FinnGen GWAS via https://github.com/FINNGEN/saige-pipelines/; UK Biobank GWAS via http://cnsgenomics.com/software/gcta/#fastGWA and QTL colocalization analysis via https://github.com/Moritz-JD-Krueger/Colocalization-Analysis/.

References

1. Clegg, A., Young, J., Iliffe, S., Rikkert, M. O. & Rockwood, K. Frailty in elderly people. *Lancet* **381**, 752–762 (2013).
2. Kojima, G., Iliffe, S. & Walters, K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing* **47**, 193–200 (2018).
3. Theou, O., Brothers, T. D., Mitnitski, A. & Rockwood, K. Operationalization of frailty using eight commonly used scales and comparison of their ability to predict all-cause mortality. *J. Am. Geriatr. Soc.* **61**, 1537–1551 (2013).
4. Gilbert, T. et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet Lond. Engl.* **391**, 1775–1782 (2018).
5. Searle, S. D., Mitnitski, A., Gahbauer, E. A., Gill, T. M. & Rockwood, K. A standard procedure for creating a frailty index. *BMC Geriatr.* **8**, 24 (2008).
6. Fried, L. P. et al. Frailty in older adults: evidence for a phenotype. *J. Gerontol. A. Biol. Sci. Med. Sci.* **56**, M146–M156 (2001).
7. Young, A. C. M., Glaser, K., Spector, T. D. & Steves, C. J. The identification of hereditary and environmental determinants of frailty in a cohort of UK twins. *Twin Res. Hum. Genet.* **19**, 600–609 (2016).
8. Mak, J. K. L. et al. Sex differences in genetic and environmental influences on frailty and its relation to body mass index and education. *Aging* **13**, 16990–17023 (2021).
9. Mak, J. K. L. et al. Genetic and environmental influences on longitudinal frailty trajectories from adulthood into old age. *J. Gerontol. A. Biol. Sci. Med. Sci.* **78**, 333–341 (2023).
10. Atkins, J. L. et al. A genome-wide association study of the frailty index highlights brain pathways in ageing. *Aging Cell* **20**, e13459 (2021).
11. Ye, Y. et al. A genome-wide association study of frailty identifies significant genetic correlation with neuropsychiatric, cardiovascular, and inflammation pathways. *Geroscience* **45**, 2511–2523 (2023).
12. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
13. Rosoff, D. B. et al. Multivariate genome-wide analysis of aging-related traits identifies novel loci and new drug targets for healthy aging. *Nat. Aging* **3**, 1020–1035 (2023).
14. Xie, Z. et al. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).

15. Aslam, M. M. et al. Genome-wide analysis identifies novel loci influencing plasma apolipoprotein E concentration and Alzheimer's disease risk. *Mol. Psychiatry* **28**, 4451–4462 (2023).
16. Yeh, K. -H. et al. Genetic variants at the *APOE* Locus predict cardiometabolic traits and metabolic syndrome: a Taiwan Biobank study. *Genes* **13**, 1366 (2022).
17. Kemper, K. E. et al. Genetic influence on within-person longitudinal change in anthropometric traits in the UK Biobank. *Nat. Commun.* **15**, 3776 (2024).
18. Ambrose, R. L. et al. Molecular characterisation of ILRUN, a novel inhibitor of proinflammatory and antimicrobial cytokines. *Heliyon* **6**, e04115 (2020).
19. PubChem. CHST9 - carbohydrate sulfotransferase 9 (human). https://pubchem.ncbi.nlm.nih.gov/gene/CHST9/human
20. NCBI Gene. CGREF1 cell growth regulator with EF-hand domain 1 (human). https://www.ncbi.nlm.nih.gov/gene/10669
21. Ziembik, M. A., Bender, T. P., Larner, J. M. & Brautigan, D. L. Functions of protein phosphatase-6 in NF-κB signaling and in lymphocytes. *Biochem. Soc. Trans.* **45**, 693–701 (2017).
22. Tin, A. et al. Identification of circulating proteins associated with general cognitive function among middle-aged and older adults. *Commun. Biol.* **6**, 1117 (2023).
23. Lim, S. M. et al. Differential proteomic profiles between cognitive frail and robust older adults from the MELoR cohort. *Geroscience* https://doi.org/10.1007/s11357-024-01462-z (2024).
24. Turcotte, L. A. et al. External validation of the Hospital Frailty Risk Score among hospitalised home care clients in Canada: a retrospective cohort study. *Age Ageing* **52**, afac334 (2023).
25. Mak, J. K. L., Kuja-Halkola, R., Wang, Y., Hägg, S. & Jylhävä, J. Frailty and comorbidity in predicting community COVID-19 mortality in the UK Biobank: the effect of sampling. *J. Am. Geriatr. Soc.* **69**, 1128–1139 (2021).
26. Zhang, S. et al. A metabolomic profile of biological aging in 250,341 individuals from the UK Biobank. *Nat. Commun.* **15**, 8081 (2024).
27. Mak, J. K. L. et al. Unraveling the metabolic underpinnings of frailty using multicohort observational and Mendelian randomization analyses. *Aging Cell* **22**, e13868 (2023).
28. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
29. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
30. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
31. Pärn, K. et al. Genotype imputation workflow v3.0. dx.doi.org/10.17504/protocols.io.xbgfijw (2018).
32. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
33. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
34. Loh, P. -R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
35. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
36. GitHub. R package for processing of GWAS output. https://lcpilling.github.io/gwasRtools/
37. Jiang, L. et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).
38. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
39. EMBL-EBI. GWAS Catalog. https://www.ebi.ac.uk/gwas/
40. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
41. Dai, Y. et al. WebCSEA: web-based cell-type-specific enrichment analysis of genes. *Nucleic Acids Res.* **50**, W782–W790 (2022).
42. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
43. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
44. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
45. Ge, T., Chen, C. -Y., Ni, Y., Feng, Y. -C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
46. Chou, W. -C. et al. A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. *Sci Rep.* **6**, 39313 (2016).
47. Bai, G. et al. Is frailty different in younger adults compared to old? Prevalence, characteristics, and risk factors of early-life and late-life frailty in samples from Sweden and UK. *Gerontology* **69**, 1385–1393 (2023).

Acknowledgements

This work was supported by the Swedish Research Council (grant nos. 2018-02077 to J.J., 2019-01272, 2020-06101, 2022-01608), the Research Council of Finland to J.J. (grant no. 3493358), the Sigrid Jusélius Foundation to J.J., the Yrjö Jahnsson Foundation to J.J. (grant no. 20217416), Instrumentarium Science Foundation to J.J. and Signe and Ane Gyllenberg Foundation to J.J. (grant no. 6226). This research was conducted using the UK Biobank resource, as part of the registered project 22224. The analyses of UK Biobank genotypes were enabled by resources in project sens2017519 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie, AstraZeneca, Biogen, Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech, Merck Sharp & Dohme, Pfizer, GlaxoSmithKline Intellectual Property Development, Sanofi US Services, Maze Therapeutics, Janssen Biotech and Novartis. The following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (https://www.auria.fi/biopankki/), THL Biobank (https://www.thl.fi/biobank/), Helsinki Biobank (https://www.helsinginbiopankki.fi/), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx/), Finnish Clinical Biobank Tampere (https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere/), Biobank of Eastern Finland (https://www.ita-suomenbiopankki.fi/en/), Central Finland Biobank (https://www.ksshp.fi/fi-FI/Potilaalle/Biopankki/), Finnish Red Cross Blood Service Biobank (https://www.veripalvelu.fi/verenluovutus/biopankkitoiminta/) and Terveystalo Biobank (https://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (https://www.bbmri.fi/). The Finnish Biobank Cooperative-FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be

accessed through the Fingenious services (https://site.fingenious.fi/en/) managed by FINBB. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.J. conceived the study plan and designed the proof outline. J.K.L.M., C.Q., M.K., J.L. and A.K. performed the analyses. J.J., J.L. and S.H. were responsible for data acquisition. All authors contributed to the writing of the paper and interpretation of the results. All authors listed under FinnGen contributed to the generation of the primary data of the FinnGen data release 12. FinnGen authors are listed in Supplementary Table 19.

Funding

Open access funding provided by Karolinska Institute.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43587-025-00925-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43587-025-00925-y.

Correspondence and requests for materials should be addressed to Juulia Jylhävä.

Peer review information *Nature Aging* thanks Zuyun Liu, Julian Mutz and Luke Pilling, for their contribution to the peer review of this work.

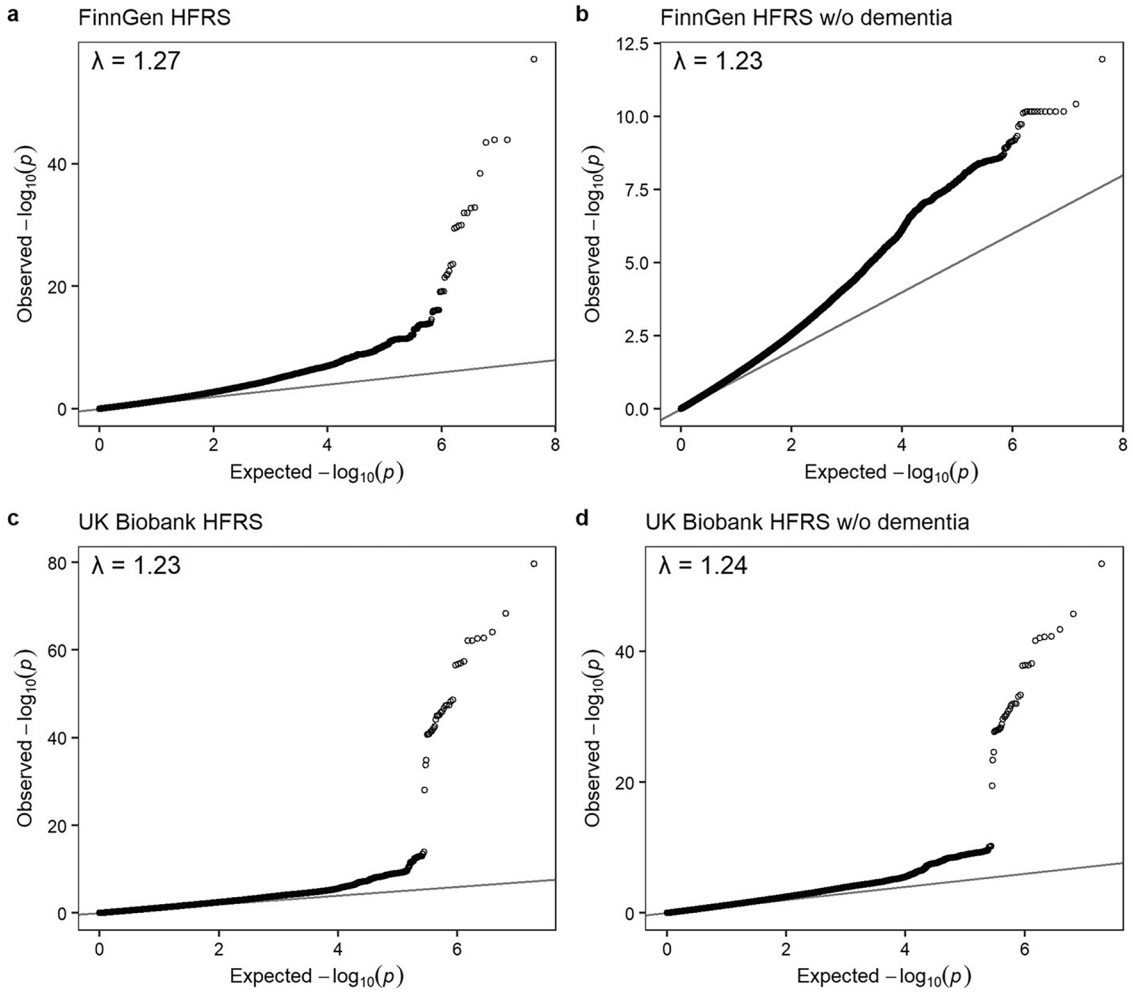
FinnGen

Aarno Palotie[5,6], Mark Daly[5,6], Bridget Riley-Gillis[7], Howard Jacob[7], Dirk Paul[8], Slavé Petrovski[8], Heiko Runz[9], Sally John[9], George Okafo[10], Robert Plenge[11], Joseph Maranville[11], Mark McCarthy[12], Margaret G. Ehm[13], Kirsi Auro[14], Simonne Longerich[15], Anders Mälarstig[16], Katherine Klinger[17], Clément Chatelain[17], Matthias Gossel[17], Karol Estrada[18], Robert Graham[18], Dawn Waterworth[19], Chris O´Donnell[20], Nicole Renaud[20], Tomi P. Mäkelä[21], Jaakko Kaprio[5], Petri Virolainen[22], Antti Hakanen[22], Terhi Kilpi[23], Markus Perola[23], Jukka Partanen[24], Anne Pitkäranta[25], Taneli Raivio[25], Jani Tikkanen[26], Raisa Serpi[26], Tarja Laitinen[27], Veli-Matti Kosma[28], Jari Laukkanen[29], Marco Hautalahti[30], Outi Tuovila[31], Raimo Pakkanen[31], Jeffrey Waring[7], Fedik Rahimov[7], Ioanna Tachmazidou[8], Chia-Yen Chen[9], Zhihao Ding[10], Marc Jung[10], Hanati Tuoken[10], Shameek Biswas[11], Rion Pendergrass[12], David Pulford[32], Neha Raghavan[15], Adriana Huertas-Vazquez[15], Jae-Hoon Sul[15], Xinli Hu[16], Åsa Hedman[16], Manuel Rivas[18,33], Ma´en Obeidat[20], Jonathan Chung[20], Jonas Zierer[20], Mari Niemi[20], Samuli Ripatti[5], Johanna Schleutker[22], Mikko Arvas[24], Olli Carpén[25], Reetta Hinttala[26], Johannes Kettunen[26], Arto Mannermaa[28], Katriina Aalto-Setälä[34], Mika Kähönen[27], Johanna Mäkelä[30], Reetta Kälviäinen[35], Valtteri Julkunen[35], Hilkka Soininen[35], Anne Remes[36], Mikko Hiltunen[37], Jukka Peltola[38], Minna Raivio[39], Pentti Tienari[39], Juha Rinne[40], Roosa Kallionpää[40], Juulia Partanen[41], Adam Ziemann[7], Nizar Smaoui[7], Anne Lehtonen[7], Susan Eaton[9], Sanni Lahdenperä[9], Natalie Bowers[12], Edmond Teng[12], Fanli Xu[42], Laura Addis[42], John Eicher[42], Qingqin S. Li[43], Karen He[19], Ekaterina Khramtsova[19], Martti Färkkilä[39], Jukka Koskela[39], Sampsa Pikkarainen[39], Airi Jussila[38], Katri Kaukinen[38], Timo Blomster[36], Mikko Kiviniemi[35], Markku Voutilainen[40], Tim Lu[12], Linda McCarthy[42], Amy Hart[19], Meijian Guan[19], Jason Miller[15], Kirsi Kalpala[16], Melissa Miller[16], Kari Eklund[39], Antti Palomäki[40], Pia Isomäki[38], Laura Pirilä[40], Oili Kaipiainen-Seppänen[35], Johanna Huhtakangas[36], Nina Mars[5], Apinya Lertratanakul[7], Coralie Viollet[8], Marla Hochfeld[11], Jorge Esparza Gordillo[42], Fabiana Farias[15], Nan Bing[16], Margit Pelkonen[35], Paula Kauppi[39], Hannu Kankaanranta[44,45,46], Terttu Harju[36], Riitta Lahesmaa[40], Hubert Chen[12], Joanna Betts[42], Rajashree Mishra[42], Majd Mouded[47], Debby Ngo[47], Teemu Niiranen[48], Felix Vaura[48], Veikko Salomaa[48], Kaj Metsärinne[40], Jenni Aittokallio[40], Jussi Hernesniemi[38], Daniel Gordin[39], Juha Sinisalo[39], Marja-Riitta Taskinen[39], Tiinamaija Tuomi[39], Timo Hiltunen[39], Amanda Elliott[5,49,50], Mary Pat Reeve[5], Sanni Ruotsalainen[5], Audrey Chu[42], Dermot Reilly[51], Mike Mendelson[52], Jaakko Parkkinen[16], Tuomo Meretoja[39], Heikki Joensuu[39], Johanna Mattson[39], Eveliina Salminen[39], Annika Auranen[38], Peeter Karihtala[36], Päivi Auvinen[35], Klaus Elenius[40], Esa Pitkänen[5], Relja Popovic[7], Margarete Fabre[8], Jennifer Schutzman[12], Diptee Kulkarni[42], Alessandro Porello[19], Andrey Loboda[15], Heli Lehtonen[16], Stefan McDonough[16], Sauli Vuoti[53], Kai Kaarniranta[35,54], Joni A. Turunen[55,56], Terhi Ollila[39], Hannu Uusitalo[38], Juha Karjalainen[5], Mengzhen Liu[7], Stephanie Loomis[9], Erich Strauss[12], Hao Chen[12], Kaisa Tasanen[36], Laura Huilaja[36], Katariina Hannula-Jouppi[39], Teea Salmi[38], Sirkku Peltonen[40], Leena Koulu[40], David Choy[12], Ying Wu[16],

Pirkko Pussinen[39], Aino Salminen[39], Tuula Salo[39], David Rice[39], Pekka Nieminen[39], Ulla Palotie[39], Maria Siponen[35], Liisa Suominen[35], Päivi Mäntylä[35], Ulvi Gursoy[40], Vuokko Anttonen[36], Kirsi Sipilä[57,58], Hannele Laivuori[5], Venla Kurra[38], Laura Kotaniemi-Talonen[38], Oskari Heikinheimo[39], Ilkka Kalliala[39], Lauri Aaltonen[39], Varpu Jokimaa[40], Marja Vääräsmäki[36], Outi Uimari[36], Laure Morin-Papunen[36], Maarit Niinimäki[36], Terhi Piltonen[36], Katja Kivinen[5], Elisabeth Widen[5], Taru Tukiainen[5], Niko Välimäki[59], Eija Laakkonen[60], Jaakko Tyrmi[61], Heidi Silven[61], Eeva Sliz[61], Riikka Arffman[61], Susanna Savukoski[61], Triin Laisk[62], Natalia Pujol[62], Janet Kumar[13], Iiris Hovatta[59], Erkki Isometsä[39], Hanna Ollila[5], Jaana Suvisaari[48], Antti Mäkitie[63], Argyro Bizaki-Vallaskangas[38], Sanna Toppila-Salmi[64,65], Tytti Willberg[40], Elmo Saarentaus[5], Antti Aarnisalo[39], Elisa Rahikkala[36], Kristiina Aittomäki[66], Fredrik Åberg[67], Mitja Kurki[5,49], Aki Havulinna[5,48], Juha Mehtonen[5], Priit Palta[5], Shabbeer Hassan[5], Pietro Della Briotta Parolo[5], Wei Zhou[49], Mutaamba Maasha[49], Susanna Lemmelä[5], Aoxing Liu[5], Arto Lehisto[5], Andrea Ganna[5], Vincent Llorens[5], Henrike Heyne[5], Joel Rämö[5], Rodos Rodosthenous[5], Satu Strausz[5], Tuula Palotie[68], Kimmo Palin[59], Javier Garcia-Tabuenca[69], Harri Siirtola[69], Tuomo Kiiskinen[5], Jiwoo Lee[5,49], Kristin Tsuo[5,49], Kati Kristiansson[23], Kati Hyvärinen[70], Jarmo Ritari[70], Katri Pylkäs[61], Minna Karjalainen[61], Tuomo Mantere[26], Eeva Kangasniemi[27], Sami Heikkinen[37], Nina Pitkänen[22], Samuel Lessard[17], Lila Kallio[22], Tiina Wahlfors[23], Eero Punkka[25], Sanna Siltanen[27], Teijo Kuopio[29], Anu Jalanko[5], Huei-Yi Shen[5], Risto Kajanne[5], Mervi Aavikko[5], Helen Cooper[5], Denise Öller[5], Rasko Leinonen[5,71], Henna Palin[27], Malla-Maria Linna[25], Masahiro Kanai[49], Zhili Zheng[49], L. Elisa Lahtela[5], Mari Kaunisto[5], Elina Kilpeläinen[5], Timo P. Sipilä[5], Oluwaseun Alexander Dada[5], Awaisa Ghazal[5], Anastasia Kytölä[5], Rigbe Weldatsadik[5], Kati Donner[5], Anu Loukola[25], Päivi Laiho[23], Tuuli Sistonen[23], Essi Kaiharju[23], Markku Laukkanen[23], Elina Järvensivu[23], Sini Lähteenmäki[23], Lotta Männikkö[23], Regis Wong[23], Auli Toivola[23], Minna Brunfeldt[23], Hannele Mattsson[23], Sami Koskelainen[23], Tero Hiekkalinna[23], Teemu Paajanen[23], Shuang Luo[5], Shanmukha Sampath Padmanabhuni[5], Marianna Niemi[69], Mika Helminen[69], Tiina Luukkaala[69], Iida Vähätalo[69], Jyrki Tammerluoto[5], Sarah Smith[72], Tom Southerington[72] & Petri Lehto[72]

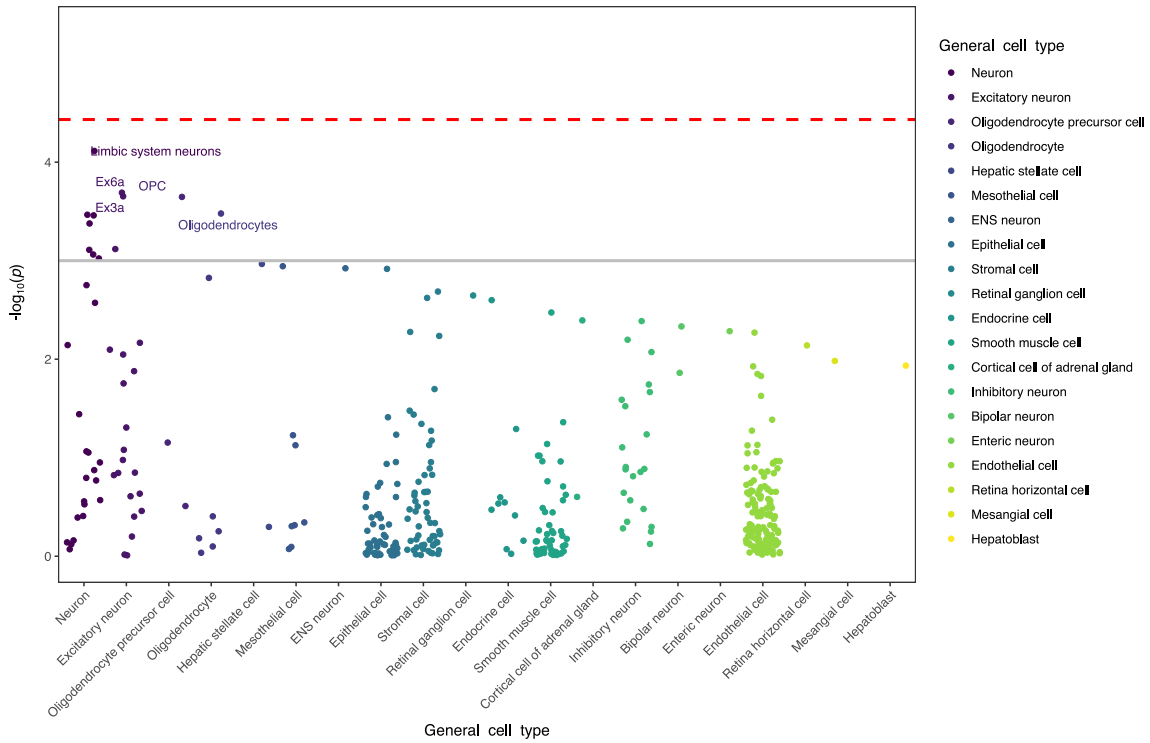
[6]Broad Institute of MIT and Harvard; Massachusetts General Hospital, Cambridge, MA, USA. [7]Abbvie, Chicago, IL, USA. [8]AstraZeneca, Cambridge, UK. [9]Biogen, Cambridge, MA, USA. [10]Boehringer Ingelheim, Ingelheim am Rhein, Germany. [11]Bristol Myers Squibb, New York, NY, USA. [12]Genentech, San Francisco, CA, USA. [13]GlaxoSmithKline, Collegeville, PA, USA. [14]GlaxoSmithKline, Espoo, Finland. [15]Merck, Kenilworth, NJ, USA. [16]Pfizer, New York, NY, USA. [17]Translational Sciences, Sanofi R&D, Framingham, MA, USA. [18]Maze Therapeutics, San Francisco, CA, USA. [19]Janssen Research & Development, LLC, Spring House, <City>, PA, USA. [20]Novartis Institutes for BioMedical Research, Cambridge, MA, USA. [21]HiLIFE, University of Helsinki, Finland, Finland. [22]Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland. [23]THL Biobank/Finnish Institute for Health and Welfare (THL), Helsinki, Finland. [24]Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland. [25]Helsinki Biobank/Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [26]Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland. [27]Finnish Clinical Biobank Tampere/University of Tampere/Pirkanmaa Hospital District, Tampere, Finland. [28]Biobank of Eastern Finland/University of Eastern Finland/Northern Savo Hospital District, Kuopio, Finland. [29]Central Finland Biobank/University of Jyväskylä/Central Finland Health Care District, Jyväskylä, Finland. [30]FINBB - Finnish Biobank Cooperative, Turku, Finland. [31]Business Finland, Helsinki, Finland. [32]GlaxoSmithKline, Stevenage, UK. [33]University of Stanford, Stanford, CA, USA. [34]Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [35]Northern Savo Hospital District, Kuopio, Finland. [36]Northern Ostrobothnia Hospital District, Oulu, Finland. [37]University of Eastern Finland, Kuopio, Finland. [38]Pirkanmaa Hospital District, Tampere, Finland. [39]Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [40]Hospital District of Southwest Finland, Turku, Finland. [41]Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsink, Finland. [42]GlaxoSmithKline, Brentford, UK. [43]Janssen Research & Development, LLC, Titusville, NJ, USA. [44]University of Gothenburg, Gothenburg, Sweden. [45]Seinäjoki Central Hospital, Seinäjoki, Finland. [46]Tampere University, Tampere, Finland. [47]Novartis, Basel, Switzerland. [48]Finnish Institute for Health and Welfare (THL), Helsinki, Finland. [49]Broad Institute, Cambridge, MA, USA. [50]Massachusetts General Hospital, Boston, MA, USA. [51]Janssen Research & Development, LLC, Boston, MA, USA. [52]Novartis, Boston, MA, USA. [53]Janssen-Cilag Oy, Espoo, Finland. [54]Department of Molecular Genetics, University of Lodz, Lodz, Poland. [55]Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [56]Eye Genetics Group, Folkhälsan Research Center, Helsinki, Finland. [57]Research Unit of Oral Health Sciences Faculty of Medicine, University of Oulu, Oulu, Finland. [58]Medical Research Center, Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland. [59]University of Helsinki, Helsinki, Finland. [60]University of Jyväskylä, Jyväskylä, Finland. [61]University of Oulu, Oulu, Finland. [62]Estonian biobank, Tartu, Estonia. [63]Department of Otorhinolaryngology - Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [64]University of Eastern Finland and Kuopio University Hospital, Department of Otorhinolaryngology, Kuopio, Finland. [65]Department of Allergy, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [66]Department of Medical Genetics, Helsinki University Central Hospital, Helsinki, Finland. [67]Transplantation and Liver Surgery Clinic, Helsinki University Hospital, Helsinki University, Helsinki, Finland. [68]University of Helsinki and Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [69]University of Tampere, Tampere, Finland. [70]Finnish Red Cross Blood Service, Helsinki, Finland. [71]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK. [72]Finnish Biobank Cooperative - FINBB, Turku, Finland.

https://doi.org/10.1038/s43587-025-00925-y



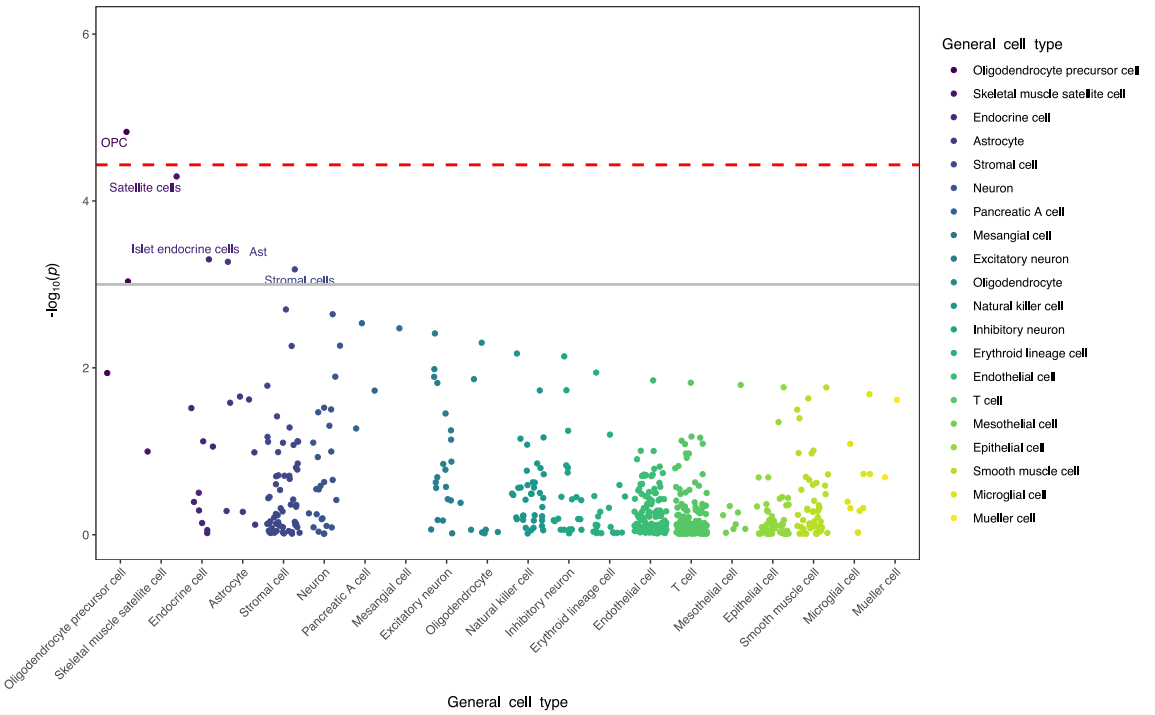
Extended Data Fig. 1 | QQ-plots of association summary statistics for the HFRS and HFRS without dementia. Panels **a** and **b** show results for FinnGen, and panels **c** and **d** for the UK Biobank, respectively.

https://doi.org/10.1038/s43587-025-00925-y



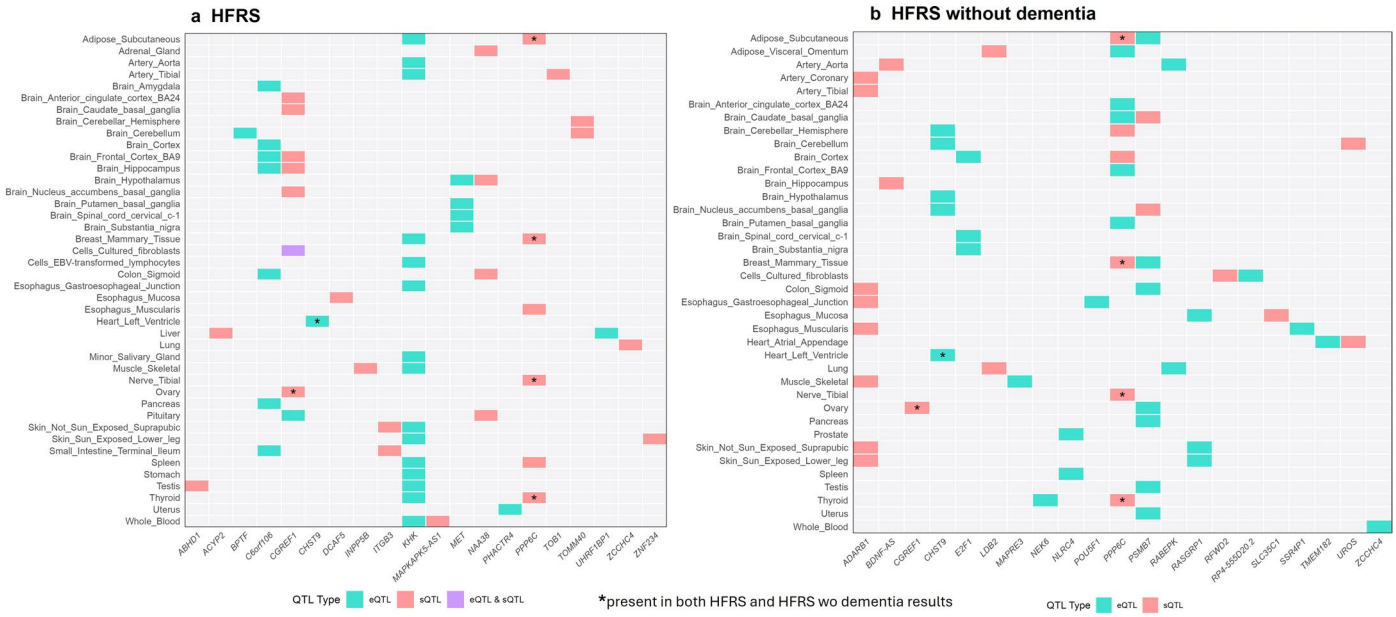
Extended Data Fig. 2 | Cell-type enrichment analysis of HFRS. Top 20 enriched cell types for the variants associated with the HFRS in FinnGen are shown.

https://doi.org/10.1038/s43587-025-00925-y



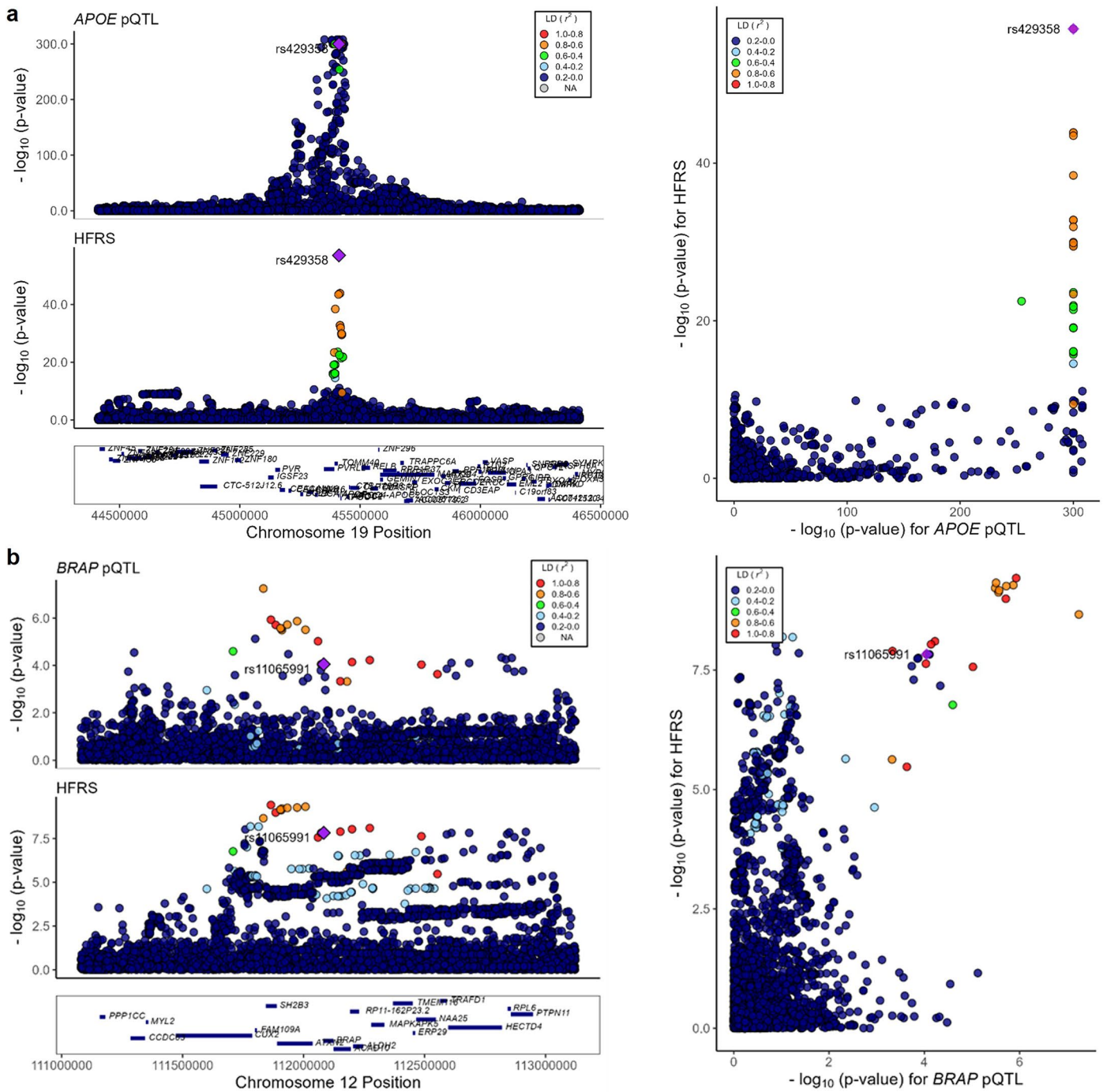
Extended Data Fig. 3 | Cell-type enrichment analysis of HFRS without dementia. Top 20 enriched cell types for the variants associated with the HFRS without dementia in FinnGen are shown.

https://doi.org/10.1038/s43587-025-00925-y



Extended Data Fig. 4 | Colocalized expression and splicing quantitative trait loci. Colocalized eQTL and sQTL by tissue with the genes associated with (**a**) HFRS and (**b**) HFRS without dementia. For each gene, the posterior probability for a shared causal variant was >80%.

https://doi.org/10.1038/s43587-025-00925-y



Extended Data Fig. 5 | Regional association plots for gene loci. Regional association plots for gene loci – panel **a** for *APOE* and panel **b** for *BRAP* – identified in the colocalization analysis of protein quantitative trait loci (pQTL) and the variants associated with the Hospital Frailty Risk Score (HFRS).

nature portfolio

Corresponding author(s): Juulia Jylhävä

Last updated by author(s): May 2, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Genotyping in FinnGen was performed on Illumina (Illumina Inc., San Diego, CA) and custom AxiomGT1 Affymetrix (Thermo Fisher Scientific, Santa Clara, CA) genome-wide arrays and imputed to 16,387,711 (INFO > 0.6) variants using a population-specific SISu v.3 imputation reference panel (available at: https://www.protocols.io/view/genotype-imputation-workflow-v3-0-e6nvw78dlmkj/v2). UK Biobank samples (v3 genotyping release) were genotyped on custom Affymetrix microarrays and imputed using the 1000 Genomes and the Haplotype Reference Consortium reference panels to ~93M variants. Detailed procedures on genotype calling, quality controls and imputation have been previously described for FinnGen (Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 613, 508–518 (2023)) and UK Biobank (Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018)).
Data analysis	All the data processing, visualization, and statistical analyses were performed using Python 3.8 (2.7 for LDSC) and R v.4.3.2 (R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org/). Venn diagrams were created using the R package ggvenn (version 0.1.10; https://cran.r-project.org/web/packages/ggvenn/index.html). Correlation plots were created using the R package corrplot (v.0.92; https://cran.r-project.org/web/packages/corrplot/index.html). Forest plots were created using the R package ggforestplot (v.0.1.0; https://nightingalehealth.github.io/ggforestplot/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Individual-level data cannot be stored in public repositories or otherwise made publicly available due to ethical and data protection restrictions. However, data are available upon request for researchers who meet the criteria for access to confidential data. Data from the UK Biobank are available to bona fide researchers upon application at https://www.ukbiobank.ac.uk/enable-your-research. The following UK Biobank–associated data were accessed through, and as part of, our UK Biobank accession: Hospital Episode Statistics for England (HES), Scottish Morbidity Record (SMR), and Patient Episode Database for Wales (PEDW).
FinnGen results, according to FinnGen consortium agreement, are subjected to one year embargo and summary statistics are then made available to the scientific community and release two times a year. Information on accessing FinnGen data can be found at https://www.finngen.fi/en/access_results. The national healthcare, population and cause of death registers were accessed through, and as part of, our FinnGen accession, implemented in the FinnGen pipelines.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Our data and results pertain to sex as a biological attribute. Information on sex was obtained from register and genotype data. Individual-level data on sex are not shared as part of this publication; details on how to access such data in FinnGen and UK Biobank are presented in the manuscript under "Data availability" . In FinnGen, 282,202 (56.4%) participants were females, whereas in the UK Biobank, 232,380 (54.1%) of the participants were females. The main analysis was adjusted for sex, whereas the PRS replication of the main findings was performed additionally stratified by sex. Sex-specific estimates are also provided for the PRS associations on mortality and hospitalizations.
Reporting on race, ethnicity, or other socially relevant groupings	The analysis samples included individuals of Finnish ancestry in FinnGen and white British individuals in the UK Biobank. No socially constructed variables were used in this study. To control for confounding, the models were adjusted for birth year, birth region, sex and the 10 first principal components.
Population characteristics	FinnGenn: - mean age (SD) was 53.1 (17.9) at baseline and 60.8 (18.0) at the end of follow-up/death - BMI was 27.35 (5.53) at baseline - 156,355 (50.9) were non-smokers at baseline (information missing in 211,792 individuals) - median Hospital Frailty Risk Score (IQR) was 5.2 (1.6–10.4) at baseline - 62,764 (12.1%) of the participants died during the follow-up - median number (IQR) of hospitalizations was 8 (4–17) UK Biobank: - mean age (SD) was 53.1 (17.9) at baseline and 60.8 (18.0) at the end of follow-up/death - BMI (SD) was 27.41 (4.76) at baseline - 232,968 (54.2) were non-smokers at baseline (information missing in 1,471 individuals) - median Hospital Frailty Risk Score (IQR) was 1.5 (0–5) at baseline - 38,636 (9.0%) of the participants died during the follow-up - median number (IQR) of hospitalizations was 1 (0–3)
Recruitment	The FinnGen participants were recruited through the following mechanisms: 1) legacy samples collected by the National Institute of Health and Welfare (THL) and 2) prospective samples that were mainly collected by Finnish hospital biobanks. The UK Biobank participants were recruited through 22 assessment centers across England, Scotland and Wales. In the UK Biobank, a healthy selection bias may exist due to volunteer-based participation, whereas in FinnGen that consists of national cohorts and biobank samples of hospitalized individuals, some diseases may be overrepresented compared to general population.
Ethics oversight	FinnGen: Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Finnish Medicines Agency (Fimea), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017. The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020,,THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020,

THL/2658/14.06.00/2021, THL/4235/14.06.00/202, Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20) TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4th July 2019.
The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 9 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, Auria Biobank AB17-5154 and amendment #1 (August 17 2020), AB20-5926 and amendment #1 (April 23 2020) and it´s modification (Sep 22 2021), Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020.

UK Biobank:
The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee (approval number: 11/NW/03820). All participants provided written informed consent for data collection, analysis, and record linkage. We have also obtained an ethical approval for the use of UK Biobank data in Sweden (Dnr 2016/1888-31/1).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We included all available Finngen (N=500,737) and UK Biobank (N=407,463) participants in the latest data releases that met our inclusion criteria.
Data exclusions	Individuals with missing minimum genotype data or ambiguous sex information based on genotype and register data were excluded. Additionally, individuals with a non-Finnish ancestry were excluded in FinnGen, whereas non-white British participants in were excluded in the UK Biobank.
Replication	The GWAS results in FinnGen (discovery sample) were replicated in the UK Biobank (replication sample) at the individual variant level and using polygenic risk scores. Proteomics integration for a funtional follow-up of the GWAS results was performed in the UK Biobank.
Randomization	Our study did not involve allocation into experimental groups, as it was an observational, hypothesis-free GWAS treating the Hospital Frailty Risk Score (HFRS) as a continuous outcome. Therefore, no experimental manipulation or group assignment was performed. To control for potential confounding, we adjusted the linear regression models for key covariates: birth year, sex, and the first ten genetic principal components, which account for population structure and ancestry differences. This is a standard approach in large-scale genetic association studies to ensure that associations between genetic variants and the outcome are not driven by non-genetic confounders.
Blinding	In a hypothesis-free GWAS, blinding is not possible/necessary because the analysis is fully automated and applies standardized statistical tests uniformly across the genome. There is no prior hypothesis guiding the selection of variants, reducing the potential for investigator bias. The role of the analyst is limited to managing predefined workflows, with no influence on the outcome. As such, the objectivity of the process is maintained without the need for blinding.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*

Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☐ Antibodies
☐	☐ Eukaryotic cell lines
☐	☐ Palaeontology and archaeology
☐	☐ Animals and other organisms
☐	☐ Clinical data
☐	☐ Dual use research of concern
☐	☐ Plants

Methods

n/a	Involved in the study
☐	☐ ChIP-seq
☐	☐ Flow cytometry
☐	☐ MRI-based neuroimaging

Antibodies

Antibodies used — *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation — *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s) — *State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*

Authentication — *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination — *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See ICLAC register) — *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance — *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*

Specimen deposition — *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods — *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals — *For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*

Wild animals — *Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Reporting on sex — *Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall*

numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.

Field-collected samples	For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.
Ethics oversight	Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.
Study protocol	Note where the full trial protocol can be accessed OR if not available, explain why.
Data collection	Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.
Outcomes	Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks — *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes — *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication — *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication. — *For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*

Files in database submission — *Provide a list of all files available in the database submission.*

Genome browser session
(e.g. UCSC) — *Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates — *Describe the experimental replicates, specifying number, type and replicate agreement.*

Sequencing depth — *Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*

Antibodies — *Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Peak calling parameters — *Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality — *Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*

Software — *Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation — *Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*

Instrument — *Identify the instrument used for data collection, specifying make and model number.*

Software — *Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See Eklund et al. 2016)

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.