medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

The Association of Total and Differential Leukocyte Counts with Incident Parkinson's Disease in UK Biobank

Melanie P Jensen (MBBS)*[1], Benjamin Meir Jacobs (MRCP)*[1], Ruth Dobson (MRCP, PhD)[1], Sara Bandres-Ciga (PhD)[2], Cornelis Blauwendraat (PhD)[2], Anette Schrag (MD, PhD)[3], The International Parkinson's Disease Genomics Consortium (IPDGC), Alastair J Noyce (MRCP, PhD)[+1,3]

* equal contribution

Affiliations

1. Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
2. Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
3. Department of Clinical and Movement Neurosciences, UCL Institute of Neurology, London, UK

Corresponding author ([+])

Dr Alastair Noyce

Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ

Email - a.noyce@qmul.ac.uk Telephone - +44 (0) 207 882 5841

Word count: 2,999

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Key points

Question: Is the leukcoyte differential count a feature of pre-diagnostic Parkinson's disease?

Findings: In the UK Biobank, a longitudinal cohort study with over 500,000 participants, lower lymphocyte count was associated with a 23% increased odds of incident PD, a significant difference.

Meaning: Pre-diagnostic Parkinson's disease is associated with lower lymphocyte counts; this could enhance positive predictivity in algorithms designed to identify groups at higher risk of future Parkinson's disease.

Abstract

Importance: Biomarkers for the early detection of Parkinson's disease (PD) are needed; emerging evidence implicates immune dysregulation. Altered leukocyte differentials and C-reactive protein (CRP), common markers of immune function, occur in patients with PD. Whether these changes drive pathogenesis remains unclear.

Objectives: We sought to identify whether peripheral immune dysregulation can be seen as a pre-diagnostic feature of PD, and whether it appears to play a causal role.

Design: We examined the relationship between differential leukocyte count and other markers of acute inflammation at enrolment, and incident cases of PD, in UK Biobank, a longitudinal cohort study. We used Mendelian randomization to establish whether differences in leukocyte subsets have a causal influence on risk of PD.

Setting: The UK Biobank; a population-based cohort with over 500,000 participants aged 40-69 recruited in the UK between 2006 and 2010.

Participants: PD cases were defined as individuals with an ICD-10 coded diagnosis of PD. Cases were defined as 'incident' if their age at diagnosis was greater than their age at recruitment to UKB. 'Controls' were defined as all other individuals in the dataset after

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

applying the above exclusions. After applying exclusion criteria for pre-existing health conditions that can influence blood counts, 507 incident PD cases and 328,280 controls were included in the analysis.

Exposure: Blood cell markers (absolute and relative counts) and other markers of inflammation (CRP and albumin) were obtained from blood tests of participants taken at the initial assessment visit.

Results: Lower lymphocyte count was associated with increased risk of incident PD. There was weaker evidence of association between lower eosinophil and monocyte counts, lower CRP, and higher neutrophil counts on risk of incident PD. The association between lymphopenia and incident PD remained robust to sensitivity analyses. Mendelian randomization did not reveal a clear causal effect of low lymphocyte count on PD risk, but a weak trend in that direction was seen.

Conclusions and relevance: In this large, prospective setting, lower lymphocyte count was associated with higher risk of subsequent PD diagnosis. The absence of a clear causal effect indicates that lymphopenia might be a consequence of prodromal PD.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Parkinson's disease (PD) affects 2% of the population over 65.[1] The diagnosis is made once motor signs appear, however by this stage ~50% of nigrostriatal neurons have been lost.[2] Pre-symptomatic identification and disease modification is critical; reversing motor disability is likely to be difficult.

Immune dysregulation may be a key component of susceptibility to PD. The white blood cell (WBC) differential is a crude marker of immune function, but is amenable to large-scale observational studies. Studies reveal lower lymphocyte counts in PD patients versus controls, driven by reductions in helper-$CD4^+$, rather than cytotoxic-$CD8^+$, T-cell and B-cells counts; hypothesized to represent a cytotoxic immune signature.[3-7] Case-control studies have identified higher neutrophil and lower lymphocyte counts in PD patients relative to controls.[8]

Genetic, epidemiological, and cytokine profiling studies have refined this area of study.[9] Human Leukocyte Antigen (HLA) variants (HLA-DRB1/DRB5) have been identified as risk loci for PD in genome-wide association studies (GWAS).[10,11] Large-scale polygenic risk score analyses suggest this pathway contributes to PD heritability.[12] *In vitro*, alpha-synuclein-derived peptides are preferentially displayed on major histocompatibility (MHC) molecules associated with PD risk, driving proinflammatory T-cell responses.[13,14] Variants in the leucine-rich repeat kinase 2 (*LRRK2*) gene, a target for pro-inflammatory signals, confer effects in the same direction on risk for PD and Crohn's disease, suggesting common genetic links.[15] Observational studies have reported reduced risk of PD and reduced penetrance in *LRRK2*-associated PD with use of immunosuppressants and non-steroidal anti-inflammatory drugs.[16,17] The prospective ICICLE-PD study found that a baseline 'pro-inflammatory'

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

cytokine serum profile in PD patients was associated with faster motor deterioration than an 'anti-inflammatory' profile.[18]

Whether immune dysregulation occurs as an early feature of PD and may yield biomarkers, or whether it drives PD risk causally, remains unclear. We studied the relationship between differential leukocyte count and biochemical markers of acute inflammation at enrolment, and incident cases of PD in the UK Biobank (UKB) (https://www.ukbiobank.ac.uk), a large longitudinal cohort with ~500,000 participants. We followed-up signals detected to determine whether differences in leukocyte subsets exerted a causal influence on PD risk using Mendelian randomization (MR).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

Population

UKB recruited ~500,000 individuals aged 40-69 between 2006-2010; prospective follow-up data, including census data, blood tests, and healthcare records, is regularly obtained.

PD cases were defined as individuals with an ICD-10 diagnosis of PD (code G20) derived from Hospital Episode Statistics or a self-reported diagnosis of PD. Date at PD diagnosis was determined using the UKB data field 'Date of Parkinson's Disease report'. Age at diagnosis was derived using this field, age at recruitment, and birth year. Cases were defined as 'incident' if their age at diagnosis was greater that at recruitment. 'Prevalent' PD cases, i.e. with a diagnosis of PD at baseline, were excluded from analyses. 'Controls' were defined as all other individuals in the dataset after applying the above exclusions.

Various pre-existing health conditions can influence blood counts. To minimize bias from unbalanced comorbidities among cases and controls we excluded individuals with ICD-10 diagnoses of malignant neoplasms, disease of the blood and blood-forming organs, autoimmune disease, thyrotoxicosis, demyelinating disease of the central nervous system, inflammatory respiratory conditions (asthma and bronchiectasis), non-infective enteritis, inflammatory dermatological conditions (atopic dermatitis and psoriasis), inflammatory polyarthropathies, spondylopathies, and eating disorders (supplementary table 1).

Blood cell markers

Blood cell markers (absolute and relative counts) and other markers of inflammation (CRP and albumin) were obtained from baseline blood tests of UKB participants taken at the initial

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

assessment. Details of data processing can be found on the UKB website (http://biobank.ndph.ox.ac.uk/showcase/showcase/docs/haematology.pdf).

Statistical analysis

We determined associations of blood cell and inflammatory markers with incident PD risk using logistic regression. As our primary analysis, we conducted multivariable logistic regression, modelling incident PD diagnosis as the outcome and adjusting for age, sex, Tonwsend deprivation score, and ethnicity (dichotomized as 'White' background vs all other ethnicities). Models were of the form: Incident PD~Age+Sex+Deprivation+Ethnicity+blood cell marker. Association strength was determined using the likelihood ratio test, comparing the full model to a null model consisting of the confounding covariates only (Incident PD~Age+Sex+Deprivation+Ethnicity). We then undertook a variety of sensitivity analyses. First, we included additional covariates in the models: body mass index (BMI) at recruitment, smoking status ("ever" vs "never"), and alcohol consumption ("ever" vs "never"). Second, we excluded individuals within serial time windows of PD diagnosis (<1, <2, <3 years from diagnosis etc.) to determine whether the effects from the primary analysis were restricted to individuals who would go on to develop PD sooner. Third, we repeated the analysis in a matched case:control analysis, individually matching controls by age and sex to PD cases 4:1. To determine associations between blood markers and time until PD diagnosis, we constructed linear models for the inverse-normal-transformed time to PD diagnosis on age, sex, Townsend score, ethnicity, and blood cell marker. Model fit was quantified using the likelihood ratio test.

Genotyping and GWAS of lymphocyte count

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

UKB participants were genotyped on the BiLEVE or Axiom arrays and densely imputed to the Haplotype Reference Consortium reference panel. Imputation procedures and quality control have been described elsewhere.[19]

To determine the association of common genetic variants with absolute lymphocyte count in a GWAS we applied the same exclusions as for the observational study. Next, we excluded individuals with incident PD, of non-European genetic ancestry, with >10% missing genotypes, and missing covariate or lymphocyte count data. We standardized the outcome (absolute lymphocyte count) by applying the inverse-normal transformation. We conducted the GWAS using mixed linear models in BOLT-LMM v2.3.4, adjusting for age, age^2, sex, age x sex, and the first 20 genetic principal components.[20,21] We performed standard single nucleotide polymorphism (SNP) quality control, removing SNPs with minor allele frequency <0.05, deviation from Hardy-Weinberg equilibrium at p-value<1e-5, genotype missingness of >10%, and INFO score <0.8. We restricted the analysis to autosomes. To calculate the relatedness matrix, we used a pruned subset of SNPs using genotyped markers only, and restricting to SNPs with a pairwise R^2<0.5 within each 10,000 variant window with a sliding increment of 1000 variants at each step (--indep-pairwise 10000 1000 0.5).

Test statistics were adjusted for genomic control and asymptotic p-values were recalculated from the adjusted statistics. Heritability estimates and genetic correlation were estimated using linkage disequilibrium (LD) score regression implemented in the command line ldsc software tool.[22,23] Post-GWAS functional annotation and gene mapping were performed using FUMA.[24]

Mendelian randomization

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

In MR, genetic variants associated with the exposure of interest are used as instrumental variables to estimate the effect of the exposure on the outcome.[25,26] The random allocation of genetic variants from parent to offspring and lifelong exposure mean there are advantages over traditional observational studies in reducing confounding and reverse causation.[27]

To determine whether lymphocyte count exerted a causal effect on PD risk, we performed two-sample MR using the TwoSampleMR R package.[28,29] For the exposure instrument, we used summary statistics from our GWAS. We applied steps to: remove SNPs not typed/imputed in the outcome GWAS; restrict to biallelic SNPs; restrict to SNPs strongly associated with standardized lymphocyte count ($\text{p-value}_{\text{adjusted}}$<5e-08); and clump SNPs using default parameters (LD window=10 000 kb, clumping r^2=0.001).

We harmonized exposure and outcome SNPs to ensure effect estimates were aligned for the same effect allele. As our primary analysis, we used the inverse-variance weighted (IVW) MR estimate, which provides a causal estimate when MR assumptions are valid.[30] As secondary sensitivity analyses, we applied the Mixture of Experts approach, which applies different MR estimators and methods for SNP instrument selection (heterogeneity and directionality filtering), and predicts which method has the highest probability of accurately estimating the true causal effect based on the data characteristics.[28] Power calculations were performed using Stephen Burgess' online calculator (https://sb452.shinyapps.io/power/).

Data and code availability

UK Biobank data are available via application (https://www.ukbiobank.ac.uk/). Code is available at https://github.com/benjacobs123456/PD_FBC_UKB. PD GWAS summary statistics are from Nalls *et al.* 2019, available via application at 23andMe

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

https://research.23andme.com/dataset-access/.[31] Generated GWAS summary statistics are available via the UK Biobank 'returns' catalogue and on request to authors.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Results

Demographics

After applying the exclusion criteria, 507 incident PD cases and 328,280 controls remained in the main (unmatched) analysis. Participant demographic data are shown in table 1 for the unmatched and matched analyses.

Association of blood cell and inflammatory traits with incident PD

In multivariable logistic regression adjusting for age, sex, deprivation score, and ethnicity, there was evidence of association (false discovery rate [FDR] Q<0.05) between lower lymphocyte count and incident PD (odds ratio [OR] 0.77, 95% confidence interval [CI] 0.65-0.90, table 2, figure 1). There was evidence of association (FDR Q<0.05) between lower eosinophil count and incident PD but with wide CIs (OR 0.33, 95% CI 0.14-0.76, table 2). There was weaker evidence (FDR Q<0.10) of associations between lower CRP and monocyte count, and higher neutrophil count on increased risk of incident PD (table 2).

To determine whether these associations could be driven by confounding, we constructed models in which we also controlled for variables which can impact both PD risk and blood cell indices: BMI, smoking, and alcohol consumption. The effect estimates from these models were less precise but of a similar magnitude to the primary analysis (supplementary table 2).

To examine the possibility that reverse causation, i.e. early PD influencing lymphocyte count, could be driving our findings, we excluded individuals who underwent blood draw within 5 years of PD diagnosis. Despite the loss of power resulting from this restriction, we obtained similar effect estimates (supplementary table 3). To examine whether the association between

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

lower lymphocyte count and higher PD risk persisted in the prodromal period, we repeated this analysis excluding individuals within serial time windows of diagnosis (within 1, 2, 3, 4, 5, 6, 7, and 8 years of diagnosis). Despite lower numbers of cases with increasing time pre-diagnosis, there was a consistent signal in all groups (supplementary table 4). We obtained similar results with the matched case-control analysis (supplementary table 5). Lymphopenia considered as a binary trait (defined as absolute lymphocyte count<$1x10^9$ cells/L), was strongly associated with a higher risk of PD (OR 1.93, 95% CI 1.26-2.97, p=0.006). Lastly, exclusion of extreme lymphocyte counts (mean ± >3SD, leaving an inclusion lymphocyte count range 0.05-$3.85x10^9$ cells/L) did not substantially alter the observed association (OR 0.76 per 1-SD increase in lymphocyte count, 95% CI 0.64-0.90).

To assess whether inflammatory and blood cell markers were associated with time until PD diagnosis, we constructed linear models adjusting for age, sex, deprivation score and ethnicity. There was no strong evidence that any blood markers were associated with time until PD diagnosis (FDR >0.05, supplementary table 6).

Lymphocyte count GWAS

We conducted a GWAS of normalized lymphocyte count using 256,596 individuals of European ancestry from UKB. 163 independent genomic risk loci were associated with lymphocyte count, mapping to 783 genes (figure 2, supplementary file). There was some evidence of test statistic inflation (lambda=1.25, figure 2c-d). Mapped genes were enriched for expression in relevant tissues (whole blood, spleen, and lung), for relevant gene ontology biological processes, and for overlap with relevant traits from the GWAS catalog (lymphocyte count, differential lymphocyte count, autoimmune diseases, cancers) (supplementary tables 7-9). The LD score regression estimate of lymphocyte count

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

heritability was 28.5% (h^2=0.285, SE=0.023), with an intercept of 1.07 (SE=0.03) and ratio of 0.0785 (0.0367), suggesting that the relatively high test statistic inflation was due to a combination of polygenic architecture and residual population stratification. Bivariate LD score regression using summary statistics from the latest PD GWAS (UKB data removed) revealed a weak negative genetic correlation between lymphocyte count and PD risk (r_g = -0.068, p = 0.031).

Mendelian randomization

To generate an instrument for lymphocyte count, we excluded SNPs not typed/imputed in both the exposure and outcome datasets, SNPs with $p \geq 5e^{-08}$, SNPs lying within the super-extended MHC (chr6:25,000,000-35,000,000 on hg19), and palindromic SNPs with intermediate effect allele frequencies. We then performed LD clumping using default parameters, yielding a genetic instrument of 152 independent non-MHC autosomal SNPs associated with lymphocyte count (supplementary table 10). Collectively these SNPs explained 2.9% of the variance in lymphocyte count in this sample, and constituted a strong instrument (F statistic=49.8).

The primary MR analysis (IVW) did not show clear evidence of a causal effect of lymphocyte count on PD risk (OR 0.94, 95% CI 0.85-1.03, p=0.19). There was no evidence that unbalanced horizontal pleiotropy (whereby variants influence the outcome via pathways other than through the exposure) was biasing the IVW result (MR-Egger intercept 0.002, p=0.63). Neither standard sensitivity analyses nor the 'mixture of experts' approach revealed strong evidence of a causal effect of genetically-determined lymphocyte count on PD risk, however the point estimates from some of the MR methods suggest the possibility of a weak, imprecise negative effect (figure 3, supplementary table 11).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Power calculations indicated that, given a case:control ratio of 1:26 in the outcome GWAS, a sample size of 1,019,060 (the number of individuals in the PD GWAS), and an instrument R^2=0.029, our study would have power of 92.8% to detect an effect of magnitude OR=0.9, and 100% power to detect an effect of magnitude OR=0.8 at alpha=5% significance threshold, where the unit of exposure measurement is SD change in lymphocyte count.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion

In this prospective UKB cohort involving >320 000 individuals, lower lymphocyte counts were associated with increased risk of incident PD; marginal associations were observed for lower eosinophil and monocyte counts, and higher neutrophil counts. The association between lymphopenia and incident PD risk remained robust to a range of sensitivity analyses.

Only one study has explored the link between leukocyte subsets and risk of incident PD. In the Swedish Apolipoprotein-Related Mortality Risk cohort, lower lymphocyte count was associated with a lower risk of incident PD (HR 0.74, 95% CI 0.59-0.94).[32] However, pre-existing health conditions which can influence blood counts were not excluded or adjusted for so the possibility of residual confounding remained; we applied stringent exclusion criteria around comorbidities which influence leukocyte counts, such as autoimmune disease and cancer.

A number of studies have assessed changes in leukocyte populations in PD post-diagnosis. Established PD is associated with lymphopenia; driven by absolute reductions in $CD4^+$ T-helper cells, $CD19^+$ B-cells and Treg cells.[4,5,33,34] In a study of 123 newly diagnosed PD patients, the percentage of neutrophils and lymphocytes had positive and negative correlations, respectively, with UPDRS motor scores.[35] It has been suggested that this points to an autoimmune/inflammatory component in PD pathogenesis.[36]

Reductions in the absolute number of naïve-T/B-cells is also a feature of age-related immune dysregulation, leading to the hypotheses that PD is a state of premature ageing.[37-39] Flow cytometric analysis has demonstrated that leukocyte apoptosis is higher in PD patients than

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

controls and is associated with dopaminergic deficits on SPECT.[40] However, studies to date have examined immune markers in established PD and cannot determine whether leukocytes causally influence PD risk or represent a downstream consequence of PD pathology. Moreover, most PD cases these studies are receiving dopaminergic medication, which could confound immunophenotypic patient/control differences.[41-43] We demonstrate that lower lymphocyte count is a feature of PD at least 8 years before diagnosis, before initiation of dopaminergic therapy, extending the observations made in patients with established PD into the prodromal phase.

Our observed association between lymphopenia and incident PD may be driven by a causal effect, residual confounding or reverse causation. MR did not reveal a clear causative effect of low lymphocyte count on PD risk, however at the more conservative bounds of the observational estimate (per 1-SD increase in lymphocyte count, OR of PD 0.84, 95% CI 0.76-0.93), our MR instrument would only have 65% power to detect an effect at a 5% type-1 error rate. To address confounding, we corrected for potential confounders in the primary analysis and undertook sensitivity analyses (addition of covariates and a matched analysis); however residual confounding may remain. It is plausible that reverse causation underlies our observed effect; i.e. prodromal PD affects lymphocyte count. Although the effect persisted after excluding individuals who underwent blood draw within 8 years of diagnosis, prodromal PD including constipation may predate clinical diagnosis by 20 years.[44,45] The gut has been proposed as the site of initiation of PD and fecal microbiome changes are also noted in prodromal PD.[46] As such, lymphopenia may be a feature of prodromal PD, and reflect early pathological changes in peripheral tissues.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

CRP was marginally associated with reduced risk of PD. A meta-analysis of 23 case-control studies found significantly higher levels of CRP in the serum, cerebrospinal (CSF), and whole blood of manifest PD subjects compared with controls.[47] CSF CRP levels tend to increase over time from PD diagnosis and the development of PD-dementia.[48,49] These findings suggest that systemic inflammation impacts or occurs as a consequence of disease progression, but does not shed light on whether it influences disease initiation.[9]

Strengths of our study include the large sample size derived from UKB. Despite low response rates potentially leading to more favorable risk factor profiles, exposure-outcome associations in the UKB seem to be generalisable.[50] The comprehensive phenotyping of individuals in the cohort allowed us to correct for multiple potential confounding factors. In contrast to previous work utilizing cohorts with manifest PD, the availability of baseline blood tests and longitudinal assessment of PD diagnosis enabled us to examine the association between leukocyte subsets and risk of incident PD.

Limitations of our study include the lack of adjustment for medication which could impact leukocyte subsets, although exclusion for comorbidity will have captured a significant proportion of this confounding. Due to lack of flow cytometric data in UKB, we could not establish whether the observed association was driven by reductions in T-cells and/or B-cells. MR analysis precludes the identification of non-linear exposure-outcome associations, although non-linear mechanisms are not seen in other conditions in which lymphopenia influences outcome.[51,52]

In conclusion we report that lower lymphocyte count was associated with higher risk of subsequent diagnosis of PD in a large UK cohort. The association remained robust to

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

multivariate and sensitivity analyses, and persisted at 8 years pre-diagnosis. MR analyses demonstrated that lymphopenia is unlikely to causally reduce PD risk. Together, these results indicate that lymphopenia could be a consequence of prodromal PD which, although lacking specificity in isolation, could enhance predictivity in PD risk-prediction algorithms.[53]

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Acknowledgements

This research has been conducted using the UK Biobank Resource. We thank the participants of the UK Biobank. We would also like to acknowledge the 23andMe Research Team (especially Dr Karl Heilbron) and the 23andMe research participants.

Conflicts of Interest: The authors have no potential conflicts of interest to report.

Data access: BMJ and AJN had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis

Sources of Funding: The Preventive Neurology Units is funded by the Barts Charity. Melanie Jensen received support from the Isaac Schapera Trust and Benjamin Jacobs is an NIHR Academic Clinical Fellow. The funding bodies had no role design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. de Rijk MC, Launer LJ, Berger K, et al. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology*. 2000;54(11 Suppl 5):S21-23.
2. Cheng H-C, Ulane CM, Burke RE. Clinical progression in Parkinson disease and the neurobiology of axons. *Ann Neurol.* 2010;67(6):715-725.
3. Gruden MA, Sewell RDE, Yanamandra K, et al. Immunoprotection against toxic biomarkers is retained during Parkinson's disease progression. *Journal of neuroimmunology*. 2011;233(1):221-227.
4. Bas J, Calopa M, Mestre M, et al. Lymphocyte populations in Parkinson's disease and in rat models of parkinsonism. *J Neuroimmunol.* 2001;113(1):146-152.
5. Baba Y, Kuroiwa A, Uitti RJ, Wszolek ZK, Yamada T. Alterations of T-lymphocyte populations in Parkinson disease. *Parkinsonism Relat Disord.* 2005;11(8):493-498.
6. Stevens CH, Rowe D, Morel-Kopp M-C, et al. Reduced T helper and B lymphocytes in Parkinson's disease. *Journal of neuroimmunology.* 2012;252(1):95-99.
7. Niwa F, Kuriyama N, Nakagawa M, Imanishi J. Effects of peripheral lymphocyte subpopulations and the clinical correlation with Parkinson's disease. *Geriatrics & Gerontology International.* 2011;12(1):102-107.
8. Akil E, Bulut A, Kaplan I, Ozdemir HH, Arslan D, Aluclu MU. The increase of carcinoembryonic antigen (CEA), high-sensitivity C-reactive protein, and neutrophil/lymphocyte ratio in Parkinson's disease. *Neurol Sci.* 2015;36(3):423-428.
9. Tan E-K, Chao Y-X, West A, Chan L-L, Poewe W, Jankovic J. Parkinson disease and the immune system — associations, mechanisms and therapeutics. *Nature Reviews Neurology*. 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

10. Saiki M, Baker A, Williams-Gray CH, et al. Association of the human leucocyte antigen region with susceptibility to Parkinson's disease. *Journal of neurology, neurosurgery, and psychiatry.* 2010;81(8):890-891.

11. Nalls MA, Plagnol V, Hernandez DG, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet (London, England).* 2011;377(9766):641-649.

12. Bandres-Ciga S, Saez-Atienzar S, Kim JJ, et al. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. *Acta Neuropathol.* 2020.

13. Sulzer D, Alcalay RN, Garretti F, et al. T cells from patients with Parkinson's disease recognize alpha-synuclein peptides. *Nature.* 2017;546(7660):656-661.

14. Bandres-Ciga S, Cookson MR. Alpha-synuclein triggers T-cell response. Is Parkinson's disease an autoimmune disorder? *Mov Disord.* 2017;32(9):1327.

15. Hui KY, Fernandez-Hernandez H, Hu J, et al. Functional variants in the LRRK2 gene confer shared effects on risk for Crohn's disease and Parkinson's disease. *Science translational medicine.* 2018;10(423).

16. Noyce AJ, Bestwick JP, Silveira-Moriyama L, et al. Meta-analysis of early nonmotor features and risk factors for Parkinson disease. *Ann Neurol.* 2012;72(6):893-901.

17. San Luciano M, Tanner CM, Meng C, et al. Nonsteroidal Anti-Inflammatory Use and LRRK2 Parkinson's Disease Penetrance. *Movement Disorders.* 2020;n/a(n/a).

18. Williams-Gray CH, Wijeyekoon R, Yarnall AJ, et al. Serum immune markers and disease progression in an incident Parkinson's disease cohort (ICICLE-PD). *Movement disorders : official journal of the Movement Disorder Society.* 2016;31(7):995-1003.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

19. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.

20. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet.* 2015;47(3):284-290.

21. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet.* 2018;50(7):906-908.

22. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-1241.

23. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295.

24. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1826.

25. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):R89-98.

26. Bandres-Ciga S, Noyce AJ, Traynor BJ. Mendelian Randomization-A Journey From Obscurity to Center Stage With a Few Potholes Along the Way. *JAMA Neurol.* 2019.

27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* 2015;44(2):512-525.

28. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 2017;13(11):e1007081.

29. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 2018;7:e34408.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

30. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* 2013;178(7):1177-1184.

31. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 2019;18(12):1091-1102.

32. Yazdani S, Mariosa D, Hammar N, et al. Peripheral immune biomarkers and neurodegenerative diseases: A prospective cohort study with 20 years of follow-up. *Ann Neurol.* 2019;86(6):913-926.

33. Charlett A, Dobbs RJ, Dobbs SM, et al. Blood profile holds clues to role of infection in a premonitory state for idiopathic parkinsonism and of gastrointestinal infection in established disease. *Gut Pathog.* 2009;1(1):20.

34. Stevens CH, Rowe D, Morel-Kopp MC, et al. Reduced T helper and B lymphocytes in Parkinson's disease. *J Neuroimmunol.* 2012;252(1-2):95-99.

35. Umehara T, Oka H, Nakahara A, Matsuno H, Murakami H. Differential leukocyte count is associated with clinical phenotype in Parkinson's disease. *Journal of the Neurological Sciences.* 2020;409:116638.

36. Sommer A, Winner B, Prots I. The Trojan horse - neuroinflammatory impact of T cells in neurodegenerative diseases. *Molecular Neurodegeneration.* 2017;12(1):78.

37. Valiathan R, Ashman M, Asthana D. Effects of Ageing on the Immune System: Infants to Elderly. *Scandinavian Journal of Immunology.* 2016;83(4):255-266.

38. Sansoni P, Cossarizza A, Brianti V, et al. Lymphocyte subsets and natural killer cell activity in healthy old people and centenarians. *Blood.* 1993;82(9):2767-2773.

39. Costantini E, D'Angelo C, Reale M. The Role of Immunosenescence in Neurodegenerative Diseases. *Mediators of Inflammation.* 2018;2018:6039171.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

40. Lin W-C, Tsai N-W, Huang Y-C, et al. Peripheral leukocyte apoptosis in patients with Parkinsonism: correlation with clinical characteristics and neuroimaging findings. *Biomed Res Int.* 2014;2014:635923-635923.

41. McKenna F, McLaughlin PJ, Lewis BJ, et al. Dopamine receptor expression on human T- and B-lymphocytes, monocytes, neutrophils, eosinophils and NK cells: a flow cytometric study. *Journal of neuroimmunology.* 2002;132(1):34-40.

42. Besser MJ, Ganor Y, Levite M. Dopamine by itself activates either D2, D3 or D1/D5 dopaminergic receptors in normal human T-cells and triggers the selective secretion of either IL-10, TNFα or both. *Journal of neuroimmunology.* 2005;169(1):161-171.

43. Carr L, Tucker A, Fernandez-Botran R. In vivo administration of L-dopa or dopamine decreases the number of splenic IFN gamma-producing cells. *Journal of neuroimmunology.* 2003;137(1-2):87-93.

44. Mahlknecht P, Seppi K, Poewe W. The Concept of Prodromal Parkinson's Disease. *J Parkinsons Dis.* 2015;5(4):681-697.

45. Adams-Carr KL, Bestwick JP, Shribman S, Lees A, Schrag A, Noyce AJ. Constipation preceding Parkinson's disease: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry.* 2016;87(7):710-716.

46. Heinzel S, Aho VTE, Suenkel U, et al. Gut microbiome signatures of risk and prodromal markers of Parkinson’s disease. *bioRxiv.* 2019:2019.2012.2011.872481.

47. Qiu X, Xiao Y, Wu J, Gan L, Huang Y, Wang J. C-Reactive Protein and Risk of Parkinson's Disease: A Systematic Review and Meta-Analysis. *Front Neurol.* 2019;10:384-384.

48. Hall S, Janelidze S, Surova Y, Widner H, Zetterberg H, Hansson O. Cerebrospinal fluid concentrations of inflammatory markers in Parkinson's disease and atypical parkinsonian disorders. *Sci Rep.* 2018;8(1):13276.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

49. Lindqvist D, Hall S, Surova Y, et al. Cerebrospinal fluid inflammatory markers in Parkinson's disease--associations with depression, fatigue, and cognitive impairment. *Brain Behav Immun.* 2013;33:183-189.

50. Batty GD, Gale CR, Kivimaki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ.* 2020;368:m131.

51. Britanova OV, Putintseva EV, Shugay M, et al. Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling. *The Journal of Immunology.* 2014;192(6):2689.

52. Zidar DA, Al-Kindi SG, Liu Y, et al. Association of Lymphopenia With Risk of Mortality Among Adults in the US General Population. *JAMA Network Open.* 2019;2(12):e1916526-e1916526.

53. Noyce AJ, Bestwick JP, Silveira-Moriyama L, et al. PREDICT-PD: identifying risk of Parkinson's disease in the community: methods and baseline results. *J Neurol Neurosurg Psychiatry.* 2014;85(1):31-37.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.13.20189530; this version posted September 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

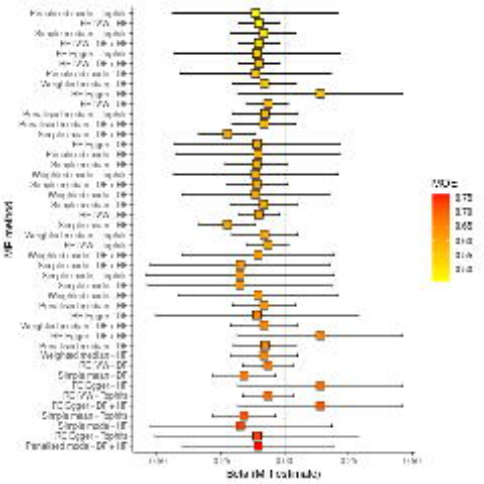
It is made available under a CC-BY-NC-ND 4.0 International license .

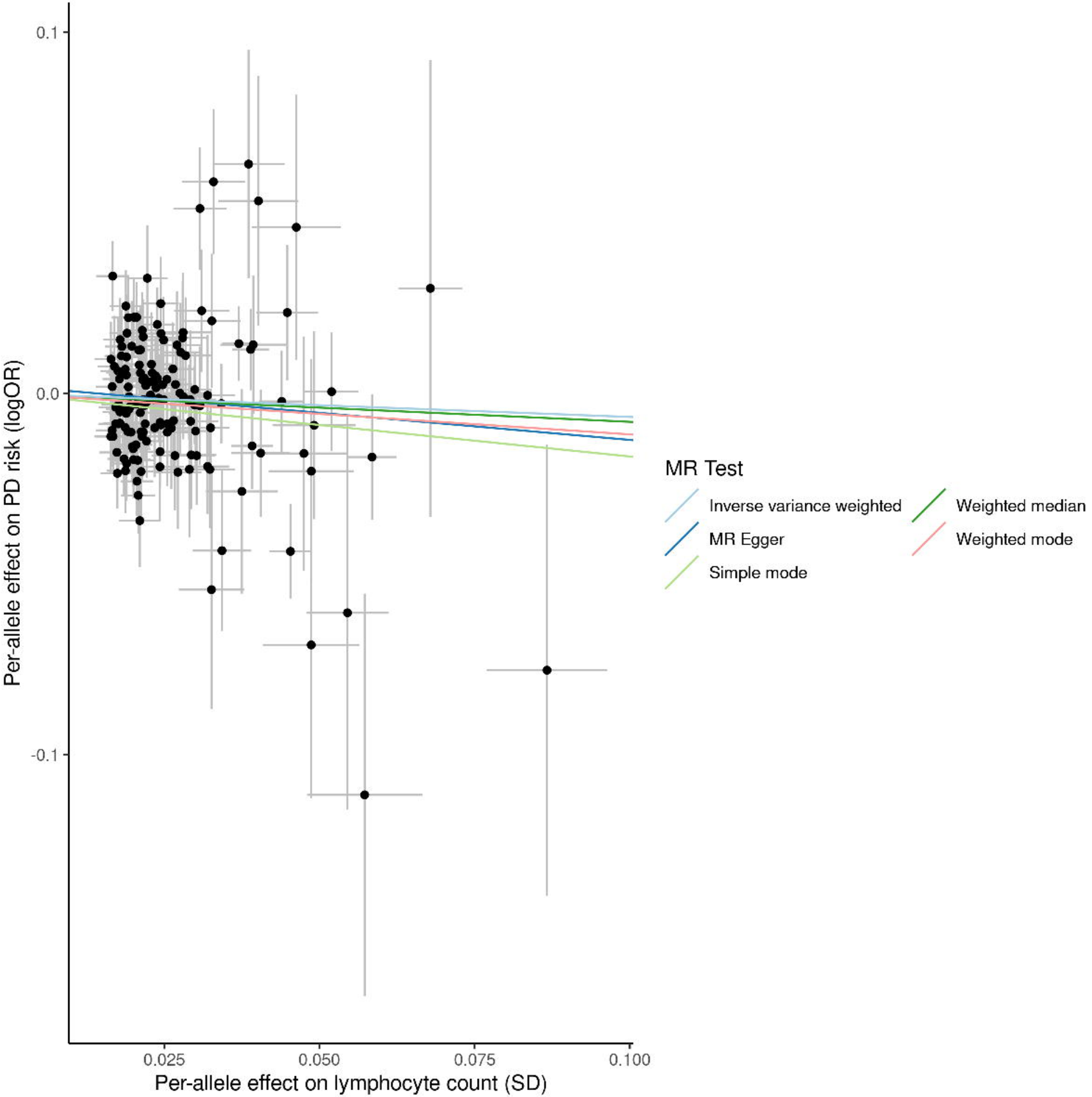
Figure legends

Figure 1: Association of blood cell marker and inflammatory markers with incident PD in UKB. Betas and 95% CIs are shown from multivariable logistic regression models of the form PD ~ Age + Sex + Ethnicity + Townsend deprivation score + blood biomarker. Estimates shown here are for Z-score standardised variables (x-mean / sd) to allow comparison between effect sizes.

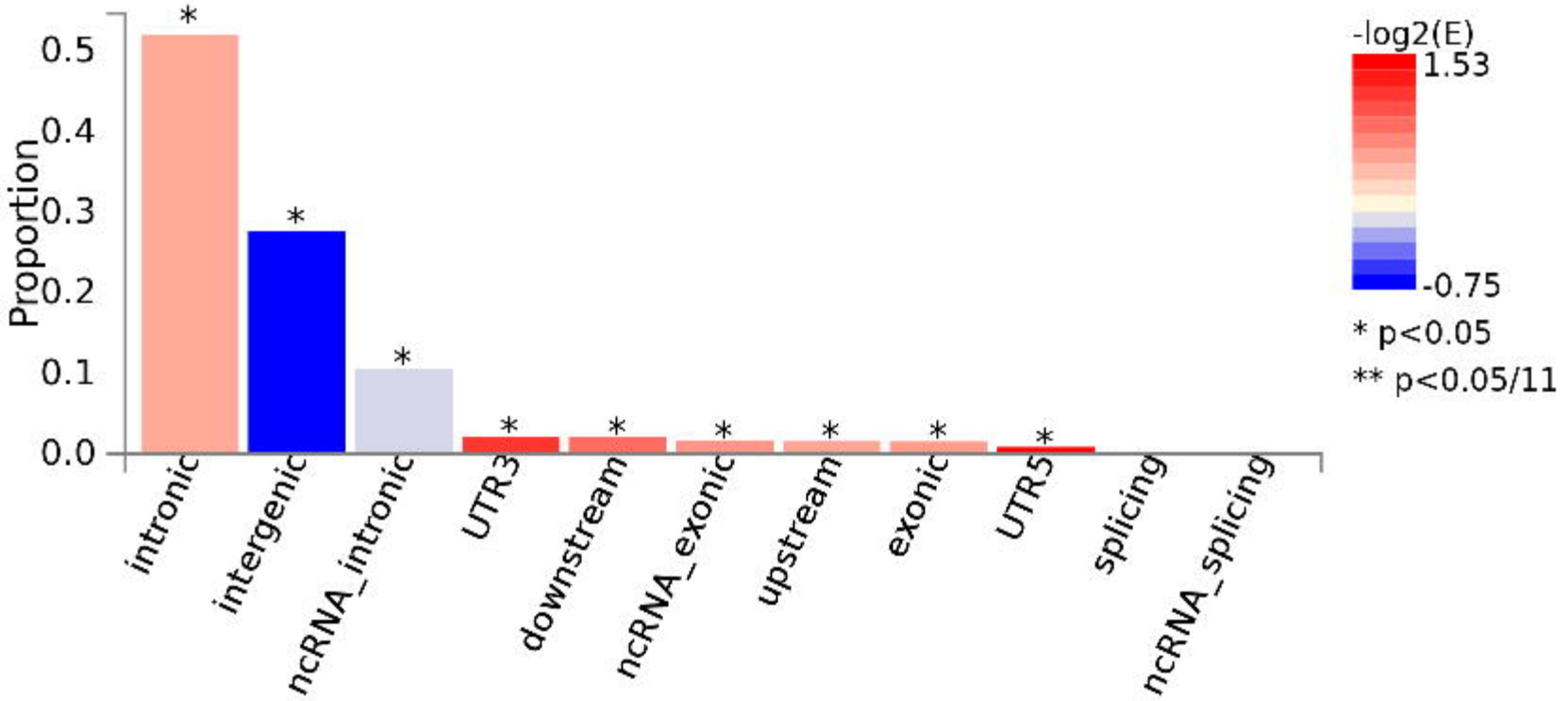
Figure 2: Genome-wide association study of lymphocyte count. A: Manhattan plot showing -log10 p values for association of SNPs with inverse-normal-transformed lymphocyte count in UK Biobank (methods). B: ANNOVAR annotations of SNPs associated with lymphocyte count at genome-wide significance. Figure produced by FUMA (methods). C: Differential expression of mapped genes from GWAS in GTEX v8 tissues.

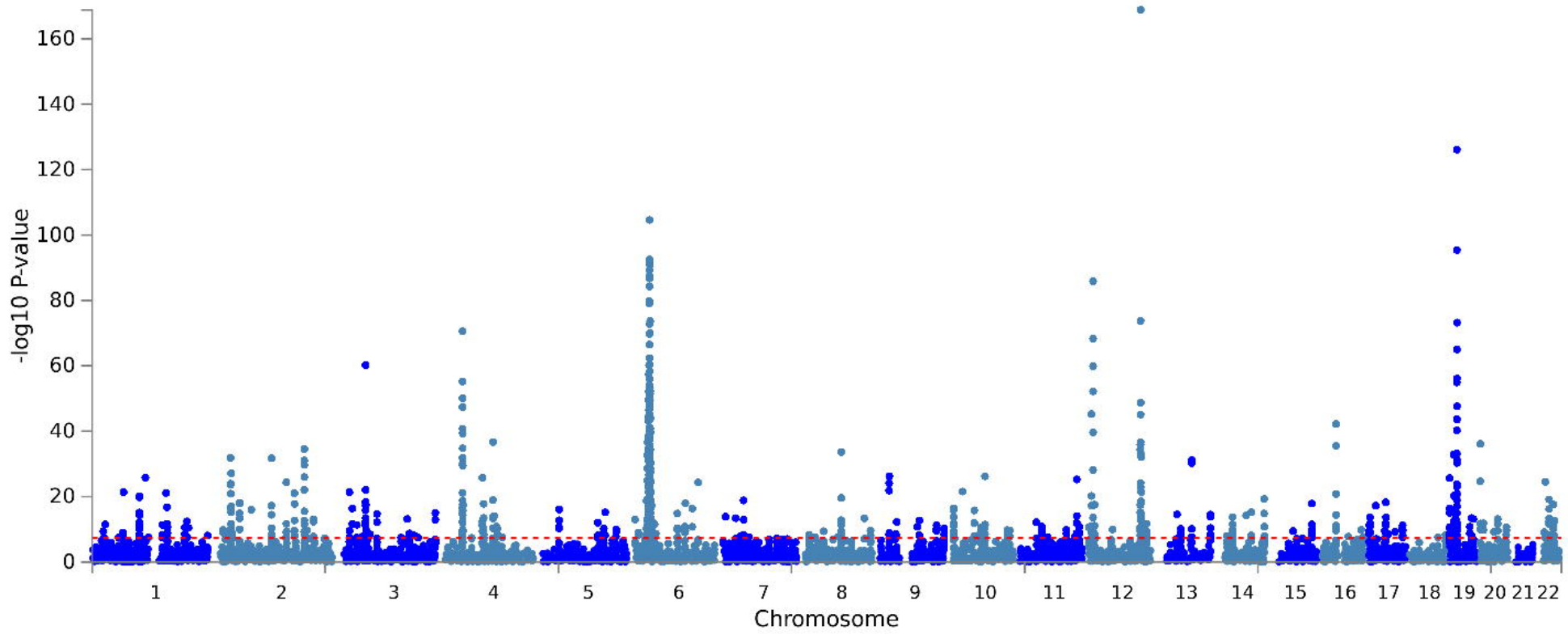
Figure 3: Mendelian randomisation analysis of the effect of lymphocyte count on PD risk. A: scatter plot showing SNP associations with lymphocyte count from our GWAS (x) and with the per-allele log(OR) for PD (estimates from the largest PD GWAS published, Meta5, excluding UKB participants). The model fit lines indicate MR estimates (of the 'causal effect) from different MR methods. B: MR estimates from various methods using the 'Mixture of Experts approach'. The y axis shows different MR methods and different approaches for filtering SNPs to be included in the genetic instrument (heterogeneity filtering, HF; directionality filtering, DF). Estimates are coloured and ordered by the 'MOE' statistic, which is similar to an area under the curve statistic in that it quantifies that ability of a given MR method to distinguish a true effect from the absence of a true effect. MOE statistics closer to 1 indicate a higher likelihood that the given MR method will give an accurate estimate for the given dataset.

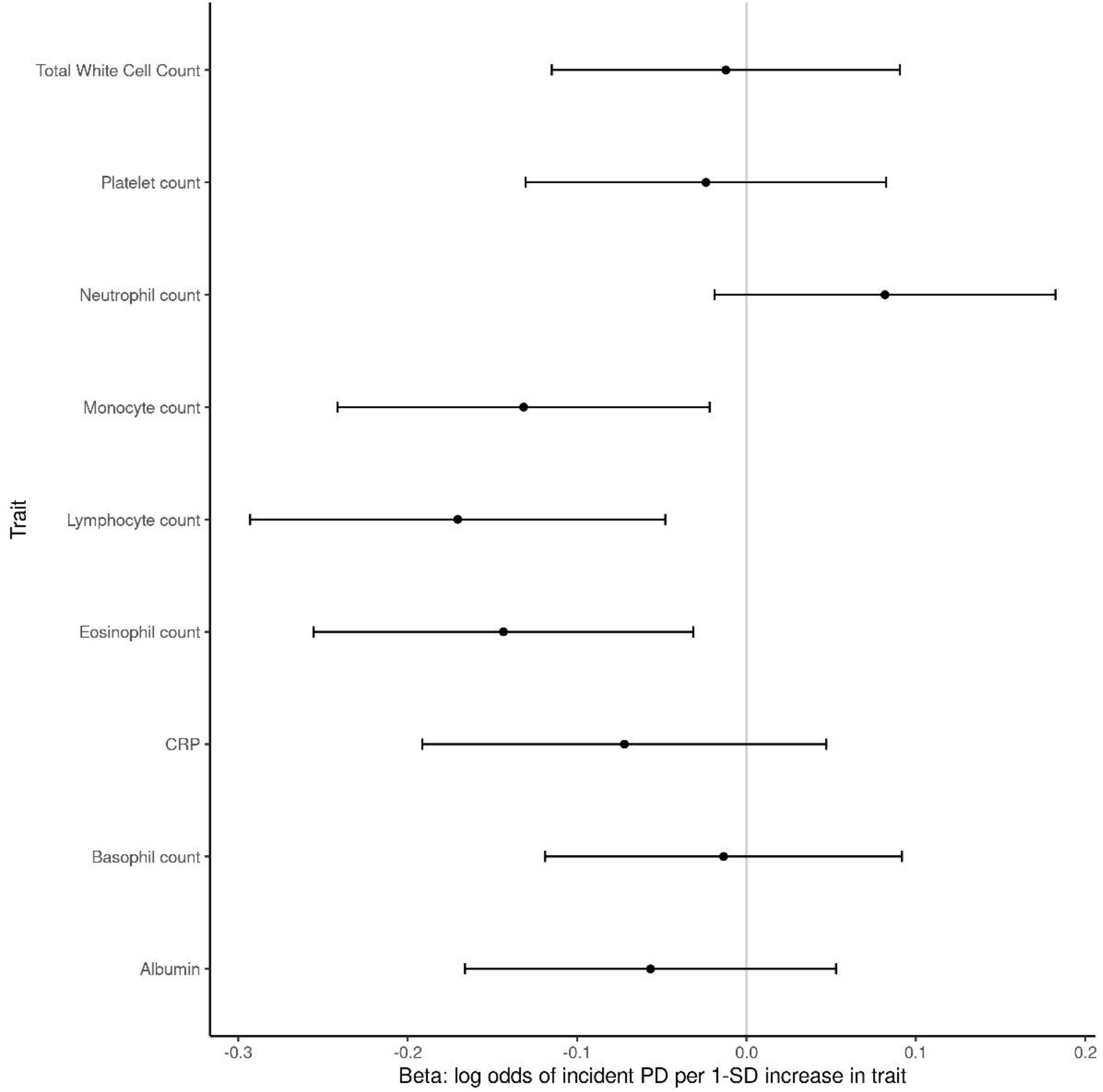

Weighted median - HF
Simple mean - DF
Simple mode - HF
Simple mean - Tophits
Simple mode - HF
Parallised mode - DF + HF


0.1
0.0
-0.1
Per-allele effect on PD risk (logOR)
0.025
0.050
0.075
0.100
Per-allele effect on lymphocyte count (SD)
MR Test
Inverse variance weighted
MR Egger
Simple mode
Weighted median
Weighted mode

-log 10 P-value
Up-regulated DEG
Down-regulated DEG
DEG (both side)
0
2
4
6
8
10
Spleen
Whole_Blood
Cells_EBV-transformed_lymphocytes
Lung
Small_Intestine_Terminal_Ileum
Pancreas
Adrenal_Gland
Heart_Left_Ventricle
Kidney_Cortex
Cervix_Ectocervix
Liver
Minor_Salivary_Gland
Skin_Not_Sun_Exposed_Suprapubic
Heart_Atrial_Appendage
Brain_Substantia_nigra
Ovary
Brain_Spinal_cord_cervical_c-1
Brain_Caudate_basal_ganglia
Fallopian_Tube
Brain_Cerebellum
Muscle_Skeletal
Skin_Sun_Exposed_Lower_leg
Cervix_Endocervix
Thyroid
Brain_Hippocampus
Pituitary
Stomach
Esophagus_Mucosa
Brain_Putamen_basal_ganglia
Colon_Sigmoid
Bladder
Adipose_Visceral_Omentum
Esophagus_Gastroesophageal_Junction
Brain_Cerebellar_Hemisphere
Brain_Nucleus_accumbens_basal_ganglia
Kidney_Medulla
Prostate
Colon_Transverse
Artery_Tibial
Testis
Brain_Hypothalamus
Brain_Amygdala
Brain_Frontal_Cortex_BA9
Adipose_Subcutaneous
Uterus
Vagina
Brain_Cortex
Brain_Anterior_cingulate_cortex_BA24
Esophagus_Muscularis
Cells_Cultured_fibroblasts
Artery_Coronary
Breast_Mammary_Tissue
Artery_Aorta
Nerve_Tibial


Proportion
0.0
0.1
0.2
0.3
0.4
0.5
*
intronic
intergenic
ncRNA_intronic
UTR3
downstream
ncRNA_exonic
upstream
exonic
UTR5
splicing
ncRNA_splicing
-log2(E)
1.53
-0.75
* p<0.05
** p<0.05/11


-log10 P-value
0
20
40
60
80
100
120
140
160
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome


Total White Cell Count
Platelet count
Neutrophil count
Monocyte count
Lymphocyte count
Eosinophil count
CRP
Basophil count
Albumin
Trait
-0.3
-0.2
-0.1
0.0
0.1
0.2
Beta: log odds of incident PD per 1-SD increase in trait