SUPPLEMENTARY APPENDIX

Identification of a novel genetic locus associated with immune-mediated thrombotic thrombocytopenic purpura

Matthew J. Stubbs,[1,2] Paul Coppo,[3] Chris Cheshire,[2] Agnès Veyradier,[4] Stephanie Dufek,[2] Adam P. Levine,[2] Mari Thomas,[1,5] Vaksha Patel,[2] John O. Connolly,[2] Michael Hubank,[6] Ygal Benhamou,[3] Lionel Galicier,[3] Pascale Poullin,[3] Robert Kleta,[2#] Daniel P. Gale,[2#] Horia Stanescu[2#] and Marie A. Scully[1,5#]

[1]Hemostasis Research Unit, ULC, London, UK; [2]Department of Renal Medicine, UCL, London, UK; [3]Centre de Référence des Microangiopathies Thrombotiques, Hôpital Saint-Antoine, Paris, France; [4]Department d'Hematologie, Centre de Référence des Microangiopathies Thrombotiques, Hôpital Lariboisière, Paris, France; [5]National Institute for Health Research Cardiometabolic Programme, UCLH/UCL Cardiovascular BRC, London, UK and [6]Clinical Genomics, Royal Marsden Hospital, London, UK.

[#]RK, DPG, HS and MAS contributed equally as co-senior authors.

 doi:10.3324/haematol.2020.274639

Received: October 21, 2020.

Accepted: February 2, 2021.

Pre-published: February 18, 2021.

Correspondence: *MATTHEW JAMES STUBBS* - m.stubbs@doctors.org.uk

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY METHODOLOGY

COHORTS

TTP cases were collected as described in the main paper between 2012 and 2017, with UK TTP cases (discovery cohort) and French TTP cases (replication cohort).

- UK TTP cases, n=413
- French TTP cases, n=200

The control cohorts include the 1958 British Birth Cohort and National Blood Service control samples, in addition to reference genotypes from the Illumina reference panel (HapMap Ethnicity controls) and Oxford controls.[1–5]

- Illumina Ethnicity Cohort, n=90
- Oxford Cohort, n=432
- British Birth Cohort, n=2867
- National Blood Service Cohort, n=2737

GENOTYPING

Samples were genotyped on the following SNP chips;

- UK TTP cases - HumanOmniExpress-12v1_H and 24v102_A1
- French TTP cases – HumanOmniExpress- 24v102_A1

Controls were genotyped on the following SNP chips;

- Illumina Reference - Illumina HumanOmniExpress-12v1_C
- Oxford controls - Illumina HumanOmniExpress-12v1_J
- British Birth Cohort - Human1-2M0DuoCustom_v1_A.
- National Blood Service Cohort - Human1-2M0DuoCustom_v1_A

Genotypes were re-encoded using in-house software to genomic forward for further analysis.[6]

QUALITY CONTROL

Quality control was performed using SNP & Variation Suite[7] PLINK version 1.90[8] and PRIMUS.[9]

Strict quality control per sample was performed, excluding individuals with call rate (CR) <0.90, duplicated samples/related individuals (sample identity by state (IBS) >0.1875), sample heterozygosity rate >3SD, in addition to excluding individuals not of European ancestry by principal component analysis (PCA) filtering.

Case Sample Quality control is summarised below;

- **UK TTP Cohort -** 241 UK TTP patients were included for subsequent analysis, from 413 samples genotyped.
- **French TTP Cohort** - 112 French TTP patients were included for subsequent analysis, from 200 samples genotyped.

Control Sample Quality control is summarised below;

- **Illumina Ethnicity Controls** - 58 individuals were included for subsequent analysis, from 90 samples genotyped.
- **Oxford Controls** - 381 individuals were included for subsequent analysis, from 432 samples genotyped.
- **British Birth Cohort** - 2761 individuals were included for subsequent analysis, from 2867 samples genotyped.
- **National Blood Service Cohort** - 2603 individuals were included for subsequent analysis, from 2737 samples genotyped.

Quality control was performed per SNP, and SNPs were excluded with that had a CR<0.99, an allele count (AC) >2, minor allele frequency (MAF) <0.05, and Hardy Weinberg Equilibrium (HWE) p<0.001, non-autosomal markers, in addition to ambiguous SNPs.

Case SNP Quality Control is summarised below;

- **UK TTP Cohort** - QC was performed on 675,533 SNPs, and post QC 521,046 SNPs remained.

- **French TTP Cohort** - QC was performed on 675,533 SNPs, and post QC 490,032 SNPs remained.

Control SNP Quality Control is summarised below;

- **Illumina Ethnicity Controls** - QC was performed on 711,320 SNPs, and post QC 531,093 SNPs remained.
- **Oxford Controls** - QC was performed on 712,878 SNPs, and post QC 567,947 SNPs remained.
- **British Birth Cohort** - QC was performed on 1,066,003 SNPs, and post QC 722,672 SNPs remained.
- **National Blood Service Cohort** - QC was performed on 1,066,003 SNPs, and post QC 736,251 SNPs remained.

The UK and French datasets were combined with separate control datasets, and the above per-SNP QC performed on the merged datasets

- **UK Discover Cohort** - The UK TTP cohort (n=241) was combined with control datasets (Illumina Ethnicity, Oxford and British Birth cohorts) (n=3200) for overlapping SNPs (n=337,088).
- **French Replication Cohort** - The French TTP cohort (n=112) was combined with control datasets (National Blood Service cohort) (n=2603) for overlapping SNPs (n=334,756).

IMPUTATION

Genotype data was imputed using Beagle version 5.0, utilising the 1000 Genome European CEU reference population (Supplemental Figure 1 for QC).[10] Cases and controls were imputed together using individuals and markers that had previously passed stringent QC. Following imputation filtering was performed using bcftools[11] (https://samtools.github.io/bcftools/bcftools.html), and markers with a Dosage R-squared (DR2) less than 0.80 were removed, and imputed genotype data was also re-filtered per SNP using SNP & Variation Suite,[7] details listed below:

- The UK TTP cohort and control data sets were imputed, with indels and SNPs with DR2<0.80 excluded. Post QC 3,649,349 remained for analysis, and

further QC SNP's were excluded, CR<0.99, AC>2 (n=0), MAF<0.05, HWE p<0.001.

- The French TTP cohort and control data was imputed, and indels and SNPs with DR2<0.80 were excluded. Post QC n=3,649,546 remained for analysis, and further QC SNP's were excluded, CR<0.99, AC>2, MAF<0.05 and HWE p<0.001.

GENOME WIDE ASSOCIATION TESTING

Genome wide association testing was performed using SNP & Variation Suite, using logistic regression with correction of 10 principal components.[12–14] The logistic regression p-values, odds ratios were calculated in addition to Lambda inflation factors. A standardised genome wide significance level of $5x10^{-8}$ was applied.[15] Meta-analysis was performed by combining the independent cohorts and subsequently undertaking analysis by logistic regression with 10 principal component correction.

CONDITIONAL ANALYSIS

To investigate for independent signal conditional analysis was undertaken using a full versus reduced regression model in SVS. Lead SNPs were used as conditional inputs to determine independence, with results plotted using Locus Zoom software.[16]

HLA IMPUTATION

HLA imputation was performed utilising SNP2HLA to impute HLA types using previously genotyped markers.[17] Imputed HLA types were excluded if the DR2 (confidence) was <0.80. Conditional analysis was subsequently performed, using the previously described method. To validate our HLA imputation, we compared imputed HLA types in a subset of serologically HLA typed individuals (n=17), and found a concordance of >80%.

EXPRESSION QUANTITATIVE TRAIT LOCUS

Expression quantitative trait locus analysis was performed subsequently to associate identified SNPs with differential gene expression.[18] Reduced POGLUT1 expression associated with our haploblock was the most significant, and frequently reported

association, and the only gene with expression reduced across different platforms.[18,19]

LD-LINK

Additional markers in linkage disequilibrium with our lead SNP were identified by LD-link (https://ldlink.nci.nih.gov).[20] LD between variants rs71767581 and rs9884090 was supported using the NIHR BioResource-Rare Diseases dataset, comprising whole genome sequences from 6,588 European individuals.[21]

FUNCTIONAL ANNOTATION

Functional annotation of the haploblock was performed using the UCSC genome browser (https://genome.ucsc.edu)[22,23], to identify functional important variants. Functional annotations, Chip-Seq data and expression data to identify functionally important variants such as missense variants or regulatory variants.

FACTOR BOOK

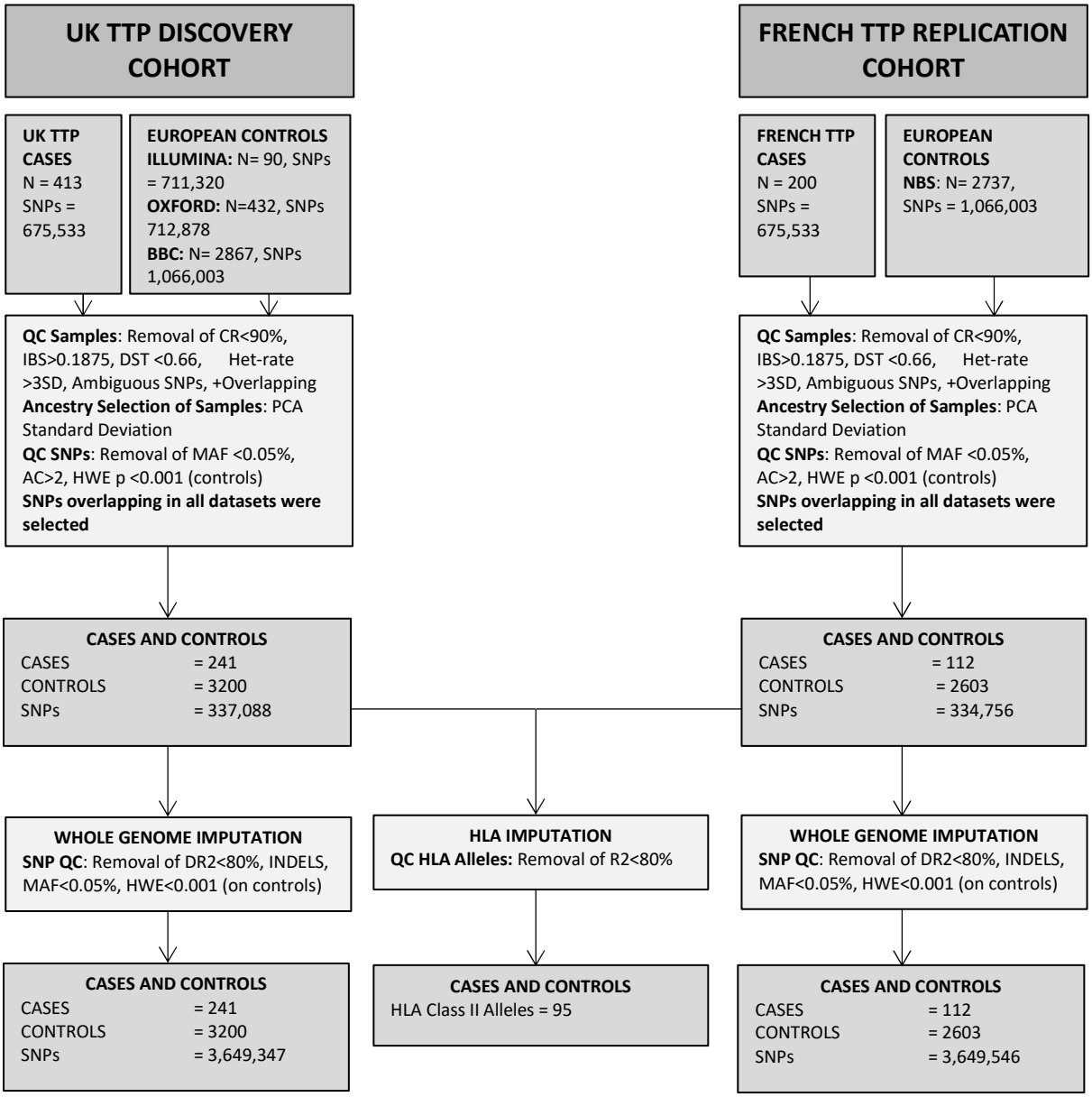
Binding sites of transcription factors that were identified through functional annotation in the region of interest, with potential functional importance were obtained from FactorBook.[24] Searching for specific cells lines (HEPG2) the position weight matrix (PWM) binding motifs of transcription factors of interest were identified, to be analysed alongside genetic variants derived from UCSC genome browser.

MAST/MEME

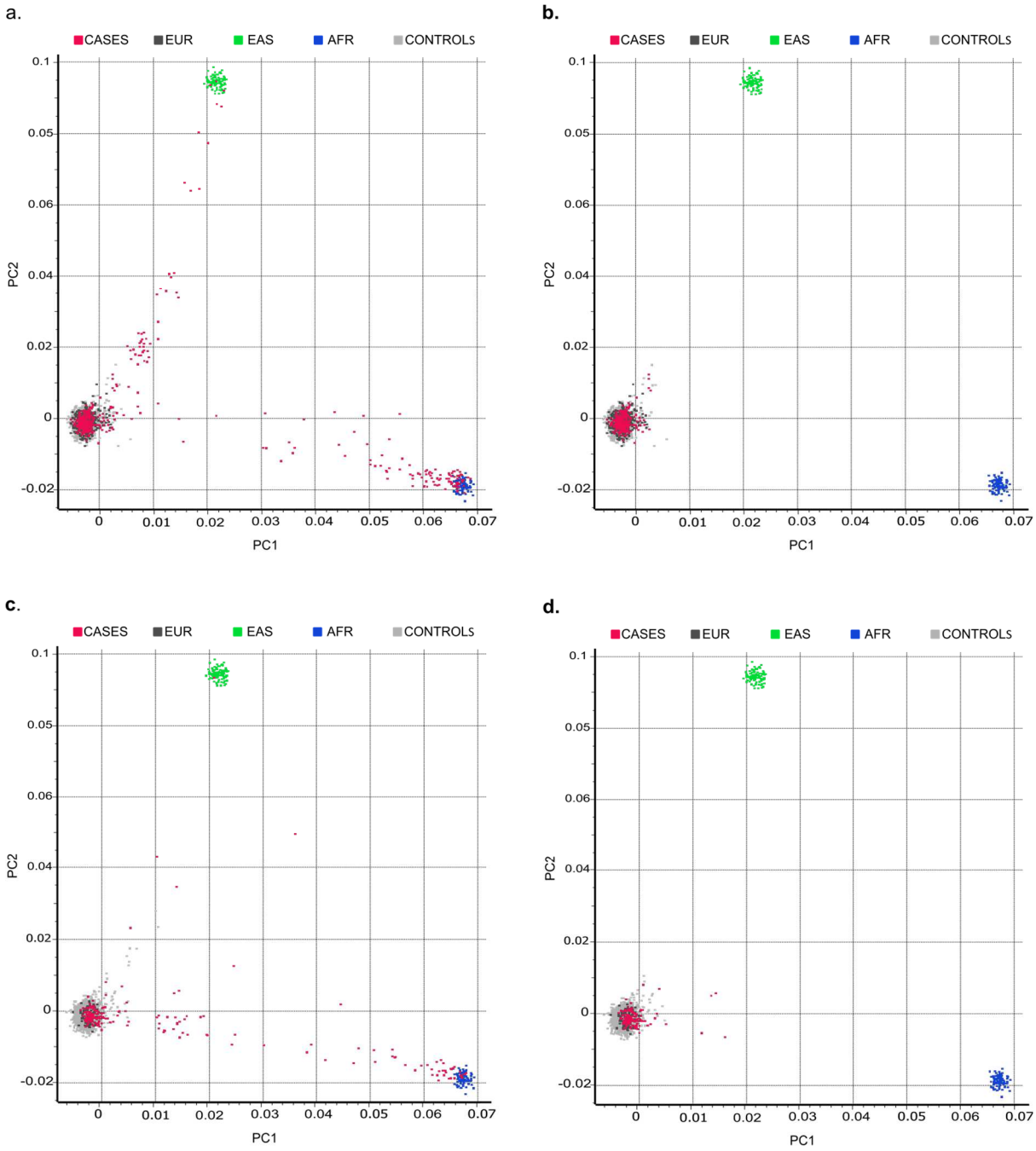
PWM binding motifs that were obtained from factor book were analysed along-side haploblock genetic variants, obtained from UCSC. A 80bp DNA sequence (40bp flanking, listed below) were analysed for potential DNA-transcription factor binding).[25]

- Wide Type Sequence: TATGCTAATTGCAACCTTTGGGGTCTAACCGTGC TGTACACACACACACACACACACACACACACACACACACACATGCTCA
- Variant Sequence: TATGCTAATTGCAACCTTTGGGGTCTAACCGTGCTGT -/- ACACACACACACACACACACACACACACACACACACACATGCTCA

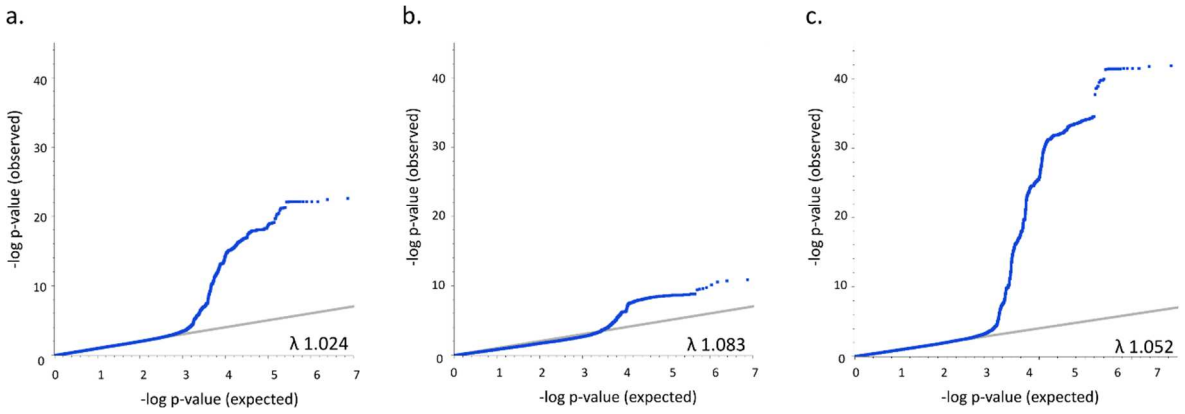
SUPPLEMENTARY FIGURES / RESULTS



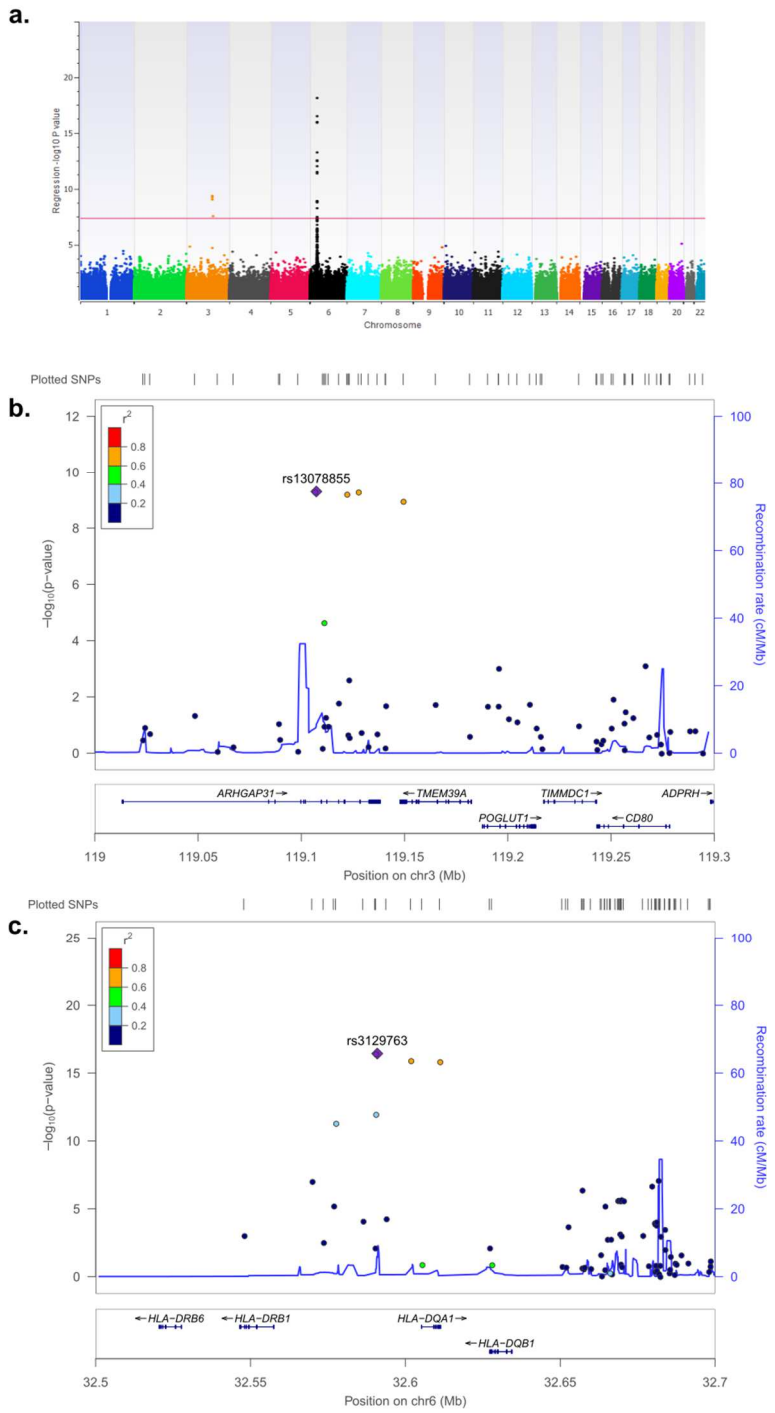
Supplemental Figure 1 – Summary of Quality Control in UK Discovery Cohort and French Replication Cohort.



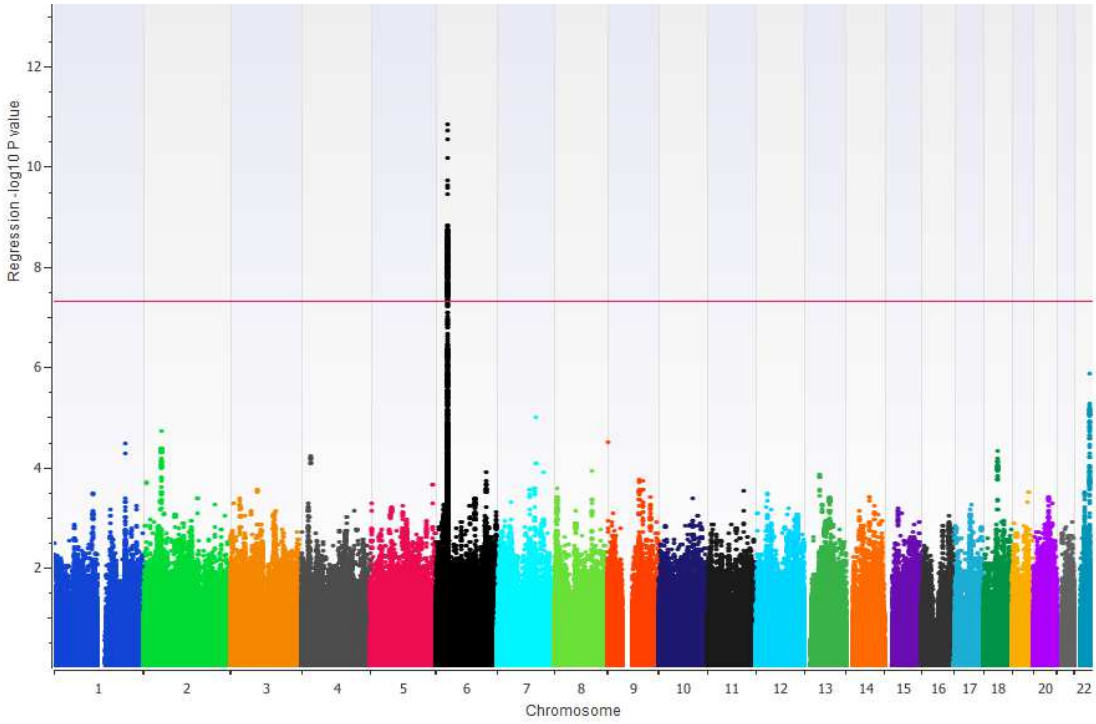
Supplemental Figure 2 – Principal Component Analysis in UK and French Cohorts. Cases are shown in red, and control genotypes in grey, and controls with known genetic ancestry are shown in black (EUR, European), blue (AFR, African) and green (EAS, East Asian). a. UK discovery cohort (without ethnicity ancestry filtering) and b. UK discovery cohort following ethnicity ancestry filtering applying 8.0 standard deviations to the principal component data to select cases with European ancestry. c. French replication cohort (without ethnicity ancestry filtering) and d. French replication cohort following ethnicity ancestry filtering applying 8.0 standard deviations to the principal component data to select cases with European ancestry.



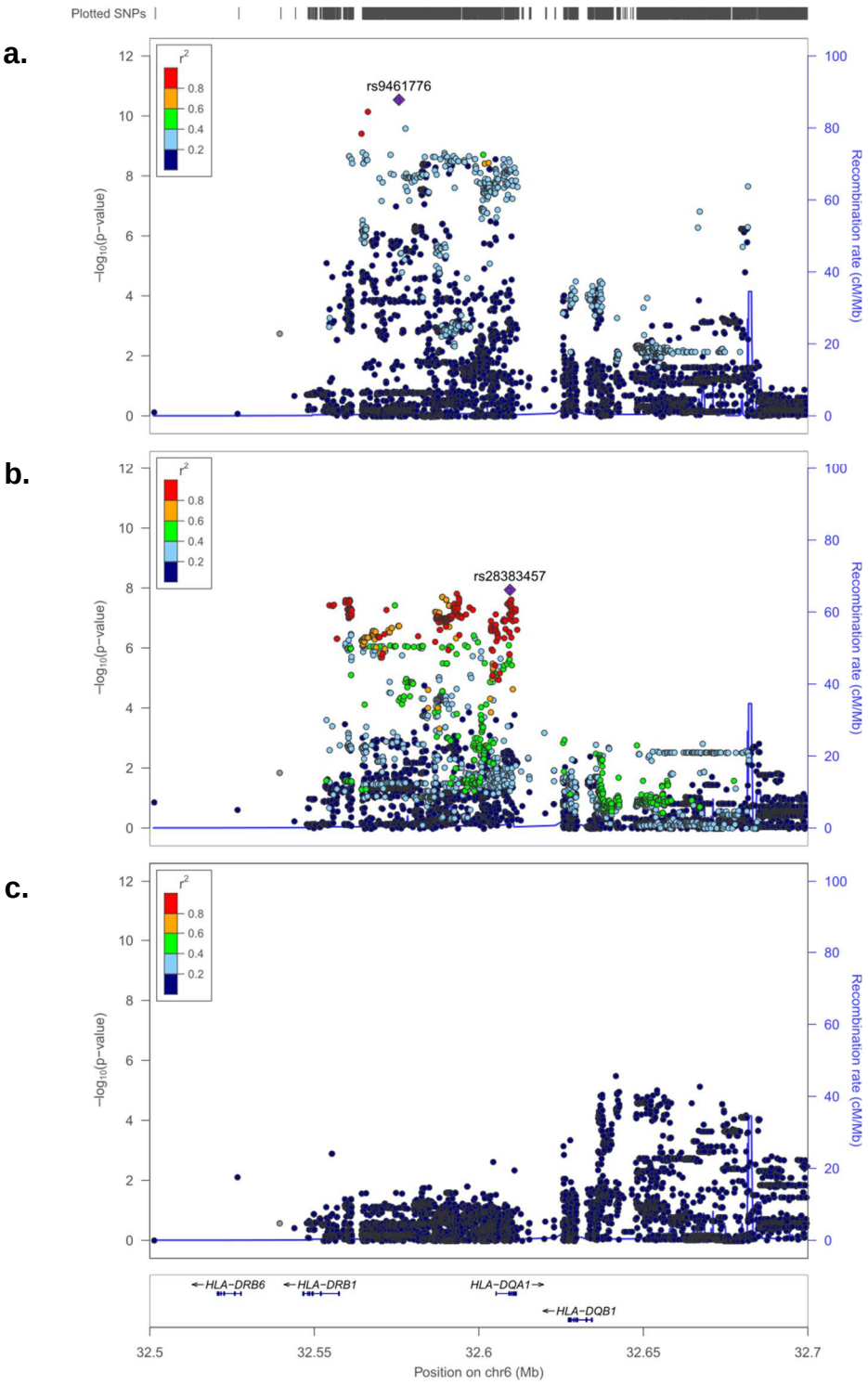
Supplemental Figure 3 – QQ plots, observed against expected p-values, for a. UK discovery population, b. French replication cohort, and c. Combined Analysis.



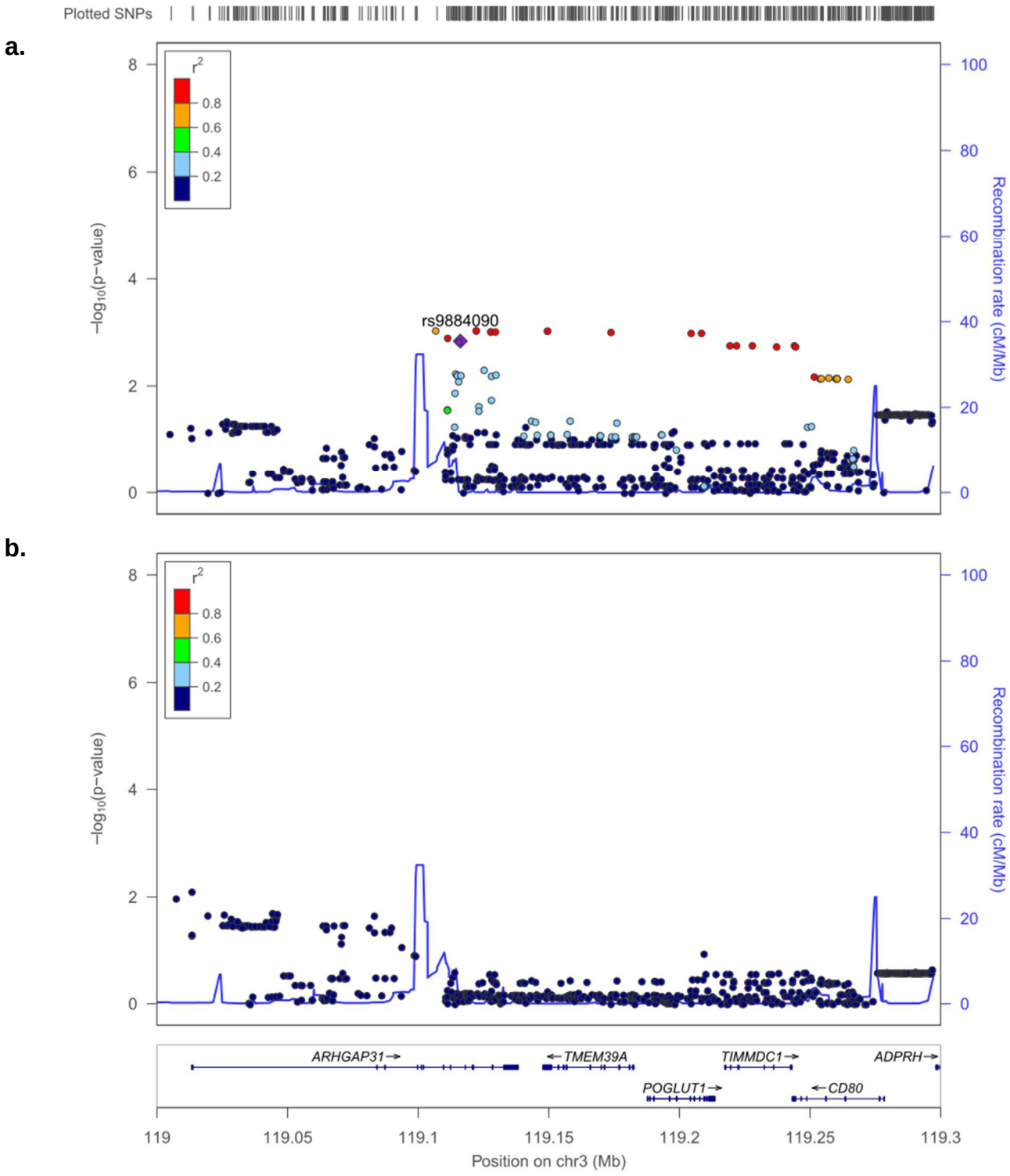
Supplemental Figure 4 – UK Cohort Directly Genotyped GWAS - (a) Manhattan plot of genome wide association test for directly genotyped SNPs only, comparing UK cases against controls, utilising the logistic regression method, corrected for top 10 principal components for stratification. The X axis shows chromosome location, and the Y axis shows logarithmic p-values. Standardised genome wide significant $5x10^{-8}$ is depicted by the red line. Locus zoom plot for the UK discovery cohort (visualising only directly genotyped SNPs) are shown for (b) chromosome 3 and (c) and chromosome 6 peak. The X axis shows chromosome location, and the left Y axis shows logarithmic p-values (logistic regression), and the right Y axis shown the recombination rate (shown as the blue line) (Human Assembly GRCh37/hg19).



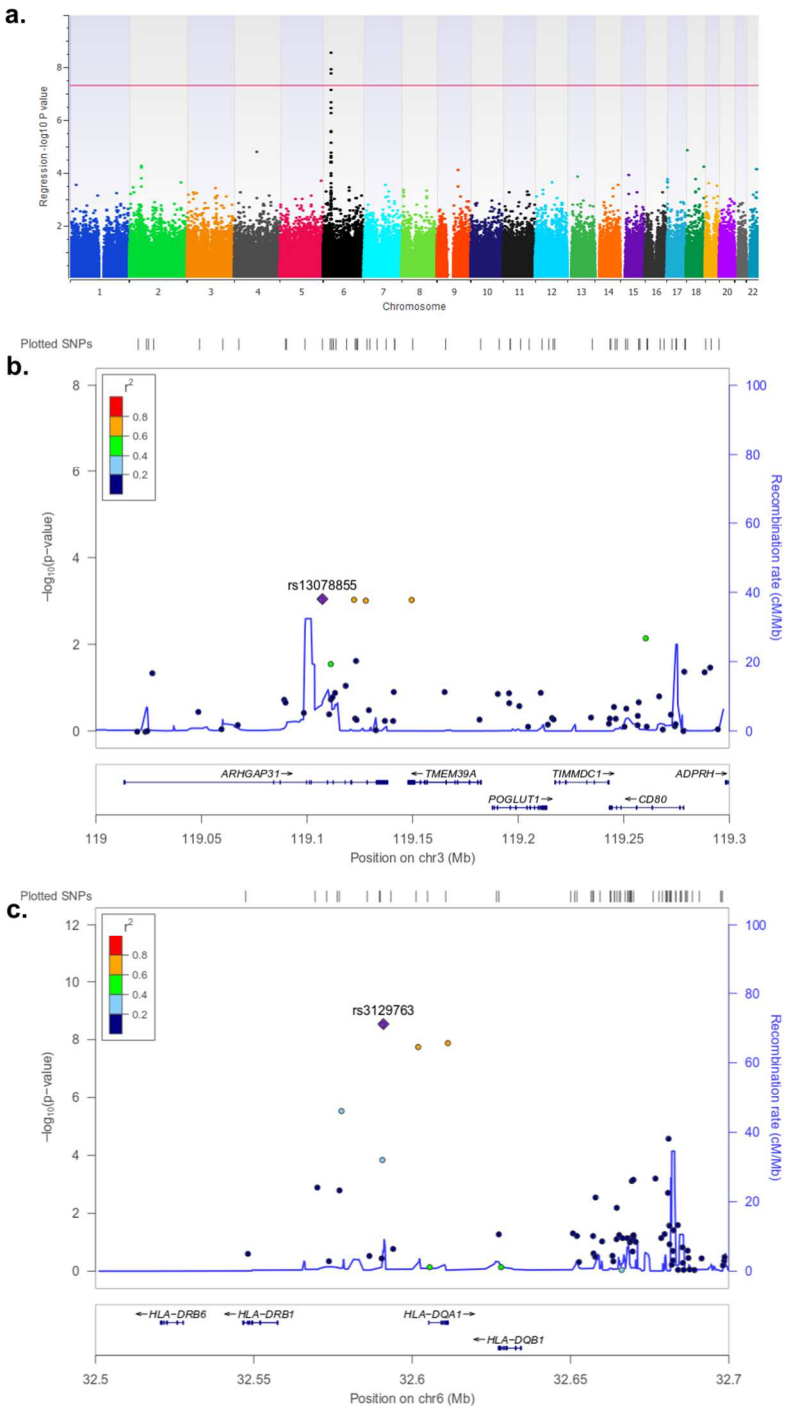
Supplemental Figure 5 – French Cohort Imputed GWAS – Manhattan plot of genome wide association test comparing French Replication cohort compared against controls, utilising the logistic regression method, corrected for top 10 principal components for stratification, in all imputed SNPs. The X axis shows chromosome location, and the Y axis shows logarithmic p-values. Standardised genome wide significant $5x10^{-8}$ is depicted by the red line. The HLA peak is visualised on chromosome 6 (black).



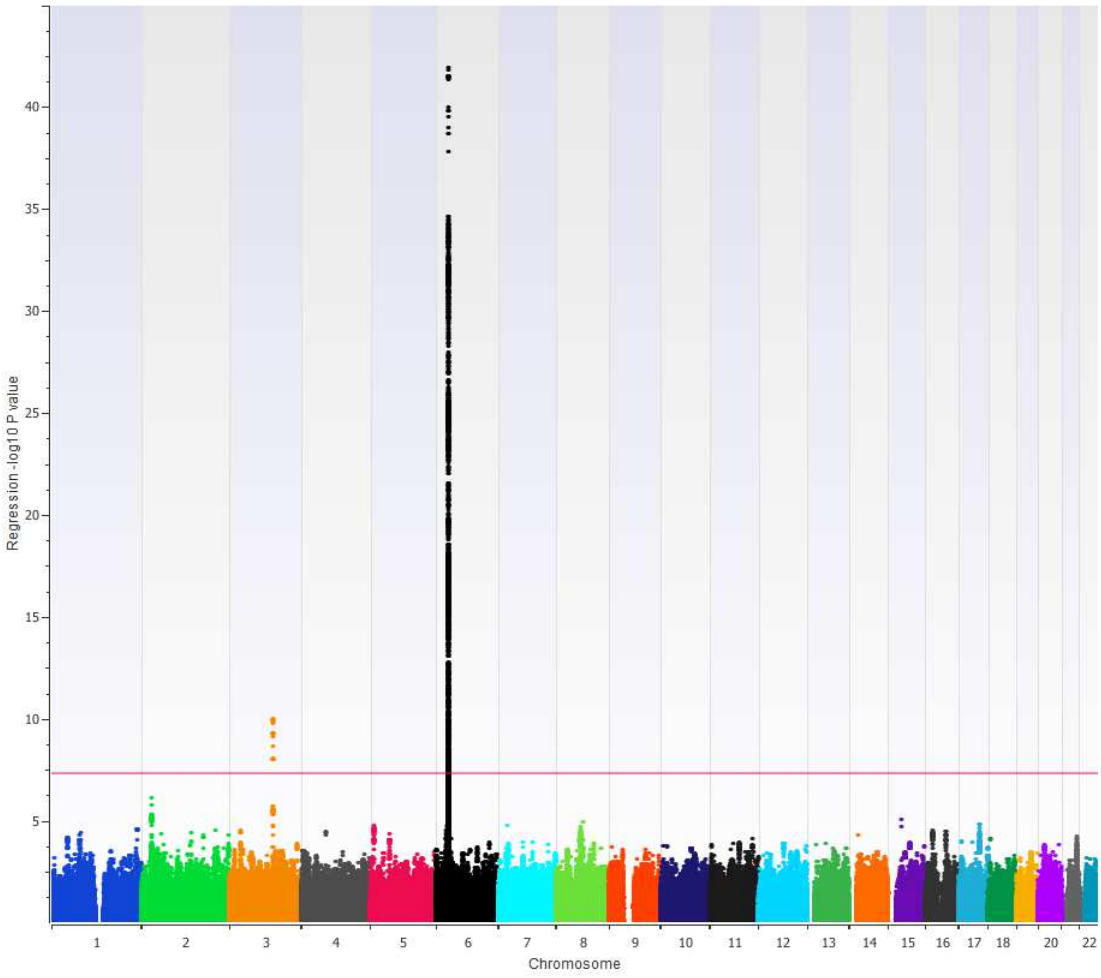
Supplemental Figure 6 – Locus zoom plot of the chromosome 6 peak in the French replication cohort. Genomic area displayed is 32.5 Mb to 32.7 Mb on chromosome 3 (Human Assembly GRCh37/hg19). The X axis shows chromosome location, and the left Y axis shows logarithmic p-values (logistic regression), and the right Y axis shown the recombination rate (shown as the blue line). a. shows the unconditional analysis with the lead SNP rs9461776, b. shows conditional analysis conditioned on rs9461776, revealing an independent association signal, lead SNP rs28383457, and c. shows analysis conditioned on conditioned on rs9461776 and rs28383457.



Supplemental Figure 7 – Locus zoom plot of the chromosome 3 peak in the French Replication cohort. Genomic area displayed is 119.0 Mb to 119.3 Mb on chromosome 3 (Human Assembly GRCh37/hg19). The X axis shows chromosome location, and the left Y axis shows logarithmic p-values (logistic regression), and the right Y axis shown the recombination rate (shown as the blue line). a. shows the unconditional analysis with the lead SNP rs9884090. b shows the same region containing the same markers, following conditioning on rs9884090 (identified from the UK discovery cohort analysis).



Supplemental Figure 8 – French Cohort Directly Genotyped GWAS - (a) Manhattan plot of genome wide association test for directly genotyped SNPs only, comparing French cases against controls, utilising the logistic regression method, corrected for top 10 principal components for stratification. The X axis shows chromosome location, and the Y axis shows logarithmic p-values. Standardised genome wide significant $5x10^{-8}$ is depicted by the red line. Locus zoom plot for the French replication cohort (visualising only directly genotyped SNPs) are shown for (b) chromosome 3 and (c) and chromosome 6 peak. The X axis shows chromosome location, and the left Y axis shows logarithmic p-values (logistic regression), and the right Y axis shown the recombination rate (shown as the blue line) (Human Assembly GRCh37/hg19).



Supplemental Figure 9 – GWAS Meta-Analysis of UK and French Cohorts - Manhattan Plot of Genome wide association tests for the meta-analysis of UK and French combined cohorts, utilising the logistic regression method, corrected for top 10 principal components for stratification. The X axis shows chromosome location, and the Y axis shows logarithmic p-values. Standardised Genome wide significant $5x10^{-8}$ is displayed in as the red line. The HLA peak is visualised on chromosome 6 (black, grey), in addition to the novel chromosome 3 association (orange, grey).

RS_ID	Location (GRCh37/hg19)	Alleles	MAF	Distance	D'	R2	Functional Annotation
rs9884090	chr3:119116150	(G/A)	0.1364	0	1	1	ARHGAP31 INTRON
rs9834901	chr3:119111870	(T/C)	0.1364	-4280	1	1	ARHGAP31 INTRON
rs12494314	chr3:119122820	(T/C)	0.1364	6670	1	1	ARHGAP31 INTRON
rs2305249	chr3:119128398	(G/A)	0.1364	12248	1	1	ARHGAP31 EXON (SYNON)
rs9855065	chr3:119130141	(G/A)	0.1364	13991	1	1	ARHGAP31 INTRON
rs3732421	chr3:119150089	(A/G)	0.1364	33939	1	1	TMEM39A 3'UTR
rs7650774	chr3:119205050	(T/C)	0.1364	88900	1	1	POGLUT1 INTRON
rs12695386	chr3:119209027	(T/C)	0.1364	92877	1	1	POGLUT1 CTCF
rs12636784	chr3:119174383	(A/G)	0.1313	58233	1	0.95	TMEM39A INTRON
rs2293370	chr3:119219934	(G/A)	0.1515	103784	1	0.88	TIMMDC1 TFBS
rs1131265	chr3:119222456	(G/C)	0.1515	106306	1	0.88	TIMMDC1 EXON (SYNONYMOUS)
rs9843355	chr3:119228508	(G/A)	0.1515	112358	1	0.88	TIMMDC1 INTRON
rs144104218	chr3:119237726	(AAC/-)	0.1515	121576	1	0.88	TIMMDC1 INTRON
rs62264485	chr3:119237798	(C/A)	0.1515	121648	1	0.88	TIMMDC1 INTRON
rs35264490	chr3:119238753	(A/-)	0.1515	122603	1	0.88	TIMMDC1 INTRON
rs57271503	chr3:119244593	(G/A)	0.1515	128443	1	0.88	CD80 EXON ENHANCER
rs13092998	chr3:119245044	(G/T)	0.1515	128894	1	0.88	CD80 INTRON
rs3830649	chr3:119246385	(G/-)	0.1515	130235	1	0.88	CD80 INTRON
rs71767581	***chr3:119187433***	***(AC/-)***	***0.1364***	***71283***	***0.91***	***0.84***	***POGLUT1 TFBS***
rs1132200	chr3:119150836	(C/T)	0.1162	34686	1	0.83	TMEM39A EXON (MISSENSE)

Supplemental Table 1 – Additional SNP's identified from LD-Link, found to be in Linkage disequilibrium with rs9884090 (lead chromosome 3 haploblock 3, identified through GWAS). SNP's with R^2 and D' >0.80 are shown. Functional annotations (derived from UCSC) are also included.

Transcription Factors
TCF12
GATA1
JUND
CHD1
MYBL2
TEAD4
STAT5A
POLR2A
NR3C1
RELA
REST
YY1
E2F6
PHF8

Supplemental Table 2 – Transcription factors, identified from UCSC that have ChipSeq tracts overlaying the proposed functional variant rs71767581.

Supplemental Results

Several other SNPs within the haploblock containing rs9884090 have also been implicated with other autoimmune disease. Previously published SNPs were analysed for linkage with rs9884090 (D' and R^2) using LD-link (with 1000G European CEU reference panel)[20] and also searched for any evidence of eQTL using GTEX, particularly indicating any evidence of altered POGLUT1 expression. The LD (D' and R2 shown) and also eQTL data is shown below for different autoimmune disease. Notably the eQTL data was not included in the for the majority of the initial studies.

SNPs associated with Multiple Sclerosis

- rs1132200, D' 1.0 R^2 0.83, Reduced POGLUT1 expression on eQTL analysis.[26,27]

SNPs associated with Systemic Lupus Erythematosus

- rs1132200; D' 1.0 R^2 0.83, Reduced POGLUT1 expression on eQTL analysis.[28]
- rs12494314, D' 1.0, R^2 1.0, Reduced POGLUT1 expression on eQTL analysis (in addition to TIMMDC1).[29]
- rs12493175, D' 1.0, R^2 0.04, Reduced POGLUT1 on eQTL analysis.(30)
- rs13062955,D' 1.0, R^2 0.04, No eQTL data available.[30]

SNPs associated with Autoimmune Thyroid Disease

- rs12492609, D' 1.0, R^2 0.036, No eQTL data available.[31]
- rs7629750, D' 1.0, R^2 0.27, No eQTL data available.[31]

SNPs associated with Primary Biliary Cholangitis

- rs2293370, D' 1.0, R^2 0.88, Reduced POGLUT1 expression on eQTL analysis, which was reported in the published paper.[32]

The above results demonstrate that the majority of SNPs associated with different autoimmune disease within this haploblock have evidence of strong linkage with rs9884090, and where available eQTL demonstrates altered POGLUT1 expression.

SUPPLEMENTAL WEBLINKS

- WTCCC[1]
  WTCCC Available from the European Genome Archive,
  http://www.wtccc.org.uk

- Golden Helix, SNP and Variation Suite (SVS)[7]
  Details of SVS available from http://www.goldenhelix.com (Bozeman)

- PLINK[8]
  Software available from http://www.cog-genomics.org/plink/1.9

- PRIMUS[9]
  Primus available from http://primus.gs.washington.edu/primusweb/index.html

- Beagle[10]
  Software available from http://faculty.washington.edu/browning/beagle/

- Locus Zoom[16]
  Locus Zoom web platform access via http://locuszoom.org

- SNP2HLA[17]
  SNP2HLA software available from
  http://software.broadinstitute.org/mpg/snp2hla

- LD-LINK[20]
  LD-link online platform available from https://ldlink.nci.nih.gov

- UCSC genome browser[22,23]
  UCSC Genome browser available from https://genome.ucsc.edu

- FACTOR BOOK[24]
  Factor book online software available at www.factorbook.org

- MASTMEME[25]
  Mast-Meme online software available at https://meme-suite.org

REFERENCES

1. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun 7;447(7145):661–78.

2. Illumina. “TOP/BOT” Strand and “A/B” Allele - Technical Note. Illumina SNP Genotyping. 2006;

3. Gibbs RA, Belmont JW, Hardenbol P, et al. The International HapMap Project. Nature. 2003 Dec;426(6968):789–96.

4. Fairfax BP, Makino S, Radhakrishnan J, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet. 2012 Mar 25;44(5):502–10.

5. Fairfax BP, Humburg P, Makino S, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science. 2014 Mar 7;343(6175):1246949.

6. Dufek S, Cheshire C, Levine AP, et al. Genetic Identification of Two Novel Loci Associated with Steroid-Sensitive Nephrotic Syndrome. J Am Soc Nephrol. 2019;30(8):1375–84.

7. Golden Helix, Inc., Bozeman, MT www.goldenhelix.com. SNP & Variation Suite ™ (Version 8.8.1). Golden Helix, Inc.

8. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

9. Staples J, Qiao D, Cho MH, Silverman EK, Nickerson DA, Below JE. PRIMUS: Rapid Reconstruction of Pedigrees from Genome-wide Estimates of Identity by Descent. Am J Hum Genet. 2014 Nov 6;95(5):553–64.

10. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet. 2018 06;103(3):338–48.

11. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 2016 01;32(11):1749–51.

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

13. Sillanpää MJ. Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. Heredity (Edinb). 2011 Apr;106(4):511–9.

14. Zhang Y, Pan W. Principal component regression and linear mixed model in association analysis of structured samples: competitors or complements? Genet Epidemiol. 2015 Mar;39(3):149–55.

15. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet. 2016;24(8):1202–5.

16. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010 Sep 15;26(18):2336–7.

17. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE. 2013;8(6):e64683.

18. Lonsdale, J, Thomas, J, Salvatore, M, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580–5.

19. Westra H-J, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013 Oct;45(10):1238–43.

20. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015 Nov 1;31(21):3555–7.

21. Levine AP, Chan MMY, Sadeghi-Alavijeh O, et al. Large-Scale Whole-Genome Sequencing Reveals the Genetic Architecture of Primary Membranoproliferative GN and C3 Glomerulopathy. J Am Soc Nephrol. 2020 Feb;31(2):365–73.

22. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996–1006.

23. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. 2013 Jan;41(Database issue):D56-63.

24. Wang J, Zhuang J, Iyer S, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. 2012 Sep;22(9):1798–812.

25. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015 Jul 1;43(Web Server issue):W39–49.

26. International Multiple Sclerosis Genetics Consortium (IMSGC). Comprehensive follow-up of the first genome-wide association study of multiple sclerosis identifies KIF21B and TMEM39A as susceptibility loci. Hum Mol Genet. 2010 Mar 1;19(5):953–62.

27. D'Cunha MA, Pandit L, Malli C. CD6 gene polymorphism rs17824933 is associated with multiple sclerosis in Indian population. Ann Indian Acad Neurol. 2016 Dec;19(4):491–4.

28. Lessard CJ, Adrianto I, Ice JA, et al. Identification of IRF8, TMEM39A, and IKZF3-ZPBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. Am J Hum Genet. 2012 Apr 6;90(4):648–60.

29. Sheng Y, Xu J, Wu Y, et al. Association analyses confirm five susceptibility loci for systemic lupus erythematosus in the Han Chinese population. Arthritis Res Ther. 2015 Mar 28;17:85.

30. Cai X, Huang W, Liu X, Wang L, Jiang Y. Association of novel polymorphisms in TMEM39A gene with systemic lupus erythematosus in a Chinese Han population. BMC Med Genet. 2017 20;18(1):43.

31. Yao Q, Wang B, Qin Q, Jia X, Li L, Zhang J-A. Genetic Variants in TMEM39A Gene Are Associated with Autoimmune Thyroid Diseases. DNA Cell Biol. 2019 Nov;38(11):1249–56.

32. Hitomi Y, Ueno K, Kawai Y, et al. POGLUT1, the putative effector gene driven by rs2293370 in primary biliary cholangitis susceptibility locus chromosome 3q13.33. Sci Rep. 2019 14;9(1):102.