**Supplementary Notes**

**Serology measures**

As in our previous work(1)we used pre-derived serostatus variables from the UK Biobank (UKB) (2), and then derived serostatus in the MRC National Survey of Health and Development (NSHD) and Southall and Brent Revisited (SABRE) using validated cut-offs supplied by the developers of the multiplex serology platform (German Cancer Research Center in Heidelberg) (2,3). UKB variables can be extracted using field codes, which are included below, alongside details of serostatus derivation NSHD and SABRE. Analyses were restricted to the 14 pathogens common to all three studies with a seroprevalence >5%. As such, measures of antigens for the following pathogens were available but not investigated: Kaposi’s sarcoma-associated virus (all studies), human papillomavirus 16 (all studies), human papillomavirus 18 (all studies), human immunodeficiency virus-1 (UKB only), hepatitis B (UKB only), hepatitis C (UKB only), human T-cell leukemia virus-1 (UKB only), *F.nucleatum* (NSHD and SABRE only).

|  |  |  |
| --- | --- | --- |
|  | **UKB field codes** | **NSHD and SABRE** |
| Herpes simplex virus-1  | 23050 | 1gG (170) |
| Herpes simplex virus-2 | 23051 | 2mgG unique (180) |
| Varicella zoster virus  | 23052 | gE / gI (100) |
| Epstein-Barr virus  | 23053 | ≥2 positive out of EBNA (411), EA-D (110), VCAp18 (2526), Zebra (74) |
| Cytomegalovirus  | 23054 | ≥2 positive out of pp150 (100), pp52 (854), pp28 (100) |
| Human herpesvirus-6A | 23056 | IE1A (100) AND/OR p100 (75) |
| Human herpesvirus-6B  | 23057 | IE1B (100) AND/OR p101K (100)  |
| Human betaherpesvirus 7  | 23058 | U14 (225) |
| BK virus | 23065 | VP1 (250) |
| JC virus  | 23066 | VP1 (100) |
| Merkel Cell virus  | 23067 | VP1 (250) |
| *T.gondii* | 23062 | p22 (150) AND/OR sag1 (150) |
| *H.pylori* | 23073 and 23074 | ≥3 positive out of 8: HP 10 (GroEL) (100)HP1098 (Hcp C) (200)HP 1564 (400)HP 887/2 (VacA-C) (250)HP547/1 (CagA-N) (400)HP305 (150)HP73 (UreA) (250)HP875 (Catalase) (400) |
| *C.trachomatis* | 23070 and 23071 | pGP3 (95) |

In addition, in some cases, there were multiple antigens recommended by the developers of the multiplex serology platform for use in seroreactivity analyses - these are displayed below. In these cases, one antigen was randomly selected using the sample function in base R. The other recommended antigens were then followed up in sensitivity analyses (see table below).

|  |  |  |  |
| --- | --- | --- | --- |
| **Infectious agent group** | **Pathogen** | **N Antigens available** | **Recommended antigen** |
| **UKB** | **NSHD/SABRE** |
| Herpesviruses | Herpes simplex virus-1  | 1 | 1gG  | 1gG  |
| Herpes simplex virus-2  | 1 | 2mgG unique  | 2mgG unique |
| Varicella zoster virus  | 1 | gE/gI  | gE/gI |
| Epstein Barr virus  | 4 | VCA p18 or EBNA  | VCA p18 or EBNA  |
| Cytomegalovirus  | 3 | pp150 or pp52 or pp28 | p150 or pp52 or pp28  |
| Human herpesvirus-6A | 1-2 | 1E1A | 1E1A |
| Human herpesvirus-6B  | 2 | 1E1B  | 1E1B |
| Human betaherpesvirus 7  | 1 | U14  | U14 |
| Polyomaviruses | BK virus | 1 | BK VP1  | BK VP1 |
| JC virus  | 1 | JC VP1 | JC VP1 |
| Merkel Cell virus  | 1 | MCV VP1 | MCV VP1  |
| Bacteria/Protozoa | *T.gondii* | 2 | p22 or sag-1  | p22 or sag-1  |
| *H.pylori* | 6-8 | CagA-N  | HP 10 (GroEL)HP1098 (Hcp C)HP 1564HP 887/2 (VacA-C)HP547/1 (CagA-N)HP305HP73 (UreA)HP875 (Catalase) |
| *C.trachomatis* | 1 | pGP3 | pGP3 |

 **Genetic data quality control (QC)**

Details of genotyping can be found for NSHD (4), SABRE (5), and UKB (6) elsewhere. Briefly, blood samples were collected at the age 53 visit for NSHD and baseline (visit 1) for SABRE, and participants were genotyped using the DrugDev array. Two genotyping arrays (BiLEVE Axiom and Affymetrix UK Biobank Axiom) were used to genotype UKB participants. NSHD and SABRE genetic QC were performed in house, and we used UKB provided QC criteria and genetic principal components (PCs) (following their own central QC) for UKB. For SABRE, QC were applied separately by self-reported ethnicity. QC for SABRE and NSHD included filtering for common (minor allele frequency >0.01) biallelic autosomal variants that did not exhibit significant missingness (call rate >98%) or deviate from Hardy-Weinberg equilibrium (p<1e-4), as well as samples that did not exhibit significant heterozygosity (mean ±3sd) or missingness (call rate >95%), Similar QC steps were performed and supplied by the UKB (6), filtering for those with concordant genetic and self-reported sex that did not exhibit significant heterozygosity or missingness.

**Genetic principal components and relatedness**

**Mixed models**

We applied PC-AiR and PC-Relate to calculate genetic PCs and a kinship matrix (7,8) using the directly genotyped data. PC-AiR calculates genetic PCs using an unrelated and ancestrally representative subset, defined based on cut-offs of genetic relatedness and ancestral divergence. These PCs are then projected onto the related subset. PC-Relate estimates relatedness accounting for genetic PCs. First, following initial QC, genetic data were further filtered for independent common variants (linkage disequilibrium threshold = sqrt(0.1), max sliding window = 1x10-6, minor allele frequency>0.05). We then performed two rounds of PC-AiR and PC-Relate with the genetic PCs and kinship estimates from the second iteration taken forward for statistical analyses. In round one, kinship was first estimated using KING-robust (9) and genetic PCs were computed by PC-AiR (kinship threshold=2^(-9/2) corresponding to 3rd degree relatives; divergence threshold=-2^(-9/2)). Kinship was then re-estimated using PC-Relate, accounting for these genetic PCs. In round two, genetic PCs were recomputed, this time using the unrelated subset defined using PC-Relate. A kinship matrix was then derived from a second iteration of PC-Relate accounting for these genetic PCs.

To test associations between *APOE* genotypes and serology outcomes, mixed models were implemented in GENESIS, including genetic PCs, age, sex, and genotyping batch (for UKB only) as fixed-effects, and the PC-Relate kinship matrix as a random-effect. The number of genetic PCs included was chosen based on visual inspection of PC plots.

**Stratified analyses**

For stratified analyses, we followed our previous QC steps (1). In brief, all samples were additionally filtered for unrelated participants (KING>0.088). NSHD and SABRE samples closely clustering with reference panel populations (defined visually using genetic PCs) were additionally retained. Genetic PCs were generated in each study population following pruning and removal of variants in long range LD regions (10). QC steps were performed using plink1.9 and plink2 (11), and PCs using smartpca in EIGENSOFT (12). In UKB, participants that self-identified as “White-British” and were not PC outliers (field id: 21000) were retained. UKB provided PCs which we used.

To test associations between *APOE* genotypes and serology outcomes, analyses were conducted using logistic regression including genetic PCs, age, sex, and genotyping batch (for UKB only) as covariates.

**References**

1. Green RE, Sudre CH, Warren-Gash C, Butt J, Waterboer T, Hughes AD, et al. Common infections and neuroimaging markers of dementia in three UK cohort studies. Alzheimer’s & Dementia. 2024 Jan 22.

2. Mentzer AJ, Brenner N, Allen N, Littlejohns TJ, Chong AY, Cortes A, et al. Identification of host–pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. Nat Commun. 2022 Apr 5;13(1):1818.

3. Waterboer T, Sehr P, Michael KM, Franceschi S, Nieland JD, Joos TO, et al. Multiplex Human Papillomavirus Serology Based on In Situ–Purified Glutathione S-Transferase Fusion Proteins. Clinical Chemistry. 2005 Oct 1;51(10):1845–53.

4. Bann D, Wright L, Hardy R, Williams DM, Davies NM. Polygenic and socioeconomic risk for high body mass index: 69 years of follow-up across life. PLOS Genetics. 2022 Jul 14;18(7):e1010233.

5. Tillin T, Forouhi NG, McKeigue PM, Chaturvedi N. Southall And Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. Int J Epidemiol. 2012 Feb;41(1):33–42.

6. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.

7. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019 Dec 15;35(24):5346–8.

8. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. The American Journal of Human Genetics. 2016 Jan 7;98(1):127–48.

9. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010 Nov 15;26(22):2867–73.

10. Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, et al. Long-Range LD Can Confound Genome Scans in Admixed Populations. Am J Hum Genet. 2008 Jul 11;83(1):132–5.

11. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

12. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904–9.