

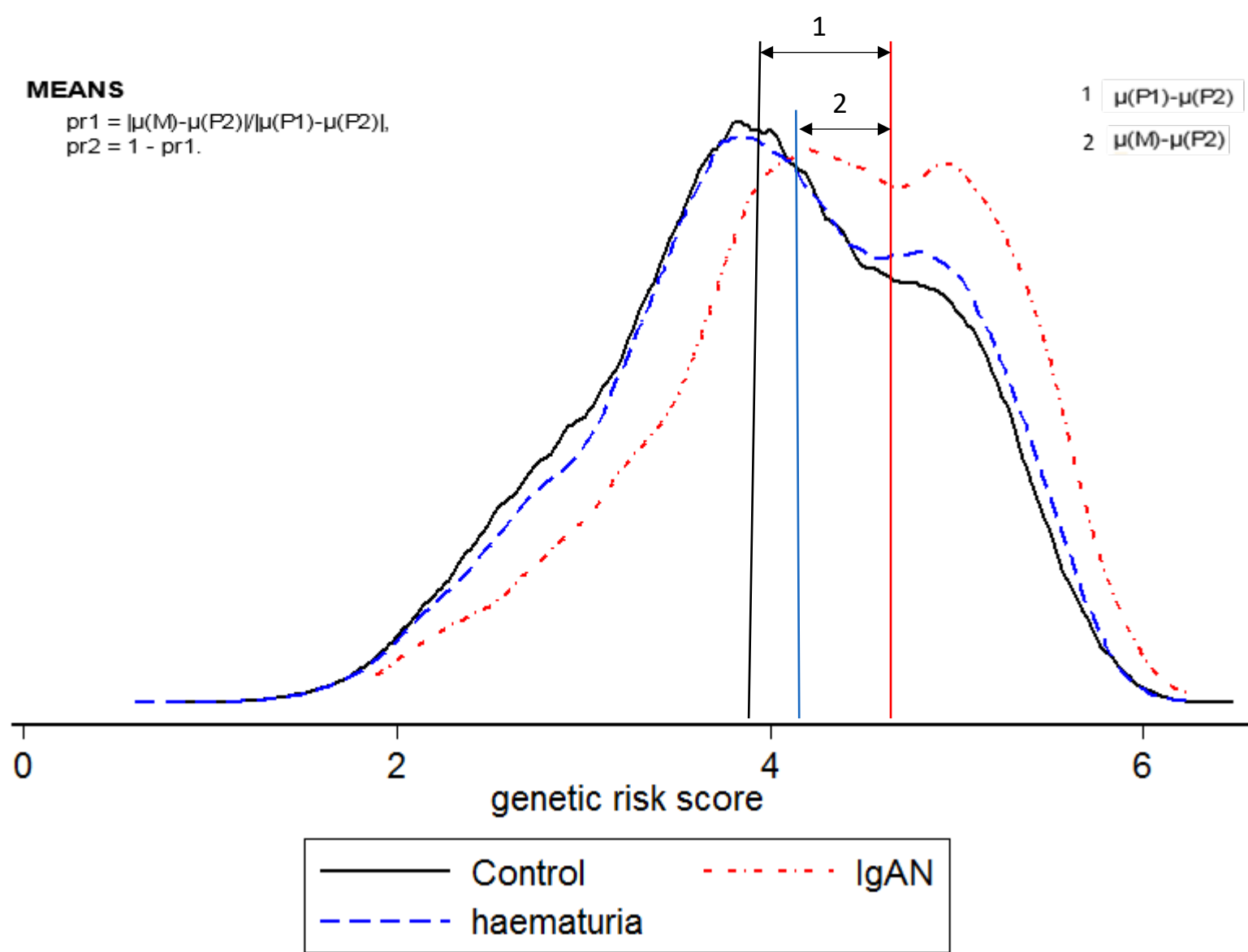
# SNP list

rs ID	Chr	Location	Risk allele	OR	Locus name	P value	Freq in European controls
rs17019602	1	107990381	G	1.17	VAV3	$6.8 \times 10^{-9}$	0.19
rs6677604	1	194953541	G	1.35	CFHR3-CFHR1 deletion	$4.8 \times 10^{-14}$	0.8
rs7763262	6	32532860	C	1.41	HLA-DR-HLA-DQ	$1.8 \times 10^{-38}$	0.69
rs9275224	6	32767856	G	1.36	HLA-DR-HLA-DQ	$5.9 \times 10^{-30}$	0.51
rs2856717	6	32778286	G	1.27	HLA-DR-HLA-DQ	$1.1 \times 10^{-15}$	0.62
rs9275596	6	32789609	T	1.44	HLA-DR-HLA-DQ	$2.5 \times 10^{-31}$	0.65
rs2071543	6	32919607	G	1.15	AP2-PSMB9	$1.5 \times 10^{-4}$	0.87
rs1883414	6	33194426	G	1.22	HLA-DP	$1.5 \times 10^{-11}$	0.68
rs2738048	8	6810195	T	1.10	DEFA	$1.6 \times 10^{-4}$	0.69
rs4077515	9	138386317	T	1.16	CARD9	$1.2 \times 10^{-9}$	0.4
rs11150612	16	31265261	A	1.18	ITGAM-ITGAX	$1.3 \times 10^{-11}$	0.36
rs11574637	16	31276375	T	1.32	ITGAM-ITGAX	$8.1 \times 10^{-13}$	0.82
rs3803800	17	7403693	A	1.12	TNFSF13	$9.3 \times 10^{-6}$	0.2
rs2412971	22	28824371	G	1.20	HORMAD2	$4.8 \times 10^{-12}$	0.54

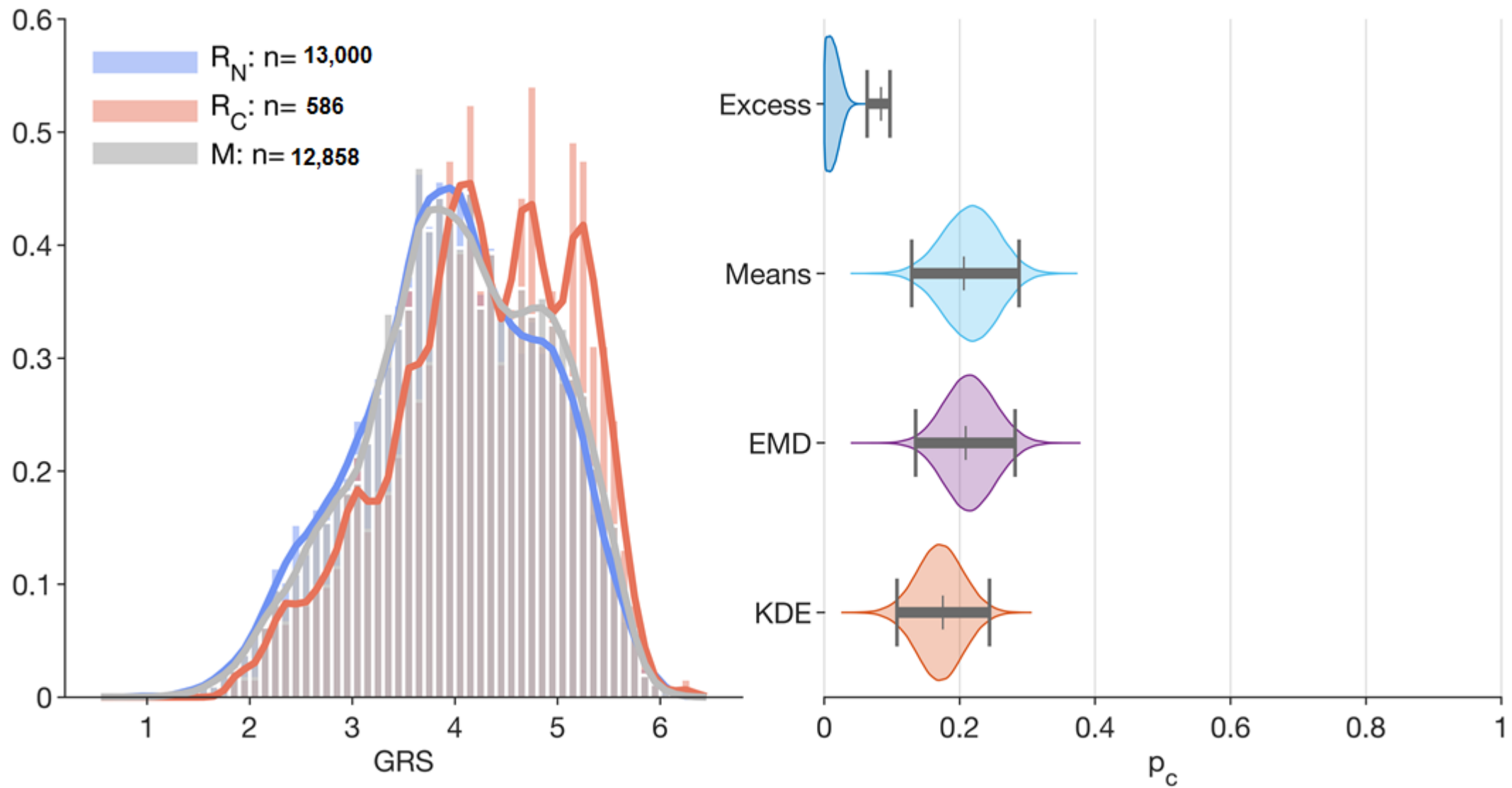
Supplementary Table S1. Table containing SNPs used to generate IgAN genetic risk score and their basic information.

<b>rsID</b>	<b>R2</b>
rs17019602	0.97
rs6677604	1
rs7763262	0.94
rs9275224	0.97
rs2856717	1.0
rs9275596	0.93
rs2071543	1.0
rs1883414	1.0
rs2738048	0.96
rs4077515	1.0
rs11150612	1.0
rs11574637	0.73
rs3803800	1.0
rs2412971	1
rs10086568	0.44

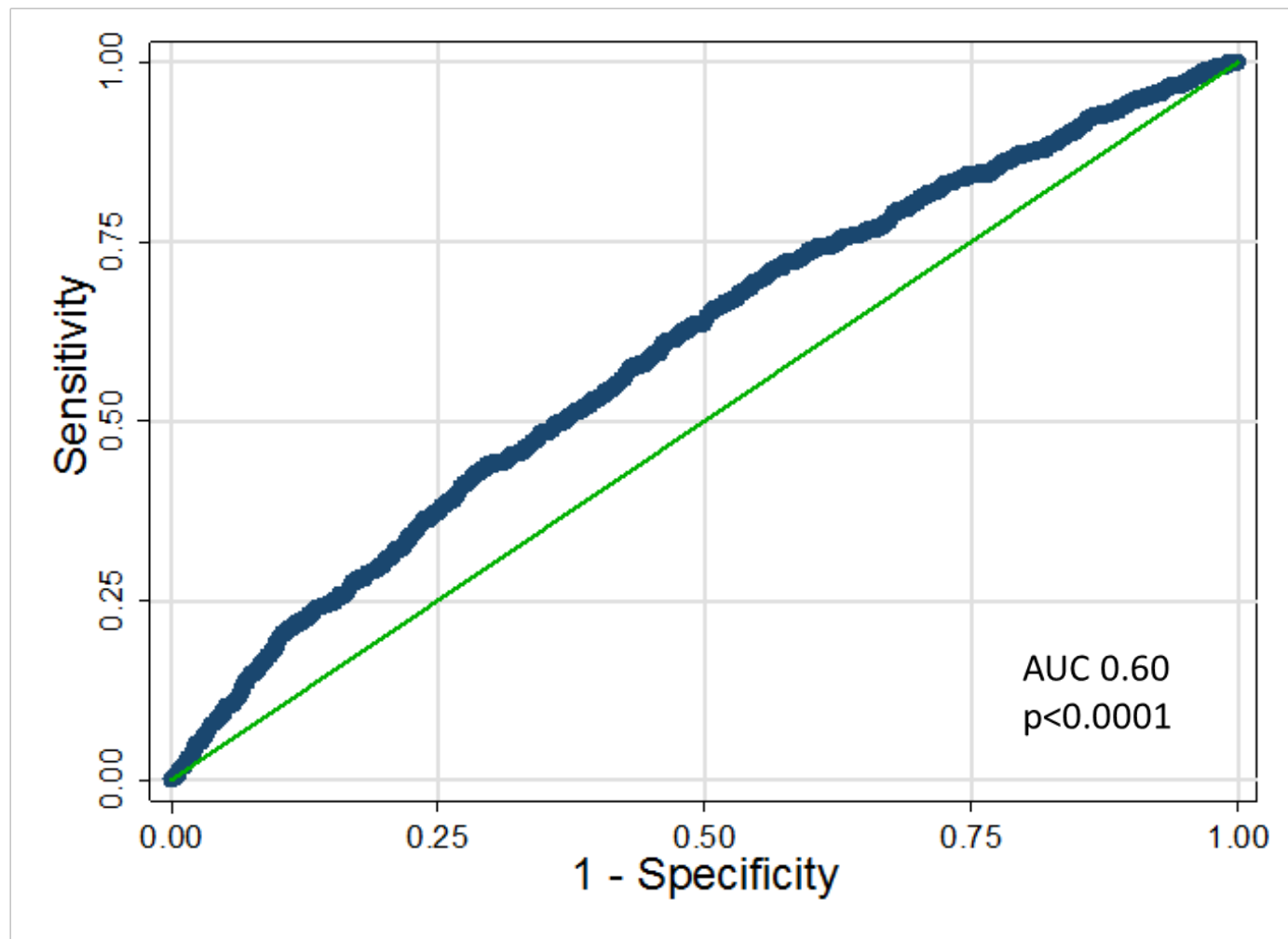
Supplementary Table S2. Table containing SNPs used to generate IgAN genetic risk score and their imputation  $r^2$ .



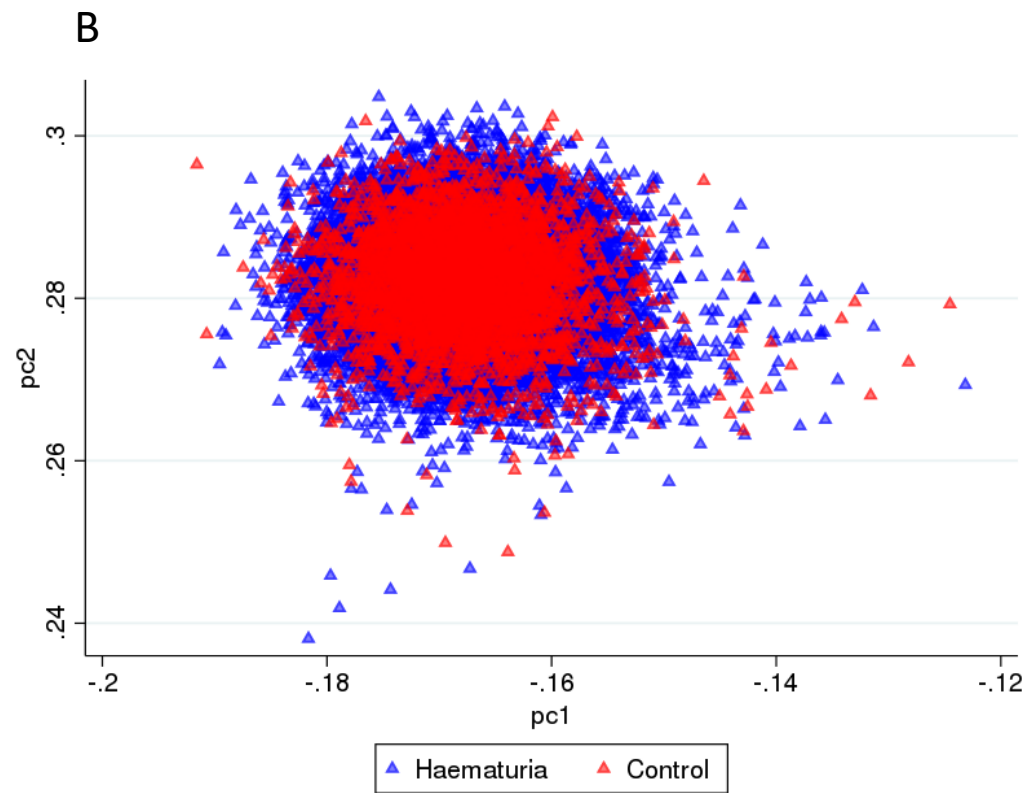
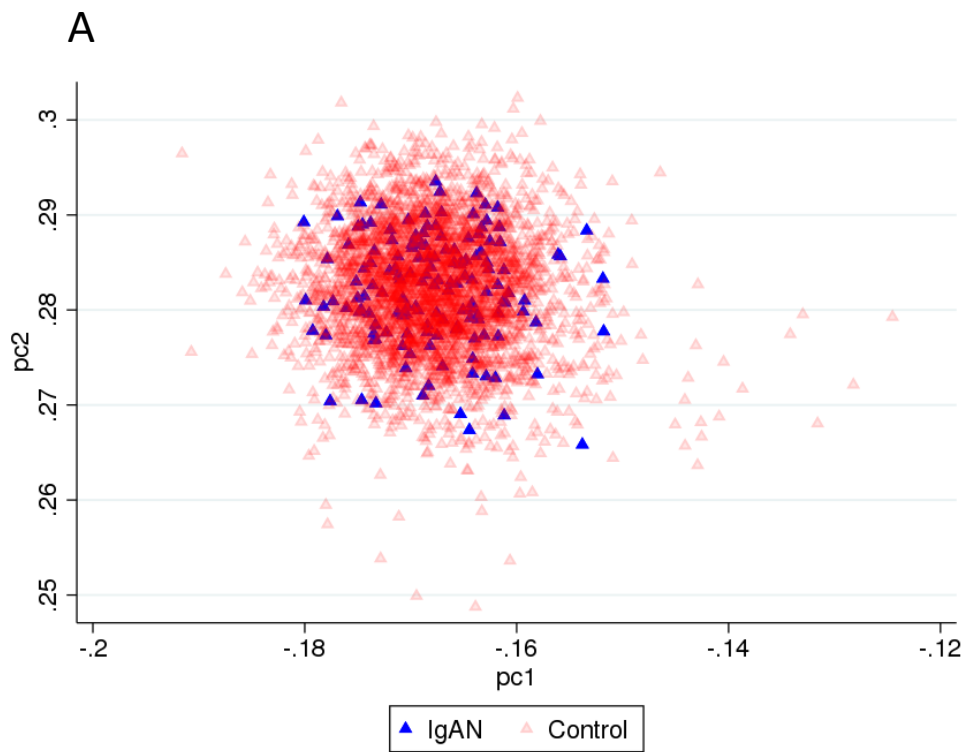
Supplementary Figure S1. Density plot demonstrating the distribution of controls, haematuria and IgAN cases. The arrows demonstrate the use of the means method to calculate the potential proportion if IgAN.



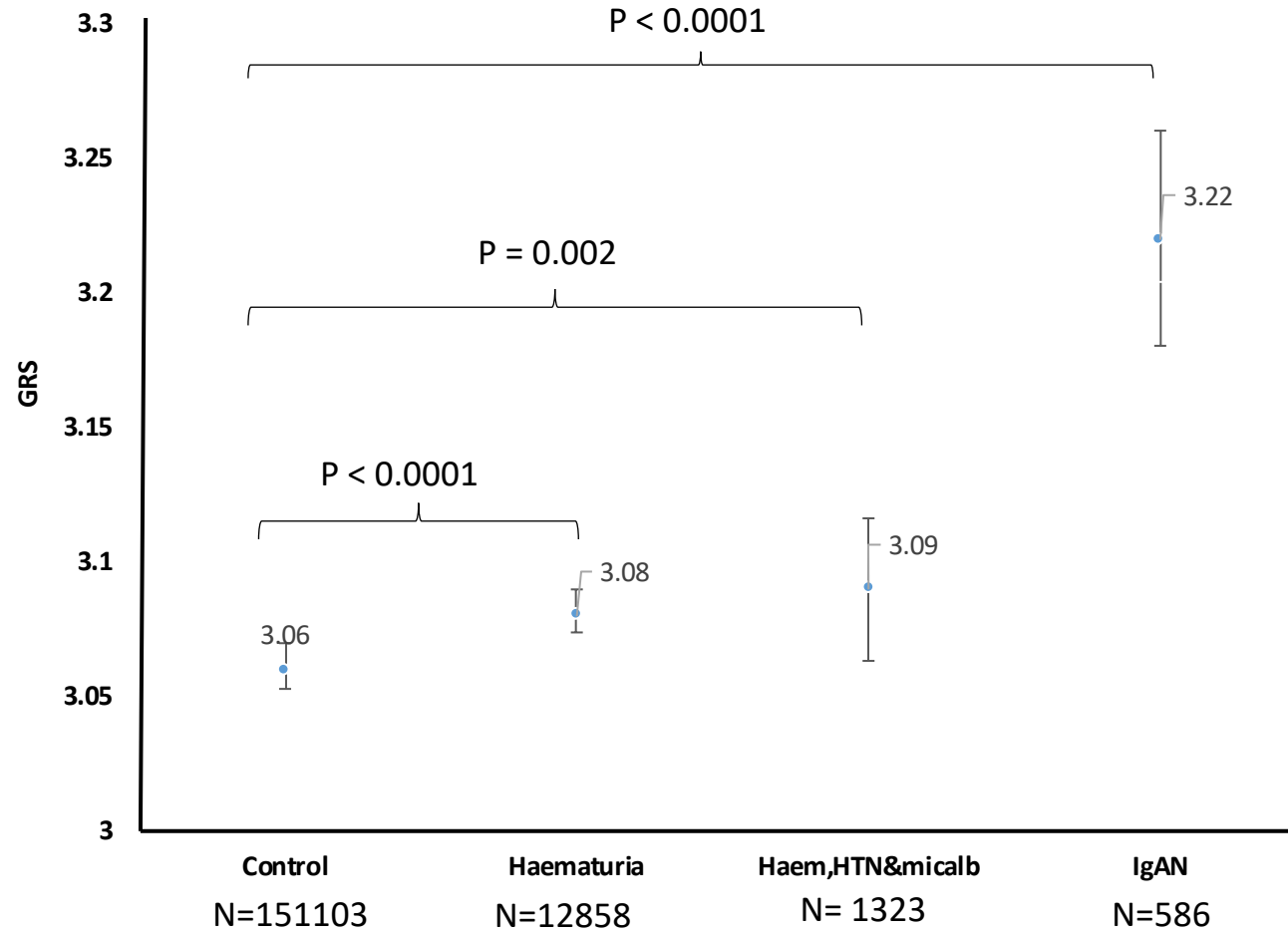
Supplementary Figure S2. Comparison of Earth Mover's Distance (EMD) between distributions; a linear combination of kernel density estimates (KDE) of distributions; a published Excess method and Means methods. (mixture population  $\tilde{M}$ ). The reference and the mixture distributions are plotted on the left ( $R_C$ , shaded red,  $R_N$ , shaded blue,  $\tilde{M}$ , shaded grey, respectively). Estimated values of prevalence  $p_c$  and 95% confidence intervals (grey dots and lines with bars at the ends) are plotted on the right. The violin plots show the distribution of the 100,000 estimates of prevalence ( $p'_c$ ) in the bootstrap samples. The proportion of individual with IgAN is shown as a dashed vertical line. Calculations were based on the following participants: non-cases UK Biobank, UKGDB IgAN, ICD10 IgAN in UK Biobank.



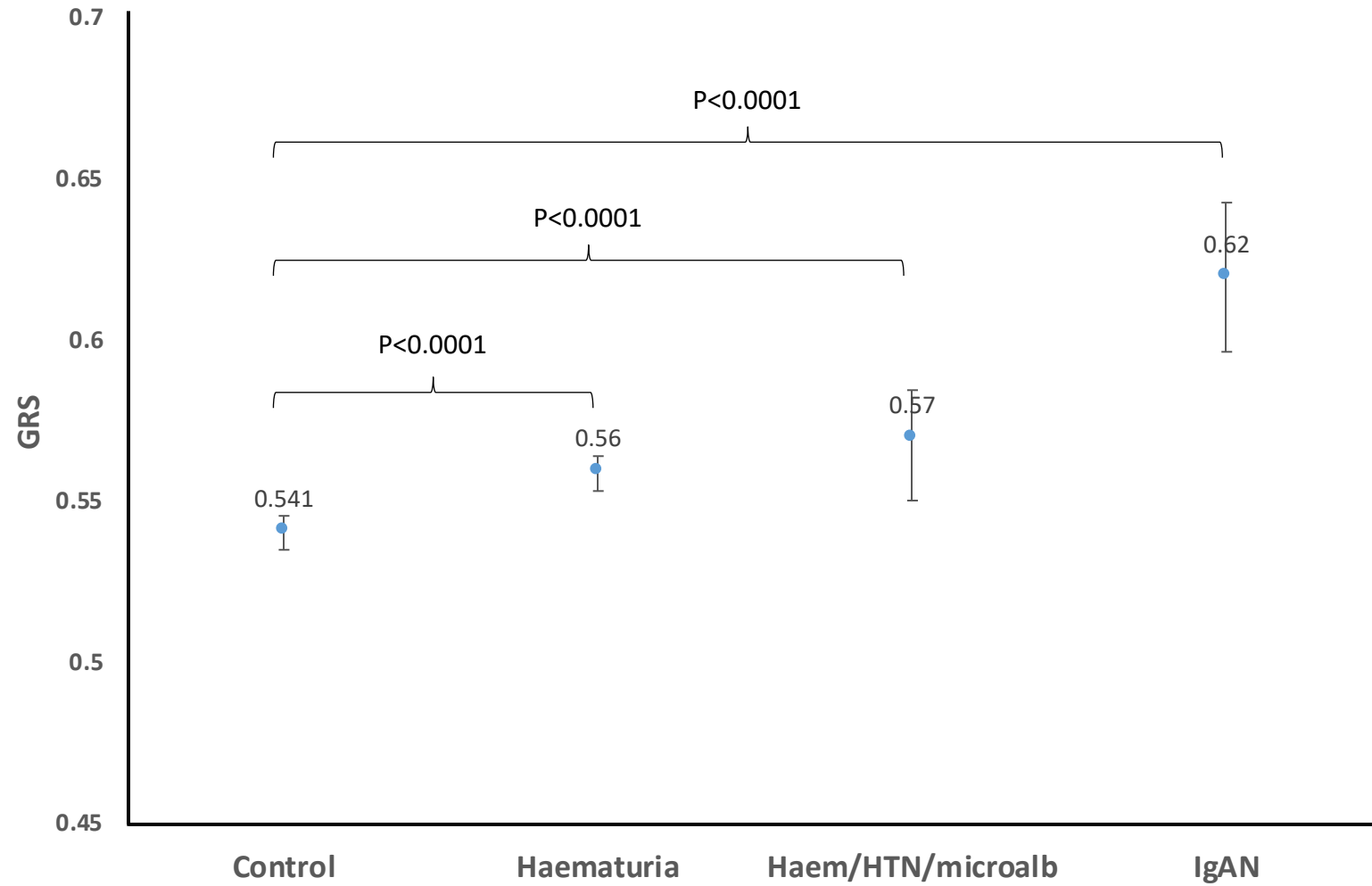
Supplementary figure S3. ROC curve assessing the of the IgAN-GRS to discriminate IgAN (from UKBB and UKGDB) from UKBB controls. ROC AUC was 0.6 (CI 0.57 to 0.62,  $p<0.0001$ ).



Supplementary Figures S4A and S4B. PCA plots (PC1 V PC2). S4A - PC1VPC2 between IgAN cases and controls. S4B – PC1VPC2 between UKBB Haematuria and controls. The graphs demonstrate that there were no differences by principal components.

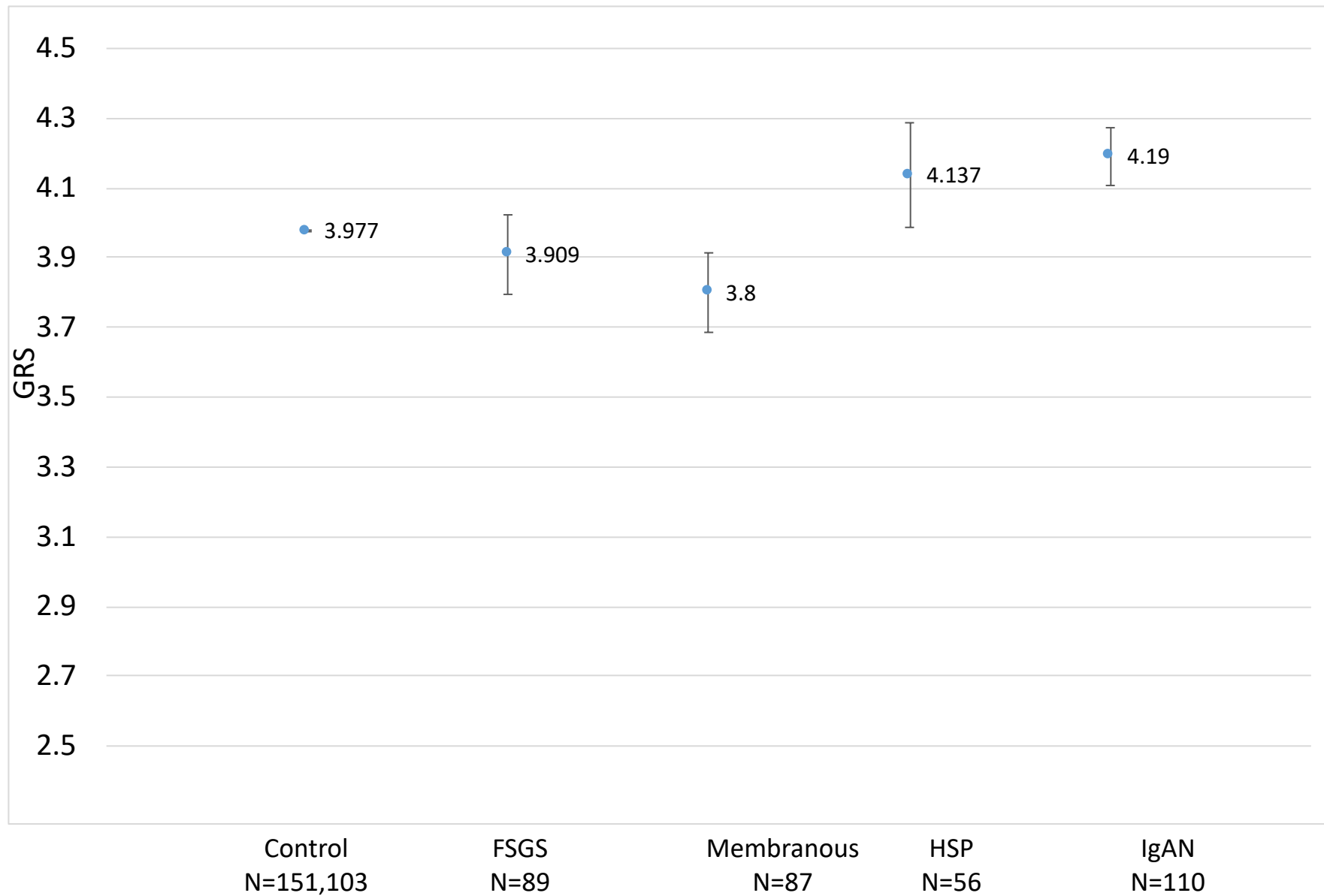


Supplementary Figure S5. 11 SNP IgAN-GRS. A regenerated 11 SNP score which included one of four HLA SNP (with 95% confidence intervals represented by error bars).



Supplementary Figure S6. 1 SNP IgAN-GRS. A regenerated 1 SNP score. (with 95% confidence intervals represented by error bars).





Supplementary Figure S7. Analysis of individuals with different glomerulonephritis as (ICD10 codes) in UKBB. (with 95% confidence intervals represented by error bars).