

# Uncovering the multivariate genetic architecture of frailty with genomic structural equation modelling

Isabelle F Foote<sup>1,2</sup>, Jonny P Flint<sup>3,4,5,6,7</sup>, Anna E Fürtjes<sup>3,4</sup>, Donncha S Mullin<sup>7,8</sup>, John D Fisk<sup>9,10</sup>, Tobias K Karakach<sup>11</sup>, Andrew Rutenberg<sup>12</sup>, Nicholas G Martin<sup>6</sup>, Michelle K Lupton<sup>6,13,14</sup>, David J Llewellyn<sup>15,16</sup>, Janice M Ranson<sup>15</sup>, Simon R Cox<sup>3,4</sup>, Michelle Luciano<sup>3,4</sup>, Kenneth Rockwood<sup>9,10</sup>, Andrew D Grotzinger<sup>1,17</sup>

## Author Affiliations:

<sup>1</sup> Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, Colorado, USA

<sup>2</sup> Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary University of London, London, UK

<sup>3</sup> Lothian Birth Cohorts, University of Edinburgh, Edinburgh, UK

<sup>4</sup> Department of Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

<sup>5</sup> Advanced Care Research Centre School of Engineering, College of Science and Engineering, University of Edinburgh, Edinburgh, UK

<sup>6</sup> Brain and Mental Health Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

<sup>7</sup> Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK

<sup>8</sup> Centre for Clinical Brain Sciences, Division of Psychiatry, University of Edinburgh, Edinburgh, UK

<sup>9</sup> Nova Scotia Health, Halifax, Nova Scotia, Canada

<sup>10</sup> Division of Geriatric Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>11</sup> Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>12</sup> Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>13</sup> School of Biomedical Sciences, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia

<sup>14</sup> School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Queensland, Australia

<sup>15</sup> University of Exeter Medical School, University of Exeter, Exeter, UK

<sup>16</sup> Alan Turing Institute, London, UK

<sup>17</sup> Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, Colorado, USA

**Corresponding author:** Isabelle F Foote

**Full address:** Institute for Behavioral Genetics, University of Colorado Boulder, 1480 30th St, Boulder, CO 80303, USA

**E-mail:** [isabelle.foote@colorado.edu](mailto:isabelle.foote@colorado.edu)

**Running title:** Multivariate GWAS of frailty

**Keywords:** Frailty, multivariate genomics, aging, Genomic SEM

## Abstract

Frailty is a multifaceted clinical state associated with accelerated aging and adverse health outcomes. Informed etiological models of frailty hold promise for producing widespread health improvements across the aging population. Frailty is currently measured using aggregate scores, which obscure etiological pathways that are only relevant to subcomponents of frailty. Therefore, we performed the first multivariate genome-wide association study of the latent genetic architecture between 30 frailty deficits, which identified 408 genomic risk loci. Our model included a general factor of genetic overlap across all deficits, plus six novel factors indexing shared genetic signal across specific groups of deficits. Follow-up analyses demonstrated the added clinical and etiological value of the six factors, including predicting frailty in external datasets, divergent genetic correlations with clinically relevant outcomes, and unique underlying biology linked to aging. This suggests nuanced models of frailty are key to understanding its causes and how it relates to worse health.

Frailty is a complex clinical state that affects more than 40% of adults aged over 65 years(1). It is defined as a state of progressive, multisystem physiological decline that reduces an individual's ability to withstand external stressors(1). This deterioration can lead to both physical and mental impairment and is strongly associated with adverse health outcomes, including earlier mortality(2) and increased levels of disability and hospitalization(3). Global population aging means that frailty represents a growing public health concern(4). Family-based studies indicate a substantial genetic component to frailty, with heritability estimates of ~45%(5). Therefore, genetic methods offer a promising tool for better understanding the risk pathways to this critical health state. Nevertheless, the etiology of frailty remains largely unknown, limiting our potential to identify effective therapeutic or preventive treatments. We hypothesized that what has limited our etiological understanding is that we have traditionally only considered aggregate measurements of frailty in genetic studies.

The two most common methods for measuring frailty are the Frailty Index (FI) and Fried Frailty Score (FFS)(6, 7). The FI quantifies frailty by calculating the proportion of 'deficits' that are present within an individual from a set of  $\geq 30$  phenotypes associated with poor health outcomes in older adults(6). The FFS uses an aggregate score across five physical frailty deficits (weight loss, weakness, exhaustion, slow walking speed and physical inactivity), where the presence of  $\geq 3$  of these deficits indicates frailty(7). Although the FFS is easier to measure in large samples, it is intended to capture pathways of physical frailty and does not provide sufficient information to assess the more nuanced subgroupings that may occur within the broader frailty construct(8). By contrast, the deficits that form the FI span many levels of functional, psychological, and social aspects of health, allowing frailty to be measured across a broad spectrum of traits. Yet, FI deficits are heterogeneous and vary in their underlying etiology.

Phenotypic work clearly indicates that the deficits comprising the FI are not always strongly correlated and are driven by diverse biological mechanisms(9-11). Their combination into a single aggregate score is, therefore, likely to obscure causal pathways of frailty. For example, previous work that applied principal component analysis to phenotypic data of FI deficits demonstrated that additional informative variance associated with frailty was captured when three clusters were modelled instead of one cluster(12). Information that is lost by aggregation could be identified by a more detailed genetic analysis. Recent advances in multivariate genomics, such as the development of Genomic Structural Equation Modelling (Genomic SEM)(13), offer the opportunity to model the genetic basis of frailty at a multi-dimensional level.

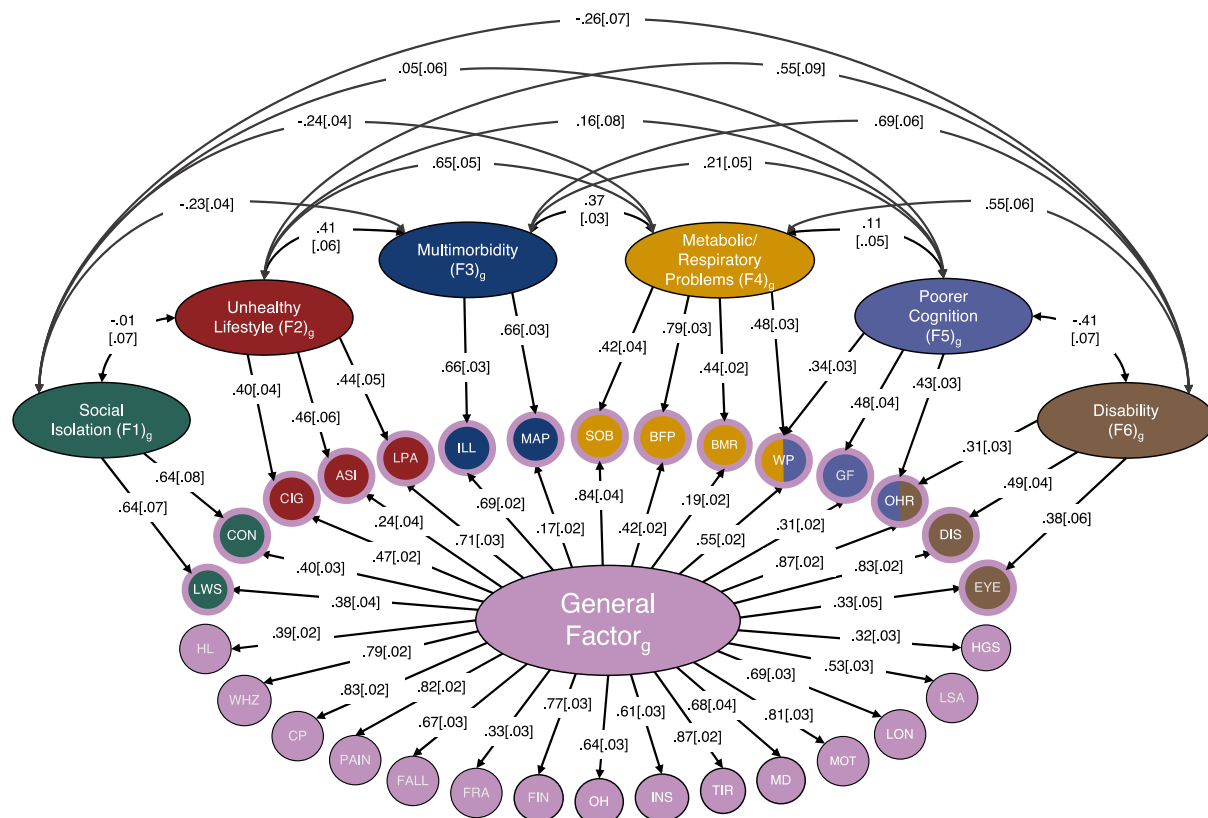
We used Genomic SEM to identify novel groupings of genetic overlap between 30 frailty deficits using publicly available GWAS summary statistics. We identified seven distinct latent factors underpinning frailty that displayed unique genetic overlap with clinically relevant health outcomes and were defined by divergent sets of genomic risk loci and biological pathways, which provided empirical support for a majority of the 12 'hallmarks of aging'(14). In combination, these latent constructs yielded enhanced prediction of frailty status in external cohorts and a dramatic improvement in genomic locus discovery compared to aggregate measures.

## Results

### Multivariate Genetic Architecture of Frailty

Following careful quality control and selection of frailty deficit phenotypes (**Online Methods**), we used Genomic SEM to model the genome-wide genetic overlap from GWAS data for 30 frailty deficits(15-19) (**Figure 1** and **Tables S1-S2**). As Genomic SEM can model genetic overlap across even mutually exclusive participant samples, this allowed us to bring together the most well-powered genomic studies currently available for the 30 frailty deficits and produce estimates with the highest possible precision. Using a combination of exploratory and confirmatory factor analysis (**Online Methods**), this modeling procedure yielded a bifactor model that provided a good fit to the data (**Figure 1** and **Table S3**; comparative fit index (CFI) = 0.93; standardized root mean squared residual (SRMR) = 0.07). This model included a General Factor that indexed genetic overlap across all 30 deficits and so is conceptually similar to prior aggregate measures of frailty. In addition, the model produced six residual group factors that were orthogonal (i.e. uncorrelated) to the General Factor, which were defined by additional genetic overlap within subsets of frailty deficits. These six factors captured distinct, albeit

inter-correlated, frailty pathways related to social isolation (Factor 1), unhealthy lifestyle (Factor 2), multimorbidity (Factor 3), respiratory/metabolic problems (Factor 4), poorer cognition (Factor 5), and disability (Factor 6). The remaining analyses evaluated the divergent validity and clinical utility of these six factors by examining their ability to uniquely capture frailty-relevant pathways at increasingly granular levels of biological analysis.



**Figure 1: A path diagram of the standardized results for our bifactor model of frailty.** All 30 frailty deficits load onto the General Factor of frailty (large pink oval), which is orthogonal to Factors 1 to 6 (i.e. residual factors). The small circles represent the 30 measured frailty indicators (i.e. genetic variance captured in the univariate GWAS for that phenotype), whereas the medium-size ovals represent latent factors (i.e. unmeasured constructs representing genetic overlap between the indicators that load onto them). Single-headed arrows represent a directional genetic correlation between a latent factor and an indicator (i.e. factor loadings), whereas curved double-headed arrows represent inter-factor correlations between Factors 1 to 6. Standard errors of the correlation coefficients are reported in square brackets. SOB = shortness of breath walking on flat ground; BFP = body fat percentage; BMR = basal metabolic rate; WP = slow walking pace; GF = low fluid intelligence score; OHR = poorer overall health rating; DIS = long-standing illness, disability or infirmity; EYE = eye disorder/problem; HGS = low hand grip strength; LSA = low social/leisure activity; LON = loneliness, isolation; MOT = feelings of unenthusiasm/disinterest; MD = major depressive disorder; TIR = tiredness/lethargy; INS = insomnia; OH = poor oral health; FIN = financial difficulties; FRA = fracture in last 5 years; FALL = number of falls in past year; PAIN = pain experienced in past month; CP = chest pain; WHZ = wheeze/whistling in chest in past year; HL = age-related hearing loss; LWS = not living with spouse/partner; CON = unable to confide; CIG = number of cigarettes smoked per day; ASI = pulse wave arterial stiffness index; LPA = physical inactivity; ILL = number of non-cancer illnesses; MAP = mean arterial pressure.

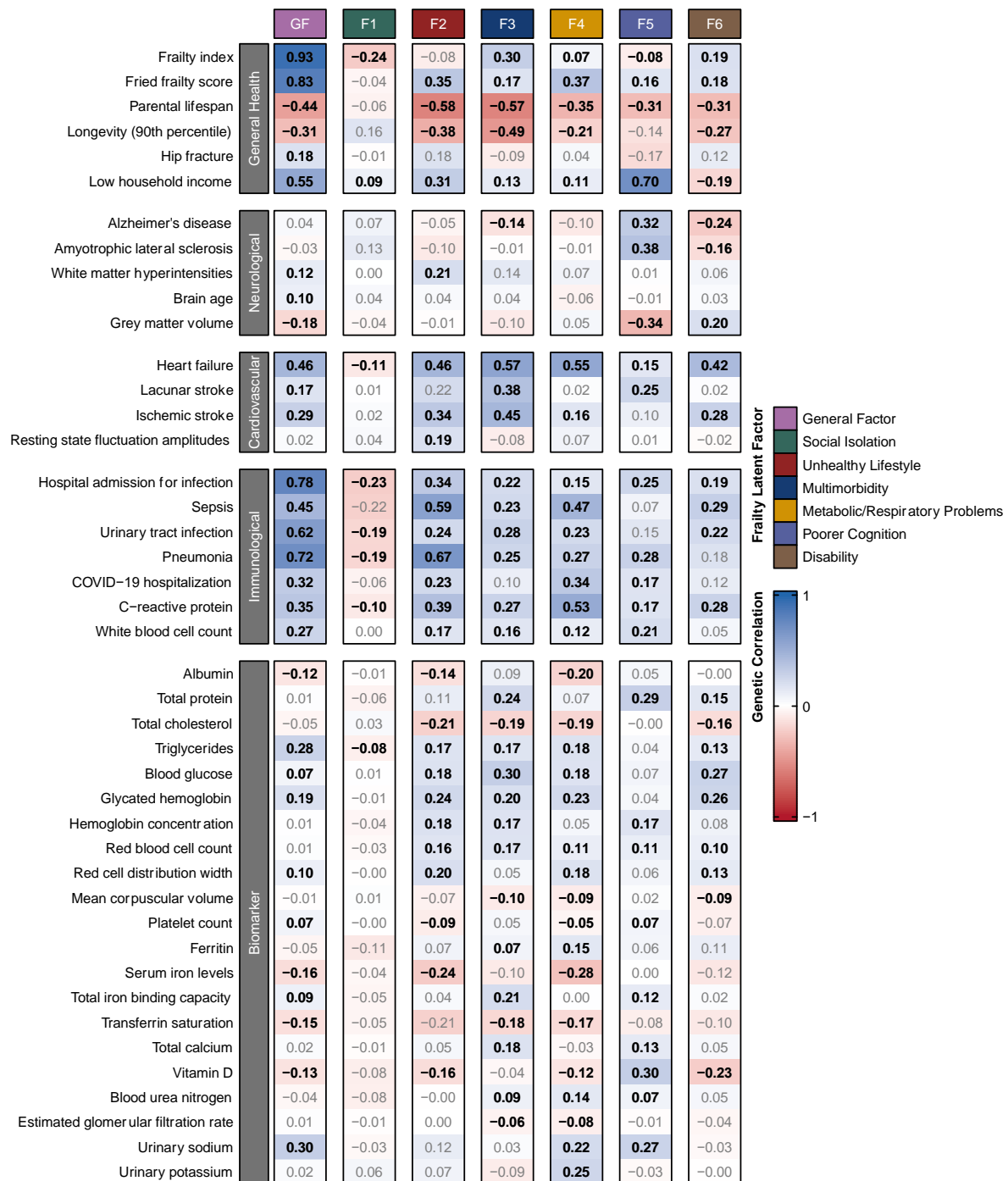
## Genetic Overlap between Latent Frailty Factors and Aging-Related Health Traits

To validate whether the latent factors in the bifactor model reflect different aspects of frailty, we measured the level of genetic correlation between each of the latent factors and 52 aging-related health outcomes and established frailty measures (**Figure 2** and **Tables S4-S5**). The General Factor displayed high positive genetic correlations with the FI ( $r_g = 0.93$  [SE = 0.02],  $P_{FDR} = 9.1 \times 10^{-298}$ ) and FFS ( $r_g = 0.83$  [0.02],  $P_{FDR} = 9.1 \times 10^{-298}$ ), indicating that the General Factor closely approximates both of these frailty phenotypes in line with expectations, given that this latent factor represents genetic overlap between all 30 included deficits. These findings were corroborated by the fact that all the latent factors (except Factor 1 (social isolation)) were associated with shorter parental lifespan ( $r_g < -0.31$ ,  $P_{FDR} < 3.2 \times 10^{-6}$ ), reduced longevity ( $r_g < -0.21$ ,  $P_{FDR} < 0.01$ ), increased risk of common aging-related infections ( $r_g > 0.22$ ,  $P_{FDR} < 0.01$ ), hospitalization due to infection ( $r_g > 0.15$ ,  $P_{FDR} < 0.008$ ), and heart failure ( $r_g > 0.15$ ,  $P_{FDR} < 0.005$ ), all of which reflect key clinical correlates of frailty.

Importantly, our findings support the inclusion of additional subgroups when measuring frailty accumulation by highlighting divergent genetic correlations with aging-related health outcomes. For example, Factor 5 (poorer cognition), was the only latent factor to demonstrate positive genetic correlations with Alzheimer's disease ( $r_g = 0.32$  [SE = 0.07],  $P_{FDR} = 2.26 \times 10^{-05}$ ) and amyotrophic lateral sclerosis ( $r_g = 0.38$  [0.06],  $P_{FDR} = 1.07 \times 10^{-08}$ ). Factor 5 also captured genetic variance associated with smaller grey matter volume ( $r_g = -0.34$  [0.05],  $P_{FDR} = 9.53 \times 10^{-12}$ ) and lacunar stroke ( $r_g = 0.25$  [0.09],  $P_{FDR} = 9.80 \times 10^{-03}$ ). By contrast, Factor 2 (unhealthy lifestyle) demonstrated genetic correlations with brain-related vascular changes, including increased white matter hyperintensities ( $r_g = 0.21$  [0.09],  $P_{FDR} = 0.04$ ) and resting state fluctuation amplitudes ( $r_g = 0.19$  [0.08],  $P_{FDR} = 0.03$ ), whereas Factor 3 (multimorbidity) was correlated with ischemic and lacunar stroke ( $r_g = 0.45$  [0.05],  $P_{FDR} = 6.21 \times 10^{-18}$  and  $r_g = 0.38$  [0.09],  $P_{FDR} = 3.00 \times 10^{-03}$ ), but not cerebrovascular markers. Finally, all the latent frailty factors displayed distinct patterns of genetic correlations with routinely collected blood and urinary biomarkers (**Figure 2**), which may represent potential endophenotype profiles for these frailty subgroups. By definition, the identified relationships are independent of shared risk pathways captured by the General Factor of frailty. It follows that our model of frailty has potential clinical utility for targeted prevention and therapeutic intervention in patients that present with elevated risk within a subgroup of frailty deficits.

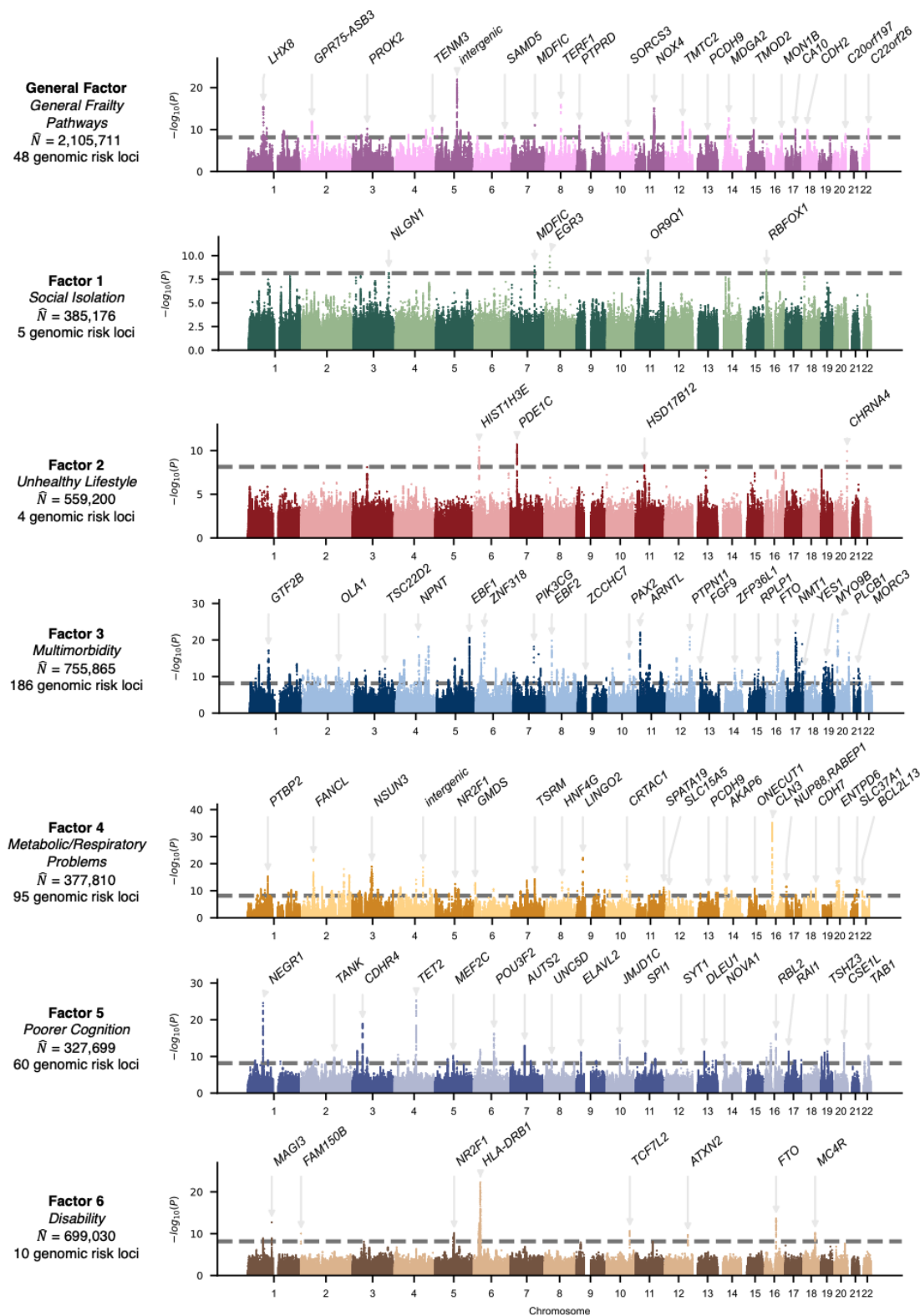
## Multivariate GWAS Identifies 408 Genomic Risk Loci Associated with Frailty

We subsequently performed a multivariate GWAS of our frailty bifactor model using Genomic SEM to uncover genomic risk loci that were associated with each latent frailty factor ( $P_{BONF} < 7.14 \times 10^{-09}$ ). We pruned out any significantly heterogeneous genetic signal ( $Q_{SNP}$ ) from our GWAS results to ensure that we were only measuring genetic effects that were shared between the deficits that defined that particular latent factor (**Online Methods**). From this shared signal, we identified a total of 408 genomic risk loci across the seven latent frailty factors (**Figure 3** and **Tables S6-12**). We integrated the results from our GWAS with the lead single nucleotide polymorphisms (SNPs) of the risk loci identified in the previously published GWAS of the FI(20) (**Table S13**) and FFS(21) (**Table S14**). Of the overlapping lead SNPs (12/14 for the FI and 18/37 for the FFS), we replicated 33.3% of the previously identified genomic risk loci for the FI and 22.2% for the FFS with the General Factor. 100% of the risk loci for both frailty phenotypes displayed nominal significance with at least one of our latent frailty factors, supporting validation of our model in measuring frailty. However, 33.3% of the FI loci and 25% of the FFS loci were significant for the SNP-level heterogeneity metric ( $Q_{SNP} P < 7.14 \times 10^{-09}$ ; **Tables S13-14**). As  $Q_{SNP}$  identifies SNPs that are likely to be deficit-specific or that show heterogeneous effects between the included frailty deficits, these findings indicate that existing aggregate frailty measures are sensitive to large effects driven by a single indicator that are not part of more general frailty pathways. Furthermore, 33.3% and 27.8% of the FI and FFS loci were replicated by the residual factors (i.e. Factors 1 to 6) instead of the General Factor. This demonstrates that our model can comprehensively define subgroups within the frailty construct, which one aggregate measure overlooks.



**Figure 2: A heatmap of the genetic correlations between aging-related health outcomes and each of the latent factors from the frailty bifactor model.** Genetic correlations with an FDR-corrected  $P$ -value  $<0.05$  are highlighted in bold. Blue shading represents a positive genetic correlation whereas red shading represents a negative genetic correlation. For visualization purposes, only health outcomes that demonstrated at least one FDR-corrected  $P$ -value  $<0.05$  with one or more of the latent factors are included in this figure (full results can be found in Table S5).

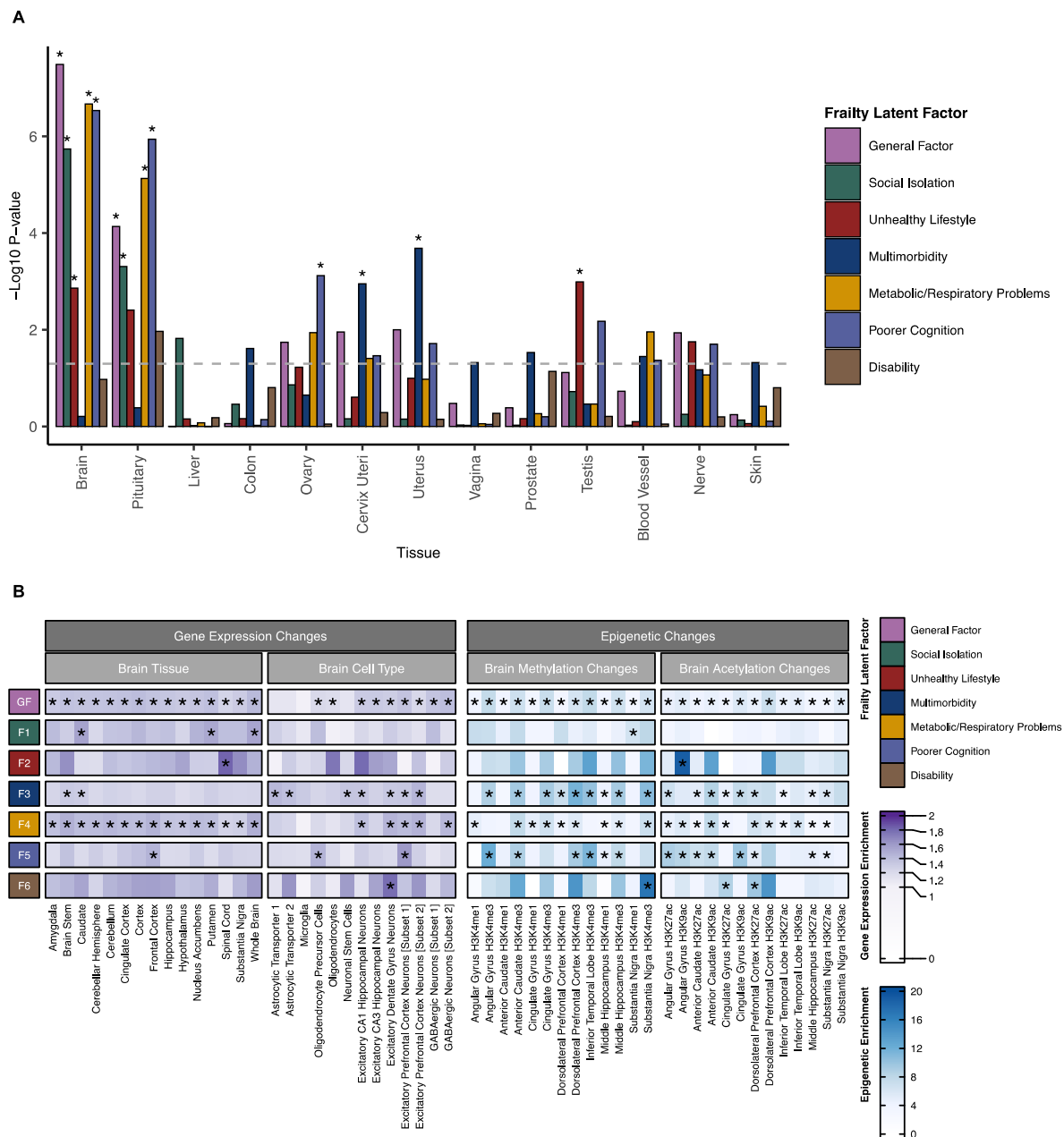




**Figure 3: Manhattan plots of the shared genetic signal for each of the latent factors in the frailty model.** The x-axis depicts the chromosomes and the y-axis represents the  $-\log_{10} P$ -values of the association between each individual SNP and each latent factor. The closest gene to the lead SNP are annotated for top loci for each latent factor. The dashed grey line denotes the genome-wide significance threshold adjusted for multiple testing (i.e.  $P_{\text{BONF}} < 7.14 \times 10^{-9}$ ).  $\tilde{N}$  is the expected sample size of each latent factor implied by the GWAS summary statistics of that factor, which are influenced by the power of the factor loadings of the indicators (i.e. frailty deficits) that define it.

## Gene Expression and Epigenetic Changes Associated with Latent Frailty Factors

We completed a series of post-GWAS analyses to explore how the genetics underpinning each of the frailty latent factors may influence underlying biology. We applied Multi-marker Analysis of GenoMic Annotation (MAGMA) gene property analysis to test for enriched gene expression changes in 54 body tissues and Stratified Genomic SEM to test for enrichment in 146 functional annotations linked to gene expression and epigenetic changes in tissues and cell subtypes within the brain (Figure 4 and Online Methods).



**Figure 4: Results from the MAGMA gene property analysis (Panel A) and the Stratified Genomic SEM (Panel B).** In Panel A, the y-axis denotes the  $-\log_{10} P$ -value of the enrichment between each latent frailty factor and body tissues from GTEx version 8 (only tissues with significant enrichment are displayed). The light grey dashed line denotes the cut-off for nominal significance (i.e.  $P$ -value  $< 0.05$ ) and bars that are marked with an asterisk indicate enrichment that remained significant after adjusting for multiple testing (i.e.  $P_{FDR} < 0.05$ ). Panel B displays heatmaps of the enrichment values calculated using Stratified Genomic SEM to test for differences in gene expression and epigenetic marks associated with each latent frailty factor in a selection of brain-relevant tissues and cell types. Significant enrichment values are marked with an asterisk ( $P_{FDR} < 0.05$ ). GF = General Factor; F1 = Factor 1; F2 = Factor 2; F3 = Factor 3; F4 = Factor 4; F5 = Factor 5; F6 = Factor 6).



MAGMA gene property analysis showed significant enrichment in the brain and pituitary gland for all the latent frailty factors except for those underlying multimorbidity (Factor 3) and disability (Factor 6) pathways (**Figure 4A**). The only other tissues that demonstrated significantly enriched changes in gene expression were the reproductive organs, whereby Factor 2 (unhealthy lifestyle) showed significant gene expression changes in the testis, Factor 3 (multimorbidity) showed significant changes in gene expression in the cervix and uterus, and Factor 5 (poorer cognition) showed significant gene expression changes within the ovaries (**Table S15**).

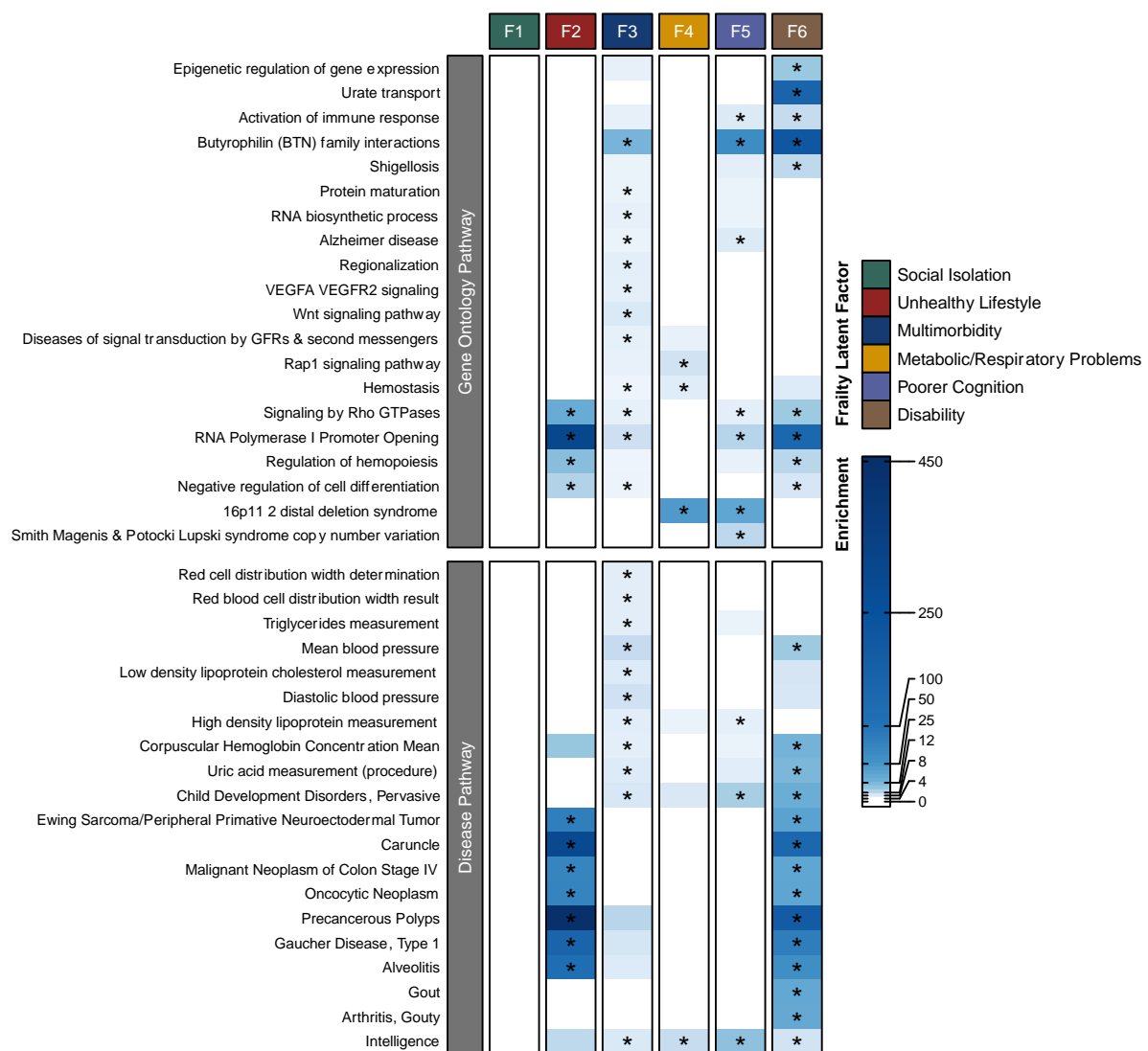
Stratified Genomic SEM provided a more in-depth picture of enrichment in the brain (**Table S16**). We found widespread enrichment in gene expression and epigenetic changes throughout the brain regions in oligodendrocytes and neurons for the General Factor of frailty. However, the widespread nature of this enrichment demonstrates that using an aggregate measure of frailty is less likely to provide a fine-tuned picture of the underlying mechanisms of frailty due to its generalized impact on brain function. In contrast, the residual factors provided a more detailed understanding of pathways implicated by different frailty deficits, which could present future therapeutic targets within the broad spectrum of frailty (**Figure 4B**). For instance, Factor 1 (social isolation) only showed significant gene expression changes in the dorsal striatum (caudate and putamen) and methylation changes in the substantia nigra, whereas Factor 2 (unhealthy lifestyle) showed enriched gene expression in the spinal cord, but not any of the tested brain regions. In addition, Factor 5 (poorer cognition) showed gene expression enrichment in excitatory prefrontal cortex neurons and oligodendrocyte precursor cells, as well as epigenetic changes in the angular gyrus, cingulate gyrus, anterior caudate, dorsolateral prefrontal cortex, hippocampus and substantia nigra.

### Gene Prioritization and Pathway Analysis

We used five methods to map potentially causal genes to each latent frailty factor to assess the biological pathways that might be associated with each frailty subgroup. These methods included mapping SNPs to genes based on their position, whether they were known expression quantitative trait loci (eQTLs) or if they were located in promoter regions known to regulate chromatin interactions (**Tables S17-S23**). In addition, we performed a genome-wide gene-based test using MAGMA (**Tables S24-S30**) and applied summary data-based Mendelian randomization (SMR) to identify SNPs that demonstrated evidence of having a pleiotropic effect on expression, splicing or methylation changes in gene function (**Tables S31-S37**). We triangulated the results from these five gene mapping techniques and prioritized the most likely candidate genes based on whether they were mapped by three or more of the methods. This resulted in 1195 genes being prioritized, which we took forward for pathway analysis (54 for the General Factor; 4 for Factor 1 (social isolation); 20 for Factor 2 (unhealthy lifestyle); 585 for Factor 3 (multimorbidity); 194 for Factor 4 (metabolic/respiratory problems); 266 for Factor 5 (poorer cognition) and 72 for Factor 6 (disability)) (**Supplementary Results**). Using METASCAPE we performed enrichment analysis to identify gene ontology pathways and disease pathways that were significantly associated with the prioritized genes mapped to each latent factor(22). As there can be extensive redundancy between gene sets, we combined highly correlated enriched pathways into clusters, named according to the gene ontology pathway that had the strongest enrichment with the latent frailty factors (**Online Methods** and **Figure 5**). Since the General Factor was orthogonal to the other latent frailty factors, we conducted pathway analysis separately for that factor, but performed a combined analysis for Factors 1 to 6 to account for the potential overlap in implicated gene pathways owing to the presence of inter-factor correlations between these latent residual factors.

Pathway analysis of the prioritized genes for the General Factor identified only two significantly enriched disease pathways for intelligence and scoliosis (**Tables S38-40**). In contrast, we found high levels of significant enrichment (i.e.  $P_{FDR} < 0.05$ ) for all residual latent frailty factors, except Factor 1 (**Figure 5** and **Tables S41-43**). The most strongly enriched pathway cluster (lead gene ontology term = RNA polymerase I promoter opening) included multiple gene sets linked to known aging-related processes including telomere function, amyloid fiber formation and oxidative stress, and was significantly enriched across Factors 2 (unhealthy lifestyle), 3 (multimorbidity), 5 (poorer cognition), and 6 (disability). Other pervasive cross-factor enrichment implicated immune function, epigenetic regulation, and cancer as key pathways involved in frailty pathogenesis. In addition to the shared enrichment in aging-related pathways observed across the factors, we also found evidence for the discriminant validity of the frailty factors. For example, only Factors 3 (multimorbidity) and 5 (poorer cognition) demonstrated significant enrichment for pathways linked to Alzheimer's disease and general neurodegeneration. Factor 3 (multimorbidity) was also enriched in protein maturation and folding pathways, providing consistent evidence that aspects of frailty related to multimorbidity and cognition may be more highly linked to

dementia and neurodegenerative pathways compared to other aspects of frailty. Factor 4 (metabolic/respiratory problems) genes were enriched in gene sets linked to cell signaling (particularly Rap1 pathways) and 16p11.2 distal deletion syndrome. This is a rare syndrome that results from the partial deletion of the short arm of chromosome 16 leading to symptoms including intellectual disability, developmental delay and autism spectrum disorder. This syndrome can be caused by unmasked recessive mutations in the *CLN3* gene(23), which was where the most significant risk locus for our frailty GWAS was located (lead SNP = rs27741; Factor 4  $P$ -value =  $1.09 \times 10^{-35}$ ) (**Figure 3**). Enrichment analysis of disease pathways from the DisGeNet database further demonstrated that the frailty factors display distinct underlying biology. Factor 3 (multimorbidity) genes were strongly enriched in pathways linked to red blood cell and lipid biomarkers, whereas Factor 2 (unhealthy lifestyle) and 6 (disability) genes were significantly enriched in cancer pathways, and additionally, in gout and arthritis pathways for Factor 6.



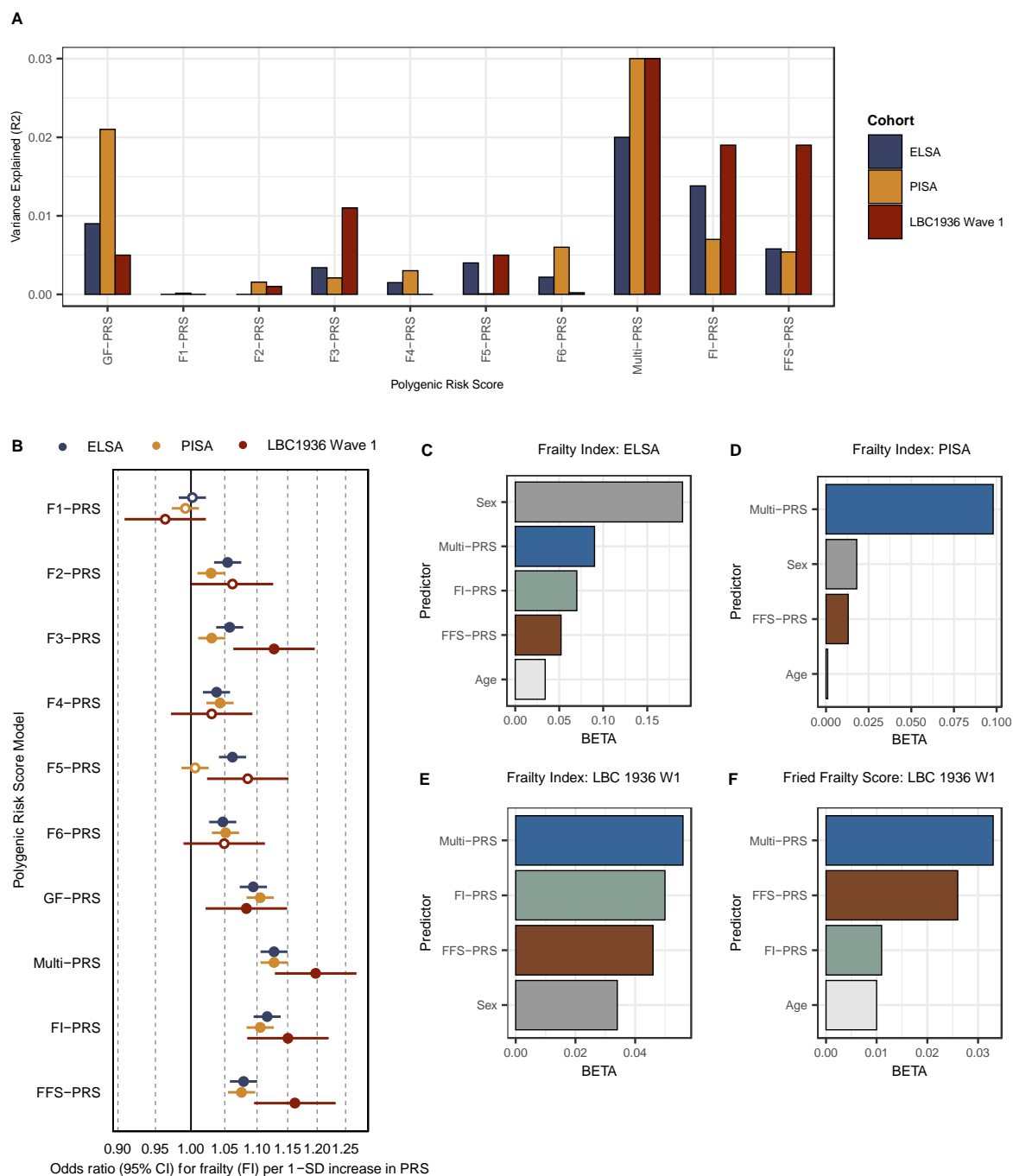
**Figure 5: Results from pathway enrichment analysis of the prioritized genes for the residual frailty factors (Factors 1-6).** The upper heatmap shows the enrichment for the top 20 enriched gene ontology pathway clusters. The displayed values represent the results for the most significant gene ontology term in the cluster (as named on the y-axis). The lower heatmap displays the enrichment for the top 20 most significantly enriched disease pathways from the DisGeNet database. There were no significantly enriched pathways for Factor 1 (social isolation) as only 4 genes (*CTNND1*, *TMX2*, *MED19* and *EGR3*) were mapped to that latent factor. Significant enrichment values are marked with an asterisk ( $P_{FDR} < 0.05$ ). GFRs = Growth Factor Receptors; VEGFA/R2 = vascular endothelial growth factor A/receptor 2; F1 = Factor 1; F2 = Factor 2; F3 = Factor 3; F4 = Factor 4; F5 = Factor 5; F6 = Factor 6.

## Polygenic Risk Scores of Frailty Factors Predict Frailty and Health in External Cohorts

To validate the latent frailty factors as phenotypes that capture frailty-specific variance, we created polygenic risk scores (PRSs) for each latent frailty factor and used regression models to test how well they predicted frailty and frailty-related outcomes in three external older adult cohorts (the Lothian Birth Cohort 1936 [LBC1936] ( $N = 1,005$ ; mean age = 69.60), the English Longitudinal Study of Aging [ELSA] ( $N = 7,181$ , mean age = 68.45) and the Prospective Imaging Study of Aging [PISA] ( $N = 3,265$ , mean age = 60.34)) (**Online Methods** and **Supplementary Methods**). To measure the cumulative predictive capacity of our frailty model, we also created a PRS phenotype that combined the polygenic signal of all seven frailty factors using multiple regression (herein referred to as Multi-PRS). This allowed us to compare the performance of our overarching multivariate model in predicting frailty status relative to PRSs created from existing aggregate frailty GWAS measures (i.e. the FI-PRS(20) and FFS-PRS(21)). The combined Multi-PRS provided the strongest prediction of the FI across all three tested cohorts (**Table S44** and **Figure 6A-B**).

We additionally measured the association of the latent frailty PRSs with other frailty-related health outcomes, which helped to provide a more detailed picture of how these frailty subgroupings may differentially impact aging processes (**Table S45** and **Supplementary Methods**). We found that the PRS for Factor 5 (poorer cognition) and the Multi-PRS significantly predicted lower cognitive ability in LBC1936 ( $\beta = -0.65$ ,  $SE = 0.10$ ,  $P_{FDR} = 1.43 \times 10^{-08}$  and  $\beta = -0.62$ ;  $SE = 0.10$ ,  $P_{FDR} = 3.34 \times 10^{-08}$ ), but not cognitive change. The Multi-PRS was also significantly associated with dementia ( $\beta = 0.2$ ,  $SE = 0.07$ ,  $P_{FDR} = 9.60 \times 10^{-03}$ ) and stroke ( $\beta = 0.18$ ,  $SE = 0.05$ ,  $P_{FDR} = 6.40 \times 10^{-04}$ ) in ELSA.

We used elastic net regression to jointly model the PRS's for the seven frailty factors so that we could rank the order that each contributed to predicting frailty status (**Supplementary Results** and **Table S46**). The General Factor of frailty was ranked as the highest contributor to FI prediction in ELSA and PISA. By contrast in LBC1936, the Factor 3 (multimorbidity) PRS was ranked highest in FI prediction while the General Factor PRS provided the most prediction of FFS status. We also used elastic net regression to rank the performance of the full latent frailty model (Multi-PRS) against the previously derived aggregate frailty GWAS measures (FI-PRS and FFS-PRS). We found that the Multi-PRS outperformed the FI-PRS and the FFS-PRS when predicting FI and FFS status in all tested datasets (**Figure 6C-F** and **Table S47**). Sensitivity analyses that grouped samples by age demonstrated that the predictive contribution of our Multi-PRS was stable across different age groups, whereas the contributions from the PRS of the individual latent frailty factors changed by age, suggesting that preventive or therapeutic interventions targeting specific frailty subgroups may be more successful if provided at specific points in the life course (**Supplementary Results**). Together, these findings validated our model as representing a novel genetic measure that captures frailty-relevant pathways, which explained more genetic variance than aggregate GWAS measures that have been used in the field so far. They also underline the importance of modeling subgroups within the frailty construct to better understand what drives frailty onset across the life course.



**Figure 6: Results from the polygenic risk score (PRS) analysis of the latent frailty factors.** *Panel A* displays a bar plot of the variance explained ( $R^2$ ) by each PRS that we estimated for prediction of the Frailty Index (FI) in each external cohort. *Panel B* depicts a forest plot of the odds ratios (95% Confidence Intervals [CI]) for each standard deviation (SD) increase in frailty (measured by the Frailty Index) associated with our frailty PRS phenotypes in each of the external cohorts. These values were calculated using linear regression models. Significant predictions (i.e.  $P_{FDR} < 0.05$ ) are depicted as filled in dots, whereas non-significant predictions are depicted as an empty circle. *Panels C-F* represent the results from the elastic net regression analyses, which rank the performance of our Multi-PRS (i.e. combined latent frailty factor score) when modelled with the aggregate Frailty Index PRS (FI-PRS), the aggregate Fried Frailty Score PRS (FFS-PRS), age, sex and ancestral principal components as covariates. Only the predictors that explain significant additional unique variance over and above the other covariates are included in these graphs. All analyses represent standardized results. LBC1936 = Lothian Birth Cohort 1936; PISA = Prospective Imaging Study of Aging; ELSA = English Longitudinal Study of Aging; F1-F6 = Factors 1-6; GF = General Factor; W1 = Wave 1.

## Discussion

We report the first genomic factor analysis of frailty. We introduce seven novel latent constructs of the shared genetics between 30 frailty deficits, including a General Factor of frailty and six additional residual factors representing genetic overlap between distinct subsets of frailty deficits related to social isolation, unhealthy lifestyle, multimorbidity, metabolic/respiratory problems, poorer cognition, and disability. We identify 408 genomic risk loci for these latent constructs that are enriched for pathways related to accelerated aging, including epigenetic modifications and immune regulation. This demonstrates a substantial advance in genomic locus discovery for frailty compared to prior GWAS of aggregate frailty measures, which only identified 14 genomic loci for the FI(20) and 37 genomic loci for the FFS(21). We further validate the latent constructs as being of relevance to frailty and related health outcomes at multiple levels of biology and in the prediction of frailty status in external data.

Our findings support previous phenotypic studies that highlight the merit, relative to single aggregate scores, of using data reduction methods to improve our understanding of frailty etiology (12, 24, 25). However, by taking a multivariate genomic approach we were able to integrate theoretical knowledge with biological evidence to better define the underlying pathways of frailty and to differentiate generalized pathogenic pathways from more nuanced pathways that are specific to a subset of deficits, both of which are fundamental to understanding this complex clinical construct. For example, our genetic correlation and pathway analyses implicate immune function and epigenetic modifications as being key drivers of frailty pathogenesis across multiple deficit groupings. This is in line with findings linking frailty and elevated CRP levels, red blood cell distribution width, and white blood cell count (26-29). Our frailty factors were also significantly genetically correlated with health complications associated with infection, including hospitalization and sepsis. The associations between frailty and common viral infections, such as pneumonia(30), COVID-19(31) and urinary tract infections(32) are well documented. Furthermore, our findings consistently demonstrated evidence for widespread epigenetic changes in frailty, supporting previous work suggesting that epigenetic biomarkers, such as epigenetic clocks(33) or epigenetic risk scores(34) could be effective predictors of frailty.

The seven frailty factors displayed discriminant validity across multiple levels of biological analysis, indicating that existing aggregate measures of frailty are likely to miss clinically relevant distinctions. For example, we find that the poorer cognition factor (Factor 5) was uniquely associated with dementia pathways, particularly Alzheimer's disease. In addition, our GWAS and gene prioritization findings implicated *SP11* as a key locus for Factor 5, which is a well-replicated Alzheimer's disease risk locus(35-37). Interestingly, Factor 5 had similar factor loadings from lower fluid intelligence and poor self-reported overall health rating, indicating that subjective health reports as well as cognitive testing could be indicative of subsequent heightened dementia risk in individuals who present with these frailty deficits. In fact, subjective cognitive decline has been widely supported as a potential early marker of cognitive impairment(38). The other loading onto this latent factor was slow walking pace, which is independently associated with heightened dementia risk(39). In addition, slow gait and subjective cognitive decline are used to measure motoric cognitive risk, a syndrome strongly associated with subsequent dementia(40).

Furthermore, our genetic correlation and PRS analyses found that the multimorbidity factor (defined by number of illnesses and high mean arterial pressure) is a strong driver of frailty over and above the variance captured by a general aggregate measure of all frailty deficits. Prevalence of multimorbidity and associated polypharmacy is a global public health concern, with rates as high as 90% in certain populations(41). This latent factor produced by far the highest number of genomic risk loci and PRS analyses demonstrated that its predictive power was substantial. Gene prioritization and pathway analysis indicated enrichment in a wide array of aging-related pathways, including *VEGFA* signaling, which was recently identified in a multivariate GWAS of aging(42) and has been shown to be important in longevity(43). Taken together, our findings suggest that this latent factor comprises a broad set of disease-related biological pathways that are associated with the most common diseases found within populations that lead to a heightened risk for developing frailty and accelerated aging. This provides empirical support for the 'geroscience hypothesis', which theorizes that manipulating aging physiology will prevent associated diseases (44).

Our findings should be viewed in light of several limitations. We did not explore the impact of sex differences, which are important in aging as evidenced by significant prevalence differences in frailty

across all age groups(45). Our tissue enrichment analyses alluded to this with significant enrichment identified for the sex-specific reproductive organs. Future work should be performed to interrogate the sex-specific multivariate genetic architecture of frailty. Furthermore, our analyses were restricted to samples of European genetic ancestry since the methods rely on linkage disequilibrium information that can vary across ancestral populations. Unfortunately, despite advances in collecting genomic data from multiple populations, it was not possible to identify publicly available GWAS data for the frailty deficits to conduct a trans-ancestry analysis, but this should be a major focus in the future to make these results more generalizable globally.

In conclusion, we introduce the first genomic latent model of frailty. We demonstrate the added potential of modelling frailty as multiple latent factors, representing both a generalized pathway of frailty and distinct subgroups of deficits that share additional underlying biology. This can be contrasted to previous studies that have relied solely upon aggregate measures of frailty. This more nuanced model offered unique etiological insights into frailty and may aid in refining risk stratification of patients. Our genomic model of frailty may also help to develop novel preventive and therapeutic strategies that minimize the broad range of adverse frailty-related health outcomes.



## Online Methods

### Phenotype Selection

Phenotypes were selected based on deficits described by the FI(46). Phenotype selection was further guided by choosing traits that reflect systemic pathways and health behaviors (e.g. number of diagnosed illnesses) as opposed to specific clinical diagnoses (e.g. type 2 diabetes). This allowed for modeling genetic variance for general aspects of frailty rather than disease-specific pathways. Traits were included if they had well-powered GWAS summary statistics that were publicly available in a sample of  $\geq 10,000$  individuals of European ancestry. Analyses were restricted to European ancestry individuals as the methods used to estimate and model genetic overlap rely on ancestry-specific patterns of linkage disequilibrium (LD) and GWAS of sufficient sample size in other ancestry groups are not yet available across all the frailty traits. When possible, we prioritized the use of GWAS data from consortium-based studies since these tend to pool the largest sample sizes and have more rigorous phenotypic definitions (15-19, 47-52). When consortia data was not available, we used European ancestry GWAS summary statistics downloaded from the PanUKB (<https://pan.ukbb.broadinstitute.org>). **Table S1** displays a summary of the 52 traits included in our initial analysis. The effect estimates for each trait were formatted in a direction that reflected the 'risk-inducing' phenotype for frailty (i.e. slow walking pace as opposed to fast walking pace).

### Genetic Correlation Estimation

We conducted multivariable linkage disequilibrium (LD) score regression (LDSC) using the *GenomicSEM* R package along with publicly available LD scores and weights from the 1000 Genomes Phase 3 European reference panel that excluded the major histocompatibility complex (MHC) owing to its complex patterns of LD(53, 54). LDSC was applied to estimate a genetic covariance matrix with SNP-based heritability ( $h^2_{SNP}$ ) of each phenotype on the diagonal and the genetic covariance across each pairwise combination of traits on the off-diagonal. Multivariable LDSC also produces a sampling covariance matrix that includes the sampling variances (i.e., the squared standard errors) of the estimates on the diagonal and the sampling dependencies on the off diagonal that can arise due to participant sample overlap. This sampling covariance matrix is estimated directly from the GWAS data and is what allows Genomic SEM to produce appropriate estimates for traits with varying levels of precision (e.g., with different participant sample sizes) and varying and unknown levels of sample overlap. This was important for producing interpretable results within the current study, wherein our frailty deficits varied widely with respect to statistical power. Prior to running LDSC, each of the GWAS summary statistic datasets underwent a uniform formatting and quality control step using the *munge* function in the *GenomicSEM* R package. This included removing SNPs with a minor allele frequency (MAF) below 0.01 and an imputation score (INFO) below 0.9, restricting SNPs to HapMap 3 SNPs, and aligning all GWAS effects to the same reference allele (13). For all genome-wide analyses, we used a false discovery rate (FDR) corrected  $p$ -value threshold of  $<0.05$  to account for multiple testing.

In line with the recommendations from the original LDSC developers, traits were removed from further analyses if they had a  $h^2_{SNP}$  Z-statistic  $\leq 4$  as these can yield unstable estimates of genetic overlap(55). We converted these estimates to the liability scale for binary traits using population prevalence values based on recent epidemiological estimates for European ancestral populations (**Table S1**)(56-65). The sum of effective sample sizes across the cohorts that had contributed to that GWAS was used to adjust for cohort-specific ascertainment(66).

The results from the genetic correlation analysis (**Table S2**) were used to perform additional quality control, where deficits were brought forward for subsequent analysis providing they met the following criteria: (i) in instances where two traits were highly multicollinear ( $r_g \geq 0.9$ ), we removed the trait in the pair that had the lowest mean genetic correlation ( $r_g$ ) with the other included traits; (ii) we removed traits that had low mean genetic correlations with the other measured traits (mean  $r_g \leq 0.10$ ) as these traits are not empirically indicated to be part of the broader multivariate genetic architecture of frailty; (iii) for phenotypically similar traits (e.g., number of cigarettes smoked per day vs smoking initiation), we retained the deficit with the highest mean  $r_g$  with all other deficits. This resulted in a final list of 30 traits that were brought forward for all subsequent analyses (**Table S1**). A frailty index constructed with 30 or more deficits has been shown to sufficiently capture frailty in phenotypic literature(8).

## Genomic Factor Analysis

**Exploratory Factor Analysis.** Since the latent pattern of the shared genetic architecture between frailty deficits has not been assessed previously, we initially ran an exploratory factor analysis (EFA) using the *stats* R package to identify a plausible latent structure that describes the genetic overlap across the included frailty deficits. To avoid model overfitting, we used the genetic covariance matrix estimated in odd autosomes as the input to the EFA and the genetic covariance matrix estimated in even autosomes as the input for the subsequent confirmatory factor analysis (CFA). We used the Kaiser rule(67) and the number of optimal coordinates test(68) to determine the number of factors to extract in the EFA, which both suggested that seven factors were appropriate. We additionally extracted a 6-factor model (**Table S48**) as there were a high number of cross-loadings in the 7-factor specification, indicating a more parsimonious structure may be appropriate, and the 5<sup>th</sup> factor in the 7-factor model only captured genetic variance related to mean arterial pressure (**Table S49**). Owing to highly pervasive correlations between our frailty deficits (**Table S2**), we applied promax factor rotation, which allows for inter-factor correlations.

**Confirmatory Factor Analysis.** We then conducted a CFA using the genetic covariance matrix from the even autosomes as input. The CFA model was guided using the EFA in odd autosomes, where a frailty deficit was specified to load on a factor when standardized loadings were  $\geq 0.30$ . Diagonally weighted least squares (DWLS) was used for model estimation and any frailty deficits that had negative residual variances were constrained to have a residual variance  $> 0.001$ . The 6-factor (Comparative Fit Index [CFI] = 0.92; Standardized Root Mean Square Residual [SRMR] = 0.07) and 7-factor model specification (CFI= 0.89; SRMR= 0.07) both provided good fit to the even autosome data (**Table S50**). The 6-factor model was selected over the 7-factor model as it: (i) provided improved fit to the data while offering a more parsimonious representation of the data relative to the 7-factor structure; (ii) it produced more theoretically interpretable factors of latent genetic architecture between distinct groups of multiple frailty deficits compared to the 7-factor model; and (iii) the 6-factor model continued to provide good fit to the data in all autosomes (CFI= 0.92; SRMR = 0.06) (**Table S51**).

**Bifactor Model.** While the 6-factor model produced theoretically meaningful latent factors, the first latent factor displayed strong factor loadings for 16 out of the 30 frailty deficits and the model included pervasive cross-loadings of frailty deficits on multiple factors. Taken together, these findings indicated that a bifactor model was an appropriate way to capture the general frailty pathways across all included deficits, as well as the genetic variance specific to distinct subsets of deficits.

Therefore, we estimated the fit of a bifactor model that included loadings for all 30 frailty deficits onto a general factor of frailty (herein termed General Factor), in addition to the loadings on the 6 latent factors from the CFA model (herein named Factors 1-6). A key benefit to this approach is that the General Factor is orthogonal (i.e. uncorrelated) to the additional residual group factors, which enabled us to interpret the General Factor as general genetic pathways of frailty that are distinct from the more focused subsets of genetic variance that underlie potential subgroups within the frailty spectrum. A bifactor model thereby provided a more direct test of our hypothesis that aggregate scores of frailty (e.g., the FI and FFS) miss unique risk pathways that are only shared between smaller subsets of frailty deficits.

Owing to the inclusion of the General Factor, some of the original factor loadings for the six CFA factors became nonsignificant. We iteratively removed any loadings from Factors 1-6 that were below our 0.30 threshold to ensure that we only retained stable loadings in the final model specification. In cases where an indicator displayed loadings above our cut-off for multiple residual factors, we retained these cross-loadings because a previous simulation study found that omitting substantial cross-loadings from a bifactor model based on a prior CFA model can upwardly bias the general factor loadings and downwardly bias residual group factor loadings, which cannot be picked up using standard model fit measures(69).

We allowed the residual group factors (Factors 1-6) to be correlated (but orthogonal to the General Factor). This form of bifactor model is known as a bifactor(S-1) model and is sufficiently identified if a subset of the indicators only load onto the bifactor (in our case 50% of the frailty deficits solely loaded onto the bifactor)(70). To ensure the model was locally identified, factor loadings were constrained to

be equal when there were only two indicators that loaded onto a factor (i.e., for Factors 1 and 3). The final bifactor(S-1) model (**Figure 1** and **Table S3**) continued to provide good fit to the data (CFI = 0.93; SRMR = 0.07) and was brought forward for all subsequent analyses.

### Genetic Correlations with Related Health Traits

Frailty is known to increase the risk of many adverse health outcomes, but it is unclear whether this is owing to shared genetics between the more general frailty pathways or whether some outcomes are only associated with certain deficits within the frailty state. Furthermore, since this represented the first time that frailty has been measured in this latent framework, we wanted to validate our factors as reflecting frailty since well-powered independent GWAS samples for all 30 deficits are not currently available to do a full replication GWAS. Therefore, we performed genetic correlation analyses to assess the associations between 52 aging-related health outcomes and frailty phenotypes to assess the different patterns of shared genetics between these outcomes and each of the latent factors (**Table S4-S5**) (20, 21, 71-90). We used the same quality control procedures and data curation steps on the GWAS summary statistics for these 52 health outcomes as described for the main frailty deficits using the *munge* function in the *GenomicSEM* R package. However, in the case of pneumonia there were no prior available GWAS summary statistics with a SNP-based heritability estimate high enough to be included in LDSC. Therefore, we conducted a fixed-effect meta-analysis using METAL software, which comprised publicly available GWAS summary statistics data from the European ancestry sample in PanUKB (<https://pan.ukbb.broadinstitute.org>;  $N_{\text{Cases}} = 14,054$  and  $N_{\text{Controls}} = 405,999$ ) and FinnGen release 10 (<https://r10.finnngen.fi>;  $N_{\text{Cases}} = 63,377$  and  $N_{\text{Controls}} = 348,804$ ). This resulted in a total GWAS sample of 832,234 individuals (77,431 pneumonia cases and 754,803 healthy controls), which produced a reasonable SNP-based heritability Z-statistic for LDSC (**Table S4**; Z-statistic = 4.1). We then constructed a separate genome-wide genetic variance and covariance matrix for each of the 52 external traits combined with the 30 frailty deficits using multivariable LD score regression. We re-specified our latent frailty model with an additional correlation between the external trait and each latent factor to get the genetic correlation and standard error estimates for each of the external traits with each latent factor. We used an FDR-corrected  $p$ -value threshold  $<0.05$  to correct for multiple testing.

### Multivariate GWAS of Latent Frailty Factors

**Data Standardization.** We performed a multivariate GWAS within the *GenomicSEM* R package that estimated the individual SNP associations with each of the latent factors in our bifactor(S-1) model. We initially used the *sumstats* function to create a standardized data file that converted the SNP effect estimates from each univariate GWAS into covariances between each individual SNP and the overall phenotypic variance of each deficit and standardized the SNP effects relative to the phenotypic variance of the trait (13). This allowed these SNP effects to be added to the genetic covariance matrix as they were then on the same scale as the LDSC estimates. All phenotypes were based on genome build 37 (GRCh37/hg19) or were converted to build 37 using the *MungeSumstats* R package(91). We removed any SNPs that had a MAF  $<0.01$  or an imputation score (INFO)  $<0.6$  according to recommended defaults within the package. We used the 1000 Genomes Phase 3 European ancestry dataset as our reference genome and aligned SNPs to have the same effect (A1) and non-effect (A2) allele across phenotypes. SNPs were removed if either the effect or non-effect allele did not match the reference genome. As *sumstats* performs listwise deletion across the included traits only SNPs that had been measured in all the contributing univariate GWASs were included, resulting in a final harmonized dataset of 5,849,452 SNPs for multivariate GWAS analysis.

**Multivariate GWAS Estimation.** The standardized GWAS data and LDSC results were used as input for calculating multivariate GWAS of the frailty latent factors using the *userGWAS* function in the *GenomicSEM* R package. We fixed the measurement model (i.e. the genome-wide factor loadings and factor correlations) for all SNPs. This improved computational tractability and model interpretability as the SNP-specific estimates were scaled according to the same measurement model across all SNPs (as opposed to the entire model being re-estimated for each SNP). We removed any SNPs that required a high level of smoothing (i.e. Z-statistic change pre- and post-smoothing  $>1.96$ ) or that produced lavaan warnings for negative observed variable or latent variable variances or non-positive definite covariance matrices (293 SNPs removed in total).

**$Q_{\text{SNP}}$  Heterogeneity Index.** As previously described, not all the genetic signal captured within the latent factor GWAS results represent genuinely shared genetic variance. For example,

strong signal from a single indicator (i.e. the *FTO* locus for body fat percentage) can lead to false positives if not properly accounted for(13). Likewise, some of the non-significant genetic signal within the multivariate GWAS results may represent areas of the genome that have highly heterogeneous magnitudes and directions of effects on the different univariate indicators(13). For this reason, it is necessary to calculate the  $Q_{SNP}$  heterogeneity statistics for each SNP, which reflects a  $\chi^2$  distributed statistic that is more significant for SNPs whose effects deviate strongly from the patterning of effects implied by the factor model. As part of the current project, we introduce and validate a more computationally efficient way of calculating  $Q_{SNP}$ . While the prior formulation of  $Q_{SNP}$  required estimating a series of follow-up models to calculate the heterogeneity statistic, our new formulation is automatically calculated for each factor predicted by a SNP in the model. This change thereby greatly reduced the run-time of our analysis. The new  $Q_{SNP}$  equation starts by calculating the residual covariance matrix for the subset of the matrix that reflects the SNP-phenotype covariances for the phenotypes that load on a given factor ( $R_{SNP}$ ) as:

$$R_{SNP} = S_{SNP} - \sum \theta_{SNP} \quad (1)$$

where  $S_{SNP}$  is the vector of SNP-phenotype covariances and  $\sum \theta_{SNP}$  reflects the model implied SNP-phenotype covariances. These model implied estimates reflect the product of the estimated SNP-effect on a given factor and the factor loadings for each trait. The precision of those SNP-phenotype estimates is indexed by taking the eigen decomposition of the portion of the sampling covariance matrix ( $V$ ) that indexes those SNP-phenotype effects ( $V_{SNP}$ ):

$$V_{SNP} = (P_1 P_0) \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} P_1' \\ P_0' \end{pmatrix} \quad (2)$$

where  $P_1$  is the matrix of principal components (eigenvectors) of  $V_{SNP}$ ,  $P_0$  is the null space of  $V_{SNP}$ , and  $E$  is a diagonal matrix of the non-zero eigenvalues of  $V_{SNP}$ . These eigenvalues and eigenvectors can then be used to weight the residual covariance matrix of the SNP-phenotype estimates to obtain a  $\chi^2$  distributed test statistic given as:

$$Q_{SNP}(df) \sim R_{SNP}' P_1 E^{-1} P_1' R_{SNP} \quad (3)$$

where  $df$  reflects the degrees of freedom, which will be one less than the number of indicators for the factor. We note that this equation is iteratively applied for each factor that is predicted by a SNP, such that a separate  $R_{SNP}$ ,  $V_{SNP}$ , and factor-specific  $Q_{SNP}$  are produced.

We demonstrate via simulation that this new approach continues to produce a  $\chi^2$  distributed statistic that is statistically equivalent to the previously described formulation of  $Q_{SNP}$ . We used the *simulateData* function within the *lavaan* R package to simulate data for 3 different factor models each with 50,000 observations for 1000 SNPs. We tested a 2-factor model with 3 indicators on each factor (2 degrees of freedom [df]), a 2-factor model with 4 indicators on each factor (3 df) and a 2-factor model with 6 indicators on each factor (5 df). We confirmed across all 3 examples that the new method remained  $\chi^2$ -distributed and that they did not differ significantly from the estimates calculated using the old method in terms of the mean  $Q_{SNP}$  (**Supplementary Results**). In addition, the new method consistently demonstrated a well-calibrated type 1 error rate ( $p < 0.05$ ) (**Supplementary Results**).

For our empirical frailty application, we pruned out the  $Q_{SNP}$  significant signal from our GWAS summary statistics for each latent frailty factor to ensure that we only measured shared genetic variance operating via each latent factor in our subsequent post-GWAS analyses. We did this by removing SNPs that had a Bonferroni-corrected  $Q_{SNP}$  p-value  $< 7.14 \times 10^{-09}$  (i.e.  $5 \times 10^{-08}/7$ ) and any SNPs that were within a 1 megabase window upstream or downstream of this location to ensure that variants that were in LD with these heterogeneous regions were removed.

Once the  $Q_{SNP}$  signal had been pruned from the latent factor summary statistics, we used the method developed by Mallard and colleagues to calculate the expected sample size ( $\hat{N}$ ) of each latent factor(92) using the following equation:

$$\hat{N} \approx \frac{1}{m} \sum_{MAF=a}^b n_j$$



Where the expected sample size ( $\hat{N}$ ) is approximately equal to the mean SNP expected sample size ( $n_j$ ) for  $m$  number of SNPs that have a MAF between  $a$  and  $b$  (in this case we used a MAF range of SNPs between 10-40% as these provide more stable estimates)(92). This value quantifies the amount of error-free genetic variance that is being captured by each latent factor, so this value can also act as an indicator of how well-powered each latent factor is within the model.

*Identification of Genomic Risk Loci for Latent Frailty Factors.* We used FUMA v1.5.2 to identify genomic risk loci for each latent factor in our model using the default software parameters(93). We used a Bonferroni corrected genome-wide significance threshold of  $p < 7.14 \times 10^{-09}$  (i.e.  $5 \times 10^{-08}/7$  factors) to identify significant SNPs within our pruned GWAS summary statistics for each latent factor (i.e.  $Q_{SNP}$  significant variants removed). A genomic risk locus was defined as the region around a genome-wide significant SNP that included all SNPs that were in LD ( $R^2 < 0.6$ ) with that variant based on LD patterns in the 1000 Genomes Phase 3 European ancestry reference genome(93). If there were additional independently significant SNPs in LD with the lead SNP ( $R^2 > 0.1$ ) or if loci were located within 250 kilobases (kb) of one another, these were merged into a single locus(93).

### Gene Mapping and Functional Annotation

As an initial step to understand the biological implications of the loci underlying each latent frailty factor, we applied the SNP2GENE function in FUMA v1.5.2 to map potentially causal genes to each locus and to functionally annotate the candidate SNPs by integrating our results with various external datasets(93). This involved positionally mapping candidate SNPs to their closest protein-coding gene (within a 10kb window upstream or downstream of the SNP location). We conducted expression quantitative trait loci (eQTL) mapping to ascertain whether the candidate SNPs in our latent factor GWAS loci are known regulators of gene expression, which we defined as any eQTL-gene interaction displaying a  $P_{FDR} \leq 0.05$ (93). Since frailty represents a multi-system clinical state, we integrated publicly available eQTL data for multiple body tissues and cell types with our GWAS results to provide a broad overview of the potential impacts that the risk loci for frailty have on gene expression. This included integration of eQTL data from 15 repositories including the Blood eQTL Browser(94), BIOS QTL browser(95), BRAINEAC(96), MuTHER(97), xQTLServer(98), CommonMind Consortium(99), eQTLGen(100), PsychENCODE(101), DICE(102), scRNA eQTLs(103), eQTL Catalogue(104-120), GTEx v8(121), EyeGEx(122), InsPIRE(123) and TIGER(124). Finally, we conducted chromatin interaction mapping, which involved annotating candidate SNPs that displayed a significant 3D chromatin interaction ( $P_{FDR} \leq 1 \times 10^{-6}$ ) with a known gene region based on pre-processed chromatin loop data and enhancer-promoter link data for multiple body tissues from 4 publicly available databases(101, 125-127). We restricted successful mappings of SNPs to those where the candidate SNP overlapped with the enhancer region within one of the selected epigenomes and the promoter region of the mapped gene overlapped with the promoter region of the associated epigenome. Enhancer and promoter information was obtained for 111 epigenomes from the Roadmap Epigenomics Project(128).

We additionally used MAGMA v1.08(129) to conduct a gene-based analysis that identified genes that were significantly associated with each latent factor. Any SNPs that were present in the GWAS summary statistics for the latent factor and that were located within one of the protein-coding genes in the Ensembl database (excluding the MHC region) were analyzed. We used the 1000 Genomes Phase 3 European dataset as our LD reference panel and applied a Bonferroni correction for number of genes tested within each latent factor GWAS(93). Using gene expression data from GTEx v8 for 54 body tissues(121), we also used MAGMA to perform a gene property analysis to ascertain whether the genes that were significantly associated with the latent factor in the gene-based test were more likely to produce gene expression changes in particular body tissues(130).

### Stratified Genomic SEM

Since frailty has been consistently linked to increased risk of poorer brain health and dementia, which represents a key burden on health services in aging populations(8), we explored whether there was evidence for brain-relevant functional enrichment in the genetic variance captured by our latent frailty factors. We applied Stratified Genomic SEM(56) to test whether there was evidence for enrichment in functional annotations (groups of genetic variants combined due to having a shared biological characteristic) that are known to influence tissue-specific gene expression in different brain regions, histone modifications, neuronal cell types, or the interaction between these neuronal cell types and protein-truncating-variant-intolerant (PI) genes. We used previously constructed functional annotations based on data from 1000 Genomes Phase 3 Baseline LD Version 2.2(131), GTEx(132), DEPICT(133),

the Roadmap Epigenetics Project(128) and the Genome Aggregation Database (gnomAD)(134), which consisted of a total of 172 functional annotations. This included 5 randomly selected non-brain control regions for the gene expression and histone modifications(135).

We performed multivariable stratified LDSC to estimate the zero-order genetic covariance matrices and the corresponding sampling covariance matrices that were partitioned across the genomic regions of each functional annotation using the *s\_ldsc* function in the *GenomicSEM* R package(56, 131). We subsequently used these matrices as the input data to the *enrich* function to calculate an enrichment ratio for the genetic variance captured within each latent factor in our bifactor model for each functional annotation(136). We removed 26 functional annotations from our analyses owing to high degrees of smoothing (defined as a Z-score difference  $>1.96$  pre and post-smoothing), as this indicates low power to detect meaningful enrichment(135). This resulted in 146 functional annotations being retained in our analyses. We used an FDR corrected *p*-value threshold of  $<0.05$  to account for multiple testing.

### Summary Data-Based Mendelian Randomization

We applied multi-SNP summary data-based Mendelian randomization (SMR-multi) to prioritize genes for each latent frailty factor by identifying SNP-outcome (i.e. variant-frailty) associations that demonstrated strong evidence for being driven via intermediary pleiotropic effects on gene expression (eQTLs), splicing ratios (sQTLs) or methylation status (mQTLs; (137, 138). We used pre-processed data curated by the original method developers, which included eQTL data from BrainMeta v2(139), CAGE(140), GTEx v8(121), GEUVADIS(111), PsychENCODE (corrected for 100 hidden covariate factors)(141) and Westra et al(94); sQTL data from BrainMeta v2(139) and GTEx v8(121); and mQTL from Brain-mMeta(142) and McRae et al(138, 143). All data was based on genome build 37 (GRCh37/hg19) and we aligned the GWAS data and QTL data to the same effect and non-effect alleles using the 1000 Genomes phase 3 European data as our reference. We converted the MAF column in the GWAS summary statistics into the effect allele frequency (EAF) by integrating data on the major allele for each SNP from dbSNP v155. Any SNPs that had a discrepant allele frequency  $>0.20$  between the GWAS and QTL data or that were located within the extended MHC region (chr6:28477797-chr6:33448354) were omitted from the analysis(138). SNPs that had a strong association ( $P_{QTL} < 5 \times 10^{-8}$ ) with the QTL region (i.e. expression, splicing or methylation) in each QTL dataset were selected as instrumental variables for SMR-multi analysis(138). Variants in high LD with the lead SNP in a region ( $R^2 > 0.9$ ) were removed using the 1000 Genomes phase 3 European data as a reference. We then applied FDR correction across all the results (i.e. all tests conducted across all latent factors and all QTL datasets) and filtered our results to only include SMR associations with  $P_{SMR-FDR} < 0.05$ . To distinguish pleiotropy (i.e., association between exposure and outcome due to shared causal variant) from linkage (i.e., association between exposure and outcome due to independent causal SNPs being in LD with one another) we used the heterogeneity in dependent instruments (HEIDI) test using the default parameters to measure whether the association patterns within an instrumental variable were homogenous or not, and we rejected any associations that demonstrated significant heterogeneity ( $P_{HEIDI} < 0.01$ )(138).

### Gene Prioritization and Pathway Analysis

In order to better understand the underlying biological pathways of each latent frailty factor, we conducted pathway analysis using the genes that had been prioritized by the aforementioned gene mapping analyses as input. However, since not all genes that are mapped to risk loci represent truly causal genes, we triangulated our results to only include genes that presented sustained evidence for being a potential causal candidate for each latent factor. We defined this as any gene that was mapped by  $\geq 3$  of our gene mapping methods (i.e. positional mapping, eQTL mapping, chromatin interaction mapping, MAGMA or SMR-multi). We used METASCAPE (22) to perform a pathway enrichment analysis to identify gene sets that were significantly overrepresented in the prioritized genes for each latent factor. Since the General Factor was orthogonal to the six latent residual group factors in our model, we conducted a single standalone pathway enrichment analysis for the General Factor prioritized genes because, by definition, the genetics underpinning this latent factor are uncorrelated with the genetics driving the other latent factors. We then performed a separate combined enrichment analysis of the prioritized genes for Factors 1-6 as this enabled us to explore what biology might be driving the inter-factor correlations that we saw in our model. The prioritized genes were assigned to any curated gene set they belonged to using data from Gene Ontology (GO) Biological Processes (1495 gene sets)(144), the Human Molecular Signatures Database (MSigDB) Canonical Pathways (17 gene sets)(145), Reactome (262 gene sets)(146), Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways (113 gene sets)(147), WikiPathways (89 gene sets)(148) and DisGeNET (2050 gene



sets)(149). The default hypergeometric test method was used to calculate enrichment for each gene set that included  $\geq 1$  of the prioritized genes and FDR correction was applied to correct for multiple testing(22). However, many of the gene ontology terms overlap, which leads to high levels of redundancy in the results, so we performed the recommended clustering analysis that combined related gene sets into groups by calculating the pairwise similarities between all enriched terms based on their Kappa-test score and hierarchically ordered them within a similarities matrix (Kappa similarities  $> 0.3$  were combined)(22). In the case of the multi-gene-list enrichment test for the analysis of Factors 1-6, this enrichment test was initially performed on each of the latent factor gene lists separately, followed by an additional test that combined the gene lists across all six latent factors to determine whether there were enriched gene ontology terms that were shared across the latent factors. This was appropriate since our initial results demonstrated that there was significant correlation between certain latent factors in our model.

## Polygenic Risk Scores

*Polygenic Risk Score Construction.* To externally validate our frailty latent factors, we constructed polygenic risk scores (PRS) of each latent frailty factor and tested whether they predicted frailty and related health outcomes in three external cohort datasets, including the Lothian Birth Cohort 1936 (LBC1936), the English Longitudinal Study of Ageing (ELSA), and the Prospective Imaging Study of Ageing (PISA) (see **Supplementary Methods** for sample descriptions). We used the GWAS summary statistics of the shared genetic signal for each of the latent frailty factors (i.e., the summary statistics that had removed significantly heterogeneous signal), as well as publicly available GWAS summary statistics (downloaded from GWAS Catalog(150)) from previously published studies of aggregate measures for the FI(20) and FFS(21) to construct a separate PRS for each of these predictor phenotypes. This enabled us to compare the prediction of the latent frailty factors with routinely used aggregate frailty measures. We performed routine quality control on each of the base datasets (i.e. the GWAS summary statistics) and the target datasets (i.e. individual-level cohort genetic data). We aligned the effect and non-effect alleles across all datasets to ensure that the direction of effect was concordant across analyses. We removed SNPs with a MAF  $< 0.01$ , as well as duplicate and ambiguous SNPs.

Following quality control, PRS were calculated for the individuals in each of the three cohorts using PRSice-2 for LBC1936 and ELSA, and SBayesR for the PISA cohort. We followed default procedures for each methodology that have been described in detail by the respective developers(151, 152). Briefly, PRSice-2 uses a  $P$ -value based clumping and thresholding (P+T) approach in which the GWAS effect size estimates are used as SNP weights in the cohort data and clumping is applied to remove SNPs in high LD so that the final PRS only includes independent significant SNPs(151). In contrast, SBayesR is a Bayesian-based method, which estimates joint SNP effects across the genome using multiple linear regression whilst assuming a finite mixture of normally distributed priors(152).

*Prediction of Frailty and Related Phenotypes in the External Cohorts.* We subsequently performed a series of analyses to explore how well the latent frailty factors predicted routinely measured frailty phenotypes and related traits in external data. First, we used linear regression models to assess how well each individual frailty latent factor PRS predicted the FI in LBC1936 (based on 30 deficits), ELSA (based on 62 deficits) and PISA (based on 69 deficits) and logistic regression models to measure how well they each predicted the FFS in LBC1936 (see **Supplementary Methods and Tables S52-54** for details on the outcomes used). Since the latent frailty factors each represented distinct genetic variance that can contribute to frailty, we also used multiple regression to calculate a PRS of the combined scores for all 7 latent factors (i.e. Multi-PRS). Finally, to enable us to compare the performance of the latent factor PRS with previously published frailty GWAS aggregate measures, we also tested how well the FI GWAS PRS (FI-PRS) and the FFS GWAS PRS (FFS-PRS) predicted the same frailty outcomes in each dataset. All models included age, sex and ancestry principal components ([PCs]; for ELSA and PISA we used 10 PCs and for LBC1936 we used 4 PCs) as covariates. These models allowed us to calculate the amount of incremental phenotypic variance explained ( $R^2$ ) by each PRS, which was calculated by subtracting the covariate-only model  $R^2$  from the  $R^2$  of the full PRS and covariate model(153).

To assess the association of the latent factor PRS's with frailty-related health outcomes, we also conducted regression analyses to test the association of each frailty PRS with cognitive ability (LBC1936), cognitive change (LBC1936), dementia (ELSA), motoric cognitive risk (LBC1936), mortality (LBC1936), stroke (LBC1936, ELSA and PISA) and memory complaints (PISA) (see **Supplementary Methods** for phenotype and analysis descriptions).

*Elastic Net Regression to Rank Performance of Frailty PRS.* Finally, since conventional linear regression models can be upwardly biased due to model overfitting, we performed elastic net regularized regression models in all three cohorts to rank the polygenic contributions to frailty whilst minimizing bias from model overfitting and multicollinearity between predictors(154). This method allows highly genetically correlated variables to be grouped and the final coefficients returned in the model allow the predictors to be ranked by their contribution of prediction to the outcome(154). Therefore, we ran an initial model predicting the FI in each cohort using the seven individual latent frailty factor PRSs and covariates (age, sex and ancestry PCs) as predictors to rank the latent factors in order of their strength in predicting frailty. We then performed elastic net regression that included the Multi-PRS, FI-PRS and FFS-PRS and covariates (age, sex, and 10 ancestry PCs) to rank prediction of the different genetic measures of frailty (i.e. multiple latent factors versus aggregate measures for the FI and FFS).

## Data Availability

The latent frailty factor and pneumonia GWAS summary statistics that were created by this study will be made publicly available on [GWAS Catalog](#) upon publication. All of the GWAS summary statistics that were used in this study for the Genomic SEM and genetic correlation analyses are publicly available via the original articles or from [PanUKB](#) and [FinnGen](#).

## Code Availability

This code was developed using publicly available software that is available via the following links:

Genomic SEM (including our new  $Q_{SNP}$  extension): <https://github.com/GenomicSEM/GenomicSEM>  
 METAL: <https://genome.sph.umich.edu/wiki/METAL>  
 FUMA: <https://fuma.ctglab.nl>  
 SMR: <https://yanglab.westlake.edu.cn/software/smr/>  
 METASCAPE: <https://metascape.org/gp/index.html#/main/step1>  
 SBayesR: <https://cns.genomics.com/software/gctb/#Overview>  
 PRSice-2: <https://choishingwan.github.io/PRSice/>  
 Lavaan: <https://lavaan.ugent.be>  
 MungeSumstats: <https://github.com/neurogenomics/MungeSumstats>

The specific code for the analyses in this study will be made available at <https://github.com/IsyFoote> upon publication. The code we used to create the latent growth curve models of cognitive ability and cognitive change in LBC1936 can be found here: [https://lothianbirthcohorts.github.io/longitudinal-g-models/longitudinal\\_g\\_models](https://lothianbirthcohorts.github.io/longitudinal-g-models/longitudinal_g_models)

## Author Contributions

IFF, ADG, JDF, AR, JMR, DJL, TTK and KR conceived the study. IFF, JPF and AEF performed the analyses. SRC, ML, NGM and MKL provided key data access and support for the study analyses. JPF and DSM performed the data curation of phenotypes in the external individual-level datasets. IFF, ADG and JPF wrote the initial manuscript. All authors contributed to the design and conceptual aspects of the study and critically reviewed the final manuscript.

## Declaration of Interests

Although not directly related to the submitted work, KR has asserted copyright over the Clinical Frailty Scale and (with Olga Theou) the Pictorial Fit-Frail Scale, which are made freely available for non-commercial education and research, as well as non-profit health care with completion of a permission agreement stipulating users will not change, charge for or commercialize the scales. For-profit entities (including pharma) pay a licensing fee, 15% of which is retained by the Dalhousie University Office of Commercialization and Innovation Engagement. After taxes, the remainder of the license fees are donated to the Dalhousie Medical Research Foundation. In the past three years licenses have been negotiated with Rebibus Therapeutics Inc., Cook Research Incorporated, W.L. Gore Associates Inc, Pfizer Inc, Cellcolabs AB, AstraZeneca UK Limited, Qu Biologics Inc, Biotest AG, BioAge Labs Inc, Congenica, Icosavax, Inc. All other authors have no conflicts of interest to declare.

## Acknowledgements

This study was facilitated by the Deep Dementia Phenotyping (DEMON) Network through the Frailty and Dementia Special Interest Group and is an outcome of a workshop entitled “Frailty and Precision Dementia Medicine” funded by the University of Exeter Global Partnerships Fund (JMR). This work was funded by grants from the National Institute on Aging (IFF, AEF and ADG; RF1AG073593) and the National Institute of Mental Health (ADG; R01MH120219). The DEMON Network is supported by Alzheimer’s Research UK (DJL and JMR). DJL is also supported by the National Institute for Health and Care Research Applied Research Collaboration South West Peninsula. The Lothian Birth Cohorts 1936 (LBC 1936) is supported by the Biotechnology and Biological Sciences Research Council, the Economic and Social Research Council [BB/W008793/1], Age UK (The Disconnected Mind Project), the Milton Damerel Trust, and The University of Edinburgh. SRC is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (221890/Z/20/Z). The English Longitudinal Study of Aging (ELSA) is funded by the National Institute on Aging (R01AG017644), and by UK Government Departments coordinated by the National Institute for Health and Care Research (NIHR). The Prospective Imaging Study of Ageing (PISA) is funded by the National Health and Medical Research Council (Grant ID: APP1095227). JPF is funded by the Legal & General Group (research grant to establish the independent Advanced Care Research Centre at the University of Edinburgh) - legal & General had no role in the conduct of the study, interpretation, or the decision to submit for publication - the views expressed are those of the authors and not necessarily those of legal and general. JDF receives research grant support from the Canadian Institutes of Health Research, the National Multiple Sclerosis Society, MS Canada, and consultation and distribution royalties from MAPI Research Trust. ADR thanks the Natural Sciences and Engineering Research Council of Canada (NSERC) for operating grant RGPIN-2019-05888. KR is supported by Dalhousie University Faculty of Medicine Advancement funding as Clinical Research Professor of Frailty and Aging.

We would like to thank the authors of the original GWAS publications that we used in this study for making their data publicly available, including the Neale lab and FinnGen (for unpublished by freely available GWAS data from UK Biobank and FinnGen). We would also like to thank the participants of the contributing cohorts and the research teams for providing their data for this work.

## References

1. Fogg C, Fraser SDS, Roderick P, de Lusignan S, Clegg A, Brailsford S, et al. The dynamics of frailty development and progression in older adults in primary care in England (2006–2017): a retrospective cohort profile. *BMC Geriatrics*. 2022;22(1):30.
2. Kojima G, Iliffe S, Walters K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age and Ageing*. 2017;47(2):193-200.
3. Park CM, Kim W, Rhim HC, Lee ES, Kim JH, Cho KH, Kim DH. Frailty and hospitalization-associated disability after pneumonia: A prospective cohort study. *BMC Geriatrics*. 2021;21(1):111.
4. O’Caoimh R, Sezgin D, O’Donovan MR, Molloy DW, Clegg A, Rockwood K, Liew A. Prevalence of frailty in 62 countries across the world: a systematic review and meta-analysis of population-level studies. *Age and Ageing*. 2020;50(1):96-104.
5. Young ACM, Glaser K, Spector TD, Steves CJ. The Identification of Hereditary and Environmental Determinants of Frailty in a Cohort of UK Twins. *Twin Research and Human Genetics*. 2016;19(6):600-9.
6. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *TheScientificWorldJournal*. 2001;1:323-36.
7. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in Older Adults: Evidence for a Phenotype. *The Journals of Gerontology: Series A*. 2001;56(3):M146-M57.
8. Howlett SE, Rutenber AD, Rockwood K. The degree of frailty as a translational measure of health in aging. *Nature Aging*. 2021;1(8):651-65.
9. Ritchie SJ, Tucker-Drob EM, Starr JM, Deary IJ. Do Cognitive and Physical Functions Age in Concert from Age 70 to 76? Evidence from the Lothian Birth Cohort 1936. *The Spanish Journal of Psychology*. 2016;19:E90.
10. Zhang X, Tan SS, Franse CB, Alhambra-Borrás T, Durá-Ferrandis E, Bilajac L, et al. Association between physical, psychological and social frailty and health-related quality of life among older people. *European Journal of Public Health*. 2019;29(5):936-42.
11. Lachmann R, Stelmach-Mardas M, Bergmann MM, Bernigau W, Weber D, Pischon T, Boeing H. The accumulation of deficits approach to describe frailty. *PLOS ONE*. 2019;14(10):e0223449.
12. Pridham G, Rockwood K, Rutenber A. Efficient representations of binarized health deficit data: the frailty index and beyond. *GeroScience*. 2023.
13. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav*. 2019;3(5):513-25.
14. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. *Cell*. 2023;186(2):243-78.
15. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-81.
16. Trpchevska N, Freidin MB, Broer L, Oosterloo BC, Yao S, Zhou Y, et al. Genome-wide association meta-analysis identifies 48 risk variants and highlights the role of the stria vascularis in hearing loss. *The American Journal of Human Genetics*. 2022;109(6):1077-91.
17. Watanabe K, Jansen PR, Savage JE, Nandakumar P, Wang X, Agee M, et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nature Genetics*. 2022;54(8):1125-32.
18. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*. 2019;51(2):237-44.
19. Jones G, Trajanoska K, Santanasto AJ, Stringa N, Kuo C-L, Atkins JL, et al. Genome-wide meta-analysis of muscle weakness identifies 15 susceptibility loci in older men and women. *Nature Communications*. 2021;12(1):654.
20. Atkins JL, Jylhävä J, Pedersen NL, Magnusson PK, Lu Y, Wang Y, et al. A genome-wide association study of the frailty index highlights brain pathways in ageing. *Aging Cell*. 2021;20(9):e13459.
21. Ye Y, Noche RB, Szejko N, Both CP, Acosta JN, Leasure AC, et al. A genome-wide association study of frailty identifies significant genetic correlation with neuropsychiatric, cardiovascular, and inflammation pathways. *GeroScience*. 2023;45(4):2511-23.
22. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523.



23. Pebrel-Richard C, Debost-Legrand A, Eymard-Pierre E, Greze V, Kemeny S, Gay-Bellile M, et al. An unusual clinical severity of 16p11.2 deletion syndrome caused by unmasked recessive mutation of CLN3. *European Journal of Human Genetics*. 2014;22(3):369-73.
24. Liotta G, Lorusso G, Madaro O, Formosa V, Gialloreti LE, Donnoli C, et al. Exploratory Factor Analysis (EFA) of the Short Functional Geriatric Evaluation (SFGE) to Assess the Multidimensionality of Frailty in Community-Dwelling Older Adults. *International Journal of Environmental Research and Public Health* [Internet]. 2023; 20(5). Available from: [https://mdpi-res.com/d\\_attachment/ijerph/ijerph-20-04129/article\\_deploy/ijerph-20-04129.pdf?version=1677317939](https://mdpi-res.com/d_attachment/ijerph/ijerph-20-04129/article_deploy/ijerph-20-04129.pdf?version=1677317939).
25. Johnson L, Guthrie B, Kelly PA, Anand A, Marshall A, Seth S. Frailty or Frailties: Exploring Frailty Index Subdimensions in the English Longitudinal Study of Ageing. *arXiv preprint arXiv:240300472*. 2024.
26. Soysal P, Stubbs B, Lucato P, Luchini C, Solmi M, Peluso R, et al. Inflammation and frailty in the elderly: A systematic review and meta-analysis. *Ageing Research Reviews*. 2016;31:1-8.
27. Pothier K, Gana W, Bailly N, Fougère B. Associations Between Frailty and Inflammation, Physical, and Psycho-Social Health in Older Adults: A Systematic Review. *Front Psychol*. 2022;13:805501.
28. Cheng Z, He D, Li J, Wu Q, Liu Z, Zhu Y. C-reactive protein and white blood cell are associated with frailty progression: a longitudinal study. *Immunity & Ageing*. 2022;19(1):29.
29. Li C-M, Chao C-T, Chen S-I, Han D-S, Huang K-C. Elevated Red Cell Distribution Width Is Independently Associated With a Higher Frailty Risk Among 2,932 Community-Dwelling Older Adults. *Frontiers in Medicine*. 2020;7.
30. Iwai-Saito K, Shobugawa Y, Aida J, Kondo K. Frailty is associated with susceptibility and severity of pneumonia in older adults (A JAGES multilevel cross-sectional study). *Sci Rep*. 2021;11(1):7966.
31. Sablerolles RSG, Lafeber M, van Kempen JAL, van de Loo BPA, Boersma E, Rietdijk WJR, et al. Association between Clinical Frailty Scale score and hospital mortality in adult patients with COVID-19 (COMET): an international, multicentre, retrospective, observational cohort study. *The Lancet Healthy Longevity*. 2021;2(3):e163-e70.
32. Chao C-T, Lee S-Y, Wang J, Chien K-L, Huang J-W. Frailty increases the risk for developing urinary tract infection among 79,887 patients with diabetic mellitus and chronic kidney disease. *BMC Geriatrics*. 2021;21(1):349.
33. Verschoor CP, Lin DTS, Kobor MS, Mian O, Ma J, Pare G, Ybazeta G. Epigenetic age is associated with baseline and 3-year change in frailty in the Canadian Longitudinal Study on Aging. *Clinical Epigenetics*. 2021;13(1):163.
34. Li X, Delerue T, Schöttker B, Holleczer B, Grill E, Peters A, et al. Derivation and validation of an epigenetic frailty risk score in population-based cohorts of older adults. *Nature Communications*. 2022;13(1):5269.
35. Bellenguez C, Küçükali F, Jansen IE, Kleindam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics*. 2022;54(4):412-36.
36. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nature Genetics*. 2021;53(9):1276-82.
37. Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet*. 2021.
38. Parfenov VA, Zakharov VV, Kabaeva AR, Vakhnina NV. Subjective cognitive decline as a predictor of future cognitive decline: a systematic review. *Dement Neuropsychol*. 2020;14(3):248-57.
39. Gao P-Y, Ma L-Z, Wang X-J, Wu B-S, Huang Y-M, Wang Z-B, et al. Physical frailty, genetic predisposition, and incident dementia: a large prospective cohort study. *Translational Psychiatry*. 2024;14(1):212.
40. Mullin DS, Gadd D, Russ TC, Luciano M, Muniz-Terrera G. Motoric cognitive risk syndrome trajectories and incident dementia over 10 years. *Cerebral Circulation - Cognition and Behavior*. 2023;5:100178.
41. Nicholson K, Liu W, Fitzpatrick D, Hardacre KA, Roberts S, Salerno J, et al. Prevalence of multimorbidity and polypharmacy among adults and older adults: a systematic review. *The Lancet Healthy Longevity*. 2024;5(4):e287-e96.
42. Rosoff DB, Mavromatis LA, Bell AS, Wagner J, Jung J, Marioni RE, et al. Multivariate genome-wide analysis of aging-related traits identifies novel loci and new drug targets for healthy aging. *Nature Aging*. 2023.

43. Grunewald M, Kumar S, Sharife H, Volinsky E, Gileles-Hillel A, Licht T, et al. Counteracting age-related VEGF signaling insufficiency promotes healthy aging and extends life span. *Science*. 2021;373(6554):eabc8479.
44. Sierra F, Kohanski R. Geroscience and the trans-NIH Geroscience Interest Group, GSIG. *GeroScience*. 2017;39(1):1-5.
45. Gordon EH, Peel NM, Samanta M, Theou O, Howlett SE, Hubbard RE. Sex differences in frailty: A systematic review and meta-analysis. *Exp Gerontol*. 2017;89:30-40.
46. Sepehri K, Braley MS, Chinda B, Zou M, Tang B, Park G, et al. A Computerized Frailty Assessment Tool at Points-of-Care: Development of a Standalone Electronic Comprehensive Geriatric Assessment/Frailty Index (eFI-CGA). *Frontiers in Public Health*. 2020;8.
47. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*. 2019.
48. Purves KL, Coleman JRI, Meier SM, Rayner C, Davis KAS, Cheesman R, et al. A Major Role for Common Genetic Variation in Anxiety Disorders. *Biorxiv*. 2019.
49. Wang H, Lane JM, Jones SE, Dashti HS, Ollila HM, Wood AR, et al. Genome-wide association analysis of self-reported daytime sleepiness identifies 42 loci that suggest biological subtypes. *Nature Communications*. 2019;10(1):3503.
50. Trajanoska K, Seppala LJ, Medina-Gomez C, Hsu Y-H, Zhou S, van Schoor NM, et al. Genetic basis of falling risk susceptibility in the UK Biobank Study. *Communications Biology*. 2020;3(1):543.
51. Kemp JP, Morris JA, Medina-Gomez C, Forgetta V, Warrington NM, Youlten SE, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nature Genetics*. 2017;49(10):1468-75.
52. Bonfiglio F, Liu X, Smillie C, Pandit A, Kurilshikov A, Bacigalupe R, et al. GWAS of stool frequency provides insights into gastrointestinal motility and irritable bowel syndrome. *Cell Genom*. 2021;1(3):None.
53. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics*. 2015;47(11):1236-41.
54. Bulik-Sullivan B, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*. 2015;47(3):291-5.
55. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2016;33(2):272-9.
56. Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM, et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics*. 2022;54(5):548-59.
57. Chawla K, Kunonga TP, Stow D, Barker R, Craig D, Hanratty B. Prevalence of loneliness amongst older people in high-income countries: A systematic review and meta-analysis. *PLOS ONE*. 2021;16(7):e0255088.
58. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*. 2020;396(10248):413-46.
59. Lisan Q, Goldberg M, Lahlou G, Ozguler A, Lemonnier S, Jouven X, et al. Prevalence of Hearing Loss and Hearing Aid Use Among Adults in France in the CONSTANCES Study. *JAMA Network Open*. 2022;5(6):e2217633-e.
60. Evans JR, Fletcher AE, Wormald RPL, Ng ES-W, Stirling S, Smeeth L, et al. Prevalence of visual impairment in people aged 75 years and older in Britain: results from the MRC trial of assessment and management of older people in the community. *British Journal of Ophthalmology*. 2002;86(7):795.
61. Kerkhof GA. Epidemiology of sleep and sleep disorders in The Netherlands. *Sleep Medicine*. 2017;30:229-39.
62. Pratt J, De Vito G, Narici M, Segurado R, Dolan J, Conroy J, Boreham C. Grip strength performance from 9431 participants of the GenoFit study: normative data and associated factors. *GeroScience*. 2021;43(5):2533-46.
63. Salari N, Darvishi N, Ahmadipanah M, Shohaimi S, Mohammadi M. Global prevalence of falls in the older adults: a comprehensive systematic review and meta-analysis. *Journal of Orthopaedic Surgery and Research*. 2022;17(1):334.



64. Verschoor CP, Cakmak S, Lukina AO, Dales RE. Activity-related dyspnea in older adults participating in the Canadian Longitudinal Study on Aging. *Journal of General Internal Medicine*. 2022;37(13):3302-9.
65. Kotwal AA, Cenzer IS, Waite LJ, Covinsky KE, Perissinotto CM, Boscardin WJ, et al. The epidemiology of social isolation and loneliness among older adults during the last years of life. *Journal of the American Geriatrics Society*. 2021;69(11):3081-91.
66. Grotzinger AD, Fuente Jdl, Privé F, Nivard MG, Tucker-Drob EM. Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution. *Biological Psychiatry*. 2023;93(1):29-36.
67. Kaiser HF. The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*. 1960;20(1):141-51.
68. Ruscio J, Roche B. Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychol Assess*. 2012;24(2):282-92.
69. Ximénez C, Revuelta J, Castañeda R. What are the consequences of ignoring cross-loadings in bifactor models? A simulation study assessing parameter recovery and sensitivity of goodness-of-fit indices. *Frontiers in Psychology*. 2022;13.
70. Eid M, Krumm S, Koch T, Schulze J. Bifactor Models for Predicting Criteria by General and Specific Factors: Problems of Nonidentifiability and Alternative Solutions. *J Intell*. 2018;6(3).
71. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414-30.
72. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*. 2019;18(12):1091-102.
73. van Rheenen W, van der Spek RAA, Bakker MK, van Vugt JJFA, Hop PJ, Zwamborn RAJ, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nature Genetics*. 2021;53(12):1636-48.
74. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nature Communications*. 2020;11(1):2175.
75. Bell S, Tozer DJ, Markus HS. Genome-wide association study of the human brain functional connectome reveals strong vascular component underlying global network efficiency. *Scientific Reports*. 2022;12(1):14938.
76. Kaufmann T, van der Meer D, Doan NT, Schwarz E, Lund MJ, Agartz I, et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*. 2019;22(10):1617-23.
77. Fürtjes AE, Arathimos R, Coleman JRI, Cole JH, Cox SR, Deary IJ, et al. General dimensions of human brain morphometry inferred from genome-wide association data. *Human Brain Mapping*. 2023;44(8):3311-23.
78. Wainberg M, Forde NJ, Mansour S, Kerrebijn I, Medland SE, Hawco C, Tripathy SJ. Genetic architecture of the structural connectome. *Nature Communications*. 2024;15(1):1962.
79. Traylor M, Persyn E, Tomppo L, Klasson S, Abedi V, Bakker MK, et al. Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *The Lancet Neurology*. 2021;20(5):351-61.
80. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-23.
81. Timmers PRHJ, Mounier N, Lall K, Fischer K, Ning Z, Feng X, et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife*. 2019;8:e39856.
82. Deelen J, Evans DS, Arking DE, Tesi N, Nygaard M, Liu X, et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nature Communications*. 2019;10(1):3669.
83. Nethander M, Coward E, Reimann E, Grahnemo L, Gabrielsen ME, Wibom C, et al. Assessment of the genetic and clinical determinants of hip fracture risk: Genome-wide association and Mendelian randomization study. *Cell Reports Medicine*. 2022;3(10):100776.
84. Butler-Laporte G, Farjoun Y, Chen Y, Hultström M, Liang KYH, Nakanishi T, et al. Increasing serum iron levels and their role in the risk of infectious diseases: a Mendelian randomization approach. *International Journal of Epidemiology*. 2023;52(4):1163-74.
85. Levin MG, Tsao NL, Singhal P, Liu C, Vy HMT, Paranjpe I, et al. Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. *Nature Communications*. 2022;13(1):6914.

86. Pathak GA, Karjalainen J, Stevens C, Neale BM, Daly M, Ganna A, et al. A first update on mapping the human genetic architecture of COVID-19. *Nature*. 2022;608(7921):E1-E10.
87. Said S, Pazoki R, Karhunen V, Vösa U, Ligthart S, Bodinier B, et al. Genetic analysis of over half a million people characterises C-reactive protein loci. *Nature Communications*. 2022;13(1):2198.
88. Stanzick KJ, Li Y, Schlosser P, Gorski M, Wuttke M, Thomas LF, et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nature Communications*. 2021;12(1):4350.
89. Lagou V, Jiang L, Ulrich A, Zudina L, González KSG, Balkhiyarova Z, et al. GWAS of random glucose in 476,326 individuals provide insights into diabetes pathophysiology, complications and treatment stratification. *Nature Genetics*. 2023;55(9):1448-61.
90. Bell S, Rigas AS, Magnusson MK, Ferkingstad E, Allara E, Bjornsdottir G, et al. A genome-wide meta-analysis yields 46 new loci associating with biomarkers of iron homeostasis. *Communications Biology*. 2021;4(1):156.
91. Murphy AE, Schilder BM, Skene NG. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics*. 2021;37(23):4593-6.
92. Mallard TT, Karlsson Linnér R, Grotzinger AD, Sanchez-Roige S, Seidlitz J, Okbay A, et al. Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Cell Genomics*. 2022;2(6):100140.
93. Watanabe K. FUMA GWAS: Functional Mapping and Annotation of Genome-Wide Association Studies Amsterdam: Dept. Complex Trait Genetics at VU University Amsterdam; 2017 [
94. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics*. 2013;45(10):1238-43.
95. Zhernakova DV, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nature Genetics*. 2017;49(1):139-45.
96. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neuroscience*. 2014;17(10):1418-28.
97. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics*. 2012;44(10):1084-9.
98. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nature Neuroscience*. 2017;20(10):1418-26.
99. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience*. 2016;19(11):1442-53.
100. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature Genetics*. 2021;53(9):1300-10.
101. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464.
102. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*. 2018;175(6):1701-15.e16.
103. van der Wijst MGP, Brugge H, de Vries DH, Deelen P, Swertz MA, Franke L, et al. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nature Genetics*. 2018;50(4):493-7.
104. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nature Genetics*. 2018;50(3):424-31.
105. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell*. 2016;167(5):1398-414.e24.
106. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nature Neuroscience*. 2018;21(8):1117-25.
107. Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charlotiaux B, et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nature Communications*. 2018;9(1):2427.

108. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science*. 2014;343(6175):1246949.
109. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genetics*. 2012;44(5):502-10.
110. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife*. 2013;2:e00523.
111. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501(7468):506-11.
112. Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*. 2017;546(7658):370-5.
113. Kasela S, Kisand K, Tserel L, Kaleviste E, Remm A, Fischer K, et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLOS Genetics*. 2017;13(3):e1006643.
114. Lepik K, Annilo T, Kukuškina V, e QC, Kisand K, Kutalik Z, et al. C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLOS Computational Biology*. 2017;13(9):e1005766.
115. Naranbhai V, Fairfax BP, Makino S, Humburg P, Wong D, Ng E, et al. Genomic modulators of gene expression in human neutrophils. *Nature Communications*. 2015;6(1):7545.
116. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. 2016;167(3):657-69.e21.
117. Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, et al. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell*. 2016;167(3):643-56.e17.
118. Schwartzentruber J, Foskolou S, Kilpinen H, Rodrigues J, Alasoo K, Knights AJ, et al. Molecular and functional variation in iPSC-derived sensory neurons. *Nature Genetics*. 2018;50(1):54-61.
119. Buil A, Brown AA, Lappalainen T, Viñuela A, Davies MN, Zheng H-F, et al. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nature Genetics*. 2015;47(1):88-91.
120. van de Bunt M, Manning Fox JE, Dai X, Barrett A, Grey C, Li L, et al. Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. *PLOS Genetics*. 2015;11(12):e1005694.
121. Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-30.
122. Ratnapriya R, Sosina OA, Starostik MR, Kwicklis M, Kappahn RJ, Fritsche LG, et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nature Genetics*. 2019;51(4):606-10.
123. Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nature Communications*. 2020;11(1):4912.
124. Alonso L, Piron A, Morán I, Guindo-Martínez M, Bonàs-Guarch S, Atla G, et al. TIGER: The gene expression regulatory variation landscape of human pancreatic islets. *Cell Reports*. 2021;37(2):109807.
125. Schmitt Anthony D, Hu M, Jung I, Xu Z, Qiu Y, Tan Catherine L, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*. 2016;17(8):2042-59.
126. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455-61.
127. Giusti-Rodríguez P, Lu L, Yang Y, Crowley CA, Liu X, Juric I, et al. Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. *bioRxiv*. 2019:406330.
128. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-30.
129. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology*. 2015;11(4):e1004219.

130. de Leeuw CA, Stringer S, Dekkers IA, Heskes T, Posthuma D. Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. *Nature Communications*. 2018;9(1):3768.
131. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*. 2015;47(11):1228-35.
132. Consortium TG, Ardlie KG, Deluca DS, Segrè AV, Sullivan TJ, Young TR, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 2015;348(6235):648-60.
133. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications*. 2015;6(1):5890.
134. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-43.
135. Grotzinger AD, Mallard TT, Liu Z, Seidlitz J, Ge T, Smoller JW. Multivariate genomic architecture of cortical thickness and surface area at multiple levels of analysis. *Nature Communications*. 2023;14(1):946.
136. Grotzinger AD, de la Fuente J, Davies G, Nivard MG, Tucker-Drob EM. Transcriptome-wide and stratified genomic structural equation modeling identify neurobiological pathways shared across diverse cognitive traits. *Nature Communications*. 2022;13(1):6280.
137. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics*. 2016;48(5):481-7.
138. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nature Communications*. 2018;9(1):918.
139. Qi T, Wu Y, Fang H, Zhang F, Liu S, Zeng J, Yang J. Genetic control of RNA splicing and its distinct role in complex trait variation. *Nature Genetics*. 2022;54(9):1355-63.
140. Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, et al. The Genetic Architecture of Gene Expression in Peripheral Blood. *The American Journal of Human Genetics*. 2017;100(2):228-37.
141. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):eaat8127.
142. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nature Communications*. 2018;9(1):2282.
143. McRae AF, Marioni RE, Shah S, Yang J, Powell JE, Harris SE, et al. Identification of 55,000 Replicated DNA Methylation QTL. *Scientific Reports*. 2018;8(1):17605.
144. Consortium TGO, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1).
145. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*. 2015;1(6):417-25.
146. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Research*. 2019;48(D1):D498-D503.
147. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*. 2022;51(D1):D587-D92.
148. Agrawal A, Balci H, Hanspers K, Coort SL, Martens M, Slenter DN, et al. WikiPathways 2024: next generation pathway database. *Nucleic Acids Research*. 2023;52(D1):D679-D89.
149. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*. 2016;45(D1):D833-D9.
150. Buniello A, MacArthur JA L, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*. 2018;47(D1):D1005-D12.
151. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 2019;8(7).
152. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nature Communications*. 2019;10(1):5086.
153. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*. 2020;15(9):2759-72.

154. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. *Molecular Psychiatry*. 2018;23(5):1368-74.