

1 **Multi-ancestry genome-wide association study of neutrophil-lymphocyte ratio and**  
2 **polygenic risk score development to explore causal association with diabetic retinopathy**

3 Aravind Lathika Rajendrakumar<sup>1,2</sup>✉, Anand Thakarakkattil Narayanan Nair<sup>1</sup>✉, Mehul Kumar  
4 Chourasia<sup>1,3</sup>✉, Charvi Nangia<sup>1</sup>, Sundararajan Srinivasan<sup>1</sup>, Venkateshan Radha<sup>4</sup>, Ranjit Mohan Anjana<sup>4</sup>,  
5 Moneeza K Siddiqui<sup>5,1</sup>, Weihua Meng<sup>1,6</sup>, Viswanathan Mohan<sup>4</sup>, Colin N A Palmer<sup>1\*</sup>

6

7 **Affiliation**

8 *1. Division of Population Health and Genomics, Ninewells Hospital, University of Dundee,*  
9 *Dundee, UK, DD1 9SY*

10 *2. Biodemography of Aging Research Unit, Duke University, Durham, NC, 27708-0408, USA*

11 *3. Patient focused Real World Evidence (PRWE)- Oncology, Merck Pharmaceuticals,*  
12 *Bangalore, India*

13 *4. Madras Diabetes Research Foundation & Dr. Mohan's Diabetes Specialities Centre,*  
14 *Gopalapuram, Chennai, India*

15 *5. Wolfson Institute of Population Health, Queen Mary University of London, E1 4NS*

16 *6. Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute,*  
17 *University of Nottingham Ningbo China, Ningbo, China, 315100*

18

19 Aravind Lathika Rajendrakumar, PHD<sup>1,2</sup>

20 Anand Thakarakkattil Narayanan Nair, PHD<sup>1</sup>

21 Mehul Kumar Chourasia, PHD<sup>1,3</sup>

22 Charvi Nangia, PHD<sup>1</sup>

23 Sundararajan Srinivasan, PHD<sup>1</sup>

24 Venkateshan Radha, PHD<sup>4</sup>

25 Ranjit Mohan Anjana MD<sup>4</sup>

26 Moneeza K Siddiqui, PHD,<sup>1,5</sup>

27 Weihua Meng, PHD<sup>1,6</sup>

28 Viswanathan Mohan, MD<sup>4</sup>

29 Colin N A Palmer\*, PHD<sup>1</sup>

30 <sup>‡</sup>Equal contribution

31

32 Word count: 4,095

33 Number of tables and figures: 6

34

35 **Abstract**

36 **Background:** Neutrophil–lymphocyte Ratio (NLR) is a biomarker of inflammation and was  
37 associated with diabetic retinopathy (DR) in earlier studies.

38 **Objective:** To investigate the genetic loci influencing NLR and to estimate the heritability and  
39 causality of DR with the NLR polygenic risk score (PRS).

40 **Design:** Genome-wide association study, conditional analysis, Fine and Gray model (FGR),  
41 Mendelian Randomization (MR)

42 **Setting:** Scottish and South Indian populations drawn from population cohorts and electronic medical  
43 records.

44 **Participants:** 29,317 individuals, with a considerable proportion diagnosed with diabetes.

45 **Measurements:** Effect estimates from GWAS to compute PRS and causal association with DR.

46 **Results:** Heritability estimates for the Scottish and Indian cohorts were 35.3% and 8.7% respectively.  
47 The top Single Nucleotide Polymorphisms (SNPs) in the multi-ancestry analysis (n=29,317) were  
48 intergenic: rs1825819 (Chr4:T/C) (Beta=-0.05, p=2.00x10<sup>-9</sup>), rs2980871 (Chr8:A/G) (Beta=0.04,  
49 p=4.64x10<sup>-8</sup>), rs2227322 (Chr17:C/G) (Beta=0.07, p=4.12x10<sup>-20</sup>) and rs4808047 (Chr19:T/C) (Beta= -

50 0.07,  $p=5.93 \times 10^{-12}$ ). For the construction of best-fit PRS, we used 74,377 of 55,333,12 variants. There  
51 was a dose-response relationship between the PRS and NLR. The subhazard ratio (sHR) for NLR PRS  
52 association with DR was not statistically significant sHR=1.01 (95% CI: 0.97, 1.06,  $p=0.48$ ). Null  
53 associations were observed in both cross-sectional and time-based MR analyses for PRS with DR.

54 **Limitations:** A substantial proportion of the dataset was used for training the PRS algorithm. Due to  
55 trans-ancestry differences, PRS and subsequent analysis were conducted only in the Scottish cohorts.

56 **Conclusions:** Multiple novel intergenic SNP associations were discovered, complementing those  
57 previously identified. Of these, some SNPs were also associated with genes known to regulate white  
58 blood cells, but not specifically NLR. More studies are required to confirm the causality between  
59 systemic inflammation and DR.

60 **Primary Funding Source:**

61 National Institute for Health Research, Pioneer and Leading Goose R&D Program of Zhejiang 2023,  
62 and the Ningbo International Collaboration Program 2023.

63

64 **Introduction**

65 Microvascular abnormalities are not observed in certain individuals despite having long-term diabetes  
66 but the exact reasons for this are ill understood.<sup>1-4</sup> In complex diseases, the genetic liability of an  
67 individual, can be summarised into an informative risk score that can be used in disease prediction  
68 models.<sup>5,6</sup> Diabetes and inflammation share overlapping biological mechanisms.<sup>7</sup> The global burden  
69 of diabetic retinopathy (DR) was close to 23% and more than 6% experience vision problems.<sup>8</sup>  
70 Increased concentrations of several inflammatory molecules were noted in tissues and vitreous fluids  
71 of individuals with DR.<sup>9</sup> Genetic modification of specific inflammatory pathways was shown to  
72 arrest the progress of DR.<sup>10</sup> Leukocyte recruitment in DR was found to prolong the inflammatory  
73 activity.<sup>11</sup> Neutrophil-Lymphocyte Ratio (NLR) was shown to be more clinically informative for  
74 several clinical conditions, than a more established marker of inflammation such as C-reactive protein

75 (CRP) in interferonopathies.<sup>12</sup> Hence, modeling the joint variations in neutrophils and lymphocytes  
76 could be more informative than using them individually as markers of inflammation.<sup>13–17</sup> Indeed, the  
77 NLR has been shown to be associated with diabetes and its complications.<sup>18–21</sup> However, no studies  
78 have directly investigated the genetic causality of the observed associations. We aimed to identify  
79 new genetic loci for NLR with a strict phenotype definition. An allele score for NLR could be more  
80 powerful as it entails comprehensive information from multiple genetic loci. We hypothesized that the  
81 weighted NLR allele score could be used as an ideal instrumental variable to represent the genetic  
82 propensity for a heightened inflammatory response. We therefore performed a genome-wide  
83 association study (GWAS) for NLR and used the resulting PRS, both as an exposure and instrument  
84 variable to predict and find causal association with DR.

## 85 **Methods**

### 86 **Cohort description and sample selection**

87 The India-Scotland Partnership for Precision Medicine in Diabetes (INSPIRED) is a collaboration  
88 between Dr. Mohan's Diabetes Specialties Centre (DMDSC), and the associated Madras Diabetes  
89 Research Foundation (MDRF) at Chennai in South India and the University of Dundee, Scotland. In  
90 the discovery phase, we meta-analyzed 3 Scottish cohorts: Genetics of Diabetes Audit and Research  
91 in Tayside Scotland (GoDARTS-T2D), Scottish Health Research Register (SHARE), and GoDARTS  
92 non-type 2 diabetes controls (GoDARTS controls). However, a small proportion of GoDARTS  
93 controls developed diabetes after enrollment. These Scottish cohorts consist of individuals of  
94 predominantly white European ancestry. These cohorts and the data linkage strategies used for  
95 research have been described previously.<sup>22,23</sup> The findings were replicated in the DMDSC cohort 1  
96 and cohort 2. The main difference between the DMDSC cohorts was the sampling strategy employed  
97 for genotyping. The participants were selected retrospectively in DMDSC cohort 1. NLR PRS was  
98 computed in SHARE using GWAS estimates from the GoDARTS T2D and controls.

99

100

## 101 **Genotyping**

102 Genotyping methods for the Scottish and Indian cohorts have been described previously.<sup>24</sup> The DNA  
103 was extracted from the stored blood plasma. A trained bioinformatician performed the quality control  
104 using standard protocols. For all the cohorts, Haplotype Reference Consortium (HRC) was used to  
105 infer the missing genotypes. We utilized the GRCh37 assembly (Genome Research Consortium  
106 human build 37) to locate SNPs. Affymetrix 6.0, Illumina Human Omni Express, and Broad Institute  
107 platforms were used for genotyping GoDARTS. Both, DMDSC and SHARE participants were  
108 assayed using the Illumina GSA platform. Alleles with a minimum allele frequency (MAF) greater  
109 than 0.05 were considered for GWAS. Additionally, we excluded single nucleotide polymorphism  
110 (SNPs) with an info score  $>0.70$  and Hardy-Weinberg equilibrium (HWE)  $<1 \times 10^{-10}$ . Principal  
111 components were computed using Plink Version 1.9.<sup>25,26</sup> We have used STREGA guidelines to report  
112 our findings.<sup>27</sup>

113

## 114 **Definition of NLR and DR phenotypes**

115 The Scottish diabetic retinopathy grading scheme was used to ascertain DR status (yes/no) in the  
116 Scottish cohort.<sup>28</sup> Details of retinal screening in Tayside were described previously.<sup>29</sup> For the time-  
117 based Mendelian Randomization (MR) analysis, the first reported DR (any grade from R1 to R4) was  
118 considered. For non-DR individuals, the maximum follow-up date was considered as the censoring  
119 date. NLR was calculated as the absolute count of neutrophils divided by the absolute count of  
120 lymphocytes from the hematology dataset. All values of neutrophils and lymphocytes above and  
121 below 5 standard deviations (SD) were removed. The change in NLR above 2.5 from the previous  
122 reading was considered and hence excluded. For, individuals having a single NLR reading, values  
123 above 5 were excluded. NLR values after diagnosis or treatment for malignancies or within a window  
124 period of 31 days after the diagnosis of infectious diseases were discarded. We also did not retain an  
125 NLR reading within 28 days after the first reading, assuming that as a case of possible visit for a  
126 disease condition.

127 **Ethics Approval**

128 **Tayside, Scotland UK:** Ethical approval for the study was provided by the Tayside Medical Ethics  
129 Committee (REF:053/04) and the study has been carried out in accordance with the Declaration of  
130 Helsinki.

131 **SHARE, Scotland UK:** Ethical approval of the study was provided by the ethics committee SHARE  
132 East of Scotland (REF: NHS REC 13/ES/0020).

133 **DMDSC:** NIHR Global Health Research Unit on Global Diabetes Outcomes Research, Institutional  
134 Ethics Committee of Madras Diabetes Research Foundation, Chennai, India. IRB number  
135 IRB00002640. Granted 24<sup>th</sup> August 2017.

136 **Statistical Analysis**

137 **GWAS analysis and ascertainment of independent loci**

138 The analytical approaches are shown in Supplementary Figure 1. Continuous variables were  
139 summarized as mean/median and categorical variables as frequencies and percentages. GWAS was  
140 performed in SNPtest version 2.5.2 with 10 principal components, age, sex, and diabetes status as  
141 covariates. Genome-wide Complex Trait Analysis (GCTA) software was used to identify independent  
142 genomic loci and genomic inflation was controlled during the computation process.<sup>30</sup> We used the  
143 cojo-joint function in the GCTA suite which uses all the SNPs to calculate joint effect size without  
144 eliminating less important SNPs from the model. GWAS summary estimates were meta-analyzed  
145 with GWAMA (Genome-Wide Association Meta-Analysis) by applying the the genomic control  
146 option (-gco) in GWAMA to address inflation and assuming an inverse variance weighted model with  
147 a p value significance cut off of  $5 \times 10^{-8}$ .<sup>31</sup>

148 **Gene enrichment, functional analysis, heritability estimation, and gene network analysis**

149 Functional Mapping and Annotation of Genome-wide Association Studies (FUMA) was used to  
150 create the regional plot, annotation, enrichment statistics, and gene-based tests. We also took  
151 annotations from gnomAD (<https://gnomad.broadinstitute.org/>), UCSC browsers

152 (<https://genome.ucsc.edu/>), and genecards (<https://www.genecards.org/>) if the annotations were not  
153 available in the FUMA. GeneMANIA was used to understand the co-expression and pathway analysis  
154 (<https://genemania.org/>). GeneMANIA predicts weighted gene networks based on well-known genetic  
155 databases to visualize gene interactions.<sup>32</sup> Narrow-sense heritability ( $h^2$ ) was estimated using LDAK  
156 software.<sup>33</sup>

### 157 **PRS estimation**

158 PRSice-2 and packages in R version 3.6.1 (The R Foundation for Statistical Computing, Vienna,  
159 Austria) were used to generate PRS and perform competing risks analyses, respectively.<sup>34-36</sup> The  
160 Clumping and Thresholding method (C+T) in PRSice-2 was utilized for SNP thinning. GoDARTS  
161 T2D and GoDARTS controls were used as the base data and SHARE as the target data.

### 162 **Competing risks regression and MR analysis**

163 MR and competing risk analyses were performed using allele score as an instrument in the SHARE  
164 cohort. Fine and Gray analysis using the riskRegression package with deaths as a competing event  
165 adjusting for age, sex, HbA1c, BMI, and creatinine.<sup>37,38</sup> Subsequently, The PRS from the SHARE  
166 cohort was used to test the causal relationship between NLR and DR using the one-sample Mendelian  
167 randomization (MR) method. For the Mendelian randomization analysis, we experimented with both  
168 cross-sectional and time-based models for the same DR phenotypes using R codes, and the *ivtools*  
169 package was used for survival MR.<sup>39</sup>

### 170 **Role of the Funding Source**

171 The funder had no role in the design, conduct, or analysis of the study or the decision to submit the  
172 manuscript for publication.

### 173 **Results**

#### 174 **Baseline characteristics of the participants**

175 The characteristics of the Scottish discovery cohort (white Europeans,  $n = 21,153$ ) are shown in  
176 Supplementary Table 1. Among these, the GoDARTS had the highest sample size. Over 96% of

177 individuals had diabetes in the Affymetrix and Illumina sub-cohorts. The median NLR for all the  
178 Scottish cohorts was above 2. GoDARTS controls had the lowest median NLR (2.04), and only 7.7%  
179 had diabetes. Participants in the GoDARTS were older in comparison to the SHARE cohort. Except  
180 for GoDARTS controls, the proportion of males in all other cohorts was above 50%. In the DMDSC  
181 (Asian Indians = 8,164), the median NLR (1.93) and the median age was lower (less than 60 years)  
182 than the Scottish cohort (Supplementary Table 2).

### 183 **Discovery Phase in white European cohorts**

184 Meta-analysis of the 3 Scottish cohorts revealed 150 significant SNPs, most of which were in  
185 Chromosome 4 and Chromosome 17 (Supplementary Figures 2-3). Out of these, 128 were from  
186 Chromosome 17 and were in high linkage disequilibrium (LD). Only 3 SNPs were independently  
187 associated with the NLR trait, of which only one, rs3826331 in chromosome 17, and the rest,  
188 rs6841652 and rs16850400 in chromosome 4 (Table 1). Among these, rs3826331 is an intron near the  
189 neutrophil gene *PSMD3*, rs6841652 can be mapped either to the *LINCO2513* or *TBC1D1* genetic  
190 regions, and the gene corresponding to rs16850400 is yet to be determined. Of the 3 SNPs, the  
191 association signal for rs16850400 was suggestive ( $p=2.39 \times 10^{-6}$ ). The estimated SNP heritability was  
192 35.3% (SD: 0.03). The genomic inflation factor ( $\lambda$ ) was within reasonable limits for individual  
193 GWAS but was slightly on the higher side for the fixed effects meta-analysis ( $\lambda=1.09$ ).

194 Table 1. SNP estimates were identified as conditionally independent using GCTA in the Scottish  
195 GWAS meta-analysis

rsid	Genotype	MAF	Beta	freq	bJ	pJ GC	Gene
rs3826331	T/C	0.40	-0.076	0.40	-0.076	$6.53 \times 10^{-14}$	<i>PSMD3</i>
rs6841652	T/C	0.31	-0.065	0.31	-0.065	$1.37 \times 10^{-9}$	<i>LINCO2513, TBC1D1</i>
rs16850400	T/C	0.40	0.049	0.60	0.049	$2.39 \times 10^{-6}$	---



196 Note. MAF minor allele frequency, freq genotype frequency, bJ beta from the conditional analysis, pJ  
197 GC - p values from the conditional analysis.

### 198 **Replication phase in Asian Indian cohort**

199 The DMSDC GWAS plots are shown in Supplementary Figures 5 and 6. The genome-wide search for  
200 SNPs in the DMDSC cohorts found two highly correlated SNPs in Chromosome 6. Table 2 shows the  
201 details of these SNPs: rs17330192 ( $p=1.8 \times 10^{-8}$ ), and rs139801819 ( $p=2.4 \times 10^{-8}$ ) were introns found  
202 near the non-coding region of *LOC102724591* (the RNA gene). Heritability estimated in the Indian  
203 population was 8.7% (SD: 0.07). During the replication effort, a reversal of SNP effects was noted  
204 (Table 3), and rs3826331 (T/C) remained statistically significant but with a reversed effect as a trait-  
205 increasing allele (Beta =0.06,  $p=1.5 \times 10^{-6}$ ). We were unable to replicate rs6841652 ( $p=0.06$ ) and  
206 rs16850400 ( $p=0.08$ ) in the DMDSC cohort. Arguably, the MAF was relatively common in both  
207 ancestries and therefore would not have been the reason for non-replication.

208 Table 2. Estimates of SNP associations with NLR from the replication analysis in the DMDSC  
209 (n=8,164)

rsid	chr: Pos	Genotype	MAF	Beta	p	I <sup>2</sup>	Gene
rs3826331	17:38150492	T/C	0.39	0.06	$1.5 \times 10^{-6}$	0.77	<i>PSMD3</i>
rs6841652	4:38363262	T/C	0.39	0.02	0.06	0	<i>LINCO2513, TBC1D1</i>
rs16850400	4:74927915	T/C	0.43	-0.02	0.08	0	---

210 Note. MAF minor allele frequency, chr: pos chromosome to position, I<sup>2</sup> - heterogeneity index  
211 indicating the percentage of variation across studies

### 212 **Fixed effects trans-ancestry meta-analysis phase**

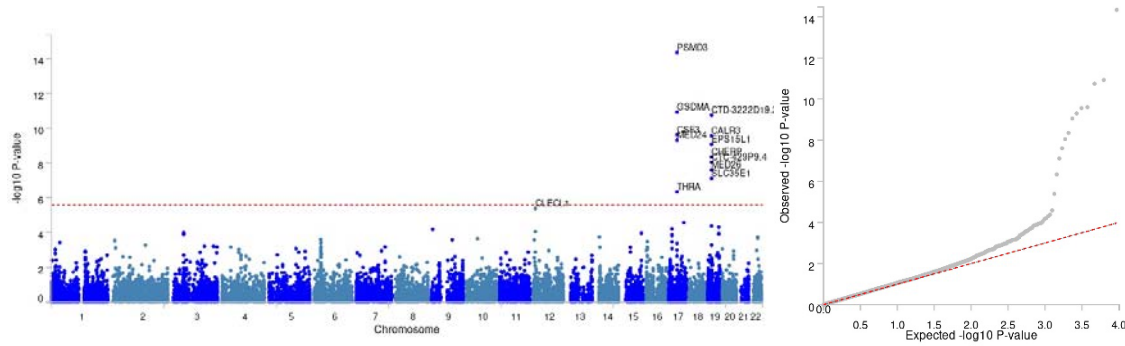
213 The final sample size for the FE trans-ancestry meta-analysis was 29,317, and the SNP associations  
214 are shown in Supplementary Figures 6-7. We identified two novel SNPs in Chromosome 4 and 8:  
215 rs1825819 ( $p=2.0 \times 10^{-9}$ ) and rs2980871 ( $p=4.6 \times 10^{-8}$ ) respectively (Table 3). The genomic inflation  
216 factor post-correction was 1.07. Regional plots show that both chromosome 17 and chromosome 19

217 had many SNPs in high linkage disequilibrium (Supplementary Figures 8-12). Conversely,  
218 chromosome 8 and chromosome 4 had very few highly correlated SNPs.

219 In the FE trans-ancestry analysis, all the lead SNPs were found to be in the non-coding regions.  
220 Four lead SNPs were found statistically to be significant, of which rs2227322 had the highest trait-  
221 increasing effect size (0.07,  $p=4.12 \times 10^{-20}$ ). This UTR5 variant was located on chromosome 17 in  
222 the *CSF3* gene. This gene was responsible for granulocyte synthesis and also cytokine-related  
223 inflammatory responses (<https://www.ncbi.nlm.nih.gov/gene/1440>). The second notable SNP was  
224 rs4808047 on Chromosome 19 ( $-0.07$ ,  $p=5.93 \times 10^{-12}$ ) in the *EPS15LI* gene and had the lowest  
225 MAF (0.28) among the lead SNPs. SNP rs2980871 in Tribbles homolog 1 (*TRIB1*) was another  
226 interesting finding. In mouse models, *TRIB1* expression affects hepatic lipid synthesis and  
227 glycogenesis.<sup>40</sup> On chromosome 4, rs1825819 with a MAF of 0.38 was found to be associated with  
228 NLR ( $-0.05$ ,  $2.00 \times 10^{-9}$ ). However, there were no known previous phenotype associations or  
229 functions identified for this specific SNP, making it a novel locus for NLR. The sentinel SNPs  
230 identified in the trans-ancestry meta-analysis generally showed a consistent direction in terms of  
231 their effects, except for the SNP rs1825819. The direction of this SNP was trait-increasing in the  
232 Broad array and reversed for other cohorts.

233 Figure 1 shows the significant genes in the trans-ancestry meta-analysis, in the gene-based test,  
234 *CLECLI* (chromosome 12), *GSDMA* (chromosome 17), *CSF3* (chromosome 17), *PSMD3*  
235 (chromosome 17), *MED24* (chromosome 17), *THRA* (chromosome 17), *CTD-3222D19.2*  
236 (chromosome 19), *EPS15LI* (chromosome 19), *CALR3* (chromosome 19), *CHERP* (chromosome 19),  
237 *CTC429P9.4* (chromosome 19), *MED26* (chromosome 19), and *SLC35E1* (chromosome 19) were  
238 identified. *CLECLI* on chromosome 12 was another important gene that was identified. Among the  
239 13 genes reported in the gene-based test, *GSDMA*, *CLECLI*, *CALR3*, and *SLC35E1* did not have any  
240 functional relationships with other genes in the network analysis. No gene interaction or pathway  
241 information was available for the genes *CTD-3222D19.2* and *CTC429P9*.

242



243

244 Figure 1. Manhattan plot and (left) Q-Q plot (right) corresponding to the gene set analysis from the  
 245 trans-ancestry GWAS summary (n= 29,317) Note: The SNPs corresponded to 9528 protein coding  
 246 genes, the genome-wide significance threshold (red dashed line in the plot) was established at P =  
 247  $0.05/9528 = 5.248e-6$ .

248 Table 3. Lead SNPs identified from trans-ancestry inverse variance meta-analysis of Scottish and  
 249 Indian Cohorts (n=29,317)

rsid	chr: Position	Genotype	MAF	Beta	p	I <sup>2</sup>	Location	Nearest Locus
rs1825819	4:38340430	T/C	0.38	-0.05	$2.00 \times 10^{-9}$	0.63	--	--
rs2980871	8:126488930	A/G	0.50	0.04	$4.64 \times 10^{-8}$	0	intron	<i>TRIB1</i> , <i>LINC00861</i>
rs2227322	17:38171668	C/G	0.45	0.07	$4.12 \times 10^{-20}$	0	UTR5	<i>CSF3</i>
rs4808047	19:16527834	T/C	0.28	-0.07	$5.93 \times 10^{-12}$	0	intronic	<i>EPS15L1</i>

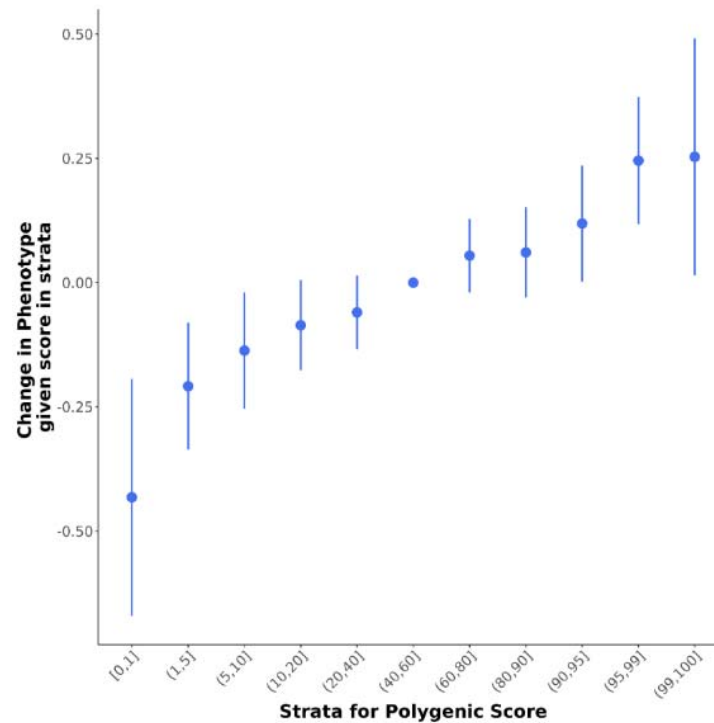
250 MAF minor allele frequency, chr: Position chromosome to position, I<sup>2</sup> heterogeneity index indicating  
 251 the percentage of variation across studies

252 The independent SNPs identified in the discovery set also remained statistically significant in the FE  
 253 tran-ancestry analysis. Supplementary Figure 13 illustrates the details of the number of SNPs and  
 254 genes in the important genomic loci in the meta-analysis. Tissue-wise differential gene expression  
 255 analysis showed that the genes were generally upregulated in the adrenal glands, whole blood, and  
 256 multiple regions of the brain (Supplementary Figure 14). However, these did not surpass the statistical  
 257 significance threshold. Gene expression for nine selected significant genes from the gene-based

258 analysis is plotted in Supplementary Figure 15. *PSMD3*, the major gene in our analysis, along with  
259 *MED24* and *KLF2*, had a wider gene expression for all the tissues explored in the *GTE*x v8 dataset.  
260 *CSF3* had an increased expression level in adipose tissue, fibroblasts, the pituitary, the prostate, the  
261 fallopian tubes, and musculoskeletal tissue. *GSMDA* expressions were significant only in two tissues,  
262 both relating to skin. Other genes, *CHERP*, *KLF2*, *MED24*, *PSMD3*, and *THRA*, showed heightened  
263 expression levels in an overlapping manner for a wide variety of tissues. Significant overlap in gene  
264 enrichment was seen in chromosomes 17 and 19 (Supplementary Figure 16). In the gene set analysis,  
265 for the 10 leading pathways, only the potassium ion transport pathway remained significant after  
266 applying the Bonferroni correction (Supplementary Table 5).

267 We also explored the network of genes closest to the significant SNPs in the trans-ancestry meta-  
268 analysis (Supplementary Figure 17). The network analysis of significant genes in the trans-ancestry  
269 meta-analysis revealed evidence of high co-expression (56.34%), physical interaction (25.35%), and  
270 common pathways (13.44%). For the construction of the best-fit PRS, we used 74,377 of 55,333,12  
271 variants using the clumping and thresholding (C+T) method for SNP thinning (Supplementary Figures  
272 18-19). The strata plot indicates a dose-response relationship between allele scores and the phenotype,  
273 although this was not discernible in the two uppermost quantiles (Figure 2). The gender-wise  
274 distribution of NLR and PRS is shown in Supplementary Figure 20.

275



276

277 Figure 2. Strata Plot for the NLR PRS in SHARE. Note: The plot illustrates the change in the  
278 computed PRS for NLR on the x-axis and the values of NLR on the y-axis. The blue dot represents  
279 the point estimates, and the lines represent the upper and lower 95% confidence intervals for the  
280 change in NLR corresponding to the PRS. In general, the NLR phenotype increases with the predicted  
281 PRS scores, indicating that PRS can adequately explain the individual variation of the NLR.

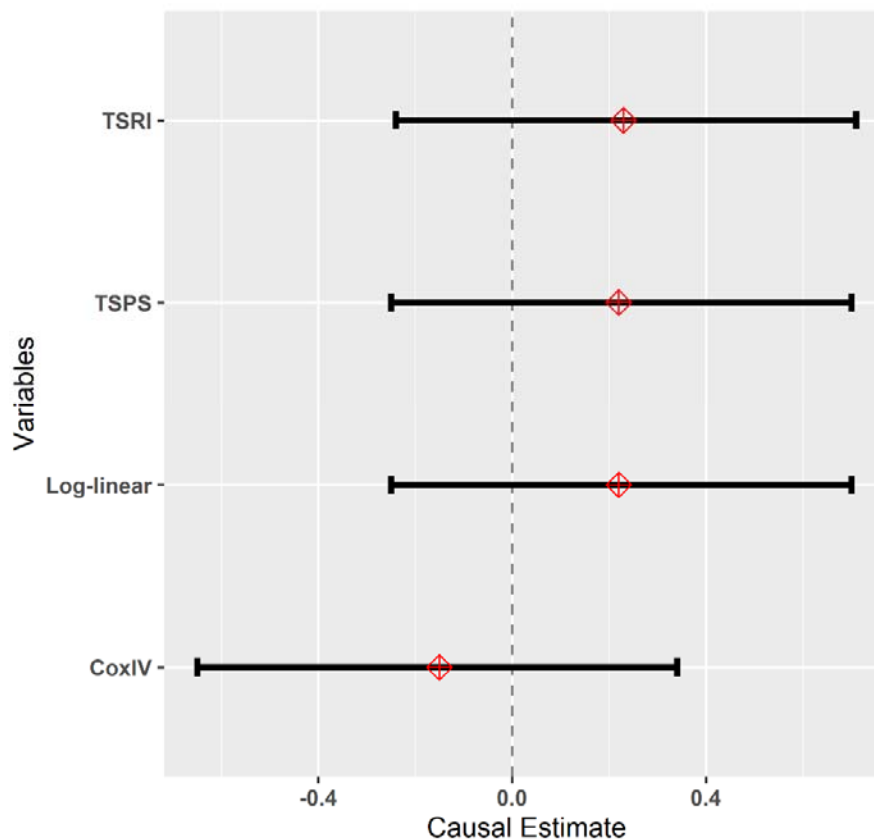
## 282 **Competing risks analyses**

283 In the SHARE cohort (n=3,892), the Fine and Gray model did not indicate any statistically significant  
284 association with DR (n=1,610) in both adjusted and unadjusted models (Supplementary Table 6).  
285 There were 184 competing events, and the mean follow-up time was 4.8 years. The adjusted  
286 subhazards for NLR PRS with DR were 1.01 (95% CI: 0.97, 1.06, p=0.48). Thus, PRS for NLR may  
287 not be useful for predicting the long-term manifestation of diabetic retinopathy.

## 288 **Mendelian randomization analysis for DR using NLR allele scores**

289 The regression coefficient between the phenotype and PRS was statistically significant at 0.11  
290 ( $p=3.41 \times 10^{-15}$ ) (Supplementary Figure 21). The PRS predicted only 1.2% of the NLR variation, which  
291 improved to 3.5% after the addition of age and sex (Supplementary Table 7). We analysed 3,561

292 Scottish individuals (Supplementary Figures 22-23) with data on PRS, NLR, and diabetic retinopathy  
293 for the MR analysis, of which 1,537 (43.1%) had DR. The F-statistic indicates a strong relationship  
294 between the PRS and NLR (close to 60), even after adjusting for the effects of age and sex. The  
295 results from the MR analysis are given in Figure 3. The estimates were very similar between the  
296 models. Logistic-linear and TSPS estimated a non-significant effect (n= 3,561, beta = 0.22, 95% CI: -  
297 0.25, 0.70, p=0.35). For the Cox IV method (n=3,081, DR=1216), The mean follow-up time was 4.8  
298 years. Again, there were no statistically relevant findings between a unit change in the PRS and  
299 hazards for DR (-0.15, 95% CI: -0.65, 0.34, p=0.55). Overall, our analysis provided no evidence of a  
300 causal association between genetically determined levels of NLR and DR.



301

302 Figure 3. MR analysis of NLR allele score with any DR. Note. Log-linear Logistic linear, TSRI two-  
303 stage residual inclusion, TSPS two-stage predictor substitution, IV Instrumental Variable, SE standard  
304 error, CI confidence interval

305 **Discussion**

306 Our study found a divergence in the genetic architecture of NLR between South Asians and white  
307 Europeans. South Asian populations are generally underrepresented in genetic studies of NLR, and  
308 the present results may cover this gap. In general, white Europeans had a much higher NLR relative  
309 to other ancestries.<sup>41</sup> The heritability estimates ( $h^2$ ) in our work also reflect this disparity. Our  
310 heritability estimates almost matched those in the previous Dutch twins' study.<sup>42</sup> White blood cells  
311 are genetically determined; an example is the Duffy-Null polymorphism among the African  
312 population arising due to selection pressures.<sup>43</sup>

313 Remarkably, NLR was significantly correlated with multiple inflammatory markers such as CRP  
314 and IL6, but not platelet-lymphocyte ratio (PLR), implying that the variation of WBC ratios could  
315 provide biological insights.<sup>42</sup> Overall, some important genes are associated with white blood cell-  
316 related traits, but not specifically with the NLR. Chromosome 17q21 contains many genes that exert  
317 substantial pleiotropic effects across different WBC subcomponents and is highly expressed in  
318 inflammatory disorders such as asthma and diabetes.<sup>44,45</sup> Given diabetes and inflammation, there was  
319 evidence of *PSMD3* involvement through insulin resistance and receptor tyrosine kinase (RET)  
320 signalling, respectively.<sup>46-49</sup> *CSF3*, *MED24*, and *PSMD3* genes were reported together to influence  
321 neutrophil counts in a Bayesian meta-analysis involving multiple ancestries.<sup>50</sup> Similarly, *GSDMA*,  
322 *THRA*, *EPS15LI*, *CALR3*, *CHERP*, and *MED24* were also associated with multiple white blood cell  
323 types, including neutrophils and lymphocytes.<sup>45</sup> Importantly, the rest of the significant genes;  
324 *CTC429P9.4*, *CTD3222D19.2*, and *SLC35E1*, do not seem to have a direct link with white blood cell  
325 physiology and are therefore novel findings that merit further mechanistic studies. Another novel  
326 NLR gene was *CLECLI*, which takes part in the T lymphocyte pathway but is also differentially  
327 expressed in clusters of inflammatory conditions such as diabetes mellitus and dyslipidemia.<sup>51</sup>

328 We replicated rs6841652 of the *LINCO2513/TBC1D* gene. This SNP was also associated with  
329 monocyte-lymphocyte ratio and multiple SNPs, that could be traced back to the same genes were  
330 significant for NLR.<sup>52</sup> We were also able to detect associations with the thyroid hormone receptor  
331 gene (*THRA*) that were previously reported for NLR in the Korean population.<sup>53</sup> The complex cross-

332 talk between thyroid hormones and immune cells is vital for the proper regulation of inflammatory  
333 activity.<sup>54</sup> Most genes from the gene-based set were related in varying degrees, as elicited in the  
334 network analysis. Moreover, the gene set analysis overwhelmingly showed affinity for the potassium  
335 ion transportation pathway.

336 One of the strengths of the study is that, we did not transform the NLR as opposed to previous studies  
337 <sup>53,55</sup> As data transformation could raise concerns of reduced statistical power and ease of beta  
338 interpretation.<sup>56</sup> A summarized NLR could override possible bias arising from health conditions,  
339 outliers, and other unusual deviations from the original phenotype for accurate genome-wide  
340 scanning. We further refined the SNP results in each ancestry to identify independent SNPs free from  
341 LD. To our knowledge, this is the first study that details the generation of an allelic summary score  
342 for NLR and uses it for a causal analysis for DR. We could not replicate the NLR association for DR  
343 in the observational study using the PRS.<sup>57-59</sup> The high F statistics for the IV-phenotype relation do  
344 not completely eliminate the possibility of bias.<sup>60</sup> It is also possible that the NLR is regulated at a  
345 more downstream level by several competing inflammatory pathways or environmental stimuli.  
346 Therefore, the current findings may not be sufficient to rule out the role of inflammation in DR. The  
347 major limitation was that a substantial proportion of the dataset was used to train the PRSice-2  
348 algorithm, leaving a much smaller sample to investigate DR. Due to ancestry differences, PRS and  
349 subsequent analysis were conducted only in the Scottish cohorts. Finally, our one sample MR analysis  
350 is relatively less powerful compared to a two-sample MR.<sup>61,62</sup> Additional studies are required to  
351 understand the genetic mechanisms influencing the NLR between ancestrally diverse populations and  
352 subpopulations.

### 353 **Conclusion**

354 In summary, we report multiple novel intergenic SNPs associated with the NLR in addition to those  
355 identified for WBC previously and most of the identified SNPs were in the intergenic regions. There  
356 was no evidence to suggest causality between systemic inflammation represented by NLR and DR. In  
357 the context of chronic diseases, our findings may provide insights into inflammatory pathways to  
358 develop better strategies to deal with inflammatory outcomes.



359 **Availability of data and materials:**

360 The Scottish and DMDSC datasets are stored electronically in a secure environment. The data is not  
361 currently available in the public domain as the participant consent restricts the public sharing of the  
362 data. Researchers will require prior approval from the HIC and DMDSC for accessing the related  
363 datasets.

364

365 **Abbreviations**

BMI	Body Mass Index
CI	Confidence Interval
CIF	Cumulative Incidence Function
chr: Position	Chromosome to Position
CURES	Chennai Urban Rural Epidemiology Study
DBP	Diastolic Blood Pressure
DMDSC	Dr. Mohan's Diabetes Specialties Clinics
DR	Diabetic Retinopathy
eGFR	Estimated Glomerular Filtration Rate
FE	Fixed Effects
FGR	Fine and Gray model
FUMA	Functional Mapping and Annotation of Genome-wide Association Studies
GCTA	Genome-wide Complex Trait Analysis
GIF	Genomic Inflation Factor
GoDARTS	Genetics of Diabetes Audit and Research in Tayside Scotland
HbA1c	Glycated Haemoglobin
HDLc	High-density lipoprotein cholesterol
HIC	Health Informatics Centre

HRC	Haplotype Reference Consortium
HWE	Hardy-Weinberg Equilibrium
LD	Linkage Disequilibrium
LDL-C	Low-density Lipoprotein -Cholesterol
MAGMA	Multi-marker Analysis of GenoMic Annotation
MDRF	Madras Diabetes Research Foundation
NLR	Neutrophil Lymphocyte Ratio
Non-HDL-C	Non-high-density Lipoprotein Cholesterol
PLR	Platelet-lymphocyte ratio
PRS	Polygenic Risk Score
RET	Receptor Tyrosine Kinase
SBP	Systolic Blood Pressure
SD	Standard Deviation
SHARE	Scottish Health Research Register
SE	Standard Error
sHR	Sub Hazard Ratio
SNP	Single Nucleotide Polymorphisms

366

367 **Grant Support:**

368 This research was funded by the National Institute for Health Research (NIHR) (INSPIRED  
369 16/136/102) using UK aid from the UK Government to support global health research. The views  
370 expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the  
371 UK Department of Health and Social Care. Weihua Meng is supported by the Pioneer and Leading  
372 Goose R&D Program of Zhejiang 2023 (2023C04049) and the Ningbo International Collaboration  
373 Program 2023 (2023H025).

374

375 **Corresponding Author:**

376 \*Corresponding author: Professor Colin N A Palmer, University of Dundee, Dundee, UK., Phone:  
377 +44 (0)1382 383155, Email: [c.n.a.palmer@dundee.ac.uk](mailto:c.n.a.palmer@dundee.ac.uk)

378

## 379 **References**

- 380 1. Klein BEK, Knudtson MD, Tsai MY, Klein R. The relation of markers of inflammation and  
381 endothelial dysfunction to the prevalence and progression of diabetic retinopathy: Wisconsin  
382 epidemiologic study of diabetic retinopathy. *Arch Ophthalmol.* 2009;127(9):1175–82.
- 383 2. Klein BEK, Myers CE, Howard KP, Klein R. Serum lipids and proliferative diabetic  
384 retinopathy and macular edema in persons with long-term type 1 diabetes mellitus: The  
385 Wisconsin epidemiologic study of diabetic retinopathy. *JAMA Ophthalmol.* 2015;133(5):503–  
386 10.
- 387 3. Klein R, Klein BEK, Moss SE, Cruickshanks KJ. Relationship of Hyperglycemia to the Long-  
388 term Incidence and Progression of Diabetic Retinopathy. *Arch Intern Med.*  
389 1994;154(19):2169-78.
- 390 4. Meng W, Shah KP, Pollack S, Toppila I, Hebert HL, McCarthy MI, et al. A genome-wide  
391 association study suggests new evidence for an association of the NADPH Oxidase 4 (NOX4)  
392 gene with severe diabetic retinopathy in type 2 diabetes. *Acta Ophthalmol.* 2018;96(7):e811–9.
- 393 5. Pollack S, Igo RP, Jensen RA, Christiansen M, Li X, Cheng CY, et al. Multiethnic genome-  
394 wide association study of diabetic retinopathy using liability threshold modeling of duration of  
395 diabetes and glycemic control. *Diabetes.* 2019;68(2):441–56.
- 396 6. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score  
397 analyses. *Nat Protoc.* 2020;15:2759–72.
- 398 7. Tsalamandris S, Antonopoulos AS, Oikonomou E, Papamikroulis GA, Vogiatzi G,  
399 Papaioannou S, et al. The Role of Inflammation in Diabetes: Current Concepts and Future  
400 Perspectives. *Eur Cardiol.* 2019;14(1):50-59.

- 401 8. Teo ZL, Tham YC, Yu M, Chee ML, Rim TH, Cheung N, et al. Global Prevalence of Diabetic  
402 Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis.  
403 *Ophthalmology*. 2021;128(11):1580–91.
- 404 9. Wakabayashi Y, Usui Y, Okunuki Y, Kezuka T, Takeuchi M, Goto H, et al. Correlation of  
405 vascular endothelial growth factor with chemokines in the vitreous in diabetic retinopathy.  
406 *Retina*. 2010;30(2):339–44.
- 407 10. Tang J, Kern TS. Inflammation in diabetic retinopathy. *Prog Retin Eye Res* [Internet].  
408 2011;30(5):343–58.
- 409 11. Noda K, Nakao S, Ishida S, Ishibashi T. Leukocyte Adhesion Molecules in Diabetic  
410 Retinopathy. *J Ophthalmol*. 2012 ;2012:279037.
- 411 12. Kourilovitch M, Galarza–Maldonado C. Could a simple biomarker as neutrophil-to-  
412 lymphocyte ratio reflect complex processes orchestrated by neutrophils? *J Transl Autoimmun*.  
413 2022;6:100159.
- 414 13. Anderson BO, Brown JM, Harken AH. Mechanisms of neutrophil-mediated tissue injury. *J*  
415 *Surg Res*. 1991;51(2):170–9.
- 416 14. Liew PX, Kubes P. The Neutrophil’s Role During Health and Disease. *Physiol Rev* [Internet].  
417 2019;99(2):1223–48.
- 418 15. WW Zhou GY An YPC. Significant difference of neutrophil-lymphocyte ratio between  
419 colorectal cancer, adenomatous polyp and healthy people. *Eur Rev Med Pharmacol Sci*.  
420 2017;21(23):5386–91.
- 421 16. Balta S, Celik T, Mikhailidis DP, Ozturk C, Demirkol S, Aparci M, et al. The Relation  
422 Between Atherosclerosis and the Neutrophil–Lymphocyte Ratio. *Clin Appl Thromb* .  
423 2016;22(5):405–11.
- 424 17. Liu CC, Ko HJ, Liu WS, Hung CL, Hu KC, Yu LY, et al. Neutrophil-to-lymphocyte ratio as a  
425 predictive marker of metabolic syndrome. *Medicine (Baltimore)*. 2019;98(43):e17537.

- 426 18. Ulu SM, Dogan M, Ahsen A, Altug A, Demir K, Acartürk G, et al. Neutrophil-to-Lymphocyte  
427 Ratio as a Quick and Reliable Predictive Marker to Diagnose the Severity of Diabetic  
428 Retinopathy. *Diabetes Technol Ther.* 2013;15(11):942–7.
- 429 19. Abdullah M. Comparative study on neutrophil-lymphocyte ratio (NLR) among diabetic cases  
430 with and without diabetic retinopathy. *IJCEO .* 2021;7(1):153–6.
- 431 20. Liu J, Liu X, Li Y, Quan J, Wei S, An S, et al. The association of neutrophil to lymphocyte  
432 ratio, mean platelet volume, and platelet distribution width with diabetic retinopathy and  
433 nephropathy: A meta-analysis. *Biosci Rep.* 2018;38(3):BSR20180172.
- 434 21. Chittawar S, Dutta D, Qureshi Z, Surana V, Khandare S, Dubey TN. Neutrophil-lymphocyte  
435 Ratio is a Novel Reliable Predictor of Nephropathy, Retinopathy, and Coronary Artery Disease  
436 in Indians with Type-2 Diabetes. *Indian J Endocrinol Metab.* 2017;21(6):864-70.
- 437 22. McKinstry B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, et al. Cohort  
438 profile: the Scottish Research register SHARE. A register of people interested in research  
439 participation linked to NHS data sets. *BMJ Open.* 2017;7(2):e013351.
- 440 23. Hébert HL, Shepherd B, Milburn K, Veluchamy A, Meng W, Carr F, et al. Cohort profile:  
441 Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). *Int J Epidemiol.*  
442 2018;47(2):380-381j.
- 443 24. Srinivasan S, Liju S, Sathish N, Siddiqui MK, Anjana RM, Pearson ER, et al. Common and  
444 distinct genetic architecture of age at diagnosis of diabetes in South Indian and European  
445 populations. *Diabetes Care.* 2023; 46(8):1515-1523.
- 446 25. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation  
447 PLINK: Rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4(1):s13742–  
448 015–0047–8.
- 449 26. Purcell S, Chang C. PLINK 1.9 [Internet]. [cited 2023 Apr 5]. Available from:  
450 <https://www.cog-genomics.org/plink/1.9/>

- 451 27. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, Von Elm E, et al. STrengthening  
452 the REporting of Genetic Association Studies (STREGA)--an extension of the STROBE  
453 statement. *Genet Epidemiol.* 2009 ;33(7):581–98.
- 454 28. Zachariah S, Wykes W, Yorston D. Grading diabetic retinopathy (DR) using the Scottish  
455 grading protocol. *Community Eye Heal.* 2015;28(92):72–3.
- 456 29. Siddiqui MK, Hall C, Cunningham SG, McCrimmon R, Morris A, Leese GP, et al. Using Data  
457 to Improve the Management of Diabetes: The Tayside Experience. *Diabetes Care.* 2022  
458 ;45(12):2828–37.
- 459 30. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and  
460 joint multiple-SNP analysis of GWAS summary statistics identifies additional variants  
461 influencing complex traits. *Nat Genet .* 2012;44(4):369-75.
- 462 31. Mägi R, Morris AP. GWAMA: Software for genome-wide association meta-analysis. *BMC*  
463 *Bioinformatics.* 2010;11(288).
- 464 32. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The  
465 GeneMANIA prediction server: biological network integration for gene prioritization and  
466 predicting gene function. *Nucleic Acids Res.* 2010;38((Web Server issue)):W214–20.
- 467 33. Speed D, Hemani G, Johnson MR, Balding DJ. Improved Heritability Estimation from  
468 Genome-wide SNPs. *Am J Hum Genet.* 2012;91(6):1011-1021.
- 469 34. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data.  
470 *Gigascience.* 2019;8(7): giz082.
- 471 35. Team RC. R: A Language and Environment for Statistical Computing [Internet]. Vienna,  
472 Austria: R Foundation for Statistical Computing; 2018. Available from: [http://www.r-](http://www.r-project.org)  
473 [project.org](http://www.r-project.org).
- 474 36. Gerds AT, Blanche P, Mortensen R, Tollenaar N, Mogensen UB, Ozenne B. Risk Regression  
475 Models and Prediction Scores for Survival Analysis with Competing Risks (riskRegression).

- 476 2020.
- 477 37. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of  
478 Competing Risks. *Circulation*. 2016;133(6):601–9.
- 479 38. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for  
480 competing risk data. *Stat Med* . 2017;36(27):4391–400.
- 481 39. Sjolander A, Martinussen T. Instrumental Variable Estimation with the R Package ivtools.  
482 *Epidemiol Methods*. 2019;8(1):20180024.
- 483 40. Ishizuka Y, Nakayama K, Ogawa A, Makishima S, Boonvisut S, Hirao A, et al. TRIB1  
484 downregulates hepatic lipogenesis and glycogenesis via multiple molecular interactions. *J Mol*  
485 *Endocrinol* . 2014 ;52(2):145–58.
- 486 41. Azab B, Camacho-Rivera M, Taioli E. Average values and racial differences of neutrophil  
487 lymphocyte ratio among a nationally representative sample of United States subjects. *PLoS*  
488 *One* . 2014;9(11):e112361.
- 489 42. Lin BD, Hottenga JJ, Abdellaoui A, Dolan C V, De Geus EJC, Kluff C, et al. Causes of  
490 variation in the neutrophil–lymphocyte and platelet–lymphocyte ratios: a twin-family study.  
491 *Biomark Med*. 2016;10(10):1061-1072.
- 492 43. Reich D, Nalls MA, Kao WHL, Akyzbekova EL, Tandon A, Patterson N, et al. Reduced  
493 neutrophil count in people of African descent is due to a regulatory variant in the Duffy  
494 antigen receptor for chemokines gene. *PLoS Genet* . 2009;5(1):e1000360.
- 495 44. Verlaan DJ, Berlivet S, Hunninghake GM, Madore AM, Larivière M, Moussette S, et al.  
496 Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with  
497 the risk of asthma and autoimmune disease. *Am J Hum Genet* . 2009 ;85(3):377-93.
- 498 45. Nalls MA, Couper DJ, Tanaka T, van Rooij FJA, Chen MH, Smith A V., et al. Multiple Loci  
499 Are Associated with White Blood Cell Phenotypes. *PLOS Genet* . 2011 ;7(6):e1002113.

- 500 46. Zheng JS, Arnett DK, Parnell LD, Lee YC, Ma Y, Smith CE, et al. Genetic variants at PSMD3  
501 interact with dietary fat and carbohydrate to modulate insulin resistance. *J Nutr* .  
502 2013;143(3):354-61.
- 503 47. Colaprico A, Olsen C, Bailey MH, Odom GJ, Terkelsen T, Silva TC, et al. Interpreting  
504 pathways to discover cancer driver genes with Moonlight. *Nat Commun*. 2020;11(1):69.
- 505 48. Szilveszter KP, Németh T, Mócsai A. Tyrosine Kinases in Autoimmune and Inflammatory  
506 Skin Diseases. *Front Immunol* . 2019 ;10:1862.
- 507 49. Zheng JS, Arnett DK, Parnell LD, Lee YC, Ma Y, Smith CE, et al. Genetic variants at PSMD3  
508 interact with dietary fat and carbohydrate to modulate insulin resistance. *J Nutr*.  
509 2013;143(3):354-61.
- 510 50. Keller MF, Reiner AP, Okada Y, van Rooij FJA, Johnson AD, Chen MH, et al. Trans-ethnic  
511 meta-analysis of white blood cell phenotypes. *Hum Mol Genet*. 2014 ;23(25):6944-60.
- 512 51. Veroneze R, Corbi SCT, Da Silva BR, De Rocha CS, Maurer-Morelli C V., Orrico SRP, et al.  
513 Using association rule mining to jointly detect clinical features and differentially expressed  
514 genes related to chronic inflammatory diseases. *PLoS One*. 2020 ;15(10):e0240269.
- 515 52. Kachuri L, Jeon S, DeWan AT, Metayer C, Ma X, Witte JS, et al. Genetic determinants of  
516 blood-cell traits influence susceptibility to childhood acute lymphoblastic leukemia. *Am J*  
517 *Hum Genet* . 2021;108(10):1823–35.
- 518 53. Park B, Choe E, Kang H, Shin E, Lee S, Won S. Genetic Polymorphisms Associated with the  
519 Neutrophil–Lymphocyte Ratio and Their Clinical Implications for Metabolic Risk Factors. *J*  
520 *Clin Med*. 2018;7(8):204.
- 521 54. Perrotta C, Buldorini M, Assi E, Cazzato D, De Palma C, Clementi E, et al. The Thyroid  
522 Hormone Triiodothyronine Controls Macrophage Maturation and Functions: Protective Role  
523 during Inflammation. *Am J Pathol*. 2014;184(1):230–47.
- 524 55. Lin BD, Carnero-Montoro E, Bell JT, Boomsma DI, De Geus EJ, Jansen R, et al. SNP



- 525 heritability and effects of genetic variants for neutrophil-to-lymphocyte and platelet-to-  
526 lymphocyte ratio. *J Hum Genet.* 2017;62(11):979-88.
- 527 56. Shen X, Rönnegård L. Issues with data transformation in genome-wide association studies for  
528 phenotypic variability. *F1000Res.* 2013;2:200.
- 529 57. Ulu SM, Dogan M, Ahsen A, Altug A, Demir K, Acartürk G, et al. Neutrophil-to-Lymphocyte  
530 Ratio as a Quick and Reliable Predictive Marker to Diagnose the Severity of Diabetic  
531 Retinopathy. *Diabetes Technol Ther.* 2013;15(11):942–7.
- 532 58. Binet F, Cagnone G, Crespo-Garcia S, Hata M, Neault M, Dejda A, et al. Neutrophil  
533 extracellular traps target senescent vasculature for tissue remodeling in retinopathy. *Science* .  
534 2020;369(6506):eaay5356.
- 535 59. He X, Qi S, Zhang X, Pan J. The relationship between the neutrophil-to-lymphocyte ratio and  
536 diabetic retinopathy in adults from the United States: results from the National Health and  
537 nutrition examination survey. *BMC ophthalmol.* 2022 ;22(1):346.
- 538 60. Burgess S, Thompson SG, Chd C, Collaboration G. Avoiding bias from weak instruments in  
539 Mendelian randomization studies. *Int J Epidemiol.* 2011 ;40(3):755–64.
- 540 61. Burgess S, Scott RA, Timpson NJ, Smith GD, Thompson SG. Using published data in  
541 Mendelian randomization: A blueprint for efficient identification of causal risk factors. *Eur J*  
542 *Epidemiol.* 2015;30(7):543–52.
- 543 62. Burgess S, Timpson NJ, Ebrahim S, Smith GD. Mendelian randomization: where are we now  
544 and where are we going? *Int J Epidemiol.* 2015;44(2):379–88.

545

546 **Acknowledgments:**

547 The authors would like to thank all participants in the GoDARTS, SHARE, and DMSDC cohorts. We  
548 would also like to thank the Health Informatics Centre (HIC), University of Dundee for providing  
549 access to the Scottish Data and staff of MDRF, Chennai.

550

551 **Author Contributions:**

552 C.N.A.P. and A.L.R were involved in the conceptualisation, analysis, interpretation and manuscript  
553 writing. A.T.N., M.C., and C.N. were involved in GWAS analysis and heritability estimation. SS  
554 made a substantial contribution toward ensuring data quality and manuscript revision. M.K.S, V.R.,  
555 R.M.A., W.M. and V.M. revised the manuscript critically for important intellectual content and  
556 provided guidance for data analysis. All authors provided approval for manuscript publication.  
557 C.N.A.P. is the guarantor of this work, had access to the database, and takes full responsibility for the  
558 integrity of the data, decision to submit and publish the manuscript.

559

560 **Competing interests:**

561 None to declare.

562

563 **Supplemental Material**

564 Supplementary File 1 - Figures

565 Supplementary File 2 – Tables

566 STREGA check list

567