

1 **Protein signatures associated with loneliness and social isolation: plasma proteome**  
2 **analyses in the English Longitudinal Study of Ageing, with causal evidence from Mendelian**  
3 **randomization**  
4

5 Jessica Gong<sup>1,2</sup>, Zohar Preminger<sup>1</sup>, Andrew Steptoe<sup>3</sup>, Daisy Fancourt<sup>3</sup>  
6

7 **Affiliations:**

8 <sup>1</sup> Department of Epidemiology and Public Health, University College London, WC1E 7HB,  
9 United Kingdom

10 <sup>2</sup> George Institute for Global Health, Imperial College London, W12 7RZ, United Kingdom

11 <sup>3</sup> Department of Behavioural Science and Health, University College London, WC1E 7HB,  
12 United Kingdom  
13

14 **Corresponding author:**

15 Dr Jessica Gong

16 Full address: Department of Epidemiology and Public Health, University College London, 1-19  
17 Torrington Pl, London, WC1E 7HB, United Kingdom

18 Email: [jessica.gong@ucl.ac.uk](mailto:jessica.gong@ucl.ac.uk)  
19

20 **Word count:** Main text: 3126 (**Oline method:** 1692)

21 **Number of references:** 60

22 **Number of figures:** 4

23 **Number of tables:** 1

1 **ABSTRACT**

2 The understanding of biological pathways related to loneliness and social isolation remains  
3 incomplete. Cutting-edge population-based proteomics offers opportunities to uncover novel  
4 biological pathways linked to social deficits. This study employed a proteome-wide and data-  
5 driven approach to estimate the cross-sectional associations between objective measures of social  
6 connections (i.e., social isolation) and subjective measures (i.e., loneliness) with protein  
7 abundance, using the English Longitudinal Study of Ageing. Greater social isolation was  
8 associated with higher levels of 11 proteins (TNFRSF10A, MMP12, TRAIL-R2, SKR3,  
9 TNFRSF11A, VSIG2, PRSS8, FGFR2, KIM1, REN, and NEFL) after minimal adjustments; and  
10 three proteins were significantly associated after full adjustments (TNFRSF10A, TNFRSF11A,  
11 and HAOX1). Findings from two-sample Mendelian randomization indicated that a lower  
12 frequency of in-person social contact with friends or family causally increased levels of  
13 TNFRSF10A, TRAIL-R2, TNFRSF11A, and KIM1, and decreased the level of NEFL. The study  
14 also highlighted several enriched biological pathways, including necrosis and cell death  
15 regulation, dimerization of procaspase-8, and inhibition of caspase-8 pathways, which have  
16 previously not been linked to social deficits. These findings could help explain the relationship  
17 between social deficits and disease, confirming the importance of continuing to explore novel  
18 biological pathways associated with social deficits.

## 1 INTRODUCTION

2 A large body of work now demonstrates the substantial impact that deficits in social connections  
3 have on health.[1, 2] Both social isolation (the objective presence or absence of other people in  
4 our lives and the frequency that which we engage with them) and loneliness (the perceived  
5 availability of others to fulfil our personal needs) are clear predictors of incidence of physical  
6 diseases (e.g. cardiovascular disease, dementia and diabetes),[3-6] psychiatric disorders  
7 (including depression, anxiety and schizophrenia),[7, 8] age-related decline,[9, 10] and mortality  
8 (both through suicide and other causes).[11-13]

9 Numerous different theories provide direct and indirect explanations for why these effects occur.  
10 In relation to **direct pathways**, according to the social neuroscience model, *homo sapiens* is an  
11 inherently social species, relying on others for species survival. If humans find themselves  
12 socially isolated and therefore vulnerable, they no longer feel safe, and experience social stress  
13 and emotional symptoms of isolation (i.e. loneliness), resulting in a cascade of bio-behavioral  
14 effects being activated, all to increase preparedness for potential assaults.[14] These effects  
15 include an increased vigilance for social threats, less settled sleep (to avoid predation), decreased  
16 impulse control, increases in depressive symptomatology (to signal the need for support and  
17 connection), and a host of physiological changes indicating a hypervigilant stress response  
18 system (e.g. elevated vascular activity, heightened hypothalamic-pituitary-adrenal (HPA) axis  
19 activity, and enhanced immune activation).[15] However, while these neural and behavioral  
20 responses may increase the likelihood of short-term survival, if isolation becomes repeated or  
21 chronic, they can carry long-term costs, including increasing the risk of chronic illness.  
22 Biologically, the mechanisms that specifically link chronic social isolation to disease include (i)  
23 an overactive HPA axis resulting in receptor cells developing glucocorticoid resistance, leading

1 to a greater susceptibility to inflammation (itself associated with adverse health conditions), (ii)  
2 dysregulation of the autonomic nervous system's ability to regular cardiovascular activity  
3 (including decreased heart rate variability), (iii) changes in immunological responses (including  
4 increased secretion of inflammatory cytokines, growth factors and antibodies) (iv) reduced repair  
5 and restorative processes (such as decreased natural killer cell activity and slower wound  
6 healing), and (v) changes in brain structure and functioning (including increased brain atrophy  
7 and reduced neurogenesis).[16-19]

8 In addition, social isolation and loneliness may influence health outcomes via **indirect**  
9 **pathways**. According to the social control hypothesis, social networks tend to discourage poor  
10 health behaviors and encourage good health behaviors, thereby supporting the health of the  
11 group as a whole by promoting the health of the individuals within it.[20] When individuals  
12 become socially isolated, they lose this beneficial influence and are more likely to engage in  
13 adverse health behaviors, including substance use (e.g. alcohol, cannabis, opiates, and tobacco),  
14 reduced physical activity, poor diet (leading to increased risk of malnutrition, eating disorders,  
15 and consumption of ultra-processed foods), reduced adherence to health guidelines (e.g. reduced  
16 medication adherence, use of preventative healthcare screening, visits to primary health care  
17 services, and following of basic hygiene procedures like hand-washing), and fewer recuperative  
18 behaviors (e.g. leisure activities).[1, 18]

19 However, while both these direct and indirect pathways provide some biological plausibility to  
20 the link between social connections and adverse health outcomes, they may not provide a  
21 complete picture. There is a risk that as certain biological pathways between social deficit and  
22 disease become better understood, they can dominate research, and attempts to identify novel  
23 pathways decline. In recent years, with the increasing availability of high-throughput

1 measurements of molecular phenotypes such as protein abundance, there are novel opportunities  
2 to substantially expand our understanding of the biological implications of social deficits.  
3 Notably, several studies have focused on the impact of social isolation on protein expression in  
4 rodents, with findings identifying some novel biological pathways previously not connected to  
5 social deficits.[21-23] But to date, there are no studies focusing on humans.  
6 Consequently, in this study, we focused specifically on identifying a protein signature for  
7 loneliness and social isolation using large-scale novel proteomics data from a representative  
8 cohort study. Literature on the relationship between social deficits and disease incidence suggest  
9 distinct risks from both loneliness and isolation,[3, 24-27] and previous research involving  
10 individual proteins relating to inflammatory pathways have corroborated the concept of distinct  
11 biological pathways for different types of social deficits.[28, 29] Therefore, we hypothesized that  
12 there would be distinct protein signatures for loneliness and social isolation. Our objectives were  
13 (i) to identify proteins associated with loneliness and social isolation, (ii) to apply two-sample  
14 Mendelian Randomisation (MR) to explore the potential causal effects of social deficits on the  
15 significantly associated proteins, and (iii) to undertake enrichment analysis of the identified  
16 proteins to identify which biological pathways are regulated by loneliness and/or isolation.

## 1 RESULTS

### 2 *Study sample*

3 At study wave 4 (2008-2009) in English Longitudinal Study of Ageing (ELSA), the mean score  
4 for loneliness was 4.3 [standard deviation (SD) = 1.6], while the mean score for social isolation  
5 was 2.7 [SD = 1.6] (Table 1), based on the pooled imputed datasets. The mean age of the study  
6 sample was 63.4 years [SD = 9.2]), 55% of the participants were female, and 97% were of White  
7 ethnicity.

8

### 9 *Associations between protein with loneliness and social isolation*

10 First, the cross-sectional associations for loneliness and social isolation with the levels of 276  
11 protein were estimated. By pooling the 30 imputed datasets, no protein was found to be  
12 significantly associated with loneliness based on the minimally-adjusted models (Figure 1). For  
13 social isolation, a total of 11 proteins were significantly associated after minimal adjustments, all  
14 indicating positive associations: **TNFRSF10A** (coefficient ( $\beta$ ) [standard error (se)]: 0.061  
15 [0.011]; False discovery rate (FDR) adjusted P (denoted as  $P_{\text{FDR}}$ ) =  $9.89 \times 10^{-6}$ ), **MMP12** ( $\beta$  [se]:  
16 0.057 [0.011];  $P_{\text{FDR}} = 4.10 \times 10^{-5}$ ), **TRAIL-R2** (also known as **TNFRSF10B**) ( $\beta$  [se]: 0.056  
17 [0.011];  $P_{\text{FDR}} = 8.34 \times 10^{-5}$ ), **SKR3** (also known as **ACVRL1**) ( $\beta$  [se]: 0.056 [0.011];  $P_{\text{FDR}} =$   
18 0.0001), **TNFRSF11A** ( $\beta$  [se]: 0.056 [0.011];  $P_{\text{FDR}} = 0.0001$ ), **VSIG2** ( $\beta$  [se]: 0.049 [0.011];  
19  $P_{\text{FDR}} = 0.0026$ ), **PRSS8** ( $\beta$  [se]: 0.045 [0.011];  $P_{\text{FDR}} = 0.013$ ), **FGFR2** ( $\beta$  [se]: 0.045 [0.011];  
20  $P_{\text{FDR}} = 0.021$ ), **KIM1** (also known as **HAVCR1**) ( $\beta$  [se]: 0.043 [0.011];  $P_{\text{FDR}} = 0.026$ ), **REN** ( $\beta$   
21 [se]: 0.041 [0.011];  $P_{\text{FDR}} = 0.045$ ), and **NEFL** ( $\beta$  [se]: 0.035 [0.009];  $P_{\text{FDR}} = 0.047$ ) (Figure 2).

22 Based on the fully-adjusted models, there were no proteins significantly associated with  
23 loneliness (Supplementary Fig. 1). For social isolation, **TNFRSF10A** ( $\beta$  [se]: 0.049 [0.011];  $P_{\text{FDR}}$

1 = 0.003), **HAOX1** ( $\beta$  [se]: 0.047 [0.012];  $P_{\text{FDR}} = 0.014$ ), and **TNFRSF11A** ( $\beta$  [se]: 0.044  
2 [0.011];  $P_{\text{FDR}} = 0.026$ ) were found to be significantly associated in the fully-adjusted model  
3 (Supplementary Fig. 2), indicating higher protein levels associating with greater level of social  
4 isolation.

#### 6 *Two-sample MR*

7 Next, two-sample MR was conducted using summary statistics from genome-wide association  
8 study (GWAS) to assess the potential causal relationship between social isolation and the protein  
9 concentrations identified as significant based on the linear regressions. Summary statistics for  
10 genetic variants associated with circulating protein levels, specifically protein quantitative trait  
11 loci (pQTL), which are also associated with dimensions of social isolation in the GWAS from  
12 the UK Biobank, were used to infer causality.

13 Based on results from two-sample MR, lower frequency of friend/family visits was causally  
14 linked to an increased level of **TNFRSF10A** ( $\beta$  [se]: 0.245 [0.117];  $P = 0.037$ , based on IVW;  $\beta$   
15 [se]: 0.324 [0.143];  $P = 0.023$ , based on weighted median;  $\beta$  [se]: 0.249 [0.103];  $P = 0.016$ , based  
16 on Maximum likelihood ) (Figure 3; Supplementary Table 1); and increased level of **TRAIL-R2**  
17 (also known as **TNFRSF10B**) ( $\beta$  [se]: 0.264 [0.102];  $P = 0.010$ , based on IVW;  $\beta$  [se]: 0.274  
18 [0.104];  $P = 0.008$ , based on Maximum likelihood); increased level of **TNFRSF11A** ( $\beta$  [se]:  
19 0.210 [0.103];  $P = 0.042$ , based on Maximum likelihood); increased level of **KIM1** (also known  
20 as **HAVCR1**) ( $\beta$  [se]: 0.195 [0.095];  $P = 0.040$ , based on Maximum likelihood); and decreased  
21 level of **NEFL** ( $\beta$  [se]: -0.257 [0.105];  $P = 0.014$ , based on IVW;  $\beta$  [se]: -0.248 [0.101];  $P =$   
22 0.014, based on Maximum likelihood). MRs of living alone or the multi-trait loneliness-isolation  
23 GWAS did not show any significant findings.

1  
2 *Enrichment analysis*  
3 The enrichment analysis of the identified proteins unveiled a spectrum of pathways and  
4 expressions significantly enriched, by searching relevant public bioinformatics databases using  
5 EnrichR [30] (Figure 4; Supplementary Table 2). These pathways include RIPK1-mediated  
6 regulated necrosis, TP53 regulation of death receptors and ligands transcription, and TRAIL  
7 signalling. Additionally, pathways regulating necrosis and cell death, such as death receptor  
8 activity and caspase activation via death receptors in the presence of ligands, and via extrinsic  
9 apoptotic signalling pathways, were prominently featured. Moreover, inhibition of caspase-8  
10 activity, and dimerization of procaspase-8 pathways was identified. These findings collectively  
11 indicate the identified proteins have pivotal role in regulating processes associated with cell  
12 death, death receptor signalling cascades, and caspase activation.



## 1 **DISCUSSION**

2 Overall, the current study utilized a data-driven approach to reveal a protein signature for social  
3 deficits. As hypothesized, we saw distinct findings for objective measures of social connections  
4 (i.e. social isolation) and subjective measures (i.e. loneliness). Specifically, greater social  
5 isolation was related to increased levels of 11 proteins cross-sectionally (**TNFRSF10A**,  
6 **MMP12**, **TRAIL-R2** (also known as **TNFRSF10B**), **SKR3** (also known as **ACVRL1**),  
7 **TNFRSF11A**, **VSIG2**, **PRSS8**, **FGFR2**, **KIM1** (also known as **HAVCR1**), **REN**, and **NEFL**),  
8 after minimal adjustments; and three proteins were significantly associated after full adjustments  
9 (**TNFRSF10A**, **TNFRSF11A**, and **HAOX1**). Findings from two-sample MR indicated that a  
10 lower frequency of in-person social contact with friends or family causally increased levels of  
11 **TNFRSF10A**, **TRAIL-R2** (also known as **TNFRSF10B**), **TNFRSF11A**, and **KIM1** (also  
12 known as **HAVCR1**), and decreased level of **NEFL**. The study also highlighted several enriched  
13 biological pathways, including necrosis and cell death regulation, dimerization of procaspase-8  
14 and inhibition of caspase-8 pathways, which have previously not been linked to social deficits  
15 and could help to explain their relationship to disease, confirming the importance of continuing  
16 to explore novel biological pathways associated with social deficits.

17 This is, to our knowledge, the first study looking at the relationship between phenotypic  
18 loneliness and social isolation and human protein expression. A previous study mapped protein  
19 expression to GWAS that had been undertaken to identify single nucleotide polymorphisms  
20 (SNPs) associated with composite indices of social deficits (loneliness, living alone, low social  
21 contact and a lack of confidants).[31] There was no overlap between the proteins identified there  
22 and in our study. However, that study explored the alternative end of the molecular cascade, i.e.  
23 how genetic risk for loneliness or isolation is related to protein expression, rather than how

1 phenotypic experiences of loneliness or isolation could moderate protein expression. What the  
2 combination of our findings with those of this previous study suggest is that the molecular  
3 influence of phenotypic experiences of loneliness/isolation may not be restricted to proteins  
4 transcribed and translated from genes specific to loneliness/isolation. Indeed, even looking  
5 across other genetics studies, none of the proteins that related to social isolation come from genes  
6 that have previously been related to social deficits.[32]

7 Instead, we identified novel proteins that may be causally influenced by social isolation, and in  
8 doing so, we identify a number of new biological pathways that have previously not been  
9 associated with social deficits. TNFRSF10A, TNFRSF10B (also known TRAIL-R2) and  
10 TNFRSF11A (also known as TRANCE receptor or RANK) are upregulated in the blood in  
11 various cancers including myeloma, several leukaemia and lung cancer, which could provide  
12 some explanation for the relationship between isolation and cancer mortality.[25] TNFRSF10A  
13 and TNFRSF10B also initiate the cascade of caspases that mediates apoptosis and promotes the  
14 activation of NF- $\kappa$ B.[33] According to the Conserved Transcriptional Response to Adversity  
15 theory (CTRA), when socially isolated, humans are evolutionarily primed to express a higher  
16 concentration of genes relating to increased inflammation and decreased antiviral activity.[34]

17 The social signal transduction process involved in this gene expression is thought to involve  
18 peripheral neural signalling via neurotransmitters such as dopamine and norepinephrine that then  
19 leads to cellular signal transduction and the activation of multiple transcription factors including  
20 CREB and NF- $\kappa$ B. So, our finding of greater TNFRSF10A and TNFRSF10B abundance in  
21 response to social isolation aligns with previous theoretical and empirical work.

22 Amongst other proteins that we identified as associated with social isolation, KIM1 (also known  
23 as HAVCR1) is a member of the TIM (T-cell immunoglobulin and mucin) gene family, which

1 plays an important role in host-virus interactions, and acts as a receptor for viruses including  
2 Ebola, Marburg, Dengue, Zika and possible SARS-CoV-2.[35] This could help to explain studies  
3 demonstrating increased viral susceptibility amongst those who are socially isolated.[36]  
4 However, it is perhaps surprising that NEFL levels were lower amongst those with social  
5 isolation. NEFL is a biomarker of axonal damage and is widely implicated in neurological  
6 disorders. We also identified diverse enriched biological pathways related to cell death, death  
7 receptor signalling cascades, and caspase activation (intrinsic both to apoptosis and inflammation  
8 responses). This is noteworthy not just in terms of alignment with the literature on  
9 immunological response to social isolation, but also given oxi-inflamm-aging theories suggesting  
10 the critical role of chronic oxidative and inflammatory stress in biological ageing  
11 specifically.[37] However, caution is important when attempting to interpret the clinical  
12 relevance of these proteins. Given the highly complex interplay between molecular biomarkers  
13 in the pathogenesis of disease, further research is required to replicate the findings here in other  
14 datasets.

15 While we identified these biological responses for social isolation, it is notable that no proteins in  
16 our analyses were associated with loneliness in fully-adjusted models. In considering why we  
17 saw results for isolation but not loneliness, there are two explanations. Taking both a social  
18 neuroscience perspective and considering the social control hypothesis, it is the act of being  
19 physically isolated that is the risk to survival and activates direct and indirect pathways.  
20 However, it is also important to remember that this study focused on protein panels that were  
21 developed to be particularly relevant to neurological and cardiovascular conditions. Loneliness  
22 may be related to protein abundance at other parts of the human proteome. As such, our results

1 do not necessarily imply that social isolation has greater effects on protein abundance than  
2 loneliness.

3 It is also of note that although we did identify a signature involving several different proteins  
4 associated with social isolation, MR analyses suggested the finding was not necessarily causal  
5 across all MR analyses. Some of the proteins identified in this study could be correlates of other  
6 molecular biomarker changes that occur due to social isolation or even of the disease states that  
7 are increased as a result of social deficits, rather than direct causal products of isolation  
8 themselves. Further, although we used large GWAS data on social isolation, none of the  
9 measures of isolation exactly replicated the social isolation phenotype we used, so there may be  
10 additional aspects of our phenotype not captured in the GWAS. Frequency of social contact (the  
11 MR that produced the findings discussed above) was the closest in measuring objective aspects  
12 of structural social connections. The “living alone” GWAS was limited in only measuring the  
13 more limited domain of domestic isolation, while the multi-trait GWAS included combinations  
14 of structural, functional and quality aspects of social connections that are much broader than the  
15 structural construct we tested phenotypically. Our GWAS summary data were also taken from  
16 UK Biobank, which contains a younger age range than ELSA, and age has been shown to affect  
17 the prediction accuracy of polygenic score (PGS).[38] As such, it is possible that alternative  
18 GWAS could have yielded different results. There are also broader acknowledged challenges  
19 around using MR for social and behavioral traits.[39] While a person’s genome can undoubtedly  
20 influence factors that have a relevance to social traits, from physical traits to personality  
21 characteristics, health factors and behaviors that could causally influence loneliness/isolation, the  
22 degree of variance that can be explained with a PGS is always limited by the true heritability a  
23 trait, which is likely smaller for complex social traits like loneliness/isolation that have large

1 environmental components.[40] If the alleles in the PGS for loneliness/isolation are not directly  
2 causally related to loneliness/isolation but in fact based on indirect pathways (e.g. assortative  
3 mating of sociable parents) or the involvement of many pleiotropic genetic variants also strongly  
4 involved in downstream health conditions related to loneliness,[18] polygenic  
5 loneliness/isolation scores may in fact not be a valid tool to enhance our understanding of  
6 causality. As a result, we encourage that the MR analyses reported are interpreted cautiously.  
7 This study has many strengths, including its use of a large, representative cohort of older adults  
8 of white European ancestry, rich validated measures of both social isolation and loneliness, and  
9 use of MR and enrichment analyses to explore issues of causality and biological plausibility with  
10 respect to disease risk. However, the study has several limitations. Because we focused on older  
11 adults, we do not know whether the biological responses identified in this study are stable across  
12 the life-course. We only had data on protein abundance for one timepoint, so we focused on  
13 cross-sectional relationships between social deficits, using MR to explore the direction of the  
14 association. But as the incorporation of high-throughput measurements of molecular phenotypes  
15 becomes more common within cohort studies, the potential inclusion of further waves of  
16 proteomics data could support the analysis of changes in protein abundance over time in relation  
17 to changing social deficits. Consequently, many pathways related to loneliness and social  
18 isolation might not have been detected. Broader proteome-wide studies are encouraged to  
19 replicate and extend our findings. Additionally, this study focused on the chronic effects of  
20 loneliness and social isolation on the proteome. But given animal research suggesting that  
21 isolation also induces acute fluctuations in protein expression, leading to different molecular  
22 signatures for each stage of response,[21] future studies are encouraged that explore the dynamic  
23 profile of isolation on the human proteome. Finally, we ran analyses that were adjusted for core

1 confounders, as well as those adjusted for health behaviors. This split model adjustment was  
2 conducted as behavioral confounders may also be partial mediators of effects, with loneliness  
3 and isolation predisposing individuals to substance use, low physical activity, and other health  
4 behaviors such as sleep disturbances and poor diet (as proposed in the social control  
5 hypothesis).[18] While it is important to acknowledge their effects on molecular biomarkers, we  
6 do not know whether social deficits lead to changes in biomarkers that then causally influence  
7 health behaviors, or whether the changes in health behaviors lead to alterations in the  
8 biomarkers. Hence, we have provided results from both models.

9 Overall, our study showed novel protein signatures for social isolation but not loneliness in a  
10 large sample of older adults, providing the first large-scale data on the relationship between  
11 social deficits and the human proteome. Our findings demonstrated novel biological pathways  
12 influenced by social deficits including (but not limited to) those involved in cell death,  
13 malignancies, and host vulnerability, which could help to provide further biological plausibility  
14 and mechanistic evidence for the impact of social deficits on human diseases. This reinforces the  
15 importance of undertaking future research into the molecular mechanisms of social deficits.

## 1 **ONLINE METHODS**

### 2 *Study population*

3 This study used data from the English Longitudinal Study of Ageing (ELSA) - a longitudinal  
4 study of individuals in England aged 50 and older, and their partners. ELSA started by using data  
5 from the Health Survey in England (1998, 1999, and 2001) and collected its first wave of data in  
6 2002/2003 through in-person interviews and self-reported surveys.[41] Blood samples were  
7 collected in the English Longitudinal Study of Aging (ELSA) starting from the second nurse visit  
8 in 2004 and then at four-year intervals thereafter.

9 The blood samples obtained during the wave 4 nurse visit in 2008 were utilized for proteomic  
10 profiling in ELSA (N = 6,271). Exclusion criteria were applied, excluding participants who  
11 either died within two years after the wave 4 nurse visit (N = 134) or were lost to follow-up  
12 (missing at least two consecutive waves) (N = 1,340). A total of 3,325 frozen plasma blood  
13 samples were sent to Olink for division into aliquots, plating, and conducting proteomics assays.  
14 A final 3,305 samples from wave 4 was viable for proteomic profiling. The analyses  
15 encompassed a final combined dataset of 3,262 participants after conducting stringent quality  
16 control of the samples (see section below). The participant selection for the proteomics assays in  
17 ELSA is depicted in Supplementary Fig. 3.

18

### 19 *Loneliness*

20 Based on the wave 4 data, ELSA measured loneliness with the three-item R-UCLA loneliness  
21 scale [42, 43]: (1) how often do you feel you lack companionship? (2) how often do you feel  
22 isolated from others? (3) how often do you feel left out? Each question asks the participant to  
23 rank their experience on a three-point Likert scale (hardly ever/never, some of the time, or often),

1 and results in a scale ranging from 3 to 9, in which a greater score corresponds with a higher  
2 level of perceived loneliness (Cronbach's  $\alpha = 0.82$ ).

3

#### 4 *Social Isolation*

5 Social isolation was measured on a 7-point scale, in which each of the following represented one  
6 point: living alone, not working, not volunteering, not belonging to a club or social organisation,  
7 and having less than monthly interaction with friends, relatives, or children, respectively. This  
8 measure was adapted from the scale proposed by Bu et al.[44] The score ranged from 0 to 7, with  
9 higher value indicating greater level of social isolation.

10

#### 11 *Proteomic data*

12 The Olink technology employs the Proximity Extension Assay (PEA), utilizing a matched pair of  
13 antibodies labelled with unique complementary oligonucleotides (proximity probes). These  
14 probes bind to their respective target proteins in a sample, bringing them into proximity. This  
15 proximity allows the probes to hybridize, enabling DNA amplification of the protein signal. The  
16 amplified signal is then quantified using next-generation sequencing.[45]

17 The curation of the proteomics data in ELSA was initially to investigate proteomic signatures  
18 related to cognitive decline and dementia specifically, resulting in a more focused selection of  
19 proteins. Three Olink™ Target 96 panels were selected for the proteomics assays:  
20 Cardiovascular II (CVDII), Neurology I (NEUI), and Neurology Exploratory (NEX). These  
21 encompass an extensive array of cardiovascular, immunological and inflammatory markers, as  
22 well as markers integral to neurological processes such as axon guidance, neurogenesis, and  
23 synapse assembly.



1 These assays incorporate an inherent quality control mechanism utilizing four internal controls  
2 added to all samples, along with external controls. The stringent quality control pipeline has been  
3 previously described,[46] and is further detailed in Supplementary Methods.  
4 After excluding those did not pass quality control and outliers, a final combined dataset of 3,262  
5 samples were included in the analyses, with a total of 276 unique proteins. All proteins were  
6 quantifiable, although a total of eight proteins (BNP, MAPT, CADM3, beta-NGF, HSP90B1,  
7 NXPH1, IKZF2, EPHA10) had  $\geq 50\%$  below Limit of detection (LOD). However, these data  
8 points were not removed in the analyses, as some of the most distinct biomarkers may be low in  
9 some groups analysed but high in other groups, and by including data under LOD does  
10 commonly not increase false positives as there is generally no significant difference between  
11 groups under LOD. Protein concentrations were quantified using Olink's standardized  
12 Normalized Protein eXpression (NPX) values, presented on a  $\text{Log}_2$  scale.

13

#### 14 *Covariates*

15 Covariates included sociodemographic factors including age, non-pension wealth quintiles, age  
16 when left formal education, sex (female vs male), and ethnicity (White vs Other ethnic group).  
17 We also included health behaviors: smoking status, alcohol consumption, physical activity, and  
18 body mass index (BMI). Participant smoking was coded to identify if an individual had never  
19 smoked, previously smoked, or currently smoked. Alcohol consumption was ordinally coded  
20 based on the number of days the participant drank alcohol in the past week (ranged from 0 to 7).  
21 Physical activity was coded as a scale ranging from participants hardly ever or never exercising,  
22 exercising monthly, exercising weekly, and exercising more than weekly. BMI was included as a  
23 continuous variable in analyses.

1 *Statistical analysis*

2 *Data normalisation and missing data handling*

3 Missing data were addressed by imputing data using K-nearest neighbours (KNN),[47] and  
4 multiple imputation by chained equations (MICE) methods.[48] Prior to the imputation, the  
5 protein levels underwent rank-based inverse normal transformation and were scaled to achieve a  
6 mean of 0 and a standard deviation of 1. All proteins had < 6% missing. KNN imputation was  
7 used to address missing data in the proteomics data, and the number of neighbouring points ( $k =$   
8  $57$ ) was based on the square root of the total sample size ( $N = 3,262$ ), using the ‘impute’ package  
9 in R.[49] For imputing missing data in the exposures and covariates, MICE was utilized (using  
10 the ‘mice’ R package.[50]), with 30 imputations and 10 iterations selected, to reduce bias and  
11 increase the efficiency and reliability of the statistical estimates based on empirical evidence.[51]

12

13 *Cross-sectional linear regression model*

14 Following this process, linear regression model was used to estimate the association between  
15 loneliness and social isolation with each protein, by pooling the estimates from all imputed  
16 datasets. Two-sided p-values were reported and displayed using a volcano plot, accompanying  
17 the coefficient, with a cut-off of 0.05 for p-value adjusted for false discovery rate (FDR; denoted  
18 as  $P_{\text{FDR}}$ ) to indicate statistical significance. Two multiple-adjusted linear regression models were  
19 constructed: a minimally-adjusted model in which sociodemographic variables (age, sex,  
20 ethnicity, wealth quintiles, and education) were included as covariates, and a fully-adjusted  
21 model in which sociodemographic and health behavior variables (alcohol consumption, smoking,  
22 BMI, and physical activity) variables were included as covariates.

23

1 *Two-sample MR*

2 Given that the limitations often associated with cross-sectional study design being unable to  
3 establish cause-and-effect relationships with temporal ambiguity, and issue with unmeasured  
4 confounding, we conducted two-sample MR on the significant proteins associated with  
5 loneliness and social isolation, based on the proteins identified from the minimally- and fully-  
6 adjusted linear regressions to infer causality,[52] by leveraging summary statistics from GWAS  
7 based on data from the UK Biobank.[53]

8 We identified three GWAS that were considered closely related to our **social isolation**  
9 phenotype: 1) frequency of contact with family and friends, 2) living alone, and 3) a multi-trait  
10 GWAS (MTAG) combining the results from three separate GWAS studies on perceived  
11 loneliness, living alone, and the ability to confide into a single analysis.[53] This meta-analytical  
12 approach enhances the discovery of genetic variants for a target trait by utilizing the statistical  
13 power from the additional traits. The pertinent information outlining the GWAS used for various  
14 measurements of social isolation is detailed in Supplementary Table 3.

15 For GWAS of protein levels, instruments representing changes in protein abundance were  
16 selected based on pQTL mapping of proteins, identifying primary genetic associations in  
17 individuals of European ancestry from the UK Biobank  
18 (<https://doi.org/10.7303/syn51364943>).[54] Standardization of the effects of protein pQTL was  
19 conducted to ensure alignment with the same effect allele.

20 Selection of instruments for exposure (various social isolation phenotypes) was carried out by  
21 considering associations at genome-wide significance ( $P < 5 \times 10^{-8}$ ) to minimize  
22 pleiotropic effects, while excluding SNPs with a minor allele frequency (MAF) < 5%. LD  
23 (linkage disequilibrium) clumping was performed with a window size of 10,000 kilobases [kb],

1 at an  $R^2 < 0.001$ . In instances where a requested SNP from the exposure GWAS was not present  
2 in the outcome GWAS (protein concentration), a proxy SNP with an LD coefficient of  $R^2 > 0.8$   
3 to the requested missing target SNP were sought as a substitute, utilizing the LDLink web server.  
4 LD proxies were determined using data from the 1000 Genomes European sample. The returned  
5 information included the effect of the proxy SNP on the outcome, along with details such as the  
6 proxy SNP itself, the effect allele of the proxy SNP, and the corresponding allele for the target  
7 SNP. The effects of SNPs on both outcome and exposure were then harmonized to be relative to  
8 the same allele. Positive strand alleles were inferred by utilizing allele frequencies for  
9 palindromes instead of eliminating palindromic variants. F-statistics were employed to assess the  
10 strength of SNP-exposure associations ( $F > 10$ ). Effects for each individual variant were  
11 calculated using a two-term Taylor series expansion of the Wald ratio. Following this, we  
12 employed the weighted delta inverse-variance weighted (IVW) method to perform a meta-  
13 analysis of individual SNP effects, aiming to estimate the combined effect of the Wald ratios.  
14 Sensitivity analysis involved employing various MR methods, such as MR-Egger, Weighted  
15 Median, Maximum Likelihood, and Weighted Mode methods.[55] All analyses were performed  
16 using the human genome reference build GRCh38. In cases where the genome build relied on  
17 GRCh37 assembly Hg19, a lift-over process was carried out to convert genome coordinates and  
18 annotations to GRCh38 using a specified alignment, using CrossMap 0.7.0 in Python (version  
19 3.8.8). MR analyses were conducted using the 'TwoSampleMR', 'MendelianRandomization', and  
20 'LDlinkR' R packages.[56]

21

22 *Enrichment analysis*

1 Enrichment analysis was conducted by searching open-source databases to further characterize  
2 the identified proteins based on the minimally- and fully- adjusted linear regression models. We  
3 employed Enrichr[30] using the full set of ELSA proteins as the background gene set to glean a  
4 deeper biological understanding, using Gene Ontology (GO)[57]: GO Molecular Function, GO  
5 Biological Process, and GO Cellular Component; Kyoto Encyclopaedia of Genes and Genomes  
6 (KEGG);[58] Reactome Pathway Database (REACTOME);[59] and Genotype-Tissue  
7 Expression (GTEx).[60] Statistical significance was indicated if  $P_{\text{FDR}} < 0.05$  for the enrichment  
8 analysis. Human Protein Atlas was also searched to further characterize the identified proteins  
9 (<https://www.proteinatlas.org/>).[35]  
10 All analyses were conducted using statistical software R Studio (version 4.4.0).

## 1 REFERENCES

- 2 1. National Academies of Sciences. Social isolation and loneliness in older adults:  
3 Opportunities for the health care system: National Academies Press; 2020.
- 4 2. Office of the Surgeon General. Our epidemic of loneliness and isolation: The US Surgeon  
5 General's Advisory on the healing effects of social connection and community [Internet]. 2023.
- 6 3. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation  
7 as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of  
8 longitudinal observational studies. *Heart*. 2016;102(13):1009-16.
- 9 4. Hodgson S, Watts I, Fraser S, Roderick P, Dambha-Miller H. Loneliness, social isolation,  
10 cardiovascular disease and mortality: a synthesis of the literature and conceptual framework.  
11 *Journal of the Royal Society of Medicine*. 2020;113(5):185-92.
- 12 5. Kuiper JS, Zuidersma M, Voshaar RCO, Zuidema SU, van den Heuvel ER, Stolk RP,  
13 Smidt N. Social relationships and risk of dementia: A systematic review and meta-analysis of  
14 longitudinal cohort studies. *Ageing research reviews*. 2015;22:39-57.
- 15 6. Lara E, Martín-María N, De la Torre-Luque A, Koyanagi A, Vancampfort D, Izquierdo  
16 A, Miret M. Does loneliness contribute to mild cognitive impairment and dementia? A  
17 systematic review and meta-analysis of longitudinal studies. *Ageing research reviews*.  
18 2019;52:7-16.
- 19 7. Erzen E, Çikrikci Ö. The effect of loneliness on depression: A meta-analysis.  
20 *International Journal of Social Psychiatry*. 2018;64(5):427-35.
- 21 8. Mann F, Wang J, Pearce E, Ma R, Schlief M, Lloyd-Evans B, et al. Loneliness and the  
22 onset of new mental health problems in the general population. *Social psychiatry and psychiatric  
23 epidemiology*. 2022;57(11):2161-78.
- 24 9. Gale CR, Westbury L, Cooper C. Social isolation and loneliness as risk factors for the  
25 progression of frailty: the English Longitudinal Study of Ageing. *Age and ageing*.  
26 2018;47(3):392-7.
- 27 10. Nummela O, Seppänen M, Uutela A. The effect of loneliness and change in loneliness on  
28 self-rated health (SRH): A longitudinal study among aging people. *Archives of gerontology and  
29 geriatrics*. 2011;53(2):163-7.
- 30 11. Holt-Lunstad J. Loneliness and social isolation as risk factors: The power of social  
31 connection in prevention. *American Journal of Lifestyle Medicine*. 2021;15(5):567-73.
- 32 12. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social  
33 isolation as risk factors for mortality: a meta-analytic review. *Perspectives on psychological  
34 science*. 2015;10(2):227-37.
- 35 13. McClelland H, Evans JJ, Nowland R, Ferguson E, O'Connor RC. Loneliness as a  
36 predictor of suicidal ideation and behaviour: a systematic review and meta-analysis of  
37 prospective studies. *Journal of affective disorders*. 2020;274:880-96.
- 38 14. Hawkey LC, Cacioppo JT. Loneliness matters: A theoretical and empirical review of  
39 consequences and mechanisms. *Annals of behavioral medicine*. 2010;40(2):218-27.
- 40 15. Cacioppo S, Capitanio JP, Cacioppo JT. Toward a neurology of loneliness. *Psychological  
41 bulletin*. 2014;140(6):1464.
- 42 16. Hawkey LC, Bosch JA, Engeland CG, Marucha PT, Cacioppo JT. Loneliness, dysphoria,  
43 stress and immunity: A role for cytokines. *Cytokines: Stress and immunity*. 2007:67-85.
- 44 17. Roddick CM, Chen FS. Effects of chronic and state loneliness on heart rate variability in  
45 women. *Annals of Behavioral Medicine*. 2021;55(5):460-75.

- 1 18. Freilich CD. How does loneliness “get under the skin” to become biologically  
2 embedded? *Biodemography and Social Biology*. 2023;68(4):115-48.
- 3 19. Xiong F, Zhang L. Role of the hypothalamic–pituitary–adrenal axis in developmental  
4 programming of health and disease. *Frontiers in neuroendocrinology*. 2013;34(1):27-46.
- 5 20. Umberson D. Family status and health behaviors: Social control as a dimension of social  
6 integration. *Journal of health and social behavior*. 1987:306-19.
- 7 21. Lavenda-Grosberg D, Lalar M, Leser N, Yaseen A, Malik A, Maroun M, et al. Acute  
8 social isolation and regrouping cause short-and long-term molecular changes in the rat medial  
9 amygdala. *Molecular psychiatry*. 2022;27(2):886-95.
- 10 22. Perić I, Costina V, Gass P, Findeisen P, Filipović D. Hippocampal synaptoproteomic  
11 changes of susceptibility and resilience of male rats to chronic social isolation. *Brain Research*  
12 *Bulletin*. 2021;166:128-41.
- 13 23. Filipović D, Novak B, Xiao J, Tadić P, Turck CW. Prefrontal Cortex Cytosolic Proteome  
14 and Machine Learning-Based Predictors of Resilience toward Chronic Social Isolation in Rats.  
15 *International Journal of Molecular Sciences*. 2024;25(5):3026.
- 16 24. Paul E, Bu F, Fancourt D. Loneliness and Risk for Cardiovascular Disease: Mechanisms  
17 and Future Directions. *Current Cardiology Reports*. 2021;23(6):68.
- 18 25. Wang F, Gao Y, Han Z, Yu Y, Long Z, Jiang X, et al. A systematic review and meta-  
19 analysis of 90 cohort studies of social isolation, loneliness and mortality. *Nature human*  
20 *behaviour*. 2023;7(8):1307-19.
- 21 26. Penninkilampi R, Casey A-N, Singh M, Brodaty H. The Association between Social  
22 Engagement, Loneliness, and Risk of Dementia: A Systematic Review and Meta-Analysis.  
23 *Journal of Alzheimer's Disease*. 2018;66:1-15.
- 24 27. Hong JH, Nakamura JS, Berkman LF, Chen FS, Shiba K, Chen Y, et al. Are loneliness  
25 and social isolation equal threats to health and well-being? An outcome-wide longitudinal  
26 approach. *SSM-Population Health*. 2023;23:101459.
- 27 28. Matthews T, Rasmussen LJH, Ambler A, Danese A, Eugen-Olsen J, Fancourt D, et al.  
28 Social isolation, loneliness, and inflammation: A multi-cohort investigation in early and mid-  
29 adulthood. *Brain, Behavior, and Immunity*. 2024;115:727-36.
- 30 29. Walker E, Ploubidis G, Fancourt D. Social engagement and loneliness are differentially  
31 associated with neuro-immune markers in older age: Time-varying associations from the English  
32 Longitudinal Study of Ageing. *Brain, behavior, and immunity*. 2019;82:224-9.
- 33 30. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr:  
34 a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*.  
35 2016;44(W1):W90-W7.
- 36 31. Gu X, Dou M, Yuan M, Zhang W. Identifying novel proteins underlying loneliness by  
37 integrating GWAS summary data with human brain proteomes. *Neuropsychopharmacology*.  
38 2023;48(7):1087-97.
- 39 32. Goossens L, Van Roekel E, Verhagen M, Cacioppo JT, Cacioppo S, Maes M, Boomsma  
40 DI. The genetics of loneliness: Linking evolutionary theory to genome-wide genetics,  
41 epigenetics, and social science. *Perspectives on Psychological Science*. 2015;10(2):213-26.
- 42 33. Chaudhary PM, Eby M, Jasmin A, Bookwalter A, Murray J, Hood L. Death receptor 5, a  
43 new member of the TNFR family, and DR4 induce FADD-dependent apoptosis and activate the  
44 NF- $\kappa$ B pathway. *Immunity*. 1997;7(6):821-30.
- 45 34. Cole SW. The conserved transcriptional response to adversity. *Current Opinion in*  
46 *Behavioral Sciences*. 2019;28:31-7.

- 1 35. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards  
2 a knowledge-based human protein atlas. *Nature biotechnology*. 2010;28(12):1248-50.
- 3 36. Cohen S, Janicki-Deverts D, Turner RB, Doyle WJ. Does hugging provide stress-  
4 buffering social support? A study of susceptibility to upper respiratory infection and illness.  
5 *Psychological science*. 2015;26(2):135-47.
- 6 37. Garrido A, de Toda IM, Del Cerro ED, Félix J, Ceprián N, González-Sánchez M, De la  
7 Fuente M. Social environment as a modulator of immunosenescence. *Expert Reviews in*  
8 *Molecular Medicine*. 2022;24:e29.
- 9 38. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable  
10 prediction accuracy of polygenic scores within an ancestry group. *elife*. 2020;9:e48376.
- 11 39. Cerdeña J, Non A. Considerations, Caveats, and Suggestions for the Use of Polygenic  
12 Scores for Social and Behavioral Traits. *Behavior Genetics: an international journal devoted to*  
13 *research in the inheritance of behavior in animals and man*. 2024;54(1).
- 14 40. Marigorta UM, Rodríguez JA, Gibson G, Navarro A. Replicability and prediction:  
15 lessons and challenges from GWAS. *Trends in Genetics*. 2018;34(7):504-17.
- 16 41. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of  
17 ageing. *International journal of epidemiology*. 2013;42(6):1640-8.
- 18 42. Lee SL, Pearce E, Ajnakina O, Johnson S, Lewis G, Mann F, et al. The association  
19 between loneliness and depressive symptoms among adults aged 50 years and older: a 12-year  
20 population-based cohort study. *The Lancet Psychiatry*. 2021;8(1):48-57.
- 21 43. Hughes ME, Waite LJ, Hawkey LC, Cacioppo JT. A short scale for measuring loneliness  
22 in large surveys: Results from two population-based studies. *Research on aging*. 2004;26(6):655-  
23 72.
- 24 44. Bu F, Zaninotto P, Fancourt D. Longitudinal associations between loneliness, social  
25 isolation and cardiovascular events. *Heart*. 2020;106(18):1394-9.
- 26 45. Wik L, Nordberg N, Broberg J, Björkesten J, Assarsson E, Henriksson S, et al. Proximity  
27 extension assay in combination with next-generation sequencing for high-throughput proteome-  
28 wide analysis. *Molecular & Cellular Proteomics*. 2021;20.
- 29 46. Gong J, Williams DM, Scholes S, Assaad S, Bu F, Hayat S, et al. Unraveling the role of  
30 plasma proteins in dementia: insights from two cohort studies in the UK, with causal evidence  
31 from Mendelian randomization. *medRxiv*. 2024:2024.06.04.24308415.
- 32 47. Batista GE, Monard MC. A study of K-nearest neighbour as an imputation method. *His*.  
33 2002;87(251-260):48.
- 34 48. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained  
35 equations in R. *Journal of statistical software*. 2011;45:1-67.
- 36 49. Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: impute: Imputation for microarray  
37 data. R package version; 2024.
- 38 50. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed?  
39 Some practical clarifications of multiple imputation theory. *Prevention science*. 2007;8:206-13.
- 40 51. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues  
41 and guidance for practice. *Statistics in medicine*. 2011;30(4):377-99.
- 42 52. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal  
43 inference in epidemiological studies. *Human molecular genetics*. 2014;23(R1):R89-R98.
- 44 53. Day FR, Ong KK, Perry JR. Elucidating the genetic basis of social interaction and  
45 isolation. *Nature communications*. 2018;9(1):2457.



- 1 54. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma  
2 proteomic associations with genetics and health in the UK Biobank. *Nature*.  
3 2023;622(7982):329-38.
- 4 55. Burgess S, Smith GD, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines  
5 for performing Mendelian randomization investigations: update for summer 2023. *Wellcome*  
6 *open research*. 2019;4.
- 7 56. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing  
8 Mendelian randomization analyses using summarized data. *International journal of*  
9 *epidemiology*. 2017;46(6):1734-9.
- 10 57. Consortium GO. The gene ontology resource: 20 years and still GOing strong. *Nucleic*  
11 *acids research*. 2019;47(D1):D330-D8.
- 12 58. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids*  
13 *research*. 2000;28(1):27-30.
- 14 59. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The  
15 reactome pathway knowledgebase 2022. *Nucleic acids research*. 2022;50(D1):D687-D92.
- 16 60. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue  
17 expression (GTEx) project. *Nature genetics*. 2013;45(6):580-5.
- 18

1 **ACKNOWLEDGMENT**

2 We would like to thank the participants in ELSA for their contribution to the research. We  
3 additionally want to acknowledge the participants and investigators of the UK Biobank study. We  
4 would like to thank Olink representatives and the Newcastle laboratory for their support for the  
5 proteomics data procurement and data pre-processing.

6

7 **AUTHOR CONTRIBUTIONS**

8 The study's conceptualization and design were undertaken by D.F., A.S., and J.G. Proteomics  
9 data curation in ELSA was a joint effort between J.G. and A.S. Z.P. handled all data processing  
10 and analysis and wrote the initial manuscript draft. D.F., A.S., and J.G. provided oversight and  
11 guidance for statistical analyses and theoretical frameworks. All authors participated in the data  
12 interpretation and subsequent manuscript's revision process and collectively accepted  
13 responsibility for its submission for publication.

14

15 **COMPETING INTERESTS STATEMENT**

16 Olink had no part in designing the study or analyzing the data. No conflicts of interest to be  
17 declared from any of the authors.

18

19 **FUNDING SOURCES**

20 The English Longitudinal Study of Ageing is funded by the National Institute on Aging (grant  
21 number R01AG17644) and the National Institute for Health and Care Research (198/1074-02).  
22 The National Institute of Aging (NIA) (grant No. [R01AG17644]) funded the proteomics data  
23 curation in ELSA. J.G. is supported by the NIA (grant No. [R01AG17644]). Z.P. is supported by

1 the Economic and Social Research Council (ESRC) and the Biotechnology and Biological  
2 Sciences Research Council (BBSRC) UCL Soc-B Doctoral Studentship programme  
3 (ES/P000347/1).

4

#### 5 **DATA AVAILABILITY**

6 The ELSA data is available on the UK Data Service. The proteomics data in ELSA will be  
7 deposited on the UK Data Service upon publication.

8 All GWAS summary statistics are available online at:

9 - <https://doi.org/10.7303/syn51364943>;

10 - <https://gwas.mrcieu.ac.uk/>;

11 - <https://doi.org/10.17863/CAM.23511>.

12

#### 13 **CODE AVAILABILITY**

14 The codes used for all analyses are available on GitHub repository:

15 [https://github.com/jgong94/ELSA\\_proteomics\\_SI\\_L](https://github.com/jgong94/ELSA_proteomics_SI_L).

## 1 **FIGURE LEGENDS**

### 2 **Figure 1. Volcano plot showing minimally adjusted coefficient (x axis) and uncorrected two-** 3 **sided P values (y axis) for the association between protein concentration with loneliness** 4 **using imputed data.**

5  
6 Coefficient from linear regression models adjusted for age, sex, education, ethnicity, education,  
7 wealth quintile.  
8 Proteins above the horizontal dotted red line were significantly associated with loneliness after  
9 false discovery rate (FDR)-correction with p-value < 0.05.

### 10 11 **Figure 2. Volcano plot showing minimally adjusted coefficient (x axis) and uncorrected two-** 12 **sided P values (y axis) for the association between protein concentration with social** 13 **isolation using imputed data.**

14  
15 Coefficient from Linear Regression models adjusted for age, sex, education, ethnicity, education,  
16 wealth quintile.  
17 Proteins above the horizontal dotted red line were significantly associated with social isolation  
18 after false discovery rate (FDR)-correction with p-value < 0.05.

### 19 20 **Figure 3. Two-sample Mendelian randomisation scatter plots for frequency of friend/family** 21 **visits and TNFRSF10A, TRAIL-R2 (TNFRSF10B), TNFRSF11A, KIM1 (HAVCR1),** 22 **NEFL.**

23  
24 Analyses were conducted using the inverse variant weighted, maximum likelihood, MR-Egger,  
25 Weighted median, Weighted mode methods. The slope of each line corresponding to the  
26 estimated MR effect per method.

### 27 28 **Figure 4. Enrichment analysis of the identified proteins.**

29 Enrichment for Gene Ontology (GO) 2023, Genotype-Tissue Expression (GTEx) 2023, Kyoto  
30 Encyclopaedia of Genes and Genomes (KEGG) 2021, and Reactome pathways 2022. Significant  
31 proteins after false discovery rate (FDR) correction derived from linear regression models in the  
32 minimally- and fully-adjusted models were fed into Enrichr (<https://maayanlab.cloud/enrichr/>)  
33 for enrichment analysis. The full list of proteins from ELSA was used as the background gene  
34 set. Terms displayed on the bar plot were filtered by a two-sided p-value < 0.05. Terms above the  
35 horizontal dotted red line were enriched after FDR-correction with p-value < 0.05, and the text  
36 were highlighted in red.

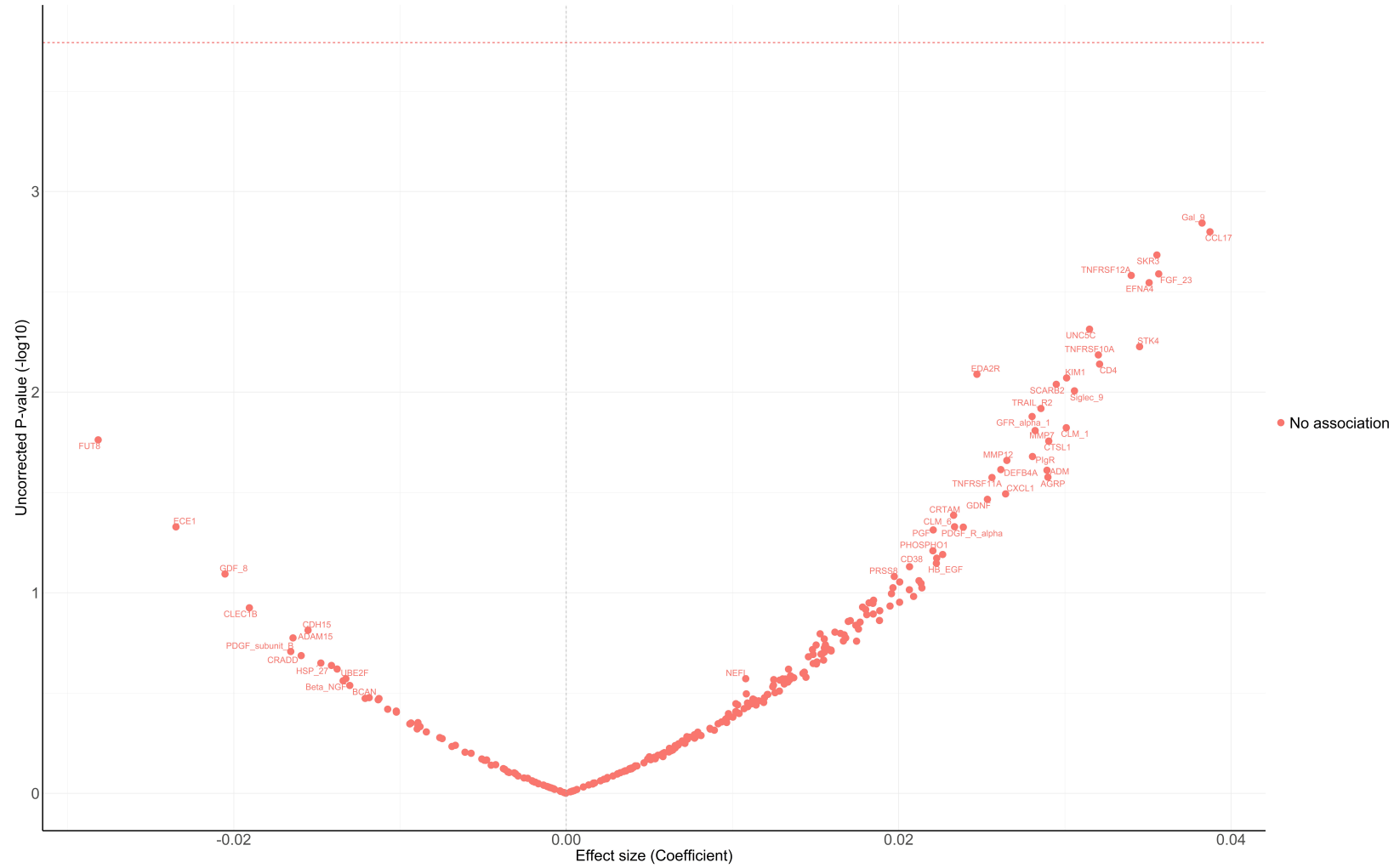
1 **TABLES**

2 **Table 1. Descriptive statistics of participants in the sample.**

	Raw dataset	Composite of 30 imputed datasets
<b>Loneliness (scale 3-9)</b>		
Mean (SD)	4.2 (1.5)	4.3 (1.6)
Missing	384 (11.8%)	0 (0.0%)
<b>Social Isolation (scale 0-7)</b>		
Mean (SD)	2.7 (1.6)	2.7 (1.6)
Missing n (%)	118 (3.6%)	0 (0.0%)
<b>Age</b>		
Mean (SD)	63.4 (9.2)	63.4 (9.2)
<b>Sex, n (%)</b>		
Female	1793 (55.0%)	1793 (55.0%)
Male	1469 (45.0%)	1469 (45.0%)
<b>Ethnicity, n (%)</b>		
White	3163 (97.0%)	3163 (97.0%)
Other ethnic groups	99 (3.0%)	99 (3.0%)
<b>Body Mass Index</b>		
Mean (SD)	28.2 (5.2)	28.2 (5.3)
Missing n (%)	81 (2.5%)	0 (0.0%)
<b>Drinking (number of days with a drink over past 7 days)</b>		
Mean (SD)	2.2 (2.4)	2.2 (2.4)
Missing n (%)	22 (0.7%)	0 (0.0%)
<b>Smoking Status, n (%)</b>		
Never smoked	1317 (40.4%)	1327 (40.7%)
Formerly smoked	1468 (45.0%)	1478 (45.3%)
Currently smoking	454 (13.9%)	457 (14.0%)
Missing	23 (0.7%)	0 (0.0%)
<b>Education, n (%)</b>		
Never went to school	16 (0.5%)	16 (0.5%)
Left school at 14 or under	273 (8.4%)	273 (8.4%)
Left school at 15	1026 (31.5%)	1026 (31.5%)
Left school at 16	743 (22.8%)	743 (22.8%)
Left school at 17	264 (8.1%)	264 (8.1%)
Left school at 18	249 (7.6%)	249 (7.6%)
Left school at 19 or over	633 (19.4%)	633 (19.4%)
Not yet finished school	58 (1.8%)	58 (1.8%)
<b>Non-Pension Wealth Quintiles, n (%)</b>		
1st	718 (22.0%)	737 (22.6%)
2nd	670 (20.5%)	686 (21.0%)
3rd	611 (18.7%)	624 (19.1%)
4th	630 (19.3%)	642 (19.7%)
5th	560 (17.2%)	573 (17.6%)
Missing	73 (2.2%)	0 (0.0%)
<b>Physical Activity, n (%)</b>		
Less than monthly moderate and vigorous exercise	424 (13.0%)	424 (13.0%)
Monthly moderate and/or vigorous exercise	196 (6.0%)	196 (6.0%)
Weekly moderate or vigorous exercise	1588 (48.7%)	1588 (48.7%)
Weekly moderate and vigorous exercise	1054 (32.3%)	1054 (32.3%)

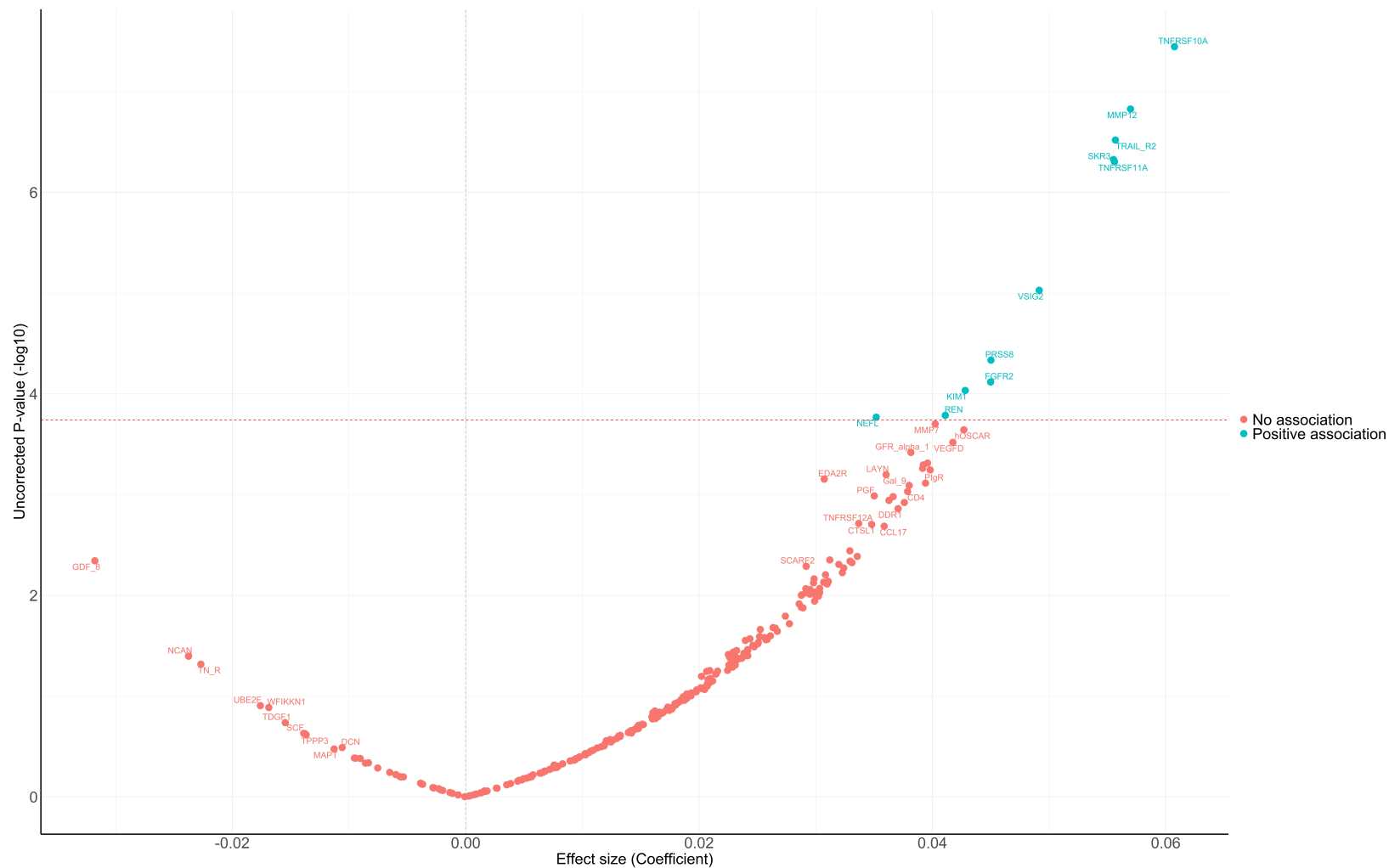
3

Figure 1. Volcano plot showing minimally adjusted coefficient (x axis) and uncorrected two-sided P values (y axis) for the association between protein concentration with loneliness using imputed data.



*Coefficient from linear regression models adjusted for age, sex, education, ethnicity, education, wealth quintile.  
Proteins above the horizontal dotted red line were significantly associated with loneliness false discovery rate (FDR)-correction with p-value < 0.05.*

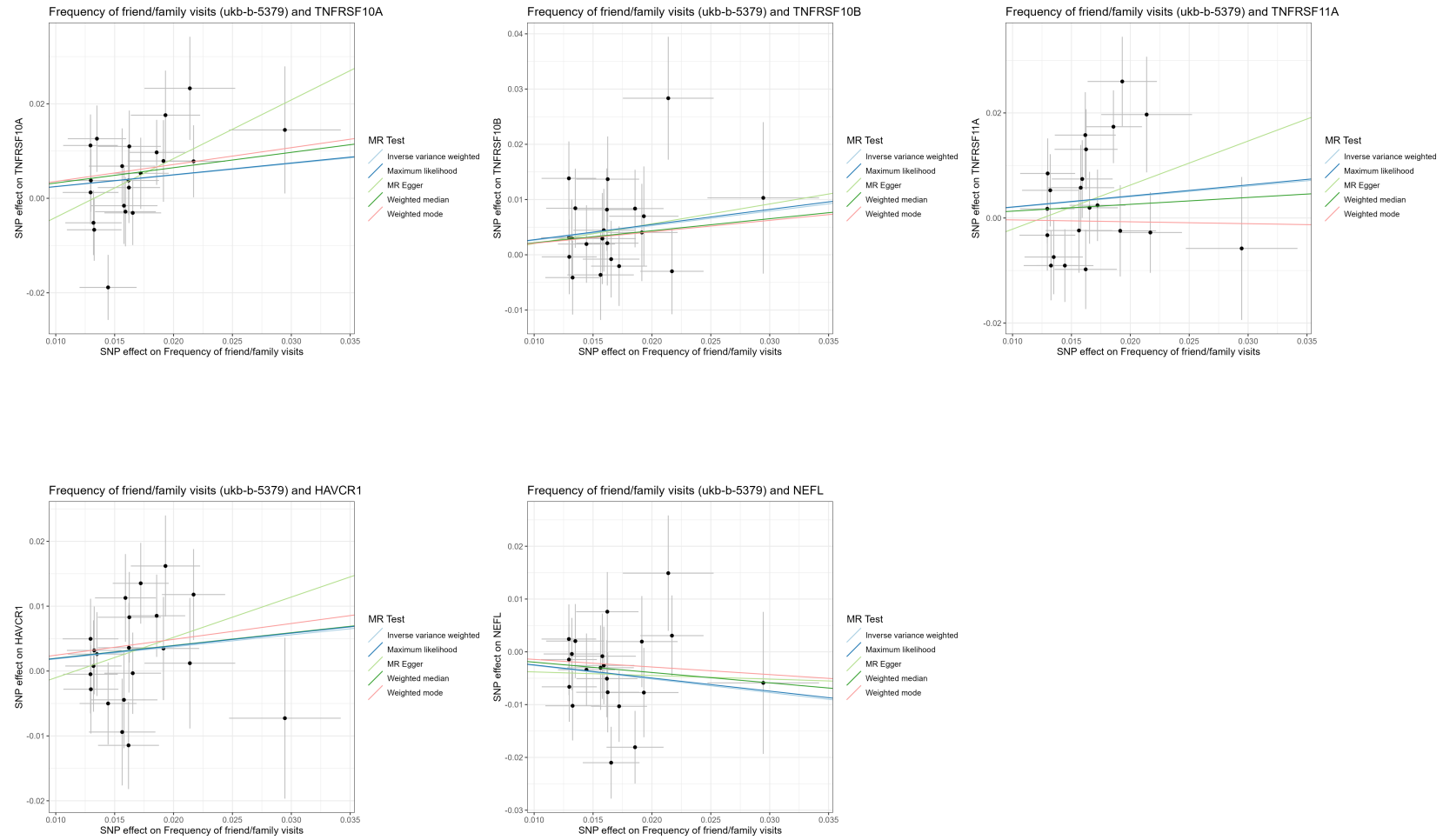
Figure 2. Volcano plot showing minimally adjusted coefficient (x axis) and uncorrected two-sided P values (y axis) for the association between protein concentration with social isolation using imputed data.



*Coefficient from Linear Regression models adjusted for age, sex, education, ethnicity, education, wealth quintile.*

*Proteins above the horizontal dotted red line were significantly associated with social isolation after false discovery rate (FDR)-correction with p-value < 0.05.*

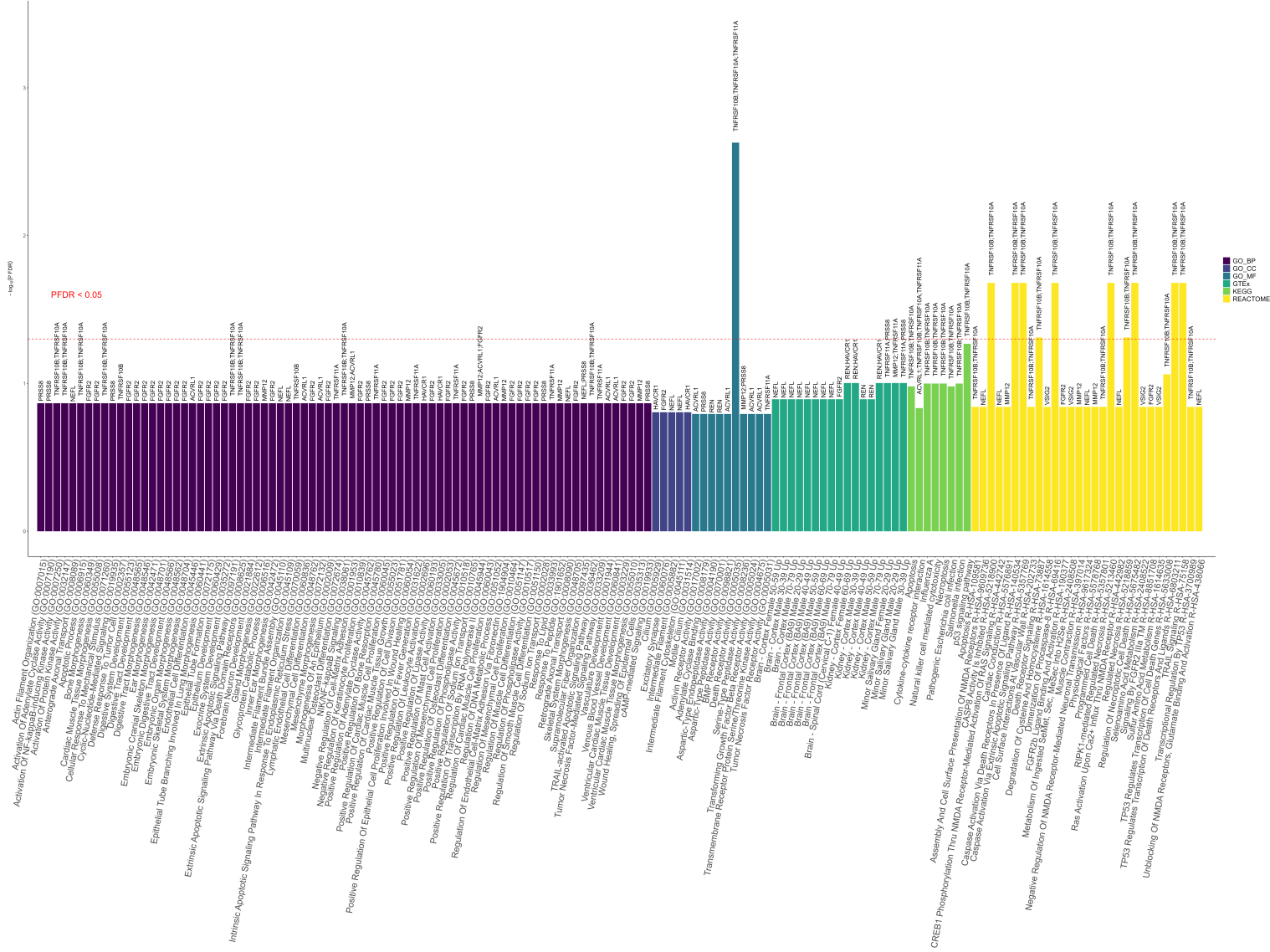
Figure 3. Two-sample Mendelian randomisation scatter plots for frequency of friend/family visits and TNFRSF10A, TRAIL-R2 (TNFRSF10B), TNFRSF11A, KIM1 (HAVCR1), NEFL.



Analyses were conducted using the inverse variant weighted, maximum likelihood, MR-Egger, Weighted median, Weighted mode methods. The slope of each line corresponds to the estimated MR effect per method.



Figure 4. Enrichment analysis of the identified proteins.



medRxiv preprint doi: <https://doi.org/10.1101/2024.07.25.24310989>; this version posted July 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Enrichment for Gene Ontology (GO) 2023, Genotype-Tissue Expression (GTEX) 2023, Kyoto Encyclopaedia of Genes and Genomes (KEGG) 2021, and Reactome pathways 2022. Significant proteins after false discovery rate (FDR) correction derived from linear regression models in the minimally- and fully-adjusted models were fed into Enrichr (<https://maayanlab.cloud/enrichr/>) for enrichment analysis. The full list of proteins from ELSA was used as the background gene set. Terms displayed on the bar plot were filtered by a two-sided  $p$ -value  $< 0.05$ . Terms above the horizontal dotted red line were enriched after FDR-correction with  $p$ -value  $< 0.05$ .