NEUROSCIENCE

Gut-brain nexus: Mapping multimodal links to neurodegeneration at biobank scale

Mohammad Shafieinouri[1]*†, Samantha Hong[1]†, Paul Suhwan Lee[1]†, Spencer M. Grant[1]†, Marzieh Khani[1]†, Anant Dadu[2], Artur F. Schumacher Schuh[3,4], Mary B. Makarious[1,2], Rodrigo Sandon[1], Emily Simmonds[5], Hirotaka Iwaki[1,2], Gracelyn Hill[1,2], Cornelis Blauwendraat[1,6], Valentina Escott-Price[5,7], Yue A. Qi[1], Alastair J. Noyce[8], Armando Reyes-Palomares[9], Hampton L. Leonard[1,2], Malu Tansey[10,11], Faraz Faghri[1,2], Andrew B. Singleton[12], Mike A. Nalls[1,2], Kristin S. Levine[1,2]†, Sara Bandres-Ciga[1]*†

Alzheimer's disease (AD) and Parkinson's disease (PD) are influenced by genetic and environmental factors. We conducted a biobank-scale study to (i) identify endocrine, nutritional, metabolic, and digestive disorders with potential causal or temporal associations with AD/PD risk before diagnosis; (ii) assess plasma biomarkers' specificity for AD/PD in the context of co-occurring gut related traits and disorders; and (iii) integrate multimodal datasets to enhance AD/PD prediction. Our findings show that several disorders were associated with increased AD/PD risk before diagnosis, with variation in the strength and timing of associations across conditions. Polygenic risk scores reveal lower genetic predisposition for AD/PD in individuals with co-occurring disorders. Moreover, the proteomic profile of AD/PD cases was influenced by comorbid gut-brain axis disorders. Last, our multimodal prediction models outperform single-modality paradigms in disease classification. This endeavor illuminates the interplay between factors involved in the gut-brain axis and the development of AD/PD, opening avenues for therapeutic targeting and early diagnosis.

INTRODUCTION

Alzheimer's disease (AD) and Parkinson's disease (PD) are the two most common neurodegenerative disorders (*1*, *2*) and cumulatively affect over 400 million individuals worldwide (*3*, *4*). Although substantial genetic risk factors for AD and PD have been identified, sporadic and late-onset forms are thought to be caused by a complex interplay between genetic (*5*, *6*) and environmental (*7*, *8*) factors. This interplay underscores the imperative to explore a multitude of variables across bodily systems to comprehend their contributions to the etiology of AD/PD (*9*, *10*).

Increasingly, research in neurodegeneration emphasizes the role of gut-brain axis health in AD and PD risk (*11*, *12*). The gut-brain axis is a complex communication network that links the gastrointestinal tract and the central nervous system. This bidirectional system, including neural pathways, hormonal signaling, and immune mechanisms, facilitates constant interactions between the brain, digestive, endocrine, metabolic systems and nutritional status. Disruptions to the gut-brain axis have been linked to various conditions, including digestive disorders (*13*), endocrine pathway dysfunctions (*14*, *15*), nutritional deficiencies (*16*, *17*), and metabolic traits (*18*).

Endocrine disorders, such as thyroid hormone imbalances, have been linked to AD and PD (*15*, *19*), with conditions such as hypothyroidism and subclinical hyperthyroidism being associated with dementia risk (*20*) and both hypo- and hyperthyroidism being associated with increasing PD risk (*15*). Metabolic disorders, particularly diabetes, are also related to neurodegenerative disease (NDD) risk. An increased severity of diabetes is associated with a higher risk of PD (*21*, *22*), and type 2 diabetes is a recognized risk factor for AD (*23*). Consequently, antidiabetic medications are being explored as potential treatments for AD and PD (*24*). In addition, nutritional deficiencies such as low vitamin D levels are more prevalent in patients with AD and PD (*25*). Digestive disorders have been observed to precede PD (*26*) or be notably associated with an increased risk for dementia (*27*). These are just a few examples of factors contributing to the gut-brain axis health and their influence on neurodegeneration.

Understanding the connection between disorders of the gut-brain axis and neurodegeneration can provide useful insights into therapeutic interventions, with major implications for prevention and disease prognosis. In this study, we conduct a large-scale biobank analysis to assess how disorders affecting the gut-brain axis—particularly those related to endocrine, nutritional, metabolic, and digestive systems—influence the risk of AD and PD. Using data from the UK Biobank (UKB), Secure Anonymized Information Linkage (SAIL), and FinnGen, we conducted a population-scale, unbiased assessment that aimed to (i) investigate the association between 155 diagnoses related to endocrine, nutritional, metabolic, and digestive system disorders and the subsequent risk of AD and PD before neurodegenerative diagnosis, while also accounting for established genetic factors known to influence the development of AD and PD; (ii) assess whether these risks were temporally dependent by

[1]Center for Alzheimer's and Related Dementias, National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA. [2]DataTecnica LLC, Washington, DC 20037, USA. [3]Departamento de Farmacologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [4]Serviço de Neurologia, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil. [5]UK Dementia Research Institute (UK DRI) at Cardiff University, Cardiff, UK. [6]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA. [7]Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK. [8]Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary University of London, London, UK. [9]Department of Molecular Biology and Biochemistry, Faculty of Sciences, University of Málaga, Málaga, Spain. [10]Department of Neuroscience, Center for Translational Research in Neurodegenerative Disease, University of Florida College of Medicine, Gainesville, FL 32608, USA. [11]Department of Neurology, Norman Fixel Institute for Neurological Diseases, University of Florida Health, Gainesville, FL 32608, USA. [12]Global Parkinson's Genetics Program.
*Corresponding author. Email: mohammad.shafieinouri@nih.gov (M.S.); sara.bandresciga@nih.gov (S.B.-C.)
†These authors contributed equally to this work.

conducting time-stratified Cox regression analyses across three pre-diagnostic intervals: 1 to 5, 5 to 10, and 10 to 15 years before AD or PD diagnosis; (iii) evaluate the specificity of plasma biomarkers associated with AD or PD with and without co-occurring conditions; (iv) develop interpretable, multimodal classification models; and (v) deploy an open-access and user-friendly web application to ensure reproducibility and transparency (Fig. 1).

RESULTS

A prior diagnosis of certain endocrine, nutritional, metabolic, and digestive system–related disorders is associated with increased risk for AD and PD

After implementing Cox proportional hazards models to examine the associations between 155 diagnoses related to endocrine, nutritional, metabolic, and digestive system disorders (table S1) and the subsequent risk for AD and PD before neurodegenerative diagnosis, we applied the Benjamini-Hochberg procedure to correct for multiple comparisons across 155 ICD-10 (International Classification of Diseases, 10th Revision) codes tested for each outcome (AD and PD). Full results, including both raw and false discovery rate (FDR)–adjusted *P* values, are provided in tables S2 and S3. Cox proportional hazards models unraveled a total of 14 ICD-10 diagnoses to be significantly associated with the risk of AD (table S2 and fig. S1) in the UKB after correction for multiple testing comparisons (discovery cohort) that were also found to be significant in SAIL, FinnGen biobanks, or both (replication cohorts) (Table 1). A total of 13 ICD-10 codes were found to be significant, with a hazard ratio (HR) of >1, suggesting that being diagnosed with these conditions increases the risk for AD; these include amyloidosis; disorders of lipoprotein metabolism and other lipidemias; gastritis and duodenitis; insulin-dependent, noninsulin-dependent, and unspecified diabetes mellitus; esophagitis; other bacterial intestinal infections; other disorders of fluid, electrolyte, and acid-base balance; other functional intestinal disorders; other noninfective gastroenteritis and colitis; vitamin D deficiency; and volume depletion. A diagnosis of hemorrhoids and perianal venous thrombosis was found to have an HR of <1 for AD in all three datasets. This observation could potentially be due to the fact that a hospitalized diagnosis of hemorrhoids and perianal venous thrombosis could be an indication of other, more serious conditions linked to a high mortality rate, thus explaining the protective observed effect (*28*).

Notably, our analyses indicate significant associations between seven disorders and the risk for PD (table S3 and fig. S2) that were replicated in either SAIL or FinnGen biobanks (Table 2). A total of four ICD-10 codes had an HR of >1 including dyspepsia, insulin-dependent and noninsulin-dependent diabetes mellitus, and other functional intestinal disorders. For PD, diverticular disease of the intestine, other diseases of the intestine, and other disorders of the peritoneum showed an HR of <1 and were replicated in two datasets. Similarly to AD, individuals diagnosed with these conditions are not representative of the entire population due to their potential severity and differential survival rates (*29*–*32*).

In addition, we conducted a time-stratified Cox proportional hazards analysis to evaluate whether the timing of diagnosis for the ICD-10 codes under study affects HR values for AD or PD. We split

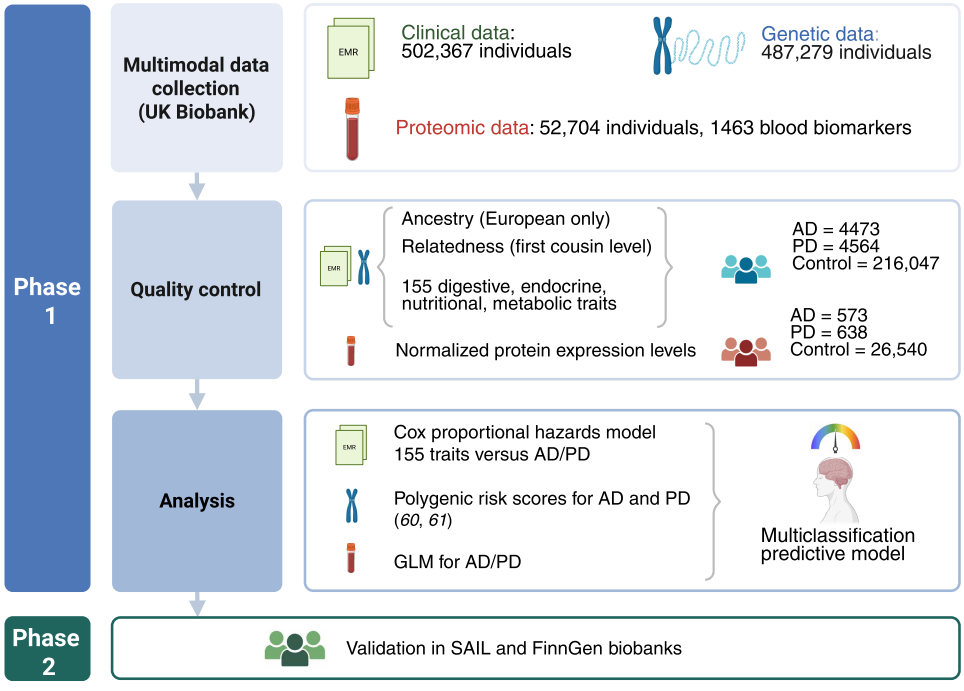

Fig. 1. Study design. The initial phase of our study used clinical data sourced from electronic medical records alongside genetic and proteomic data obtained from the UKB. Quality control procedures were rigorously applied to clinical and genetic datasets, including filtering for individuals of European ancestry, exclusion of related samples, and extraction of 155 ICD-10 codes representing diagnoses related to digestive, endocrine, nutritional, and metabolic disorders. Proteomic data underwent normalization of protein expression levels as part of quality control measures. The culmination of this phase involved the application of a Cox proportional hazards model, examination of polygenic risk scores (PRSs), and development of a generalized linear model (GLM). These analyses collectively contributed to the construction of a multimodal classification predictive model for AD and PD. Phase 2 of our study entailed validating these findings using data from the SAIL and FinnGen biobanks.

Table 1. Replicated diagnoses in endocrine, nutritional, metabolic, and digestive systems associated with AD risk. Prior ICD-10 code, initial diagnosis of endocrine, metabolic, digestive system, and nutritional disorders; CI min, CI minimum; CI max, CI maximum; *n* pairs, number of individuals identified with both ICD-10 code and NDD outcome; *n*, number of individuals identified with ICD-10 code. This table includes ICD-10 codes that were replicated across biobanks, meaning that these codes showed statistically significant associations with AD in both the UKB and at least one additional cohort.

Prior ICD-10 code	Prior ICD-10 code description	Dataset	HR	CI min	CI max	*P*	*n* pairs	*n*	FDR-corrected *P*
E85	Amyloidosis	FinnGen	3.98	1.20	13.27	2.44×10^{-2}	17	448	3.65×10^{-2}
		UKB	2.71	1.46	5.04	1.65×10^{-3}	10	191	1.07×10^{-2}
E78	Disorders of lipoprotein metabolism and other lipidemias	FinnGen	5.71	4.76	6.85	1.55×10^{-78}	2522	43,195	1.46×10^{-76}
		UKB	1.21	1.13	1.29	4.49×10^{-8}	1422	65,137	4.54×10^{-7}
K29	Gastritis and duodenitis	FinnGen	2.40	1.53	3.77	1.35×10^{-4}	416	8187	3.06×10^{-4}
		UKB	1.20	1.10	1.31	4.53×10^{-5}	625	32,770	3.17×10^{-4}
K64	Hemorrhoids and perianal venous thrombosis	SAIL	0.37	0.27	0.51	6.83×10^{-10}	39	11,389	8.64×10^{-9}
		UKB	0.78	0.69	0.88	4.25×10^{-5}	291	26,538	3.17×10^{-4}
E10	Insulin-dependent diabetes mellitus	FinnGen	2.46	1.63	3.70	1.60×10^{-5}	272	9507	4.22×10^{-5}
		SAIL	1.78	1.48	2.13	4.29×10^{-10}	118	6668	5.68×10^{-9}
		UKB	2.98	2.40	3.70	6.03×10^{-23}	84	2034	2.74×10^{-21}
E11	Noninsulin-dependent diabetes mellitus	FinnGen	4.02	3.28	4.92	2.54×10^{-41}	2641	42,593	8.54×10^{-40}
		SAIL	1.27	1.19	1.35	3.77×10^{-14}	1052	72,913	8.43×10^{-13}
		UKB	1.52	1.39	1.67	2.22×10^{-19}	558	21,278	6.72×10^{-18}
K20	Esophagitis	FinnGen	4.23	2.20	8.13	1.58×10^{-5}	77	1657	4.16×10^{-5}
		UKB	1.23	1.07	1.41	4.42×10^{-3}	208	10,362	2.24×10^{-2}
A04	Other bacterial intestinal infections	FinnGen	2.56	1.22	5.37	1.27×10^{-2}	255	6415	2.01×10^{-2}
		SAIL	1.37	1.14	1.65	6.29×10^{-4}	117	6202	2.95×10^{-3}
		UKB	1.38	1.12	1.70	2.56×10^{-3}	91	4349	1.55×10^{-2}
E87	Other disorders of fluid, electrolyte, and acid-base balance	FinnGen	10.04	7.24	13.93	2.11×10^{-43}	600	10,529	8.07×10^{-42}
		UKB	1.55	1.38	1.73	1.39×10^{-14}	350	12,397	2.54×10^{-13}
K59	Other functional intestinal disorders	FinnGen	4.37	3.30	5.80	1.02×10^{-24}	886	17,930	1.51×10^{-23}
		SAIL	1.19	1.09	1.29	9.59×10^{-5}	529	29,639	5.47×10^{-4}
		UKB	1.54	1.39	1.70	1.99×10^{-17}	445	17,716	4.52×10^{-16}
K52	Other noninfective gastroenteritis and colitis	FinnGen	2.53	1.40	4.57	2.15×10^{-3}	300	8828	3.98×10^{-3}
		SAIL	1.36	1.22	1.50	4.97×10^{-9}	373	23,128	5.56×10^{-8}
		UKB	1.38	1.23	1.56	6.06×10^{-8}	303	15,524	5.52×10^{-7}
E14	Unspecified diabetes mellitus	SAIL	1.49	1.19	1.88	6.05×10^{-4}	73	4470	2.89×10^{-3}
		UKB	1.81	1.62	2.02	1.77×10^{-26}	362	11,666	1.62×10^{-24}
E55	Vitamin D deficiency	FinnGen	13.48	4.76	38.22	9.94×10^{-7}	26	482	3.16×10^{-6}
		UKB	1.90	1.55	2.33	7.76×10^{-10}	95	2637	8.82×10^{-9}
E86	Volume depletion	FinnGen	10.65	4.09	27.71	1.24×10^{-6}	161	2913	3.88×10^{-6}
		UKB	1.77	1.52	2.08	8.97×10^{-13}	164	5323	1.17×10^{-11}

the samples from UKB into three strata: 1 to 5 years, 5 to 10 years, and 10 to 15 years before AD/PD diagnosis. We then reevaluated the HRs for AD and PD for the significant ICD-10 codes identified in the previous analysis (Tables 3 and 4).

Time-stratified Cox regression revealed distinct timing effects for several metabolic, nutritional, and gut-related conditions in relation to AD and PD. Certain diagnoses were more strongly associated with AD or PD when they were recorded at earlier time points. For example, noninsulin-dependent diabetes (E11) and unspecified diabetes (E14) were linked to greater AD risk when diagnosed 10 to 15 years before onset [HR = 1.71, 95% confidence interval (CI) = 1.45 to 2.03, FDR-adjusted $P = 2.93 \times 10^{-9}$ and HR = 3.23, 95% CI = 2.63 to 3.96, FDR-adjusted $P = 3.77 \times 10^{-28}$, respectively), suggesting cumulative metabolic effects. In contrast, insulin-dependent diabetes (E10) showed consistently elevated HRs across all time windows for AD: HR = 3.51 (95% CI = 2.21 to 5.58), HR = 2.60 (95% CI = 1.61 to 4.18), and HR = 3.62 (95% CI = 2.28 to 5.76) for diagnoses 1 to 5, 5 to 10, and 10 to 15 years before AD, respectively (FDR-adjusted $P = 2.61 \times 10^{-7}$, 4.06×10^{-4}, and 2.08×10^{-7}). Similarly, vitamin D deficiency (E55) was associated with AD both at

Table 2. Replicated diagnoses in endocrine, nutritional, metabolic, and digestive systems associated with PD risk. This table includes ICD-10 codes that were replicated across biobanks, meaning that these codes showed statistically significant associations with PD in both the UKB and at least one additional cohort.

Prior ICD-10 code	Prior ICD-10 code description	Dataset	HR	CI min	CI max	*P*	*n* pairs	*n*	FDR-corrected *P*
K57	Diverticular disease of intestine	SAIL	0.75	0.69	0.81	1.74×10^{-11}	555	70,198	2.82×10^{-10}
		UKB	0.70	0.63	0.77	6.17×10^{-12}	422	39,568	1.37×10^{-10}
K30	Dyspepsia	FinnGen	2.72	1.96	3.76	1.95×10^{-9}	86	9680	1.21×10^{-8}
		UKB	1.31	1.12	1.53	8.35×10^{-4}	164	10,647	7.44×10^{-3}
E10	Insulin-dependent diabetes mellitus	SAIL	1.59	1.31	1.93	3.36×10^{-6}	101	6668	2.48×10^{-5}
		UKB	2.69	2.10	3.43	2.20×10^{-15}	66	2016	9.78×10^{-14}
E11	Noninsulin-dependent diabetes mellitus	FinnGen	2.04	1.75	2.39	2.06×10^{-19}	421	42,593	3.38×10^{-18}
		UKB	1.30	1.17	1.44	1.61×10^{-6}	406	21,127	2.37×10^{-5}
K63	Other diseases of intestine	SAIL	0.84	0.73	0.97	1.51×10^{-2}	205	25,877	4.71×10^{-2}
		UKB	0.71	0.62	0.82	1.74×10^{-6}	212	21,069	2.37×10^{-5}
K66	Other disorders of peritoneum	SAIL	0.70	0.53	0.93	1.32×10^{-2}	48	7406	4.21×10^{-2}
		UKB	0.64	0.47	0.88	5.25×10^{-3}	41	5183	3.89×10^{-2}
K59	Other functional intestinal disorders	FinnGen	3.19	2.46	4.14	2.02×10^{-18}	178	17,930	3.02×10^{-17}
		UKB	1.50	1.34	1.68	3.74×10^{-12}	333	17,605	1.11×10^{-10}

1 to 5 years and across the full observation window, with HR = 2.11 (95% CI = 1.62 to 2.76) and HR = 1.90 (95% CI = 1.55 to 2.33), respectively (FDR-adjusted $P = 1.07 \times 10^{-7}$ and 8.82×10^{-9}).

For PD, noninsulin-dependent diabetes (E11) and unspecified diabetes (E14) showed elevated HRs across all intervals. E14 demonstrated particularly strong and stable HRs = 2.08, 3.98, and 3.02 for diagnoses at 1 to 5, 5 to 10, and 10 to 15 years, respectively (FDR-adjusted $P = 4.46 \times 10^{-3}$, 1.61×10^{-16}, and 1.05×10^{-21}). E11 was also associated with increased PD risk at 1 to 5 and 5 to 10 years: HR = 1.39 and 1.45 (FDR-adjusted $P = 4.46 \times 10^{-3}$ and 1.5×10^{-4}). Insulin-dependent diabetes (E10) showed the strongest association when diagnosed 5 to 10 years before PD (HR = 3.41, FDR-adjusted $P = 3.33 \times 10^{-7}$), with a slightly lower but still significant effect at 10 to 15 years (HR = 2.83, FDR-adjusted $P = 1.17 \times 10^{-4}$) and 1 to 5 years (HR = 2.10, FDR-adjusted $P = 4.09 \times 10^{-2}$). Deficiency of other B group vitamins (E53) also showed the highest risk when diagnosed 1 to 5 years before PD (HR = 2.21, FDR-adjusted $P = 1.95 \times 10^{-4}$). In addition, Table 3 shows that an earlier diagnosis of other disorders of fluid, electrolyte, and acid-base balance (E87) was associated with increased AD risk, while Table 4 shows that dyspepsia (K30) conferred a strong risk at all intervals before PD diagnosis. Although not all gut-brain axis–related conditions showed greater risk when diagnosed earlier, the direction of association (i.e., elevated HR) was generally consistent across time windows for significant ICD-10 codes. These findings underscore the importance of diagnosis timing in neurodegenerative risk modeling and suggest that both early-life exposures and recent comorbidities contribute to disease vulnerability.

To assess whether the associations between ICD-10 diagnoses and AD were influenced by the competing risk of death, we conducted a Fine-Gray subdistribution hazard analysis for ICD-10 codes. Among the 16 replicated ICD-10 codes from Table 1, the direction of association remained consistent between Cox and Fine-Gray models. For example, E14 (unspecified diabetes) showed an HR of 1.71 in the Cox model (95% CI = 1.52 to 1.94, FDR-adjusted $P = 3.56 \times 10^{-16}$) and an HR of 1.45 in the Fine-Gray model (95% CI = 1.26 to 1.66, FDR-adjusted $P = 2.1 \times 10^{-6}$), indicating consistent effect direction and statistical significance. However, a few diagnoses such as E85 (amyloidosis; Fine-Gray HR = 2.26, 95% CI = 1.17 to 4.33, FDR-adjusted $P = 7.67 \times 10^{-2}$), E22 (hyperfunction of pituitary gland; Fine-Gray HR = 1.57, 95% CI = 0.89 to 2.78, FDR-adjusted $P = 3.23 \times 10^{-1}$), and K92 (other diseases of digestive system; Fine-Gray HR = 1.07, 95% CI = 0.92 to 1.23, FDR-adjusted $P = 5.91 \times 10^{-1}$) lost significance in the Fine-Gray model despite being significant in the Cox model, suggesting that competing mortality may modestly bias Cox estimates for certain conditions. Full results are provided in tables S4 and S5.

Similarly, we performed Fine-Gray models for PD across ICD-10 codes that were previously significant in the Cox regression (Table 2). In all cases, the direction of effect remained consistent between models. For example, E10 (insulin-dependent diabetes) showed strong associations with PD in both Cox (HR = 2.69, 95% CI = 2.10 to 3.43, FDR-adjusted $P = 9.78 \times 10^{-14}$) and Fine-Gray (HR = 2.41, 95% CI = 1.84 to 3.17, FDR-adjusted $P = 2.3 \times 10^{-8}$) analyses, both with an FDR of <0.001. However, the code K74 (fibrosis and cirrhosis of the liver; Fine-Gray HR = 1.31, 95% CI = 0.86 to 1.99, FDR-adjusted $P = 5.22 \times 10^{-1}$) no longer met the FDR threshold in Fine-Gray models, indicating potential attenuation of effect when accounting for death as a competing risk. Full results are provided in tables S4 and S5.

Survival analysis indicates increased AD and PD incidence in individuals with significant diagnosis of endocrine, nutritional, metabolic, and digestive system–related disorders

Using UKB data, we conducted survival analyses to explore the probabilities of an AD and PD diagnosis at a certain time interval. To illustrate differences in the probability of developing AD or PD based on specific ICD-10 codes, we used Kaplan-Meier plots. The Kaplan-Meier curves demonstrate the impact of specific ICD-10 code diagnosis on the likelihood of developing these NDDs.

Table 3. Time-stratified Cox regression model for AD. ICD-10 code, initial diagnosis of endocrine, metabolic, digestive system, and nutritional disorders; ICD-10 diagnosis range, number of years from the initial diagnosis of endocrine, metabolic, digestive system, and nutritional disorders (ICD-10 code) to the occurrence of NDD outcome. This table includes all ICD-10 codes that were significant in the UKB, regardless of replication status. This analysis is stratified by time before diagnosis, allowing us to assess whether the timing of a diagnosis influences its HR for AD.

ICD-10 code	ICD-10 code description	ICD-10 diagnosis range	HR	CI min	CI max	*P*	*n* pairs	*n*	FDR-corrected *P*
A04	Other bacterial intestinal infections	1–5	2.83	1.85	4.35	1.94×10^{-6}	21	473	4.09×10^{-6}
		All	1.38	1.12	1.70	2.56×10^{-3}	91	4349	1.55×10^{-2}
E10	Insulin-dependent diabetes mellitus	1–5	3.51	2.21	5.58	1.10×10^{-7}	18	356	2.61×10^{-7}
		5–10	2.60	1.61	4.18	8.72×10^{-5}	17	436	4.06×10^{-4}
		10–15	3.62	2.28	5.76	5.19×10^{-8}	18	348	2.08×10^{-7}
		All	2.98	2.40	3.70	6.03×10^{-23}	84	2034	2.74×10^{-21}
E11	Noninsulin-dependent diabetes mellitus	1–5	1.32	1.09	1.59	5.02×10^{-3}	110	4954	6.82×10^{-3}
		5–10	1.38	1.17	1.63	1.13×10^{-4}	149	5987	4.06×10^{-4}
		10–15	1.71	1.45	2.03	5.50×10^{-10}	139	4390	2.93×10^{-9}
		All	1.52	1.39	1.67	2.22×10^{-19}	558	21,278	6.72×10^{-18}
E14	Unspecified diabetes mellitus	1–5	1.81	1.15	2.84	9.93×10^{-3}	19	687	1.26×10^{-2}
		5–10	2.06	1.38	3.08	4.23×10^{-4}	24	909	1.27×10^{-3}
		10–15	3.23	2.63	3.96	2.35×10^{-29}	95	2005	3.77×10^{-28}
		All	1.81	1.62	2.02	1.77×10^{-26}	362	11,666	1.62×10^{-24}
E16	Other disorders of pancreatic internal secretion	1–5	3.84	2.69	5.47	9.77×10^{-14}	31	407	1.67×10^{-12}
		10–15	2.81	1.46	5.40	1.98×10^{-3}	9	201	6.33×10^{-3}
		All	2.31	1.84	2.90	4.39×10^{-13}	77	1838	6.66×10^{-12}
E53	Deficiency of other B group vitamins	1–5	2.22	1.59	3.09	2.81×10^{-6}	35	759	5.35×10^{-6}
		All	1.67	1.32	2.12	1.75×10^{-5}	71	2152	1.45×10^{-4}
E55	Vitamin D deficiency	1–5	2.11	1.62	2.76	3.94×10^{-8}	55	1291	1.07×10^{-7}
		All	1.90	1.55	2.33	7.76×10^{-10}	95	2637	8.82×10^{-9}
E66	Obesity	5–10	0.76	0.61	0.95	1.72×10^{-2}	78	7362	3.86×10^{-2}
		10–15	0.61	0.45	0.84	2.67×10^{-3}	38	4451	7.12×10^{-3}
		All	0.85	0.76	0.965	5.18×10^{-3}	323	26,729	2.48×10^{-2}
E78	Disorders of lipoprotein metabolism and other lipidemias	5–10	1.57	1.38	1.79	3.30×10^{-12}	255	9808	2.97×10^{-11}
		10–15	1.72	1.54	1.91	1.19×10^{-23}	388	13,238	7.12×10^{-3}
		All	1.21	1.13	1.29	4.49×10^{-8}	1,422	65,137	4.54×10^{-7}
E85	Amyloidosis	All	2.71	1.46	5.04	1.65×10^{-3}	10	191	1.07×10^{-2}
E86	Volume depletion	1–5	1.96	1.55	2.48	2.15×10^{-8}	71	1820	6.80×10^{-8}
		All	1.77	1.52	2.08	8.97×10^{-13}	164	5323	1.17×10^{-11}
E87	Other disorders of fluid, electrolyte, and acid-base balance	1–5	1.67	1.42	1.96	6.21×10^{-10}	154	4454	2.36×10^{-9}
		All	1.55	1.38	1.73	1.39×10^{-14}	350	12,397	2.54×10^{-13}
K20	Esophagitis	1–5	2.59	2.01	3.34	1.76×10^{-13}	61	1552	1.67×10^{-12}
		All	1.23	1.07	1.41	4.42×10^{-3}	208	10,362	2.24×10^{-2}
K29	Gastritis and duodenitis	1–5	1.65	1.41	1.93	3.90×10^{-10}	163	6395	1.85×10^{-9}
		All	1.20	1.10	1.31	4.53×10^{-5}	625	32,770	3.17×10^{-4}
K52	Other noninfective gastroenteritis and colitis	1–5	1.66	1.27	2.18	2.48×10^{-4}	53	2311	3.63×10^{-4}
		5–10	2.91	2.37	3.57	1.97×10^{-24}	94	2292	3.55×10^{-23}
		All	1.38	1.23	1.56	6.06×10^{-8}	303	15,524	5.52×10^{-7}
K59	Other functional intestinal disorders	1–5	1.70	1.45	2.00	1.18×10^{-10}	154	4877	7.46×10^{-10}
		5–10	1.48	1.21	1.80	1.09×10^{-4}	102	3775	4.06×10^{-4}
		All	1.54	1.39	1.70	1.99×10^{-17}	445	17,716	4.52×10^{-16}
K64	Hemorrhoids and perianal venous thrombosis	1–5	0.60	0.46	0.78	1.51×10^{-4}	55	6399	2.40×10^{-4}
		5–10	0.67	0.51	0.90	7.14×10^{-3}	47	5081	1.84×10^{-2}
		All	0.78	0.69	0.88	4.25×10^{-5}	291	26,538	3.17×10^{-4}

(*Continued*)

(*Continued*)

ICD-10 code	ICD-10 code description	ICD-10 diagnosis range	HR	CI min	CI max	*P*	*n* pairs	*n*	FDR-corrected *P*
K66	Other disorders of peritoneum	1–5	0.55	0.33	0.92	2.21×10^{-2}	15	1821	2.62×10^{-2}
		All	0.66	0.50	0.87	3.21×10^{-3}	52	5194	1.83×10^{-2}
K92	Other diseases of digestive system	1–5	1.58	1.26	1.97	5.77×10^{-5}	80	3246	9.96×10^{-5}
		All	1.19	1.04	1.36	1.06×10^{-2}	234	13,270	4.82×10^{-2}

Notably, individuals with the significantly associated ICD-10 code diagnoses previously described exhibited a higher incidence of AD and PD. For example, individuals diagnosed with "disorders of mineral metabolism" (ICD-10 code E83) are more likely to be diagnosed with AD than those without this ICD-10 code, after 17.5 years of initial diagnosis. Similarly, for PD, a diagnosis of "deficiency of other B vitamin groups" (ICD-10 code E53) suggests a higher probability of being diagnosed with PD after 17.5 years. Other significant comorbid diagnoses can be found in figs. S1 and S2 for AD and PD, respectively. The absolute difference in survival percentages at the end of 17.5 years is shown in tables S6 and S7 for AD and PD, respectively.

The absolute difference in percentage at the end of the Kaplan-Meier analyses for significant ICD-10 codes is shown in tables S6 and S7. For instance, all three types of diabetes mellitus diagnoses ("insulin-dependent diabetes mellitus," "noninsulin-dependent diabetes mellitus," and "unspecified diabetes mellitus" with ICD-10 codes E10, E11, and E14, respectively) are shown to have a significant difference in survival probabilities in both AD and PD, corroborating previous findings that diabetes is a risk for neurodegeneration (*33*, *34*). These known associations served as internal validation for the robustness of our analytic approach. We found that there were an additional 14 significant ICD-10 codes that decreased survival probability for AD and 4 additional significant ICD-10 codes that decreased survival probability for PD.

Genetic susceptibility for AD and PD is higher in isolated cases compared to those with co-occurring endocrine, nutritional, metabolic, and digestive disorders

We calculated polygenic risk scores (PRSs) using the genome-wide association study (GWAS) summary statistics for AD and PD from the only repository with genetics data available (UKB) (see Materials and Methods for further details). A total of 22 of 23 variants [Kunkle *et al.* (*35*)] were present among AD cases and controls from the UKB data, while all the 90 variants [Nalls *et al.* (*36*)] were found to be imputed among PD cases and controls. Assessments of the accuracy of PRS in AD and PD versus ICD-10 codes were reported in tables S8 to S10. We also compared the distribution of PRS in individuals diagnosed only with AD or PD versus those having concurrent AD/PD and any ICD-10 code diagnosis under study (figs. S3 and S4). The genetic susceptibility for AD and PD was adjusted for *APOE* status in AD (by assigning "0," "1," or "2" based on the number of e4 copies present) and *LRRK2* and *GBA1* status in PD (by assigning "0" or "1" based on carrier status of each respective variant). Notably, some significant ICD-10 codes associated with AD included those pertaining to diabetes mellitus; other disorders of fluid, electrolyte, and acid-base balance; and obesity. Similarly, there were significant associations between PD and diabetes mellitus, disorders of the peritoneum, and vitamin B group deficiencies. A comprehensive summary of these results is shown in figs. S5 to S8.

For all significant associations ($P < 0.05$), a lower average PRS was observed in individuals with co-occurring AD and another condition compared to individuals with only AD. For instance, in patients with AD and gastrointestinal disorders (K59), our analysis showed a lower PRS in the AD + ICD-10 group compared to individuals with only AD, both before and after adjusting by *APOE* ($\beta = -0.37 \pm 0.07$, $P = 1.33 \times 10^{-7}$ when including *APOE* and $\beta = -0.11 \pm 0.06$, $P = 4.97 \times 10^{-2}$ when excluding *APOE*) (fig. S7, P.1 and P.2). A similar trend was observed in individuals with AD and obesity (E66), where the AD + ICD-10 group showed a lower PRS, with $\beta = -0.37 \pm 0.08$ and $P = 1.20 \times 10^{-5}$ when including *APOE* (fig. S7, J.1). Another significant association was observed in individuals with AD and esophagitis (K20), where the AD + ICD-10 group had a lower PRS, with $\beta = -0.24 \pm 0.10$ and $P = 1.78 \times 10^{-2}$ when including *APOE* and $\beta = -0.17 \pm 0.08$ and $P = 3.38 \times 10^{-2}$ when excluding *APOE* (fig. S7, K.1 and K.2). Gastritis and duodenitis (K29) were also significantly associated with lower PRS in the AD + ICD-10 group ($\beta = -0.18 \pm 0.06$, $P = 4.70 \times 10^{-3}$) (fig. S7, F.1). Individuals with AD and type 2 diabetes (E11) exhibited lower PRS, with $\beta = -0.24 \pm 0.07$ and $P = 2.54 \times 10^{-4}$ when including *APOE* and $\beta = -0.13 \pm 0.05$ and $P = 1.20 \times 10^{-2}$ when excluding *APOE* (fig. S7, I.1 and I.2).

A similar pattern of lower PRS was observed in individuals with co-occurring PD and another condition compared to individuals with only PD. For instance, patients with PD and gastrointestinal disorders (K59) exhibited a lower PRS compared to those with only PD ($\beta = -0.16 \pm 0.06$, $P = 1.24 \times 10^{-2}$) (fig. S8H). In addition, patients with PD and deficiency of other B group vitamins (E53) had a lower PRS, with $\beta = -0.32 \pm 0.15$ and $P = 3.08 \times 10^{-2}$ (fig. S8J). Furthermore, individuals with PD and insulin-dependent diabetes (E10) had a lower PRS, with $\beta = -0.32 \pm 0.14$ and $P = 2.46 \times 10^{-2}$ (fig. S8L). Similarly, individuals with PD and noninsulin-dependent diabetes (E11) had a lower PRS, with $\beta = -0.17 \pm 0.06$ and $P = 3.95 \times 10^{-3}$ (fig. S8F).

No evidence of synergistic interaction between polygenic risk scores and ICD-10 diagnoses in AD and PD outcomes

We did not identify significant synergistic interaction terms exhibiting an odds ratio (OR) of >1 at a nominal $P < 0.05$ in either AD (see tables S11 and S12) or PD (see table S13). These results suggest that the co-occurrence of endocrine, nutritional, metabolic, or digestive system–related disorders with elevated genetic risk does not produce a synergistic effect on AD or PD risk beyond the additive contribution of each factor alone.

Table 4. Time-stratified Cox regression model for PD. This table includes all ICD-10 codes that were significant in the UKB, regardless of replication status. This analysis is stratified by time before diagnosis, allowing us to assess whether the timing of a diagnosis influences its HR for PD.

ICD-10 code	ICD-10 code description	ICD-10 diagnosis range	HR	CI min	CI max	*P*	*n* pairs	*n*	FDR-corrected *P*
E10	Insulin-dependent diabetes mellitus	1–5	2.10	1.09	4.04	2.63×10^{-2}	9	347	4.09×10^{-2}
		5–10	3.41	2.19	5.29	4.76×10^{-8}	20	439	3.33×10^{-7}
		10–15	2.83	1.60	4.98	3.30×10^{-4}	12	342	1.17×10^{-4}
		All	2.69	2.10	3.43	$2.20 \times 10-^{15}$	66	2016	9.78×10^{-14}
E11	Noninsulin-dependent diabetes mellitus	1–5	1.39	1.13	1.70	1.72×10^{-3}	96	4940	4.46×10^{-3}
		5–10	1.45	1.21	1.73	4.29×10^{-5}	129	5968	1.50×10^{-4}
		All	1.30	1.17	1.44	1.61×10^{-6}	406	21,127	2.37×10^{-5}
E14	Unspecified diabetes mellitus	1–5	2.08	1.31	3.31	1.91×10^{-3}	18	686	4.46×10^{-3}
		5–10	3.98	2.90	5.46	1.15×10^{-17}	39	924	1.61×10^{-16}
		10–15	3.02	2.42	3.77	1.17×10^{-22}	81	1991	1.05×10^{-21}
		All	1.71	1.52	1.94	4.00×10^{-18}	292	11,596	3.56×10^{-16}
E16	Other disorders of pancreatic internal secretion	1–5	2.81	1.77	4.47	1.25×10^{-5}	18	394	5.82×10^{-5}
		All	1.95	1.48	2.57	1.86×10^{-6}	52	1813	2.37×10^{-5}
E53	Deficiency of other B group vitamins	1–5	2.21	1.50	3.26	5.57×10^{-5}	26	750	1.95×10^{-4}
		All	1.76	1.35	2.30	3.54×10^{-5}	55	2136	3.50×10^{-4}
K30	Dyspepsia	1–5	3.18	2.14	4.71	$8.63 \times 10^{-}9$	25	702	6.04×10^{-8}
		5–10	1.53	1.15	2.05	4.07×10^{-3}	46	2676	1.14×10^{-2}
		10–15	1.52	1.14	2.03	4.42×10^{-3}	47	2687	7.96×10^{-3}
		All	1.31	1.12	1.53	8.35×10^{-4}	164	10,647	7.44×10^{-3}
K57	Diverticular disease of intestine	5–10	0.67	0.56	0.81	3.16×10^{-5}	114	11,811	1.48×10^{-4}
		10–15	0.69	0.55	0.86	1.07×10^{-3}	80	7319	2.40×10^{-3}
		All	0.70	0.63	0.77	6.17×10^{-12}	422	39,568	1.37×10^{-10}
K59	Other functional intestinal disorders	1–5	1.81	1.51	2.16	1.26×10^{-10}	124	4847	1.77×10^{-9}
		5–10	1.34	1.05	1.69	1.68×10^{-2}	70	3744	2.94×10^{-2}
		All	1.50	1.34	1.68	3.74×10^{-12}	333	17,605	1.11×10^{-10}
K63	Other diseases of intestine	5–10	0.70	0.55	0.90	4.94×10^{-3}	66	6731	1.15×10^{-2}
		10–15	0.57	0.42	0.78	3.89×10^{-4}	40	4362	1.17×10^{-3}
		All	0.71	0.62	0.82	1.74×10^{-6}	212	21,069	2.37×10^{-5}
K64	Hemorrhoids and perianal venous thrombosis	1–5	0.69	0.53	0.90	7.35×10^{-3}	53	6397	1.29×10^{-2}
		5–10	0.68	0.49	0.93	1.58×10^{-2}	39	5073	2.94×10^{-2}
		All	0.74	0.65	0.85	1.26×10^{-5}	237	26,484	1.40×10^{-4}
K66	Other disorders of peritoneum	1–5	0.40	0.21	0.77	6.15×10^{-3}	9	1815	1.23×10^{-2}
		All	0.64	0.47	0.88	5.25×10^{-3}	41	5183	3.89×10^{-2}
K74	Fibrosis and cirrhosis of liver	5–10	2.30	1.10	4.83	2.78×10^{-2}	7	290	4.32×10^{-2}
		All	1.69	1.15	2.49	7.72×10^{-3}	26	1337	4.91×10^{-2}

The proteomic profile of individuals diagnosed with AD and PD is influenced by co-occurring endocrine, nutritional, metabolic, and digestive system conditions

We accessed Olink, which encompasses data from 52,705 individuals from the UKB (at the time of data access) and a total of 1463 proteins (see Materials and Methods for further details). Our analysis showed 22 proteomic biomarkers with notable differences in AD cases versus controls, alongside 156 proteins exhibiting significant distinctions in PD cases compared to controls after controlling for multiple comparisons (tables S14 and S15, respectively). We delved deeper into the levels of proteomic biomarkers, applying a *t* test to juxtapose average levels between standalone cases of AD/PD and cases co-occurring with specific ICD-10 codes. After implementing FDR corrections, we found significant differences in proteomic biomarker levels in individuals with only AD/PD compared to individuals with AD/PD and co-occurring ICD-10 conditions. Specifically, we identified 37 biomarkers displaying increased levels in AD cases with co-occurring ICD-10 diagnoses when compared to AD cases without additional ICD-10 diagnoses (Table 5). Using the same approach for PD, we identified five biomarkers with elevated levels when other ICD-10 diagnoses were present (Table 6). These findings hint at distinct associations between specific biomarkers and AD/PD,

Table 5. Proteomic biomarker comparison in isolated AD cases versus cases with digestive, endocrine, metabolic, and nutritional conditions.

Olink marker	ICD-10 code	ICD-10 description	Olink average (AD)	Olink average (AD + ICD-10)	*t* statistic (AD + ICD-10 versus AD)	*n* (AD)	*n* (AD + ICD-10)	*P*	FDR-corrected *P*
GFAP	E11	Noninsulin-dependent diabetes mellitus	0.735	0.45	−4.05	437	109	8.20×10^{-5}	1.45×10^{-3}
	K29	Gastritis and duodenitis	0.722	0.50	−3.46	439	107	6.74×10^{-4}	8.42×10^{-3}
	E14	Unspecified diabetes mellitus	0.713	0.42	−3.31	482	64	1.43×10^{-3}	1.47×10^{-2}
Adhesion G protein–coupled receptor G_1	E66	Obesity	0.485	1.39	4.80	497	61	9.17×10^{-6}	2.29×10^{-4}
	E14	Unspecified diabetes mellitus	0.455	1.52	6.12	491	67	4.08×10^{-8}	1.75×10^{-6}
	E11	Noninsulin-dependent diabetes mellitus	0.393	1.35	7.21	447	111	3.60×10^{-11}	5.40×10^{-9}
Calbindin	K20	Esophagitis	0.162	0.41	2.96	520	38	5.05×10^{-3}	4.10×10^{-2}
	E16	Other disorders of pancreatic internal secretion	0.167	0.38	3.07	527	31	4.06×10^{-3}	3.48×10^{-2}
	E85	Amyloidosis	0.177	0.55	12.23	556	2	1.86×10^{-3}	1.80×10^{-2}
Decorin	E87	Other disorders of fluid, electrolyte, and acid-base balance	0.118	0.20	2.86	418	136	4.68×10^{-3}	3.90×10^{-2}
	E86	Volume depletion	0.119	0.21	3.05	442	112	2.71×10^{-3}	2.54×10^{-2}
Renin	E86	Volume depletion	0.340	0.67	2.77	445	116	6.23×10^{-3}	4.67×10^{-2}
	E53	Deficiency of other B group vitamins	0.367	1.01	3.41	524	37	1.46×10^{-3}	1.47×10^{-2}
	E66	Obesity	0.349	0.89	3.71	499	62	3.89×10^{-4}	5.07×10^{-3}
	K59	Other functional intestinal disorders	0.310	0.70	3.76	420	141	2.17×10^{-4}	3.25×10^{-3}
	E10	Insulin-dependent diabetes mellitus	0.369	1.62	3.78	543	18	1.43×10^{-3}	1.47×10^{-2}
	E16	Other disorders of pancreatic internal secretion	0.349	1.46	4.30	531	30	1.60×10^{-4}	2.67×10^{-3}
	E14	Unspecified diabetes mellitus	0.293	1.25	5.73	493	68	1.86×10^{-7}	6.97×10^{-6}
	E11	Noninsulin-dependent diabetes mellitus	0.253	1.02	5.97	447	114	1.74×10^{-8}	8.71×10^{-7}
	E78	Disorders of lipoprotein metabolism and other lipidemias	0.161	0.69	6.02	296	265	3.38×10^{-9}	2.54×10^{-7}

(*Continued*)

(*Continued*)

Olink marker	ICD-10 code	ICD-10 description	Olink average (AD)	Olink average (AD + ICD-10)	*t* statistic (AD + ICD-10 versus AD)	*n* (AD)	*n* (AD + ICD-10)	*P*	FDR-corrected *P*
Growth/ differentiation factor 15	K20	Esophagitis	0.415	0.82	3.12	530	40	3.32×10^{-3}	3.01×10^{-2}
	E86	Volume depletion	0.399	0.62	3.28	453	117	1.26×10^{-3}	1.45×10^{-2}
	E83	Disorders of mineral metabolism	0.418	0.83	3.89	535	35	3.85×10^{-4}	5.07×10^{-3}
	E87	Other disorders of fluid, electrolyte, and acid-base balance	0.382	0.63	4.26	431	139	3.03×10^{-5}	6.49×10^{-4}
	E10	Insulin-dependent diabetes mellitus	0.418	1.22	4.47	552	18	3.12×10^{-4}	4.46×10^{-3}
	E66	Obesity	0.400	0.80	4.48	508	62	2.76×10^{-5}	6.36×10^{-4}
	E78	Disorders of lipoprotein metabolism and other lipidemias	0.338	0.56	4.68	302	268	3.64×10^{-6}	1.21×10^{-4}
	E16	Other disorders of pancreatic internal secretion	0.409	1.06	4.82	540	30	3.68×10^{-5}	7.36×10^{-4}
	E11	Noninsulin-dependent diabetes mellitus	0.336	0.87	7.47	455	115	7.97×10^{-12}	2.39×10^{-9}
	E14	Unspecified diabetes mellitus	0.359	1.06	7.51	501	69	9.20×10^{-11}	9.20×10^{-9}
Cation-independent mannose-6-phosphate receptor	E14	Unspecified diabetes mellitus	0.065	0.17	3.39	492	68	1.04×10^{-3}	1.24×10^{-2}
	E66	Obesity	0.064	0.19	3.94	498	62	1.74×10^{-4}	2.74×10^{-3}
	E11	Noninsulin-dependent diabetes mellitus	0.053	0.17	4.75	446	114	4.15×10^{-6}	1.24×10^{-4}
Interleukin-1 receptor type 1	E10	Insulin-dependent diabetes mellitus	0.051	0.30	3.75	540	18	1.47×10^{-3}	1.47×10^{-2}
	E14	Unspecified diabetes mellitus	0.039	0.20	4.77	491	67	7.80×10^{-6}	2.13×10^{-4}
	E11	Noninsulin-dependent diabetes mellitus	0.027	0.18	5.99	444	114	1.17×10^{-8}	7.01×10^{-7}
Interleukin-1 receptor-like 1	E11	Noninsulin-dependent diabetes mellitus	0.109	0.26	2.80	454	114	5.67×10^{-3}	4.48×10^{-2}
	E14	Unspecified diabetes mellitus	0.115	0.31	2.81	500	68	6.14×10^{-3}	4.67×10^{-2}
	E16	Other disorders of pancreatic internal secretion	0.121	0.46	3.12	538	30	3.82×10^{-3}	3.37×10^{-2}
Latent transforming growth factor–β–binding protein 2	E87	Other disorders of fluid, electrolyte, and acid-base balance	0.243	0.39	4.08	431	139	6.02×10^{-5}	1.13×10^{-3}

Table 6. Proteomic biomarker comparison in isolated PD cases versus cases with digestive, endocrine, metabolic, and nutritional conditions.

Olink marker	ICD-10 code	ICD-10 description	Olink average (PD)	Olink average (PD + ICD-10)	*t* statistic (PD + ICD-10 versus PD)	*n* (PD)	*n* (PD + ICD-10)	*P*	FDR-corrected *P*
Collagenase 3	K30	Dyspepsia	0.04	0.23	3.28	569	36	2.08×10^{-3}	4.97×10^{-2}
Peroxiredoxin-1	E10	Insulin-dependent diabetes mellitus	0.22	−0.43	−3.60	589	16	2.13×10^{-3}	4.97×10^{-2}
CD276 antigen	E11	Noninsulin-dependent diabetes mellitus	0.08	0.23	3.38	523	102	9.29×10^{-4}	4.33×10^{-2}
	E14	Unspecified diabetes mellitus	0.09	0.33	3.40	569	56	1.20×10^{-3}	4.33×10^{-2}
	E10	Insulin-dependent diabetes mellitus	0.09	0.70	4.38	609	16	5.06×10^{-4}	3.64×10^{-2}
Interleukin-1 receptor-like 1	E14	Unspecified diabetes mellitus	0.13	0.39	3.92	575	58	1.99×10^{-4}	2.87×10^{-2}
Neural cell adhesion molecule 1	E66	Obesity	0.13	−0.03	−3.12	550	80	2.42×10^{-3}	4.97×10^{-2}

while others may be influenced by concurrent comorbidities—an important consideration when developing multimodal models. Notably, we identified well-established biomarkers of AD and PD, such as amyloid-β_{42} and α-synuclein, which served as a positive control to validate the reliability of our approach.

Among the critical biomarkers identified, β-1-glycoprotein 1 (PSG1) emerged as a particularly compelling candidate. PSG1 is primarily expressed during pregnancy and is known for its role in modulating immune responses and enhancing transforming growth factor–β signaling. A recent study using UKB data identified PSG1 as one of the proteins predictive of AD risk in individuals with depression, further highlighting it as an early and underexplored contributor (*37*). Given its robust association in our model (OR = 1.19, 95% CI = 1.11 to 1.29, FDR-adjusted $P = 1.20 \times 10^{-3}$), PSG1 represents a promising blood-based biomarker that may reflect immune modulation contributing to AD risk. In addition, integrin-αV showed an inverse association with PD (OR = 0.12, 95% CI = 0.081 to 0.19, FDR-adjusted $P = 5.19 \times 10^{-20}$), suggesting a role for extracellular matrix signaling. ADGRG2 (OR = 0.44, 95% CI = 0.34 to 0.58, FDR-adjusted $P = 2.56 \times 10^{-7}$) and TNXB (OR = 0.45, 95% CI = 0.35 to 0.59, FDR-adjusted $P = 1.57 \times 10^{-6}$) were also decreased and may represent notable PD-associated proteins.

Multimodal integration models based on clinical, genetic, and proteomic data improve accuracy to predict AD and PD risk versus a single paradigm

In our analysis, the combination of genetic, clinical, proteomic (controlling for multiple comparisons using Bonferroni correction), and demographic factors (including age at recruitment and gender) exhibited superior predictive performance for AD risk when compared to the individual datasets, with a test area under the curve (AUC) of 0.90 (95% CI = 0.88 to 0.92) and a test balanced accuracy (BA) of 0.83 (0.80 to 0.86). Curiously, the model combining genetic, proteomic, and demographic features (i.e., no clinical information) had a very similar test AUC of 0.89 (0.87 to 0.91) and a test BA of 0.82 (0.78 to 0.86), suggesting that the risk information captured by clinical features may be encompassed within the information provided by genetic, proteomic, and demographic features. The best classifier using only one data attribute was the model including proteomic data for prediction with a test AUC of 0.87 (0.85 to 0.89) and a test BA of 0.79 (0.76 to 0.82). The performance of each predictive model is shown in Fig. 2, and the metrics are listed in table S16.

For PD, the integration of genetic, proteomic, and demographic factors showcased heightened predictive efficacy with a test AUC of 0.78 (0.74 to 0.82) and a test BA of 0.70 (0.68 to 0.72). For PD, addition of the clinical data did not improve the AUC and test BA of the model. Clinical features do not contain sufficient useful information, and predicting solely on clinical features performs no better than a random classifier, with a test AUC of 0.52 (0.49 to 0.55). The best classifier using only one data attribute was the model that used demographics data (age at recruitment, sex, and Townsend deprivation index), with a test AUC of 0.77 (0.74 to 0.80) and a test BA of 0.72 (0.68 to 0.76). The performance of each model for PD is shown in Fig. 3 and table S17.

Shapley additive explanations (SHAP) values provide an in-depth analysis of machine learning (ML) classifiers by highlighting the top discriminating features of AD/PD compared to healthy controls. The biological plausibility of the models is supported by feature importance plots, which emphasize known risk factors for AD and PD, such as age and PRS. Among the proteomic features, the neurofilament light (NFL) polypeptide emerges as a significant factor for both AD and PD, suggesting its potential as a biomarker for multiple forms of neurodegeneration. In AD, other top features include the glial fibrillary acidic protein (GFAP) and the growth differentiation factor from proteomic data. From clinical features, only gastritis and duodenitis appear among top features (Fig. 4A). Proteomic factors, including adhesion G protein–coupled receptor G_2, integrin-αM, and interleukin receptor, help distinguish PD from controls (Fig. 4B). The interactive website (https://gut-brain-nexus.streamlit.app/) was developed as an open-access and cloud-based platform for researchers to investigate the top features of the ML models developed and how these may influence the AD/PD risk scores.

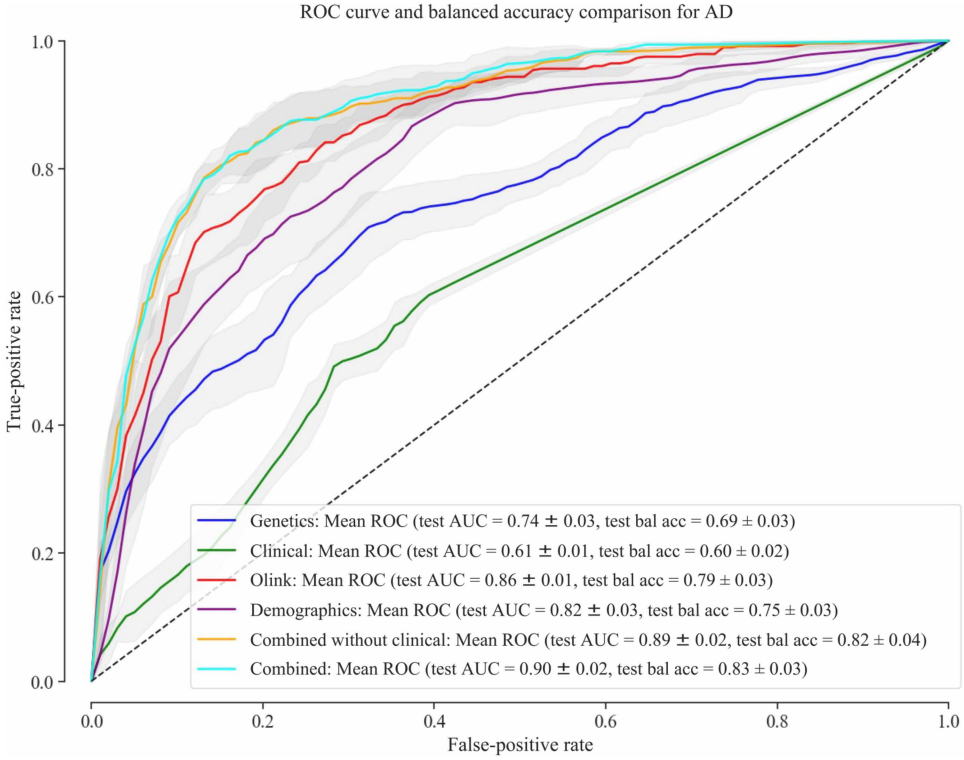

Fig. 2. Receiver operating characteristic curve and balanced accuracy comparison for AD. Performance evaluation of multiomics integration models using clinical, genetic, proteomic, and demographic data for AD.

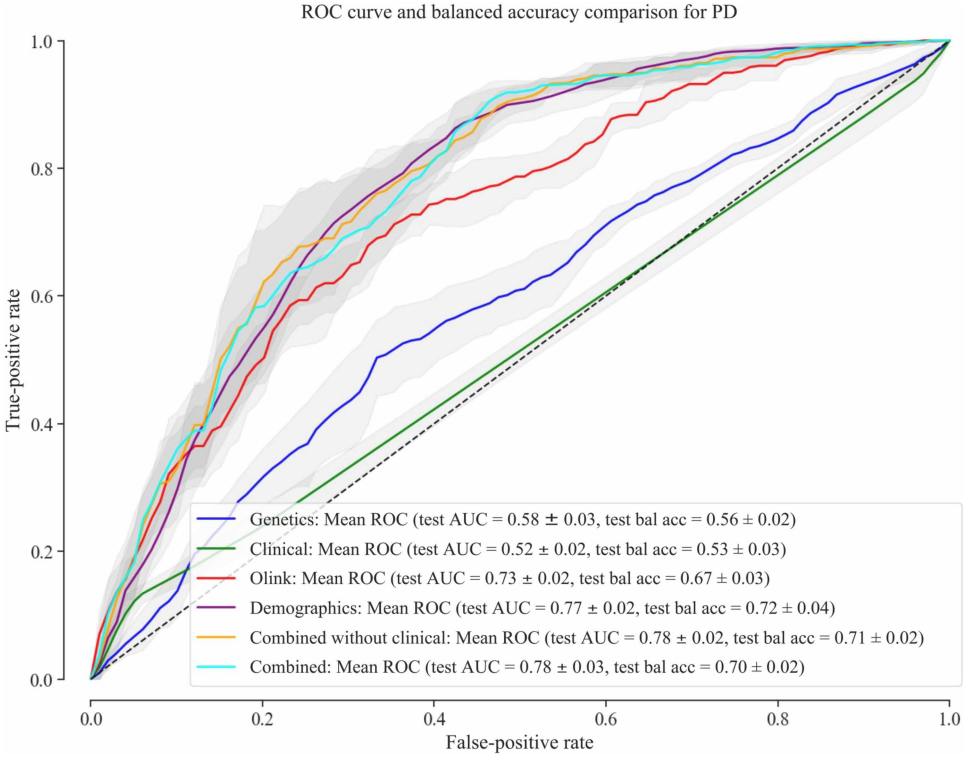

Fig. 3. Receiver operating characteristic curve and balanced accuracy comparison for PD. Performance evaluation of multiomics integration models using clinical, genetic, proteomic, and demographic data for PD.

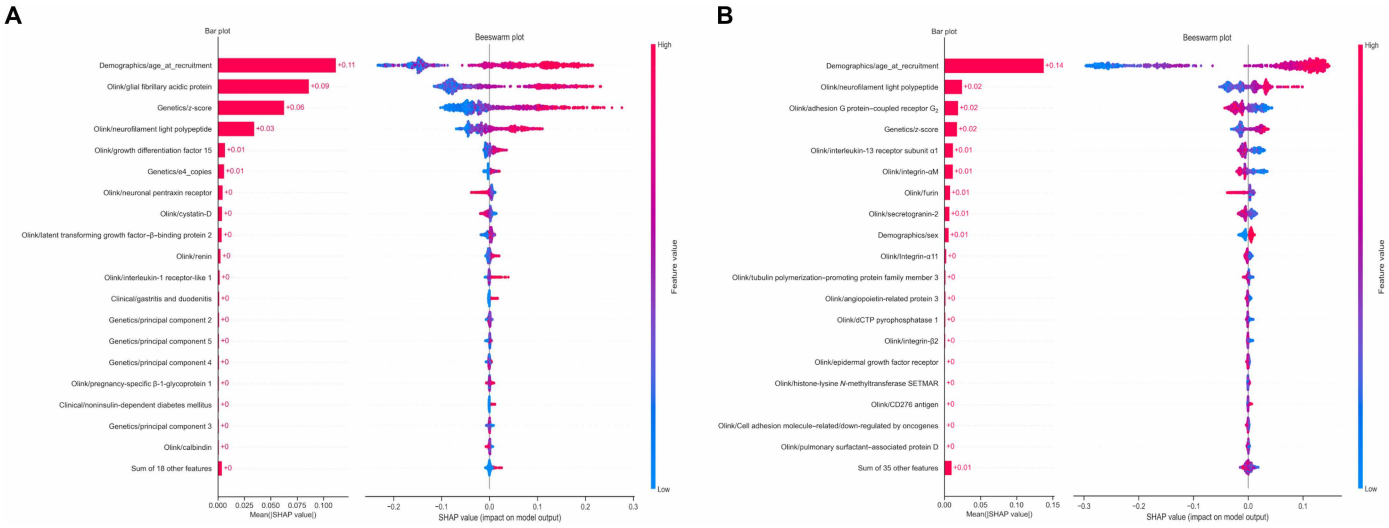

Fig. 4. Feature importance plots. (**A**) Distribution of the top 20 features that had the most substantial effect on the AD risk estimates. Each point represents a patient and the amount of effect on model output for each feature depends on its SHAP value. For example, the effect of the "Demographics/age_at_recruitment" feature on model output is large and positive (indicating a higher risk) when the patient has high values for "Demographics/age_at_recruitment" (more red points are on the right side). Similarly, (**B**) shows the top features for PD risk estimates. dCTP, 2′-deoxycytidine 5′-triphosphate.

DISCUSSION

With the rising prevalence of AD and PD, it is imperative to enhance our understanding of the determinants that increase the risk for these common NDDs and develop improved prediction models for early detection. Here, we have undertaken the most extensive biobank-scale omics study to date to assess the influence of main biological system disorders implicated in the gut-brain axis (including endocrine, nutritional, metabolic, and digestive-related conditions) preceding the diagnosis of AD and PD. The culmination of which is a multimodal classification model that combines clinical, genetic, and proteomic data enhancing the prediction accuracy of AD and PD.

In a large-scale and data-driven manner, we demonstrate that certain endocrine, nutritional, metabolic, and digestive system–related traits and disorders are notably associated with an increased risk of AD and/or PD before diagnosis. Notably, individuals with noninfective gastroenteritis and colitis; esophagitis; gastritis and duodenitis; disorders of fluid, electrolyte, and acid-base balance; pancreatic internal secretion disorders; and functional intestinal disorders showed a higher likelihood of developing AD later in life. Recent literature has suggested the amplification of AD risk from disorders affecting the gut-brain axis, such as gastritis (*38*). Our results corroborate these findings and reveal additional, notable potential disorders of interest for further study, with replication across multiple datasets. In regard to PD, substantial diagnoses associated with increased PD risk include functional intestinal disorders, disorders of pancreatic internal secretion, and deficiency of other B group vitamins. The correlation between the deficiency of B group vitamins and PD expands upon previous studies that have looked at other vitamin deficiencies, such as vitamin D (*25*), and warrants more research on the impact of nutritional deficiencies on NDDs. Previous literature has linked individual disorders, such as functional intestinal disorders (*39*), disorders of pancreatic internal secretion (*40*), and deficiency of B group vitamins (*41*, *42*), to both AD and PD. It is encouraging that our findings align with and further support these established associations. The interplay between multiple traits at once and their combined impact on AD/PD was not explored in our analysis, presenting an opportunity for future further investigation.

The number of patients with co-occurring amyloidosis and AD was low compared to other ICD-10 diagnoses analyzed. Although AD features amyloid-β pathology and is considered a form of localized cerebral amyloidosis (*43*, *44*), it is not classified as a subtype of systemic amyloidosis. The ICD-10 diagnosis of amyloidosis typically refers to systemic forms, such as light-chain or transthyretin amyloidosis, which affect organs such as the heart or kidneys. These systemic forms may occur concurrently with AD in some individuals. However, the low frequency of codiagnosis in our data may reflect underrecording, as cerebral amyloid pathology inherent to AD is not typically coded separately in clinical records.

Our study shows that risk of neurodegeneration persists up to 15 years before AD/PD onset with a co-occurring diagnosis of an endocrine, metabolic, digestive, or nutritional disorder or trait. For example, having a diagnosis of other functional intestinal disorders results in an increased HR for both AD and PD in the periods spanning 1 to 5, 5 to 10, and 10 to 15 years before AD/PD diagnosis. The time-stratified Cox regression approach allows for variations in HR across different time periods, which accounts for changes in the hazard over time. By splitting the samples according to time before diagnosis and analyzing only ICD-10 codes within specific time periods, the model is adjusting for the potential nonproportionality of hazards and is not assuming constant HRs across the entire study period. This approach helps control for the nonproportional hazards that may arise from variations in the timing of diagnoses.

In an effort to investigate genetic distinctions and potential etiological subtypes of AD and PD, we compared PRS for AD and PD in individuals diagnosed only with AD or PD and individuals with AD or PD co-occurring with other endocrine, nutritional, metabolic, and digestive-related disorders. Notably, our study confirmed that individuals diagnosed with any type of diabetes mellitus in addition to AD or PD are shown to have a notable different PRS than individuals with AD or PD alone, in concordance with previous studies

showing that diabetes is a risk factor for both AD and PD (*45*–*47*). In addition, we observed substantial relationships in AD with other bacterial intestinal infections and other functional intestinal disorders, as well as in PD with other functional intestinal disorders and other disorders of the peritoneum. Participants with either AD or PD and any of the notable disorders affecting the gut-brain axis showed lower average PRS scores. While this may suggest an inverse relationship between genetic risk for AD/PD and co-occurring gut-brain disorders, we caution that this pattern may be influenced by collider bias, where conditioning on the presence of AD or PD could induce spurious associations between genetic and nongenetic risk factors (*48*, *49*). Together, our findings promote the importance of considering both genetic and other health factors in assessing the overall risk of developing AD and PD.

There are several possible explanations for the notable differences observed in PRS distributions between individuals diagnosed solely with AD or PD and those with concurrent AD/PD and additional ICD-10 diagnoses. Differences in PRS distributions may be influenced by gene-gene interactions (epistasis), gene-environment interactions, and PRS model limitations. The PRS approach primarily captures additive genetic effects but does not account for potential synergistic or antagonistic interactions between genetic variants that may modify disease risk. These factors collectively highlight the complexity of interpreting PRS distributions in individuals with multiple diagnoses and underscore the need for careful consideration of underlying biases and confounders, even when adjusting for population structure using principal components (PCs).

In addition, our analysis revealed that the relationship between the genetic risk for AD/PD and many disorders of the endocrine, nutritional, metabolic, and digestive systems resulted in a combined impact not notably greater than the sum of their individual effects. These interactions are not synergistic, confirming the notion that known genetic risk factors included in the models under study for both AD and PD are independent from gastrointestinal and metabolic disorders, which highlights the importance of environmental factors in the development of both AD and PD. While both genetic and systemic health disorders independently influence the risk of AD and PD, many disorders may not interact in a way that notably amplifies this risk when combined.

Our findings suggest that proteomic biomarkers, such as NFL for AD and PD, could be valuable diagnostic tools in identifying individuals at risk before clinical symptoms appear. The models built on proteomic data alone, even without clinical data, performed impressively well, indicating that biological markers of disease might offer more direct and accurate insights into the neurodegenerative processes of AD and PD than traditional clinical assessments alone. The ability to predict risk with high accuracy using these biomarkers highlights the potential for early detection, personalized medicine, and better-targeted interventions, making proteomic biomarkers essential components for future multimodal models for NDDs. However, these findings suggest that while some biomarkers are distinctly associated with AD/PD, others may be influenced by co-occurring comorbidities and medications, among other influential factors—an important consideration when developing multimodal models.

Our analysis reveals that specific co-occurring conditions may influence molecular markers of PD or AD. For example, we observed that peroxiredoxin 1 (PRDX1) levels were decreased in PD cases with a co-occurring ICD-10 code for insulin-dependent diabetes mellitus. This observation reflects biomarker-level differences that may arise because of the presence of specific co-occurring conditions. While these conditions did not notably interact with genetic risk to affect AD or PD outcomes, they may still alter biological pathways relevant to disease manifestation. This highlights the importance of separating disease risk from disease manifestation, as co-occurring conditions can affect biological changes in the disease without necessarily increasing the overall risk.

Through our exploration of proteomics and AD/PD risk in UKB, we confirmed several promising candidates for AD/PD diagnosis. For instance, we found that for AD, the proteomic biomarker with the largest impact on disease was GFAP, supporting previous literature findings that GFAP can serve as an indicator of AD pathology (*50*). However, the levels of GFAP were notably decreased in samples with AD and a co-occurring ICD-10 code for gastritis and duodenitis or noninsulin-dependent diabetes mellitus. For PD, in concordance with previous studies, we identified PRDX1 as a potential biomarker for disease (*51*). Again, the levels of PRDX1 were notably decreased in PD cases and a coinciding ICD-10 code for insulin-dependent diabetes mellitus. These differences in biomarker levels between samples solely with AD/PD versus samples with additional diagnoses highlight the influence of comorbidities on disease manifestation. Although longitudinal research would need to be conducted, our findings suggest that these biomarkers could serve as valuable diagnostic or prognostic tools for AD and PD, potentially enhancing early detection and disease management. We acknowledge the limitation that the SAIL and FinnGen datasets do not include neither genetic nor proteomic data, preventing us from further validating our modeling efforts.

The inclusion of multiple features, integrating clinical data on digestive, endocrine, nutritional, and metabolic disorders, genetic risk scores, and proteomic data in our prediction model, demonstrated superior performance in predicting both AD and PD compared to single-variable paradigms. Co-occurring diagnoses for conditions influencing the gut-brain axis do not seem to influence the predictability of AD or PD as much as the other variables (demographic factors, biomarkers, and genetic status), but the fact that, for both AD and PD, the combination of data (clinical, genetic, proteomic, and demographic) produces a higher AUC for the receiver operating characteristic (ROC) curve compared to single modalities underscores the value in including multiple facets of data in predictive models. The relatively lower contribution of clinical features may be partially explained by the limited granularity of diagnostic data in biobanks and the variability in the severity of clinical conditions, which can introduce heterogeneity into the models. On the basis of SHAP values, the most influential predictors of AD and PD risk were age and molecular biomarkers. "Demographics/age at recruitment" was consistently the top-ranked feature in both models, which aligns with established knowledge that age is the strongest risk factor for AD and PD (*52*, *53*). Among molecular features, seven of the top 10 contributors were proteomic biomarkers, including GFAP and NFL—markers of astrocytic reactivity and axonal injury, respectively (*54*, *55*). Genetic factors also played a notable role, with both PRS (*z*-score) and *APOE* ε4 copy number (in AD) positively associated with disease classification. In the PD model, immune-related proteins such as interleukin-13 receptor, α1 and integrin-αM contributed modestly, suggesting an inflammatory component.

In contrast, clinical features showed limited contribution to model predictions. This may be partly explained by feature redundancy, where molecular markers already capture the pathophysiological

signals reflected in clinical diagnoses. In addition, biobanks may lack comprehensive clinical data, and diagnostic labels can vary in specificity and severity, introducing heterogeneity that diminishes the utility of clinical variables in risk prediction.

A particularly intriguing finding in our study is the lack of a consistent temporal pattern between the diagnosis of prior ICD-10 conditions and the subsequent diagnosis of AD or PD. This observation does not align with the expectations of the conventional "dual-hit" hypothesis, which posits that a secondary insult, such as a peripheral infection, can initiate or accelerate neurodegenerative processes, particularly in PD (*56*, *57*). For example, previous studies have highlighted the potential role of peripheral infections in triggering α-synuclein misfolding and aggregation, a hallmark of PD (*58*). However, our data indicate that the timing of the ICD-10 diagnoses, which were predominantly noninfectious in nature, varied in magnitude across time windows, without a consistent temporal pattern in their association with AD or PD onset. This suggests that shared vulnerabilities, such as genetic predispositions or cumulative environmental exposures, may play a more dominant role in risk of developing AD or PD than specific temporal triggers. In addition, the limited temporal resolution of diagnosis data in biobanks may reduce the sensitivity to detect more nuanced timing-dependent effects.

The absence of a clear temporal link between prior conditions and the onset of AD or PD calls for further investigation into the underlying mechanisms. It is possible that the impact of prior conditions is cumulative, with the overall burden of these comorbidities—rather than their precise timing—modulating NDD risk. Future research should focus on detailed longitudinal studies examining the temporal relationship between specific infectious or inflammatory exposures and the development of AD and PD. These studies, incorporating both clinical and biomarker data, will be critical for advancing our understanding of the complex interplay between prior conditions and the pathogenesis of NDDs.

Transparency and reproducibility are critical aspects of science. This becomes even more important for ML models due to their black-box nature. To facilitate this, we made two contributions: (i) model interpretability analysis, using SHAP values to identify the top features, corresponding to each modality, that can be validated against existing biological findings; and (ii) development of an open-access, cloud-based platform for researchers to explore the model developed in this study and investigate how different factors may influence classification (or, in some cases, misclassification). In addition, we incorporated a model perturbation analysis feature on our website, allowing researchers to manually adjust features and observe the resulting changes. These efforts enhance transparency and move the research community away from black-box predictors through interpretable modeling.

Some overall limitations of our work include the use of solely ICD-10 codes for diagnosis, and not other additional assays, which may overlook undiagnosed and misdiagnosed cases, leading to a potential underestimation of the true impact of these disorders on AD/PD risk. Across datasets, the available diagnosis codes also differ causing further limitations in comparison and validation. Because individuals with AD often present with different cognitive, behavioral, and pathological features, diagnosis of AD is often difficult and inconsistent (*59*). Although we use the ICD-10 code to filter for AD in this study, it is possible for AD to be interchanged with dementia, as diagnosing AD is a daunting challenge. The sample makeup across data sources can also be different. Since the UKB participants are volunteers who may have agreed to participate before reaching old age, the incidence of AD/PD may be different compared to the SAIL and FinnGen datasets. The focus on samples of European ancestry constrains the generalizability of our findings to other populations from other genetic backgrounds. Future work needs to be done across different ancestral backgrounds to be globally representative. In addition, the multimodal classification model is influenced by statistically significant differences in age between case and control groups for both AD ($P = 6.9 \times 10^{-75}$) and PD ($P = 3.9 \times 10^{-52}$). However, feature importance plots and other aspects of the model highlight the contributions of various features independent of age.

Our study delves into the intricate interactions between clinical, genetic, and proteomic data, culminating in the construction of a comprehensive multimodal classification model. This pioneering endeavor aims to shed light on the nuanced interplay between various physiological factors playing a role on the gut-brain axis and the development of AD and PD, offering a multifactorial systemic understanding that transcends traditional approaches. Our integrated approach serves as a proof of concept, aligning with the expanding body of evidence that underscores the intricate etiological foundations of NDDs and holds promise for refining risk prediction models and devising targeted preventive strategies. Further, we have developed an interactive resource for the scientific community (https://gut-brain-nexus.streamlit.app/). This, in turn, propels our endeavors in elucidating clinical interventions aimed at addressing these debilitating conditions. As it stands, the tool is best suited for researchers who already have access to the required genetic data and wish to implement our approach to testing other hypotheses and datasets. This was designed to allow readers to access a large volume of results in a user-friendly and interactive manner, rather than to enable users to input their own data.

MATERIALS AND METHODS

This study uses data from three biobanks: the UKB, FinnGen, and the SAIL (see Fig. 1 for a workflow rationale of this study). For genetic and proteomic modeling, we used UKB data, as it was the only biobank to provide access to both genetic and proteomic data.

UK Biobank

The UKB data, accessed via DNAnexus under application number 33601, include electronic health records of 502,367 individuals, single-nucleotide polymorphism data of 487,279 individuals, and proteomic (Olink) data of 52,705 individuals (https://ukbiobank.dnanexus.com) (accessed on May 2023). The control group for the UKB dataset consisted of individuals who have not been diagnosed with any NDD condition (table S18) and have no family history of an AD or PD diagnosis. Endocrine, nutritional, metabolic, and digestive disorders, as well as AD and PD diagnosis, were derived from ICD-10 codes. The AD cohort was obtained from the G30 and F00 ICD-10 codes in the UKB, and the PD cohort was obtained from the G20 ICD-10 code. We excluded any individuals who had received an ICD-10 diagnosis for endocrine, nutritional, metabolic, and digestive disorders before 1 January 1999 to match SAIL's records and right-censored any individuals who received an ICD-10 diagnosis after being diagnosed with AD or PD. Furthermore, any ICD-10 code with fewer than five cases was excluded from the analysis. Demographic characteristics are shown in table S19. For our

approach, related individuals were filtered out from further analyses based on a kinship coefficient greater than 0.0884, and only individuals of European descent were selected for this study, based on field ID 22006, who self-identified as "White British" and confirmed through PC analysis (PCA) of the genotyping data to avoid potential confounding effects. We lacked a sufficient number of individuals with non-European ancestry to conduct a meaningful multiancestry assessment. After all filtering steps, our dataset included 216,047 controls, 4473 AD cases, and 4564 PD cases.

ICD-10 codes were included in the Cox regression models only if at least three individuals had the condition before AD/PD onset and at least two experienced the outcome (AD/PD). With regard to the Olink proteomic data provided by UKB, we deferred to the quality control and outlier exclusion that UKB conducted before our data access (https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=4658). Quality control and outlier detection were performed on the basis of PCA and interquartile range median analysis per their documentation (https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=4658). None of the biomarker models among the top hits were flagged as potentially unreliable because of outlier participants, as assessed using Cook's distance (threshold > 0.5). Therefore, their associations with AD or PD appear stable and are unlikely to be artifacts driven by individual influential observations.

SAIL Databank

The SAIL Databank is a virtual platform providing anonymized medical data of the population in Wales (*60*). Diagnoses in SAIL are sourced from the Patient Episode Database for Wales (PEDW), records from clinicians and hospital staff, the Welsh Longitudinal General Practitioner dataset (WLGP), records from primary care physicians of diagnoses, treatments, symptoms, and referrals. Demographic information such as sex, age, address, and death were obtained from the Welsh Demographic Services Database and WLGP. Individuals with missing age or sex data and those without a Welsh address were excluded from further analysis.

Diagnoses were identified in the PEDW using ICD-10 codes and in the WLGP using National Health System (NHS) read codes (CVT2 and CVT3). Neurodegenerative disorders from the outpatient data were excluded because of their minimal representation of dementia cases (only 0.1%) and the absence of reliable diagnosis dates. Similarly, dementia diagnoses from death records were not included because of inaccurate diagnosis dates. This study covered the period from 1 January 1999 to 31 December 2018. For inclusion, individuals were required to have been alive at the start of 1999 and to have been at least 45 years old on 1 January 1999.

FinnGen Biobank

The FinnGen study is a large-scale genomics initiative that has analyzed over 500,000 Finnish biobank samples and correlated genetic variation with health data to understand disease mechanisms and predispositions. The project is a collaboration between research organizations and biobanks within Finland and international industry partners (*61*).

FinnGen provides survival analyses across numerous clinical end points. The HRs are adjusted for sex and year of birth. FinnGen bases the calculation of these HRs on a wide array of clinical end points defined through data from nationwide registries, including, but not limited to, Statistics Finland (https://finngen.gitbook.io/documentation/methods/endpoints). We downloaded the HRs for AD/PD from FinnGen's Risteys R10 platform for AD using the G6_AD_WIDE category matching more closely the UKB grouping than G6_AD (https://r10.risteys.finngen.fi/endpoints/G6_AD_WIDE) and for PD (https://r10.risteys.finngen.fi/endpoints/G6_PARKINSON) to explore the putative impact of endocrine, digestive, metabolic, and nutritional disorders on the risk of AD/PD before diagnosis. Not all ICD-10 codes used in our discovery phase in the UKB are represented in the FinnGen dataset. This discrepancy is perhaps due to the differences in health registries and data collection methodologies between the UKB and FinnGen biobanks.

Cox proportional hazards model

A Cox proportional hazards model was used to calculate the HR between risk for incident AD and PD and endocrine, nutritional, metabolic, and digestive system disorders. The disorders under study are represented as 155 ICD-10 codes (table S1). In this model, 1 January 1999 was used as the cutoff date for the diagnosis of these conditions. Consequently, any individual diagnosed with these traits before 1 January 1999 was excluded from the analysis. Logistic regressions and Cox proportional hazards models were adjusted for age, sex, the Townsend deprivation index, and five PCs to account for population stratification (precomputed in UKB only) (tables S20 to S31). In addition, we conducted time-stratified Cox proportional hazards analysis, for which cohorts were divided into three strata based on ICD-10 codes: 1 to 5 years, 5 to 10 years, and 10 to 15 years before NDD diagnosis. ICD-10 diagnoses within the specified time periods were retained, while any ICD-10 codes outside of these time frames were converted to Not a Number (NaN) (Tables 3 and 4). In our analysis, ICD-10 codes were included in the Cox regression models only if at least three individuals had the condition before AD/PD onset and at least two experienced the outcome (AD/PD). We applied the Benjamini-Hochberg procedure for FDR correction of HRs, using the fdrcorrection function from the statsmodels.stats.multitest module. The FDR correction was applied to the *P* values obtained from the survival models across 155 ICD-10 codes, allowing us to adjust for multiple comparisons and identify robust associations.

Fine-Gray model

To account for the competing risk of death, we fit a Fine-Gray subdistribution hazard model. The primary outcome in this model was the diagnosis of AD/PD. The competing event was death before AD/PD diagnosis, with the date of death obtained from UKB (data field: p40000_i0). Individuals who neither developed AD/PD nor died during follow-up were censored at the end of follow-up (1 January 2023). Follow-up started at the year of recruitment, and the end of follow-up was defined as the earliest of the following: the date of AD/PD diagnosis, date of death, or 1 January 2023. Participants whose AD/PD diagnosis occurred before recruitment were excluded. Duration of follow-up was calculated in years as the difference between start and end times. In the model, age at recruitment, sex, Townsend deprivation index, and genetic PC1 to PC5 (precalculated by UKB) were used as covariates. We used age at recruitment instead of year of birth to avoid collinearity in the model. When year of birth was included, the Fine and Gray model fitting failed because of a singular design matrix error, which was caused by collinearity among covariates. We excluded ICD-10 codes if fewer than three individuals had the diagnosis or if fewer than two individuals with the diagnosis developed AD/PD during follow-up. We applied the

FDR correction to control for multiple testing across all ICD-10 codes. Results were considered significant at an FDR of < 0.05.

Survival analysis using Kaplan-Meier plots

We generated Kaplan-Meier plots for the significantly associated ICD-10 codes as depicted in figs. S3 and S4, respectively. At the beginning of the observation period, the survival probability starts at 1.0, indicating that all individuals diagnosed with an ICD-10 code are initially free from AD and PD. Over time, this probability diminishes as more individuals are diagnosed with AD or PD. These Kaplan-Meier curves describe the proportion with and without AD/PD over the study follow-up period, stratified by exposure status for associated ICD-10 codes derived from the discovery phase. These can be interpreted as visual representations of the HR. Significant differences can be interpreted as nonoverlapping CIs in the images.

Time-to-event analysis was performed using the Python package Lifelines (version 0.29.0) for survival analysis, along with Statsmodels (version 0.14.4) for statistical modeling. These software tools were implemented in a Python environment (version 3.9).

Polygenic risk distribution for AD and PD

Risk allele loci and β values from GWAS summary statistics, specifically 23 risk predictors linked to AD risk from Kunkle *et al.* (*35*) and 90 risk predictors associated with PD from Nalls *et al.* (*36*), were used to estimate PRS for AD and PD, respectively. Of the 23 predictors from Kunkle *et al.* (*35*), 22 overlapped with the UKB data; all 90 predictors from Nalls *et al.* (*36*) were available. Each risk allele was assigned a weight based on the magnitude of its effect in the published studies, giving greater emphasis to alleles with higher risk estimates. Genetic variants were extracted for each individual from the UKB imputed data using the "bgenix" package (*62*). The extracted variants were then used to compute the PRS score for each individual using PLINK 2.0 (*63*). Estimated profiles were then normalized to *z*-scores using the UKB cohort without NDDs as the reference group.

To assess whether *z*-score differed between individuals diagnosed with AD or PD alone and those with both AD or PD and a specific ICD-10 diagnosis, we fit linear regression models with *z*-score as the dependent variable. The primary independent variable was group membership (AD/PD + ICD-10 versus AD/PD only), adjusting for year of birth, sex, and the first five genetic PCs. The β coefficient for the group variable estimates the adjusted difference in mean PRS between the two clinical subgroups for each neurodegenerative outcome.

Statistical analysis accounting for the impact of *APOE*, *LRRK2*, and *GBA1* risk variants

PRS analyses were adjusted for major genetic risk factors associated with AD and PD to determine whether observed similarities or differences between AD/PD and the assessed ICD-10 codes could be attributed to pleiotropic effects. For AD, individuals homozygous for the "C" allele at both *APOE* rs429358 and rs7412 were identified as having two copies of the *APOE*-ε4 allele, coded as 2, and one copy of the *APOE*-ε4 allele coded as 1 in our regression analysis. All other configurations were coded as 0. For PD, we examined *LRRK2* at rs76904798, where homozygous C alleles were coded as 0, and heterozygous and homozygous for "T" alleles as 1. Similarly, at rs34637584 (*LRRK2 G2019S*), we applied 0 for homozygous "G" alleles and 1 for heterozygous and homozygous "A" alleles. For *GBA1*, at rs35749011 (proxy for *GBA1* E326K), 0 was assigned to homozygous G alleles, and 1 was assigned to heterozygous and homozygous A alleles; and at rs76763715 (*GBA1* N370S), 0 was assigned to homozygous T alleles, and 1 was assigned to heterozygous and homozygous C alleles.

Interaction model for genetic risk across endocrine, metabolic, and digestive system disorders and nutritional status

We aimed to understand how the interplay between genetics underlying AD and PD risk and a clinical diagnosis for any investigated ICD-10 code could eventually influence an AD or PD diagnosis. A generalized linear model (GLM) was used to account for more complex relationships, where the impact of genetic risk for AD or PD (as measured by PRS) might interact differently across the diagnoses under study represented by ICD-10 codes. The interaction term used in this model was *z*-score * ICD-10 terms adjusted by sex, age, and Townsend deprivation index.

Proteomic biomarker data analyses

We aimed to explore differences in the levels of proteomic biomarkers associated with AD or PD between individuals with and without co-occurring ICD-10 code diagnoses related to endocrine, nutritional, metabolic, and digestive system disorders among UKB participants. For this purpose, we used data from the Pharma Proteome Project, which provides thousands of plasma protein biomarkers in blood samples (https://olink.com). In a cross-sectional analysis, we focused on baseline (instance 0) proteomic data, as it includes measurements for the largest number of individuals, encompassing data for 52,705 individuals (at the time of data access) and a total of 1463 proteins (table S32). These biomarkers span across cardiometabolics, inflammation, neurology, and oncology markers.

We applied GLM to analyze 1463 available biomarkers on AD and PD risk. The regressions were adjusted for age at recruitment, Townsend deprivation index, sex, and five PCs. Subsequently, we selected FDR-corrected significant proteomic biomarkers from the GLM results. In addition, we compared the average levels of proteomic biomarkers in isolated cases of AD/PD with cases of AD/PD co-occurring with selected ICD-10 codes (found to be significantly associated with AD or PD risk before AD/PD diagnosis). We selected FDR-corrected significant proteomic biomarkers with an OR greater than 1 and applied a *t* test to compare their average levels in the isolated cases of AD or PD and cases co-occurring with specific ICD-10 codes (tables S33 and S34).

Using proteomic data from the UKB, we examined whether specific ICD-10 diagnoses associated with AD or PD were also linked to differences in plasma biomarker levels. As proteomic data were not available in SAIL or FinnGen, this analysis was limited to UKB. We included ICD-10 codes that were significant in UKB Cox models and reported only Olink biomarkers with significant FDR-corrected *P* values and OR greater than 1. In addition, the reported *t* tests are limited to results that passed FDR correction.

Multimodal classification model for clinical, genetic, and proteomic data on AD and PD risk

To evaluate the role of endocrine, metabolic, digestive, and nutritional status–related diagnoses in predicting AD or PD status, we developed a multimodal classification model on a subset of the UKB dataset. This subset included individuals with variables encompassing clinical diagnosis for endocrine, metabolic, digestive, and nutritional

status; proteomic biomarkers; demographic factors (age at recruitment, sex, and Townsend deprivation index); genetic risk (*z*-score for the PRS); PCs; and *APOE* status for AD or *LRRK2*/*GBA1* status for the PD model. The clinical diagnoses used in these models include the ICD-10 codes found to have significant HRs in UKB. We used an ML approach to comparing the predictive performance of various feature sets on AD/PD outcomes. These included genetics, clinical, proteomic, and demographic factors (including sex and age at recruitment) and combinations thereof.

Each dataset was independently analyzed using an eXtreme Gradient Boosting Classifier, an ensemble learning method for classification tasks (https://github.com/dmlc/xgboost). We conducted hyperparameter tuning through GridSearchCV, optimizing for the number of estimators, learning rate, and maximum depth. For AD classification, the model was optimized by tuning the hyperparameters: number of estimators (*2*, *3*, *5*, *10*, *15*), learning rate (0.001, 0.01, and 0.1), and max depth (*3*–*5*). For PD, the model tuned the parameters as follows: number of estimators (*1*, *3*, *5*, *10*, *15*), learning rate (0.001, 0.01, and 0.1), and max depth (*2*, *3*, *5*). These hyperparameters were chosen according to the sample size of the dataset; consequently, some of the hyperparameters were slightly more conservative for the PD model than for the AD model. Both models used nested cross-validation, optimizing based on ROC AUC. We used fivefold cross-validation and, within each training fold, conducted an additional fivefold cross-validation for hyperparameter tuning. Feature selection was performed using the least absolute shrinkage and selection operator as a part of the hyperparameter tuning procedure. The dataset was downsampled to have an equal number of cases and controls and then z-scale normalized using StandardScaler to ensure uniformity and prevent bias due to variance in measurement scales before classification training.

Model performance was evaluated on the basis of ROC AUC and BA scores. The 95% CIs were calculated on the basis of performance metrics obtained from a fivefold cross-validation. We used the SHAP approach to assessing the impact of each feature on the ML model predictions (*64*). SHAP values are derived from game theory and approximate a feature's effect on the model. SHAP enhances understanding by creating accurate explanations for each observation. The SHAP package was used to calculate and visualize these Shapley values seen in the figures in the manuscript and the interactive website. A surrogate LightGBM regression model (https://lightgbm.readthedocs.io/) was trained on the risk estimates to calculate SHAP values.

Data balancing for multimodal approach

To address the class imbalance in our dataset, we selected an equal number of controls relative to the cases while maintaining age comparability. The AD dataset initially included 478 cases and 22,808 controls, while the PD dataset included 455 cases and 19,523 controls. To prepare the data for downstream classification modeling, we filtered the control group based on year of birth. We calculated the mean and SD of the cases' birth years and selected controls whose birth years fell within two SDs of the mean case birth year. This step ensured that the age distribution of the selected controls matched that of the cases. Next, we randomly sampled an equal number of controls to match the cases from the eligible subset. This random sampling step reduced selection bias while maintaining an equal class distribution. The final balanced dataset consisted of 956 samples for AD (478 cases and 478 controls) and 910 samples for PD (455 cases and 455 controls).

Supplementary Materials

The PDF file includes:
Figs. S1 to S8
Legends for tables S1 to S34

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S34

REFERENCES AND NOTES

1. C. A. Lane, J. Hardy, J. M. Schott, Alzheimer's disease. *Eur. J. Neurol.* **25**, 59–70 (2018).
2. E. Tolosa, A. Garrido, S. W. Scholz, W. Poewe, Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol.* **20**, 385–397 (2021).
3. A. Gustavsson, N. Norton, T. Fast, L. Frölich, J. Georges, D. Holzapfel, T. Kirabali, P. Krolak-Salmon, P. M. Rossini, M. T. Ferretti, L. Lanman, A. S. Chadha, W. M. van der Flier, Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimers Dement.* **19**, 658–670 (2023).
4. Z. Ou, J. Pan, S. Tang, D. Duan, D. Yu, H. Nong, Z. Wang, Global trends in the incidence, prevalence, and years lived with disability of Parkinson's disease in 204 countries/territories from 1990 to 2019. *Front. Public Health* **9**, 776847 (2021).
5. G. D. Rabinovici, Late-onset Alzheimer disease. *Continuum (Minneap. Minn.)* **25**, 14–33 (2019).
6. D. K. Simon, C. M. Tanner, P. Brundin, Parkinson disease epidemiology, pathology, genetics, and pathophysiology. *Clin. Geriatr. Med.* **36**, 1–12 (2020).
7. V. Bellou, L. Belbasis, I. Tzoulaki, E. Evangelou, J. P. A. Ioannidis, Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism Relat. Disord.* **23**, 1–9 (2016).
8. L. Migliore, F. Coppedè, Gene-environment interactions in Alzheimer disease: The emerging role of epigenetics. *Nat. Rev. Neurol.* **18**, 643–660 (2022).
9. J. A. Santiago, J. A. Potashkin, The impact of disease comorbidities in Alzheimer's disease. *Front. Aging Neurosci.* **13**, 631770 (2021).
10. J. A. Santiago, V. Bottero, J. A. Potashkin, Biological and clinical implications of comorbidities in Parkinson's disease. *Front. Aging Neurosci.* **9**, 394 (2017).
11. A. H. Tan, S. Y. Lim, A. E. Lang, The microbiome-gut-brain axis in Parkinson disease–from basic research to the clinic. *Nat. Rev. Neurol.* **18**, 476–495 (2022).
12. K. Kowalski, A. Mulak, Brain-gut-microbiota axis in Alzheimer's disease. *J. Neurogastroenterol. Motil.* **25**, 48–60 (2019).
13. J. Kang, M. Lee, M. Park, J. Lee, S. Lee, J. Park, A. Koyanagi, L. Smith, C. J. Nehs, D. K. Yon, T. Kim, Slow gut transit increases the risk of Alzheimer's disease: An integrated study of the bi-national cohort in South Korea and Japan and Alzheimer's disease model mice. *J. Adv. Res.* **65**, 283–295 (2024).
14. J. H. Moon, Endocrine risk factors for cognitive impairment. *Endocrinol. Metab. (Seoul)* **31**, 185–192 (2016).
15. N. Charoenngam, T. Rittiphairoj, B. Ponvilawan, K. Prasongdee, Thyroid dysfunction and risk of Parkinson's disease: A systematic review and meta-analysis. *Front. Endocrinol. (Lausanne)* **13**, 863281 (2022).
16. B. Chai, F. Gao, R. Wu, T. Dong, C. Gu, Q. Lin, Y. Zhang, Vitamin D deficiency as a risk factor for dementia and Alzheimer's disease: An updated meta-analysis. *BMC Neurol.* **19**, 284 (2019).
17. A. Pignolo, S. Mastrilli, C. Davì, V. Arnao, P. Aridon, F. A. Dos Santos Mendes, C. Gagliardo, M. D'Amelio, Vitamin D and Parkinson's disease. *Nutrients* **14**, 1220 (2022).
18. J. R. Colca, B. N. Finck, Metabolic mechanisms connecting Alzheimer's and Parkinson's diseases: Potential avenues for novel therapeutic approaches. *Front. Mol. Biosci.* **9**, 929328 (2022).
19. J. H. Kim, H. S. Lee, Y. H. Kim, M. J. Kwon, J.-H. Kim, C. Y. Min, D. M. Yoo, H. G. Choi, The association between thyroid diseases and Alzheimer's disease in a national health screening cohort in Korea. *Front. Endocrinol. (Lausanne)* **13**, 815063 (2022).
20. L.-Y. Ma, B. Zhao, Y.-N. Ou, D.-D. Zhang, Q.-Y. Li, L. Tan, Association of thyroid disease with risks of dementia and cognitive impairment: A meta-analysis and systematic review. *Front. Aging Neurosci.* **15**, 1137584 (2023).
21. K. Han, B. Kim, S. H. Lee, M. K. Kim, A nationwide cohort study on diabetes severity and risk of Parkinson disease. *NPJ Parkinsons Dis.* **9**, 11 (2023).
22. H. Chohan, K. Senkevich, R. K. Patel, J. P. Bestwick, B. M. Jacobs, S. Bandres Ciga, Z. Gan-Or, A. J. Noyce, Type 2 diabetes as a determinant of Parkinson's disease risk and progression. *Mov. Disord.* **36**, 1420–1429 (2021).
23. A. Athanasaki, K. Melanis, I. Tsantzali, M. I. Stefanou, S. Ntymenou, S. G. Paraskevas, T. Kalamatianos, E. Boutati, V. Lambadiari, K. I. Voumvourakis, G. Stranjalis, S. Giannopoulos, G. Tsivgoulis, G. P. Paraskevas, Type 2 diabetes mellitus as a risk factor for Alzheimer's disease: Review and meta-analysis. *Biomedicine* **10**, 778 (2022).
24. J. Nowell, E. Blunt, D. Gupta, P. Edison, Antidiabetic agents as a novel treatment for Alzheimer's and Parkinson's disease. *Ageing Res. Rev.* **89**, 101979 (2023).

25. Y. Zhao, Y. Sun, H.-F. Ji, L. Shen, Vitamin D levels in Alzheimer's and Parkinson's diseases: A meta-analysis. *Nutrition* **29**, 828–832 (2013).
26. B. Konings, L. Villatoro, J. Van den Eynde, G. Barahona, R. Burns, M. McKnight, K. Hui, G. Yenokyan, J. Tack, P. J. Pasricha, Gastrointestinal syndromes preceding a diagnosis of Parkinson's disease: Testing Braak's hypothesis using a nationwide database for comparison with Alzheimer's disease and cerebrovascular diseases. *Gut* **72**, 2103–2111 (2023).
27. S. Yuan, L. Dan, Y. Zhang, J. Wu, J. Zhao, M. Kivipelto, J. Chen, J. F. Ludvigsson, X. Li, S. C. Larsson, Digestive system diseases, genetic risk, and incident dementia: A prospective cohort study. *Am. J. Prev. Med.* **66**, 516–525 (2024).
28. F. Delcò, A. Sonnenberg, Associations between hemorrhoids and other diagnoses. *Dis. Colon Rectum* **41**, 1534–1541 (1998).
29. R. Cameron, M. M. Walker, M. Thuresson, B. Roelstraete, F. Sköldberg, O. Olén, N. J. Talley, J. F. Ludvigsson, Mortality risk increased in colonic diverticular disease: A nationwide cohort study. *Ann. Epidemiol.* **76**, 39–49 (2022).
30. B. C. Lucey, J. C. Varghese, J. A. Soto, Spontaneous hemoperitoneum: Causes and significance. *Curr. Probl. Diagn. Radiol.* **34**, 182–195 (2005).
31. H. van Goor, Consequences and complications of peritoneal adhesions. *Colorectal Dis.* **9**, 25–34 (2007).
32. M. Falconi, P. Pederzoli, The relevance of gastrointestinal fistulae in clinical practice: A review. *Gut* **49**, iv2–iv10 (2001).
33. J. Zhang, C. Chen, S. Hua, H. Liao, M. Wang, Y. Xiong, F. Cao, An updated meta-analysis of cohort studies: Diabetes and risk of Alzheimer's disease. *Diabetes Res. Clin. Pract.* **124**, 41–47 (2017).
34. D. Aune, S. Schlesinger, Y. Mahamat-Saleh, B. Zheng, C. T. Udeh-Momoh, L. T. Middleton, Diabetes mellitus, prediabetes and the risk of Parkinson's disease: A systematic review and meta-analysis of 15 cohort studies with 29.9 million participants and 86,345 cases. *Eur. J. Epidemiol.* **38**, 591–604 (2023).
35. B. W. Kunkle, B. Grenier-Boley, R. Sims, J. C. Bis, V. Damotte, A. C. Naj, A. Boland, M. Vronskaya, S. J. van der Lee, A. Amlie-Wolf, C. Bellenguez, A. Frizatti, V. Chouraki, E. R. Martin, K. Sleegers, N. Badarinarayan, J. Jakobsdottir, K. L. Hamilton-Nelson, S. Moreno-Grau, R. Olaso, R. Raybould, Y. Chen, A. B. Kuzma, M. Hiltunen, T. Morgan, S. Ahmad, B. N. Vardarajan, J. Epelbaum, P. Hoffmann, M. Boada, G. W. Beecham, J.-G. Garnier, D. Harold, A. L. Fitzpatrick, O. Valladares, M.-L. Moutet, A. Gerrish, A. V. Smith, L. Qu, D. Bacq, N. Denning, X. Jian, Y. Zhao, M. Del Zompo, N. C. Fox, S.-H. Choi, I. Mateo, J. T. Hughes, H. H. Adams, J. Malamon, F. Sanchez-Garcia, Y. Patel, J. A. Brody, B. A. Dombroski, M. C. D. Naranjo, M. Daniilidou, G. Eiriksdottir, S. Mukherjee, D. Wallon, J. Uphill, T. Aspelund, L. B. Cantwell, F. Garzia, D. Galimberti, E. Hofer, M. Butkiewicz, B. Fin, E. Scarpini, C. Sarnowski, W. S. Bush, S. Meslage, J. Kornhuber, C. C. White, Y. Song, R. C. Barber, S. Engelborghs, S. Sordon, D. Voijnovic, P. M. Adams, R. Vandenberghe, M. Mayhaus, L. A. Cupples, M. S. Albert, P. P. De Deyn, W. Gu, J. J. Himali, D. Beekly, A. Squassina, A. M. Hartmann, A. Orellana, D. Blacker, E. Rodriguez-Rodriguez, S. Lovestone, M. E. Garcia, R. S. Doody, C. Munoz-Fernadez, R. Sussams, H. Lin, T. J. Fairchild, Y. A. Benito, C. Holmes, H. Karamujić-Čomić, M. P. Frosch, H. Thonberg, W. Maier, G. Roshchupkin, B. Ghetti, V. Giedraitis, A. Kawalia, S. Li, R. M. Huebinger, L. Kilander, S. Moebus, I. Hernández, M. I. Kamboh, R. Brundin, J. Turton, Q. Yang, M. J. Katz, L. Concari, J. Lord, A. S. Beiser, C. D. Keene, S. Helisalmi, I. Kloszewska, W. A. Kukull, A. M. Koivisto, A. Lynch, L. Tarraga, E. B. Larson, A. Haapasalo, B. Lawlor, T. H. Mosley, R. B. Lipton, V. Solfrizzi, M. Gill, W. T. Longstreth Jr., T. J. Montine, V. Frisardi, M. Diez-Fairen, F. Rivadeneira, R. C. Petersen, V. Deramecourt, I. Alvarez, F. Salani, A. Ciaramella, E. Boerwinkle, E. M. Reiman, N. Fievet, J. I. Rotter, J. S. Reisch, O. Hanon, C. Cupidi, A. G. A. Uitterlinden, D. R. Royall, C. Dufouil, R. G. Maletta, I. de Rojas, M. Sano, A. Brice, R. Cecchetti, P. S. George-Hyslop, K. Ritchie, M. Tsolaki, D. W. Tsuang, B. Dubois, D. Craig, C.-K. Wu, H. Soininen, D. Avramidou, R. L. Albin, L. Fratiglioni, A. Germanou, L. G. Apostolova, L. Keller, M. Koutroumani, S. E. Arnold, F. Panza, O. Gkatzima, S. Asthana, D. Hannequin, P. Whitehead, C. S. Atwood, P. Caffarra, H. Hampel, I. Quintela, Á. Carracedo, L. Lannfelt, D. C. Rubinsztein, L. L. Barnes, F. Pasquier, L. Frölich, S. Barral, B. McGuinness, T. G. Beach, J. A. Johnston, J. T. Becker, P. Passmore, E. H. Bigio, J. M. Schott, T. D. Bird, J. D. Warren, B. F. Boeve, M. K. Lupton, J. D. Bowen, P. Proitsi, A. Boxer, J. F. Powell, J. R. Burke, J. S. K. Kauwe, J. M. Burns, M. Mancuso, J. D. Buxbaum, U. Bonuccelli, N. J. Cairns, A. McQuillin, C. Cao, G. Livingston, C. S. Carlson, N. J. Bass, C. M. Carlsson, J. Hardy, R. M. Carney, J. Bras, M. M. Carrasquillo, R. Guerreiro, M. Allen, H. C. Chui, E. Fisher, C. Masullo, E. A. Crocco, C. DeCarli, G. Bisceglio, M. Dick, L. Ma, R. Duara, N. R. Graff-Radford, D. A. Evans, A. Hodges, K. M. Faber, M. Scherer, K. B. Fallon, M. Riemenschneider, D. W. Fardo, R. Heun, M. R. Farlow, H. Kölsch, S. Ferris, M. Leber, T. M. Foroud, I. Heuser, D. R. Galasko, I. Giegling, M. Gearing, M. Hüll, D. H. Geschwind, J. R. Gilbert, J. Morris, R. C. Green, K. Mayo, J. H. Growdon, T. Feulner, R. L. Hamilton, L. E. Harrell, D. Drichel, L. S. Honig, T. D. Cushion, M. J. Huentelman, P. Hollingworth, C. M. Hulette, B. T. Hyman, R. Marshall, G. P. Jarvik, A. Meggy, E. Abner, G. E. Menzies, L.-W. Jin, G. Leonenko, L. M. Real, G. R. Jun, C. T. Baldwin, D. Grozeva, A. Karydas, G. Russo, J. A. Kaye, R. Kim, F. Jessen, N. W. Kowall, B. Vellas, J. H. Kramer, E. Vardy, F. M. LaFerla, K.-H. Jöckel, J. J. Lah, M. Dichgans, J. B. Leverenz, D. Mann, A. I. Levey, S. Pickering-Brown, A. P. Lieberman, N. Klopp, K. L. Lunetta, H.-E. Wichmann, C. G. Lyketsos, K. Morgan, D. C. Marson, K. Brown, F. Martiniuk, C. Medway, D. C. Mash, M. M. Nöthen, E. Masliah, N. M. Hooper, W. C. McCormick, A. Daniele, S. M. McCurry, A. Bayer, A. N. McDavid, J. Gallacher, A. C. McKee, H. van den Bussche, M. Mesulam, C. Brayne, B. L. Miller, S. Riedel-Heller, C. A. Miller, J. W. Miller, A. Al-Chalabi, J. C. Morris, C. E. Shaw, A. J. Myers, J. Wiltfang, S. O'Bryant, J. M. Olichney, V. Alvarez, J. E. Parisi, A. B. Singleton, H. L. Paulson, J. Collinge, W. R. Perry, S. Mead, E. Peskind, D. H. Cribbs, M. Rossor, A. Pierce, N. S. Ryan, W. W. Poon, B. Nacmias, H. Potter, S. Sorbi, J. F. Quinn, E. Sacchinelli, A. Raj, G. Spalletta, M. Raskind, C. Caltagirone, P. Bossù, M. D. Orfei, B. Reisberg, R. Clarke, C. Reitz, A. D. Smith, J. M. Ringman, D. Warden, E. D. Roberson, G. Wilcock, E. Rogaeva, A. C. Bruni, H. J. Rosen, M. Gallo, R. N. Rosenberg, Y. Ben-Shlomo, M. A. Sager, P. Mecocci, A. J. Saykin, P. Pastor, M. L. Cuccaro, J. M. Vance, J. A. Schneider, L. S. Schneider, S. Slifer, W. W. Seeley, A. G. Smith, J. A. Sonnen, S. Spina, R. A. Stern, R. H. Swerdlow, M. Tang, R. E. Tanzi, J. Q. Trojanowski, J. C. Troncoso, V. M. Van Deerlin, L. J. Van Eldik, H. V. Vinters, J. P. Vonsattel, S. Weintraub, K. A. Welsh-Bohmer, K. C. Wilhelmsen, J. Williamson, T. S. Wingo, R. L. Woltjer, C. B. Wright, C.-E. Yu, L. Yu, Y. Saba, A. Pilotto, M. J. Bullido, O. Peters, P. K. Crane, D. Bennett, P. Bosco, E. Coto, V. Boccardi, P. L. De Jager, A. Lleo, N. Warner, O. L. Lopez, M. Ingelsson, P. Deloukas, C. Cruchaga, C. Graff, R. Gwilliam, M. Fornage, A. M. Goate, P. Sanchez-Juan, P. G. Kehoe, N. Amin, N. Ertekin-Taner, C. Berr, S. Debette, S. Love, L. J. Launer, S. G. Younkin, J.-F. Dartigues, C. Corcoran, M. A. Ikram, D. W. Dickson, G. Nicolas, D. Campion, J. Tschanz, H. Schmidt, H. Hakonarson, J. Clarimon, R. Munger, R. Schmidt, L. A. Farrer, C. Van Broeckhoven, M. C. O'Donovan, A. L. DeStefano, L. Jones, J. L. Haines, J.-F. Deleuze, M. J. Owen, V. Gudnason, R. Mayeux, V. Escott-Price, B. M. Psaty, A. Ramirez, L.-S. Wang, A. Ruiz, C. M. van Duijn, P. A. Holmans, S. Seshadri, J. Williams, P. Amouyel, G. D. Schellenberg, J.-C. Lambert, M. A. Pericak-Vance, Alzheimer Disease Genetics Consortium (ADGC), European Alzheimer's Disease Initiative (EADI), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES), Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
36. M. A. Nalls, C. Blauwendraat, C. L. Vallerga, K. Heilbron, S. Bandres-Ciga, D. Chang, M. Tan, D. A. Kia, A. J. Noyce, A. Xue, J. Bras, E. Young, R. von Coelln, J. Simón-Sánchez, C. Schulte, M. Sharma, L. Krohn, L. Pihlstrøm, A. Siitonen, H. Iwaki, H. Leonard, F. Faghri, J. R. Gibbs, D. G. Hernandez, S. W. Scholz, J. A. Botia, M. Martinez, J.-C. Corvol, S. Lesage, J. Jankovic, L. M. Shulman, M. Sutherland, P. Tienari, K. Majamaa, M. Toft, O. A. Andreassen, T. Bangale, A. Brice, J. Yang, Z. Gan-Or, T. Gasser, P. Heutink, J. M. Shulman, N. W. Wood, D. A. Hinds, J. A. Hardy, H. R. Morris, J. Gratten, P. M. Visscher, R. R. Graham, A. B. Singleton, 23andMe Research Team, System Genomics of Parkinson's Disease Consortium, International Parkinson's Disease Genomics Consortium, Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
37. B. S. Diniz, Z. Chen, D. C. Steffens, L. Pilling, R. H. Fortinsky, G. A. Kuchel, C.-L. Kuo, Proteogenomic signature of risk of Alzheimer's disease and related dementia risk in individuals with a history of major depression disorder. medRxiv 2024.09.11.24313493 [Preprint] (2024). https://doi.org/10.1101/2024.09.11.24313493.
38. Y. A. Momtaz, T. A. Hamid, R. Ibrahim, Gastritis may boost odds of dementia. *Am. J. Alzheimers Dis. Other Demen.* **29**, 452–456 (2014).
39. M. N. Han, D. I. Finkelstein, R. M. McQuade, S. Diwakarla, Gastrointestinal dysfunction in Parkinson's disease: Current and potential therapeutics. *J. Pers. Med.* **12**, 144 (2022).
40. I. Martinez-Valbuena, R. Valenti-Azcarate, I. Amat-Villegas, I. Marcilla, G. Marti-Andres, M.-C. Caballero, M. Riverol, M.-T. Tuñon, P. E. Fraser, M.-R. Luquin, Mixed pathologies in pancreatic β cells from subjects with neurodegenerative diseases and their interaction with prion protein. *Acta Neuropathol. Commun.* **9**, 64 (2021).
41. N. S. Luthra, A. H. Marcus, N. K. Hills, C. W. Christine, Vitamin B12 measurements across neurodegenerative disorders. *J. Clin. Mov. Disord.* **7**, 3 (2020).
42. S. N. Rai, P. Singh, H. W. M. Steinbusch, E. Vamanu, G. Ashraf, M. P. Singh, The role of vitamins in neurodegenerative disease: An update. *Biomedicine* **9**, 1284 (2021).
43. J. Ghiso, B. Frangione, Amyloidosis and Alzheimer's disease. *Adv. Drug Deliv. Rev.* **54**, 1539–1551 (2002).
44. C. Ma, F. Hong, S. Yang, Amyloidosis in Alzheimer's disease: Pathogeny, etiology, and related therapeutic directions. *Molecules* **27**, 1210 (2022).
45. D. Baglietto-Vargas, J. Shi, D. M. Yaeger, R. Ager, F. M. LaFerla, Diabetes and Alzheimer's disease crosstalk. *Neurosci. Biobehav. Rev.* **64**, 272–287 (2016).
46. M. N. Haan, Therapy insight: Type 2 diabetes mellitus and the risk of late-onset Alzheimer's disease. *Nat. Clin. Pract. Neurol.* **2**, 159–166 (2006).
47. Q. Xu, Y. Park, X. Huang, A. Hollenbeck, A. Blair, A. Schatzkin, H. Chen, Diabetes and risk of Parkinson's disease. *Diabetes Care* **34**, 910–915 (2011).
48. M. Doulberis, G. Kotronis, D. Gialamprinou, S. A. Polyzos, A. Papaefthymiou, P. Katsinelos, J. Kountouras, Alzheimer's disease and gastrointestinal microbiota; impact of *Helicobacter pylori* infection involvement. *Int. J. Neurosci.* **131**, 289–301 (2021).

49. Y. Li, Y. Chen, L. Jiang, J. Zhang, X. Tong, D. Chen, W. Le, Intestinal inflammation and Parkinson's disease. *Aging Dis.* **12**, 2052–2068 (2021).
50. K. Y. Kim, K. Y. Shin, K.-A. Chang, GFAP as a potential biomarker for Alzheimer's disease: A systematic review and meta-analysis. *Cells* **12**, 1309 (2023).
51. M. Szeliga, Peroxiredoxins in neurodegenerative diseases. *Antioxidants (Basel)* **9**, 1203 (2020).
52. R. Guerreiro, J. Bras, The age factor in Alzheimer's disease. *Genome Med.* **7**, 106 (2015).
53. A. Reeve, E. Simcox, D. Turnbull, Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? *Ageing Res. Rev.* **14**, 19–30 (2014).
54. P. Sánchez-Juan, E. Valeriano-Lorenzo, A. Ruiz-González, A. B. Pastor, H. Rodrigo Lara, F. López-González, M. A. Zea-Sevilla, M. Valentí, B. Frades, P. Ruiz, L. Saiz, I. Burgueño-García, M. Calero, T. Del Ser, A. Rábano, Serum GFAP levels correlate with astrocyte reactivity, post-mortem brain atrophy and neurofibrillary tangles. *Brain* **147**, 1667–1679 (2024).
55. N. S. N. Graham, K. A. Zimmerman, F. Moro, A. Heslegrave, S. A. Maillard, A. Bernini, J.-P. Miroz, C. K. Donat, M. Y. Lopez, N. Bourke, A. E. Jolly, E.-J. Mallas, E. Soreq, M. H. Wilson, G. Fatania, D. Roi, M. C. Patel, E. Garbero, G. Nattino, C. Baciu, E. Fainardi, A. Chieregato, P. Gradisek, S. Magnoni, M. Oddo, H. Zetterberg, G. Bertolini, D. J. Sharp, Axonal marker neurofilament light predicts long-term outcomes and progressive neurodegeneration after traumatic brain injury. *Sci. Transl. Med.* **13**, eabg9922 (2021).
56. H. Braak, K. Del Tredici, U. Rüb, R. A. I. de Vos, E. N. H. Jansen Steur, E. Braak, Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**, 197–211 (2003).
57. C. H. Hawkes, K. Del Tredici, H. Braak, Parkinson's disease: A dual-hit hypothesis. *Neuropathol. Appl. Neurobiol.* **33**, 599–614 (2007).
58. E. Menozzi, J. Macnaughtan, A. H. V. Schapira, The gut-brain axis and Parkinson disease: Clinical and pathogenetic relevance. *Ann. Med.* **53**, 611–625 (2021).
59. S. Karantzoulis, J. E. Galvin, Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Rev. Neurother.* **11**, 1579–1591 (2011).
60. Secure Anonymised Information Linkage Databank (SAIL Databank), *Population Data Science* (2019); https://popdatasci.swan.ac.uk/centres-of-excellence/sail/.
61. M. I. Kurki, J. Karjalainen, P. Palta, T. P. Sipilä, K. Kristiansson, K. M. Donner, M. P. Reeve, H. Laivuori, M. Aavikko, M. A. Kaunisto, A. Loukola, E. Lahtela, H. Mattsson, P. Laiho, P. D. B. Parolo, A. A. Lehisto, M. Kanai, N. Mars, J. Rämö, T. Kiiskinen, H. O. Heyne, K. Veerapen, S. Rüeger, S. Lemmelä, W. Zhou, S. Ruotsalainen, K. Pärn, T. Hiekkalinna, S. Koskelainen, T. Paajanen, V. Llorens, J. Gracia-Tabuenca, H. Siirtola, K. Reis, A. G. Elnahas, B. Sun, C. N. Foley, K. Aalto-Setälä, K. Alasoo, M. Arvas, K. Auro, S. Biswas, A. Bizaki-Vallaskangas, O. Carpen, C.-Y. Chen, O. A. Dada, Z. Ding, M. G. Ehm, K. Eklund, M. Färkkilä, H. Finucane, A. Ganna, A. Ghazal, R. R. Graham, E. M. Green, A. Hakanen, M. Hautalahti, Å. K. Hedman, M. Hiltunen, R. Hinttala, I. Hovatta, X. Hu, A. Huertas-Vazquez, L. Huilaja, J. Hunkapiller, H. Jacob, J.-N. Jensen, H. Joensuu, S. John, V. Julkunen, M. Jung, J. Junttila, K. Kaarniranta, M. Kähönen, R. Kajanne, L. Kallio, R. Kälviäinen, J. Kaprio, N. K. FinnGen, J. Kettunen, E. Kilpeläinen, T. Kilpi, K. Klinger, V.-M. Kosma, T. Kuopio, V. Kurra, T. Laisk, J. Laukkanen, N. Lawless, A. Liu, S. Longerich, R. Mägi, J. Mäkelä, A. Mäkitie, A. Malarstig, A. Mannermaa, J. Maranville, A. Matakidou, T. Meretoja, S. V. Mozaffari, M. E. K. Niemi, M. Niemi, T. Niiranen, C. J. O. Donnell, M. E. Obeidat, G. Okafo, H. M. Ollila, A. Palomäki, T. Palotie, J. Partanen, D. S. Paul, M. Pelkonen, R. K. Pendergrass, S. Petrovski, A. Pitkäranta, A. Platt, D. Pulford, E. Punkka, P. Pussinen, N. Raghavan, F. Rahimov, D. Rajpal, N. A. Renaud, B. Riley-Gillis, R. Rodosthenous, E. Saarentaus, A. Salminen, E. Salminen, V. Salomaa, J. Schleutker, R. Serpi, H.-Y. Shen, R. Siegel, K. Silander, S. Siltanen, S. Soini, H. Soininen, J. H. Sul, I. Tachmazidou, K. Tasanen, P. Tienari, S. Toppila-Salmi, T. Tukiainen, T. Tuomi, J. A. Turunen, J. C. Ulirsch, F. Vaura, P. Virolainen, J. Waring, D. Waterworth, R. Yang, M. Nelis, A. Reigo, A. Metspalu, L. Milani, T. Esko, C. Fox, A. S. Havulinna, M. Perola, S. Ripatti, A. Jalanko, T. Laitinen, T. P. Mäkelä, R. Plenge, M. McCarthy, H. Runz, M. J. Daly, A. Palotie, FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
62. J. A. Collister, X. Liu, L. Clifton, Calculating polygenic risk scores (PRS) in UK Biobank: A practical guide for epidemiologists. *Front. Genet.* **13**, 818574 (2022).
63. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
64. S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions. arXiv:1705.07874 [cs.AI] (2017).

Acknowledgments: This work was supported in part by the Intramural Research Program of the NIH, the National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Services. This research has been conducted using the UKB Resource under application number 33601. This study makes use of anonymized data held in the SAIL Databank. We would like to acknowledge all the data providers who make anonymized data available for research. We want to acknowledge the participants and investigators of the FinnGen study. We extend our gratitude to P. Jarreau for language editing throughout the preparation of this manuscript. **Funding:** This work was supported by Intramural Research Program of the NIH, National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Services (project number ZO1 AG000534); National Institute of Neurological Disorders and Stroke. Dementia Research Institute [UKDRI supported by the Medical Research Council (UKDRI-3003), Alzheimer's Research UK, and Alzheimer's Society]; Welsh Government Joint Programming for Neurodegeneration (MRC: MR/T04604X/1); Dementia Platforms UK (MRC: MR/L023784/2); and MRC Centre for Neuropsychiatric Genetics and Genomics (MR/L010305/1). The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Author contributions: M.S.: Writing—original draft, conceptualization, investigation, Writing—review and editing, methodology, resources, funding acquisition, data curation, validation, supervision, formal analysis, software, project administration, and visualization. S.H.: Writing—original draft, writing—review and editing, and visualization. P.S.L.: Writing—review and editing, validation, formal analysis, software, investigation, and visualization. S.M.G.: Writing—review and editing, validation, formal analysis, software, and visualization. M.K.: Writing—review and editing and formal analysis. A.D.: Writing—review and editing, validation, software, and visualization. A.F.S.S.: Writing—original draft, conceptualization, writing—review and editing, methodology, and supervision. M.B.M.: Conceptualization, writing—review and editing, and software. R.S.: Writing—original draft, conceptualization, investigation, formal analysis, and software. E.S.: Writing—review and editing, methodology, validation, formal analysis, and software. H.I.: Writing—review and editing, resources, and data curation. G.H.: Methodology, data curation, and software. C.B.: Writing—review and editing and funding acquisition. V.E.-P.: Conceptualization, writing—review and editing, methodology, resources, validation, and supervision. Y.A.Q.: Conceptualization, investigation, and writing—review and editing. A.J.N.: Writing—review and editing. A.R.-P.: Writing—original draft, conceptualization, and writing—review and editing. H.L.L.: Writing—original draft, investigation, writing—review and editing, methodology, resources, data curation, and supervision. M.T.: Writing—original draft, writing—review and editing, resources, and supervision. F.F.: Conceptualization, writing—review and editing, methodology, supervision, formal analysis, and software. A.B.S.: Writing—review and editing. M.A.N.: Conceptualization, investigation, writing—review and editing, methodology, resources, data curation, validation, supervision, software, and project administration. K.S.L.: Writing—original draft, conceptualization, investigation, writing—review and editing, methodology, resources, data, curation, validation, supervision, formal analysis, software, and project administration. S.B.-C.: Writing—original draft, conceptualization, investigation, writing—review and editing, methodology, resources, funding, acquisition, data curation, validation, supervision, formal analysis, software, project administration, visualization. All authors read the paper. **Competing interests:** A.D., M.B.M., H.I., G.H., H.L.L., F.F., M.A.N., and K.S.L.'s participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research. M.A.N. also currently serves on the scientific advisory board at Clover Therapeutics and is an advisor and scientific founder at Neuron23 Inc. All other authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Notebooks containing code used in this analysis can be found in the GitHub/Zenodo link (https://github.com/NIH-CARD/Gut-Brain-Nexus and https://doi.org/10.5281/zenodo.15850226). This paper analyzes existing, publicly available data. In addition, complete summary statistics describing these data/processed datasets derived from these data have been deposited in the Supplementary Materials connected to this publication and are publicly available as of the date of publication. Further, we have developed an interactive resource for the scientific community (https://gut-brain-nexus.streamlit.app/), where researchers can investigate components of the predictive model and can investigate feature effects on a sample level.

Submitted 1 November 2024
Accepted 31 July 2025
Published 27 August 2025
10.1126/sciadv.adu2937