haematologica
Journal of The Ferrata Storti Foundation

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

by Matthew J. Stubbs, Paul Coppo, Chris Cheshire, Agnès Veyradier, Stephanie Dufek, Adam P. Levine, Mari Thomas, Vaksha Patel, John O. Connolly, Michael Hubank, Ygal Benhamou, Lionel Gallicier, Pascale Poullin, Robert Kleta, Daniel P. Gale, Horia Stanescu, and Marie A. Scully

Haematologica 2021 [Epub ahead of print]

Citation: Matthew J. Stubbs, Paul Coppo, Chris Cheshire, Agnès Veyradier, Stephanie Dufek, Adam P. Levine, Mari Thomas, Vaksha Patel, John O. Connolly, Michael Hubank, Ygal Benhamou, Lionel Gallicier, Pascale Poullin, Robert Kleta, Daniel P. Gale, Horia Stanescu, and Marie A. Scully. Identification of a novel genetic locus associated with immune mediated thrombotic thrombocytopenic purpura.
Haematologica. 2021; 106:xxx
doi:10.3324/haematol.2020.274639

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval; the final version of the manuscript will then appear in print on a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.

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*] Marie A Scully,[1,5*]

1. Haemostasis Research Unit, UCL (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),
6. Clinical Genomics, Royal Marsden Hospital (London, UK).

**Denotes equal author contribution*

Correspondence

Please address correspondence to Dr Matthew James Stubbs, email m.stubbs@doctors.org.uk

Main Manuscript - 2068 words

Clinical Trials: UK MREC: 08/H0810/54, French ClinicalTrials.gov, NCT00426686

Appendices

Supplemental materials and methods (detailing full experimental outline) and supplemental results are detailed in the Supplemental Materials appendix.

ACKNOWLEDGEMENTS

This Study was supported by research funding from Shire/Takeda to Professor Marie Scully. We would like to thank Drs Fairfax, Makino, and Knight from the University of Oxford for sharing ethnically matched control data. We would also like to acknowledge the significant contribution of our collaborators within in UK TTP registry and the French Reference Centre for Thrombotic Microangiopathies, who are listed below:

The members of the UK TTP Registry are:

Aberdeen Royal Infirmary (H Watson, M Greiss, J Dixon, S Rodwell, M Fletcher), Addenbrookes Hospital (W Thomas, S MacDonald. Bath Royal United Hospitals: S Moore, C Cox), Birmingham University Hospitals (W Lester, G Lowe, C Percy, E Dwenger, M Pope), Birmingham Children's Hospital (J Motwani), Bournemouth and Christchurch Hospitals, (S Killick, M Serrano), Bristol University Hospitals (A Clark, A Mumford, L Humphrey, S Mulligan), Cardiff & Vale University Hospitals (R Rayment, P Collins, M Norton, A Guerrero, S Cunningham), Cornwall Hospitals NHS Trust (D Beech, S Hunter, B Mills), Coventry & Warwickshire University Hospitals (O Chapman, B Bailiff, A Pearson, D Morris), Edinburgh Royal Infirmary (L Manson, N Priddee), Glasgow Royal Infirmary (K Douglas, C Tait, C Bagot), Glasgow Royal Hospital for Children (E Chalmers), Great Ormand Street Hospital (R Liesner, K Sibson, A Taylor, A Griffien), Guy's & St Thomas' NHS Trust (B Hunt, J Young), Imperial College NHS Trust (N Cooper, A Luqmani, C Vladescu, D Paul), Leeds Teaching Hospital NHS Trust (Q Hill), Leicester Hospitals (R Gooding, K Siguake), Royal Liverpool Hospital (T Dutt, C Powell), Manchester Children's Hospital (J Thachil, J Granger, S Boydell), Newcastle Hospitals (T Biss, J Wallis, J Hanley, K Talks, A Charlton), Norfolk and Norwich University Hospitals NHS Foundation Trust (H Lyall, E Malone, M Sheridan), Nottingham University Hospitals NHS Trust (G Swallow, J Hermans), Oxford University Hospital NHS Trust (S Benjamin, C Deane, A Eordogh, K Santos), Oxford Children's Hospital (N Bhatnager, S Pavord, G Hall, P Baker), Plymouth University Hospitals NHS Trust (T Nokes), Poole Hospital NHS Foundation Trust (F Jack, N Beamish, A Wandowski), Portsmouth Hospitals NHS Trust (T Cranfield, C James, S Liu), Royal Devon & Exeter NHS

Foundation Trust (J Coppell, L Ngu), Sheffield Teaching Hospital NHS Foundation Trust (J Vanveen, M Makris, R MacLean, K Harrington, S Megson, R Fretwell), South Tees Hospitals NHS Foundation Trust (R Dang, M David, J Maddox, D Winterburn), South Warwickshire NHS Foundation Trust (P Rose), St George's University Hosptials NHS Foundation Trust (S Austin, J Uprichard, J Chackathayil, A Lee), University College London Hospitals (M Scully, JP Westwood, M Thomas, R Newton, S McGuckin, I Obu, C Vendramin, L Keogh, J Shin, M Stubbs), NHSBT (K Pendry, K Slevin), Southampton University Hospital NHS Foundation Trust (S Boyce).

The members of the French Reference Center for Thrombotic Microangiopathies (CNR-MAT) are: Augusto Jean-François (Service de Néphrologie, dialyse et transplantation ; CHU Larrey, Angers); Azoulay Elie (Service de Réanimation Médicale, Hôpital Saint-Louis, Paris); Barbay Virginie (Laboratoire d'Hématologie, CHU Charles Nicolle, Rouen); Benhamou Ygal (Service de Médecine Interne, CHU Charles Nicolle, Rouen); Bordessoule Dominique (Service d'Hématologie, Hôpital Dupuytren, Limoges); Charasse Christophe (Service de Néphrologie, Centre Hospitalier de Saint-Brieuc); Charvet-Rumpler Anne (Service d'Hématologie, CHU de Dijon) ; Chauveau Dominique (Service de Néphrologie et Immunologie Clinique, CHU Rangueil, Toulouse); Choukroun Gabriel (Service de Néphrologie, Hôpital Sud, Amiens); Coindre Jean-Philippe (Service de Néphrologie, CH Le Mans); Coppo Paul (Service d'Hématologie, Hôpital Saint-Antoine, Paris); Corre Elise (Service d'Hématologie, Hôpital Saint-Antoine, Paris); Delmas Yahsou (Service de Néphrologie, CHU de Bordeaux, Bordeaux); Deschenes Georges (Service de Néphrologie Pédiatrique, Hôpital Robert Debré, Paris); Devidas Alain (Service d'Hématologie, Hôpital Sud-Francilien, Corbeil-Essonnes); Dossier Antoine (Service de Néphrologie, Hôpital Bichat, Paris); Fain Olivier (Service de Médecine Interne, Hôpital Saint-Antoine, Paris); Fakhouri Fadi (Service de Néphrologie, CHU Hôtel-Dieu, Nantes) ; Frémeaux-Bacchi Véronique (Laboratoire d'Immunologie, Hôpital Européen Georges Pompidou, Paris); Galicier Lionel (Service d'Immunopathologie, Hôpital Saint-Louis, Paris); Grangé Steven (Service de Réanimation Médicale, CHU Charles Nicolle, Rouen) ; Guidet Bertrand (Service de Réanimation Médicale, Hôpital Saint-Antoine, Paris); Halimi Jean-Michel (Service de Néphrologie Pédiatrique, Hôpital Bretonneau, Tours); Hamidou Mohamed (Service de Médecine Interne, Hôtel-Dieu, Nantes); Hié Miguel (Service de Médecine Interne, Groupe Hospitalier Pitié-Salpétrière, Paris) ; Jacobs Frédéric (Service de Réanimation Médicale, Hôpital Antoine Béclère, Clamart); Joly Bérangère (Service d'Hématologie Biologique, Hôpital Lariboisière, Paris) ; Kanouni Tarik (Unité d'Hémaphrèse, Service d'Hématologie, CHU de Montpellier) ; Kaplanski Gilles (Service de Médecine Interne, Hôpital la Conception, Marseille) ; Lautrette Alexandre (Hôpital Gabriel Montpied,

Service de Réanimation médicale, Clermont-Ferrand); Le Guern Véronique (Unité d'Hémaphérèse, Service de Médecine Interne, Hôpital Cochin, Paris) ; Mariotte Eric (Service de Réanimation, Hôpital Saint-Louis, Paris) ; Moulin Bruno (Service de Néphrologie, Hôpital Civil, Strasbourg); Mousson Christiane (Service de Néphrologie, CHU de Dijon); Ojeda Uribe Mario (Service d'Hématologie, Hôpital Emile Muller, Mulhouse); Ouchenir Abdelkader (Service de Réanimation, Hôpital Louis Pasteur, Le Coudray); Parquet Nathalie (Unité de Clinique Transfusionnelle, Hôpital Cochin, Paris); Peltier Julie (Urgences Néphrologiques et Transplantation Rénale, Hôpital Tenon, Paris) ; Pène Frédéric (Service de Réanimation Médicale, Hôpital Cochin, Paris) ; Perez Pierre (Service de Réanimation polyvalente, CHU de Nancy) ; Poullin Pascale (Service d'hémaphérèse et d'autotransfusion, Hôpital la Conception, Marseille); Pouteil-Noble Claire (Service de Néphrologie, CHU Lyon-Sud, Lyon); Presne Claire (Service de Néphrologie, Hôpital Nord, Amiens); Provôt François (Service de Néphrologie, Hôpital Albert Calmette, Lille); Rondeau Eric (Urgences Néphrologiques et Transplantation Rénale, Hôpital Tenon, Paris); Saheb Samir (Unité d'Hémaphérèse, Hôpital la Pitié-Salpétrière, Paris) ; Seguin Amélie (Service de Réanimation Médicale, centre hospitalier de Vendée) ; Servais Aude (Service de Néphrologie, CHU Necker-Enfants Malades) ; Stépanian Alain (Laboratoire d'Hématologie, Hôpital Lariboisière, Paris); Vernant Jean-Paul (Service d'Hématologie, Hôpital la Pitié-Salpétrière, Paris); Veyradier Agnès (Service d'Hématologie Biologique, Hôpital Lariboisière, Paris); Vigneau Cécile (Service de Néphrologie, Hôpital Pontchaillou, Rennes); Wynckel Alain (Service de Néphrologie, Hôpital Maison Blanche, Reims); Zunic Patricia (Service d'Hématologie, Groupe Hospitalier Sud-Réunion, la Réunion).

AUTHOR CONTRIBUTIONS

- MJ Stubbs - designed research, recruited patients, performed research, collected data, analysed and interpreted data, wrote the manuscript.
- P Coppo - designed research, recruited patients, analysed and interpreted data, wrote the manuscript.
- C Cheshire - performed research, collected data, analysed and interpreted data, wrote the manuscript.
- A Veyradier - designed research, recruited patients, analysed and interpreted data, wrote the manuscript.

- S Dufek - performed research, collected data, analysed and interpreted data, wrote the manuscript.
- AP Levine - performed research, collected data, analysed and interpreted data, wrote the manuscript.
- M Thomas - designed research, recruited patients, analysed and interpreted data, wrote the manuscript.
- V Patel - performed research, collected data, analysed and interpreted data, wrote the manuscript.
- JO Connolly - designed research, wrote the manuscript.
- M Hubank - designed research, wrote the manuscript.
- Y Benhamou - designed research, recruited patients, wrote the manuscript.
- L Galicier - designed research, recruited patients, wrote the manuscript.
- P Poullin - designed research, recruited patients, wrote the manuscript.
- R Kleta - designed research, performed research, analysed and interpreted data, wrote the manuscript.
- DP Gale - designed research, performed research, analysed and interpreted data, wrote the manuscript.
- H Stanescu - designed research, performed research, analysed and interpreted data, wrote the manuscript.
- MA Scully - designed research, recruited patients, performed research, analysed and interpreted data, wrote the manuscript.

DECLARATION OF INTERESTS

- MJ Stubbs - Shire/Takeda (Research Funding).
- P Coppo – Sanofi (advisory board and symposia fees), Alexion (advisory board and symposia fees) and Roche (advisory board and symposia fees), Octapharma (fees).
- C Cheshire - No conflicts of interest.
- A Veyradier - Ablynx-Sanofi (Advisory board), Roche-Chugai (Advisory board), Shire/Takeda (Advisory board), LFB Biomédicaments (course fees

and awards), Octapharma (course fees and awards), CSL-Behring (fees for courses and awards).

- S Dufek – No conflicts of interest.
- AP Levine - No conflicts of interest.
- M Thomas - Sanofi (Advisory board).
- V Patel - No conflicts of interest.
- JO Connolly - No conflicts of interest.
- M Hubank - No conflicts of interest.
- Y Benhamou - Sanofi (Advisory board), Octapharma (Advisory board).
- L Galicier - No conflicts of interest.
- P Poullin - Sanofi (Advisory Board).
- R Kleta - No conflicts of interest.
- DP Gale - Alexion (Honoraria, Advisory board).
- H Stanescu - No conflicts of interest.

M Scully; Novarits (Consultancy, Honoria, Advisory board, Speakers Fees), Shire/Takeda (Honoria, Advisory board, Research Funding Speakers Fees), Ablynx/Sanofi (Consultancy, Honoraria, Advisory board, Research Funding, Speakers Bureau), Shire/Takeda (Honoraria, Advisory Board, Research Funding, Speakers Bureau), Alexion (Honoraria, Advisory board, Speakers Bureau), Baxalta (Research Funding).

Abstract

Immune Thrombotic Thrombocytopenic Purpura (iTTP) is an ultra-rare, life-threatening disorder, mediated through severe ADAMTS13 deficiency causing multi-system micro-thrombi formation, and has specific HLA associations. We undertook a large genome wide association study to investigate additional genetically distinct associations in iTTP.

We compared two iTTP patient cohorts with controls, following standardised genome wide quality control procedures for SNPs and imputed HLA types. Associations were functionally investigated using expression quantitative trait loci (eQTL), and motif binding prediction software.

Independent associations consistent with previous findings in iTTP were detected at the HLA locus and in addition a novel association was detected on chromosome 3 (rs9884090, p-value of 5.22×10^{-10}, Odds Ratio (OR) = 0.40) in the UK discovery cohort. Meta-analysis, including the French replication cohort, strengthened the associations. The haploblock containing rs9884090 is associated with reduced protein O-glycosyltransferase 1 (POGLUT1) expression (eQTL $P<0.05$), and functional annotation suggested a potential causative variant (rs71767581). This work implicates POGLUT1 in iTTP pathophysiology and suggests altered post-translational modification of its targets may influence disease susceptibility.

167 words

INTRODUCTION

Thrombotic Thrombocytopenic Purpura (TTP) is an ultra-rare, life-threatening illness, with an annual incidence of approximately 6/million, and with an untreated mortality approaching 90% (10-20% with prompt intervention). It can affect patients of any age, but often affects young adults (30-40 years) and is more common in women.[1] The initial diagnosis of TTP is based on clinical suspicion, but ADAMTS13 (a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13) activity <10IU/dL confirms the diagnosis. Severe deficiency of ADAMTS13 results in failure to cleave ultra large von Willebrand Factor multimers (UL-VWF), crucial for normal haemostatic function and proteolytic regulation of VWF. ADAMTS13 deficiency in immune TTP (iTTP) is mediated through IgG autoantibodies.[2,3] The precipitant of the disease in most cases is unclear.[4]

As with many autoimmune diseases, HLA type is associated with the risk of developing iTTP, with HLA-DRB1*11, HLA-DQB1*03 and HLADRB3* increasing risk, and HLA-DRB1*04 and HLA-DRB4 (HLA-DR53) being protective in Europeans.[5,6,7] No genetic risk factors outside the HLA genes have previously been shown to be associated with iTTP.

We performed a genome wide association study in UK and French iTTP cohorts and identified association of alleles both within and beyond the HLA locus.

METHODS

COHORTS

As part of the UK TTP registry, patients were consented for DNA analysis (MREC: 08/H0810/54) (see Supplemental Materials). Patients on the UK TTP registry were screened for the clinical diagnosis, and confirmed with an ADAMTS13 level <10IU/dL at diagnosis (utilising FRETS methodology)[8] and the presence of an anti-ADAMTS13 autoantibody.[2,3] The French replication cohort TTP samples were obtained from the French Reference Centre for TMA (CNR-MAT) and informed consent was obtained from each patient with confirmed iTTP (see above criteria) (Institutional Review Board of Pitié Salpêtrière Hospital, ClinicalTrials.gov, NCT00426686). The European control genotypes were obtained from the Wellcome Trust Case Control Consortium (WTCCC), both the 1958 British Birth Cohort and National Blood Service control samples.[9] In addition, controls were used from the Illumina reference panel[10] and Oxford controls.[11,12]

GENOTYPING, QUALITY CONTROL AND IMPUTATION

TTP samples were genotyped on the Illumina Human Omni Express SNP chips and controls were genotyped on different SNP chips (see Supplementary Methodology). Pre-imputation quality control was performed in all data sets separately, and then in a combined cohort (Supplemental Figure 1). Quality control was performed for individuals and SNPs. Individuals were selected for further analysis by European ancestry principal component analysis (PCA) (see Supplemental Figure 2). Only SNPs present in all data sets were subsequently analysed.

Genome-wide imputation was performed on markers that had passed quality control, and were present in all datasets using Beagle (version 5.0) utilising the 1000 Genome Project Phase 3 as a reference panel.[13] In addition to standardised QC, only SNPs with a dosage R^2 (DR2) >0.8 were included.

GWAS AND LOCI CHARACTERISATION

Genome wide association testing was performed using SNP & Variation Suite v8, using logistic regression with principal component correction.[14,15] The logistic regression p-values, odds ratios were calculated in addition to lambda inflation factors, and QQ plots are shown (Supplemental Figure 3). A standardised genome wide significance level of $5x10^{-8}$ was applied.[15] For discovery and replication analysis meta-data please contact the authors.

Conditional analyses were undertaken using a full versus reduced regression model. Lead SNPs at each locus were used as conditional inputs to determine independence, with results plotted using Locus Zoom software.[16]

Imputation of HLA types was performed utilising SNP2HLA with previously genotyped markers.[17] Imputed HLA types were excluded if the R^2 (confidence) was <0.80. Conditional analyses were subsequently performed as described above.

Expression quantitative trait locus (eQTL) analysis was performed to associate identified SNPs with differential gene expression.[18] Additional markers in linkage disequilibrium with the lead SNP at the Chromosome 3 locus were identified by LD-link (https://ldlink.nci.nih.gov).[19] Functional annotation of the haploblock was performed using ChipSeq data via the UCSC genome browser (https://genome.ucsc.edu) . Binding sites of transcription factors (highlighted through genome annotation) were obtained from FactorBook[20], and position weight matrix (PWM) binding motifs generated. Binding motifs were generated using Mast-Meme.[21]

RESULTS

DISCOVERY COHORT

Following quality control as outlined in the methods (Supplemental Figure 1) there were 241 TTP cases and 3200 controls in the UK discovery cohort. Following imputation and quality control 3,649,347 SNPs were available for analysis. Association testing was performed using a logistic regression model with PCA correction, and the genomic inflation factor (lambda) was 1.0239 (Supplemental Figure 3).

In the UK discovery cohort two peaks were identified (Figure 1) (Supplemental Figure 4) (lead SNPs summarised in Table 1). The peak with the strongest association corresponded to the class II HLA region on chromosome 6, with 1,017 SNPs reaching genome wide significance. The lead SNP rs28383233 located in the intergenic region between *HLA-DRB1* and *HLA-DQA1* ($p=2.20x10^{-23}$, odds ratio 3.12, 95% CI 2.49-3.93) (Table 1 and Figure 2).

Conditional analysis was performed on rs28383233 and the lead SNP following this was rs1064994 (within *HLA-DQA1)*, with a p-value of $1.13x10^{-10}$ (odds ratio 2.20, 95% CI 2.06-3.37). Following conditioning on both rs28383233 and rs1064994 no further markers reached significance within the class II HLA region, indicating that there are two detectable independent genetic associations with iTTP within the HLA region.

HLA imputation was performed on the UK discovery cohort, and following quality control, 95 imputed HLA alleles remained. HLA-DRB1*11:01 was the allele most

strongly associated with iTTP, with a p-value of $3.25x10^{-17}$ (odds ratio 2.79, 95% CI 2.23-3.50). Following conditional analysis of HLA-DRB1*1101, no other HLA types reached genome wide significance, but HLA-DQA1*03:01 remained significant (with a HLA-only Bonferroni correction, $P<5.26x10^{-4}$) at $1.49x10^{-6}$ (odds ratio 0.47, 95% CI 0.33-0.65) suggesting that the protective effect of this allele is independent of HLA-DRB1*11:01.

In addition to the class II HLA peak on chromosome 6, a novel association was observed on chromosome 3. 16 markers reached genome wide significance, with the lead SNP, rs9884090(A), having a p-value of $5.22x10^{-10}$ (odds ratio 0.40, 95% CI 0.29-0.56) (Table 1 and Figure 3). Upon conditional analysis of the lead SNP no markers reached genome wide significance indicating one detectable signal at this locus. No statistical epistasis was seen between the chromosome 3 and chromosome 6 associations, with each association being independent. Five genes are annotated within this chromosome 3 haploblock: *ARHGAP31*, *TMEM39A*, *POGLUT1*, *TIMMDC1*, and *CD80*.

REPLICATION COHORT

Within the French replication cohort there were 112 cases and 2603 controls following quality control as outlined in the methods (Supplemental Figure 1 and 2). 3,649,546 SNPs were available for analysis, and association testing was performed using a logistic regression model with PCA correction, and lambda was 1.0830 (Supplemental Figure 5).

The association with the lead SNP in the chromosome 3 haploblock, rs9884090(A) was replicated with a p-value of 0.001 (odds ratio 0.52), and the two independent lead SNPs with the class II HLA peak on chromosome 6 were also replicated (Table

2). The locus zoom plots are shown (Supplemental Figures 6-8). Imputed HLA type analysis was also consistent with the UK discovery cohort with HLA-DRB1*11:01 and HLA-DQA1*03:01 representing two independent HLA signals.

In addition, a meta-analysis was performed combining the UK and French cohorts (cases 241/112, controls 3200/2603 respectively), which demonstrated strengthening of the previously observed signal (rs9884090 $p=1.60x10^{-10}$, OR 0.47, rs28383233 $p=1.22x10^{-42}$, OR 3.70, rs1064994 $p= 5.03x10^{-25}$, OR 2.89) (Table 3 and Supplemental Figure 9).

EQTL AND FUNCTIONAL DNA ANALYSIS

Expression quantitative trait loci (eQTL) data from the Genotype Tissue Expression Project and Blood eQTL Browser for the lead SNP at the chromosome 3 locus (rs9884090) demonstrated significant reduction in expression of *POGLUT1* with the protective allele in the majority of tissues tested, including blood cells ($p<0.001$).[18,22]

LD-link identified 20 markers found to be in tight linkage disequilibrium (R^2 and D' >0.80) with rs9884090 contained within the chromosomal region (see supplemental Table 1).[19] All markers were functionally annotated with information from the UCSC Genome Browser (Human Assembly GRCh37/hg19)[23,24] (see Supplemental Table 1). One variant was particularly noted, rs71767581 (Ch3, 119187422 AC/-del), which is a 2 base pair deletion in the promoter of *POGLUT1*. This may be functionally important as the haploblock identified is associated with reduced expression in POGLUT1. Upon analysis of ChipSeq data in UCSC Genome Browser 14 transcription factors were predicted to bind at this site (see Supplemental Table 2), adding further evidence that rs71767581 may be functionally important for POGLUT1 expression.

DISCUSSION

This genome wide association study, involving 2 European populations, is the first to be performed in iTTP and shows consistent evidence of association at loci on chromosome 6 and chromosome 3. The associated alleles on Chromosome 6 lie within the HLA region and imputation of HLA types and conditional analyses indicated independent association between HLA-DRB1*11:01 (OR 2.79; $p=3.25x10^{-17}$) and HLA-DQA1*03:01 (OR 0.47; $p=1.49x10^{-6}$, post conditional analysis), which are consistent, and in linkage with previously published risk and protective associations with iTTP at this locus.[5–7] A recent case-control study comparing frequency of alleles only at immune loci in 190 Italian TTP patients and 1255 controls identified the HLA variant rs6903608, (in addition to HLA-DQB1*05:03) as conferring a 2.5 fold increase of developing TTP.[25]

Here we also identified a novel association of iTTP with alleles on chromosome 3 tagged by the lead SNP rs9884090. Five genes are located within the associated haploblock: *ARHGAP31*, *TMEM39A*, *POGLUT1*, *TIMMDC1*, and *CD80. ARHGAP31* (Rho GTPase Activating Protein 31) is associated with the autosomal dominant condition Adams-Oliver Syndrome (OMIM 100300).[26] Mutations within *ARHGAP31* have been implicated with abnormal vascular development and VEGF (vascular endothelial growth factor) angiogenesis.[27] Little is understood regarding the function of *TMEM39A* (transmembrane protein 39A). While variants have been implicated in autoimmune disease such as systemic lupus erythematosus[28,29] and multiple sclerosis[30,31], understanding of its function is lacking. *TIMMDC1* is a membrane embedded mitochondrial complex factor, and is associated with mitochondrial disorders.[32] The protein encoded by the *CD80* gene functions as a membrane receptor being activated by CTLA-4 or CD28, both of which are T-cell

receptors. The downstream mechanisms are T-cell proliferation and cytokine production. CD80 and its receptors have been associated with focal segmental glomerulosclerosis[33] and systemic lupus erythematosus.[34],[35] *POGLUT1* (Protein O-Glucosyltransferase 1) is mutated in Dowling-Degos Disease-4 (an autosomal dominant genodermatosis with progressive and disfiguring reticulate hyperpigmentation and muscular dystrophy, OMIM 615696) and POGLUT1 has been shown to catalyse O-glycosylation of epidermal growth factor (EGF)-like repeats.[36,37] EGF-like repeats are well conserved structures, and highly represented with proteins involved in coagulation.[38,39] *In-vitro* work has demonstrated POGLUT1 binds and glycosylates specific coagulation factors including Factor VII and Factor IX.[37,40]

The haploblock identified in this analysis of iTTP (which is tagged by rs9884090(A)) is associated with significantly decreased *POGLUT1* expression by eQTL.[41] Several other genetic variants contained within this haploblock have been associated with other autoimmune diseases, and the majority of these variants have been shown to be in linkage with our lead variant rs9884090 (see Supplemental Results), supporting the findings described here.[28,29,31,42,43,44] eQTL analysis is a robust tool, that can associate gene expression with specific genetic variants. Our analysis found rs9884090(A) to have a reduced frequency in iTTP, and rs9884090(A) was shown to be associated with significantly decreased *POGLUT1* expression in different eQTL resources.[18,22] In order to locate the underlying genetic variant implicated in this reduced *POGLUT1* expression we used LD-link to identify additional variants, and located a 2-bp deletion with the *POGLUT1* upstream promoter region that is in tight linkage disequilibrium with the lead associated variant (R2/D'>0.80). As rs9884090(A) confers reduced risk of developing iTTP, we hypothesize that reduced

expression of *POGLUT1* leads to altered post-translational modification (O-glycosylation) of key POGLUT1 targets to reduce the risk of iTTP. The evidence we present supports POGLUT1 as the gene of interest, but we cannot exclude other genes within the associated haploblock. The pathway through which POGLUT1's effects could be mediated remains to be determined. Given there are several reported variants with this haploblock associated with different autoimmune disease, it is likely the downstream functional consequences medicated through POGLUT1 influence immune-regulatory pathways which may generally increase the risk of other autoimmune disease, in addition to iTTP, and may provide insights into potential therapies.[45–56]

In summary, we have identified a novel genetic variant, rs9884090(A), in two independent populations, which is associated with reduced risk of iTTP. Utilising linkage disequilibrium we have identified a functional variant in tight LD with the lead SNP in the POGLUT1 promoter site and eQTL demonstrates reduced POGLUT1 expression associated with this variant. We therefore hypothesise this leads to altered O-glycosylation on POGLUT1 targets. Whilst the exact role of POGLUT1 in the pathophysiology of iTTP requires further downstream functional analysis, this work represents an important step forward in our understanding of iTTP.

REFERENCES

1. Scully, M, Yarranton H, Liesner R, et al. Regional UK TTP registry: correlation with laboratory ADAMTS 13 analysis and clinical features. Br J Haematol. 2008;142(5):819-826.

2. Tsai H-M, Raoufi M, Zhou W, et al. ADAMTS13-binding IgG are present in patients with thrombotic thrombocytopenic purpura. Thromb Haemost. 2006;95(5):886-892.

3. Rieger M, Mannucci PM, Kremer Hovinga JA, et al. ADAMTS13 autoantibodies in patients with thrombotic microangiopathies and other immunomediated diseases. Blood. 2005;106(4):1262-1267.

4. Scully M, Hunt BJ, Benjamin S, et al. Guidelines on the diagnosis and management of thrombotic thrombocytopenic purpura and other thrombotic microangiopathies. Br J Haematol. 2012;158(3):323-335.

5. Scully M, Brown J, Patel R, McDonald V, Brown CJ, Machin S. Human leukocyte antigen association in idiopathic thrombotic thrombocytopenic purpura: evidence for an immunogenetic link. J Thromb Haemost. 2010;8(2):257-262.

6. Coppo P, Busson M, Veyradier A, et al. HLA-DRB1*11: a strong risk factor for acquired severe ADAMTS13 deficiency-related idiopathic thrombotic thrombocytopenic purpura in Caucasians. J Thromb Haemost. 2010;8(4):856-859.

7. John M-L, Hitzler W, Scharrer I. The role of human leukocyte antigens as predisposing and/or protective factors in patients with idiopathic thrombotic thrombocytopenic purpura. Ann Hematol. 2012;91(4):507-510.

8. Kokame K, Nobe Y, Kokubo Y, Okayama A, Miyata T. FRETS-VWF73, a first fluorogenic substrate for ADAMTS13 assay. Br J Haematol. 2005;129(1):93-100.

9. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-678.

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

11. 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;44(5):502-510.

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

13. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet. 2018;103(3):338-348.

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

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):1202-1205.

16. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-2337.

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;45(6):580-585.

19. 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;31(21):3555-3557.

20. 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;22(9):1798-1812.

21. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43(W1):W39-49.

22. Westra H-J, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45(10):1238-1243.

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

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

25. Mancini I, Ricaño-Ponce I, Pappalardo E, et al. Immunochip analysis identifies novel susceptibility loci in the human leukocyte antigen region for acquired thrombotic thrombocytopenic purpura. J Thromb Haemost. 2016;14(12):2356-2367.

26. Hassed S, Li S, Mulvihill J, Aston C, Palmer S. Adams-Oliver syndrome review of the literature: Refining the diagnostic phenotype. Am J Med Genet A. 2017;173(3):790-800.

27. Caron C, DeGeer J, Fournier P, et al. CdGAP/ARHGAP31, a Cdc42/Rac1 GTPase regulator, is critical for vascular development and VEGF-mediated angiogenesis. Sci Rep. 2016;6:27485.

28. 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;18(1):43.

29. 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;90(4):648-660.

30. Wagner M, Sobczyński M, Bilińska M, et al. Preliminary Study on the Role of TMEM39A Gene in Multiple Sclerosis. J Mol Neurosci. 2017;62(2):181-187.

31. D'Cunha MA, Pandit L, Malli C. CD6 gene polymorphism rs17824933 is associated with multiple sclerosis in Indian population. Ann Indian Acad Neurol. 2016;19(4):491-494.

32. Guarani V, Paulo J, Zhai B, Huttlin EL, Gygi SP, Harper JW. TIMMDC1/C3orf1 functions as a membrane-embedded mitochondrial complex I assembly factor through association with the MCIA complex. Mol Cell Biol. 2014;34(5):847-861.

33. Ishimoto T, Shimada M, Araya CE, Huskey J, Garin EH, Johnson RJ. Minimal change disease: a CD80 podocytopathy? Semin Nephrol. 2011;31(4):320-325.

34. Kow NY, Mak A. Costimulatory pathways: physiology and potential therapeutic manipulation in systemic lupus erythematosus. Clin Dev Immunol. 2013;2013:245928.

35. Saverino D, Simone R, Bagnasco M, Pesce G. The soluble CTLA-4 receptor and its role in autoimmune diseases: an update. Auto Immun Highlights. 2010;1(2):73-81.

36. Fernandez-Valdivia R, Takeuchi H, Samarghandi A, et al. Regulation of mammalian Notch signaling and embryonic development by the protein O-glucosyltransferase Rumi. Development. 2011;138(10):1925-1934.

37. Servián-Morilla E, Takeuchi H, Lee TV, et al. A POGLUT1 mutation causes a muscular dystrophy with reduced Notch signaling and satellite cell loss. EMBO Mol Med. 2016;8(11):1289-1309.

38. Stenflo J, Stenberg Y, Muranyi A. Calcium-binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions. Biochim Biophys Acta. 2000;1477(1-2):51-63.

39. Wouters MA, Rigoutsos I, Chu CK, Feng LL, Sparrow DB, Dunwoodie SL. Evolution of distinct EGF domains with specific functions. Protein Sci. 2005;14(4):1091-1103.

40. Li Z, Fischer M, Satkunarajah M, Zhou D, Withers SG, Rini JM. Structural basis of Notch O-glucosylation and O-xylosylation by mammalian protein-O-glucosyltransferase 1 (POGLUT1). Nat Commun. 2017;8(1):185.

41. Basmanav FB, Oprisoreanu A-M, Pasternack SM, et al. Mutations in POGLUT1, encoding protein O-glucosyltransferase 1, cause autosomal-dominant Dowling-Degos disease. Am J Hum Genet. 2014;94(1):135-143.

42. 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;19(5):953-962.

43. 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;17(1):85.

44. 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;38(11):1249-1256.

45. Lenting PJ, Christophe OD, Denis CV. von Willebrand factor biosynthesis, secretion, and clearance: connecting the far ends. Blood. 2015;125(13):2019-2028.

46. Verbij FC, Stokhuijzen E, Kaijen PHP, van Alphen F, Meijer AB, Voorberg J. Identification of glycans on plasma-derived ADAMTS13. Blood. 2016;128(21):e51-58.

47. Sorvillo N, Kaijen PH, Matsumoto M, et al. Identification of N-linked glycosylation and putative O-fucosylation, C-mannosylation sites in plasma derived ADAMTS13. J Thromb Haemost. 2014;12(5):670-679.

48. Nowak AA, O'Brien HER, Henne P, et al. ADAMTS-13 glycans and conformation-dependent activity. J Thromb Haemost. 2017;15(6):1155-1166.

49. Janghorban M, Xin L, Rosen JM, Zhang XH-F. Notch Signaling as a Regulator of the Tumor Immune Response: To Target or Not To Target? Front Immunol. 2018;9:1649.

50. Rong H, Shen H, Xu Y, Yang H. Notch signalling suppresses regulatory T-cell function in murine experimental autoimmune uveitis. Immunology. 2016;149(4):447-459.

51. Yang H, Zheng S, Mao Y, et al. Modulating of ocular inflammation with macrophage migration inhibitory factor is associated with notch signalling in experimental autoimmune uveitis. Clin Exp Immunol. 2016;183(2):280-293.

52. Kuksin CA, Minter LM. The Link between Autoimmunity and Lymphoma: Does NOTCH Signaling Play a Contributing Role? Front Oncol. 2015;5:51.

53. Bassil R, Orent W, Elyaman W. Notch signaling and T-helper cells in EAE/MS. Clin Dev Immunol. 2013;2013:570731.

54. Sandy AR, Stoolman J, Malott K, Pongtornpipat P, Segal BM, Maillard I. Notch signaling regulates T cell accumulation and function in the central nervous system during experimental autoimmune encephalomyelitis. J Immunol. 2013;191(4):1606-1613.

55. Radtke F, MacDonald HR, Tacchini-Cottier F. Regulation of innate and adaptive immunity by Notch. Nat Rev Immunol. 2013;13(6):427-437.

56. Hao X, Li Y, Wang J, et al. Deficient O-GlcNAc Glycosylation Impairs Regulatory T Cell Differentiation and Notch Signaling in Autoimmune Hepatitis. Front Immunol. 2018;9:2089.

TABLES

rsID (position)	Minor Allele / Major Allele	MAF Cases / MAF Controls	Logistic Regression p-value	Odds Ratio (95% CI)
rs9884090 (ch3:119116150)	A/G	0.08/0.19	$P = 5.22x10^{-10}$	0.40 (0.29-0.56)
rs28383233 (ch6:32584153)	G/A	0.64/0.40	$P = 2.20x10^{-23}$	3.12 (2.49-3.93)
rs1064994 (ch6:32611195)	C/T	0.25/0.11	$P = 1.13x10^{-10}$	2.20 (2.06-3.37)

Table 1 - Lead SNPs identified in the UK discovery cohort. Displayed are Minor/Major Alleles, Minor Allele Frequencies (MAF), logistic regression p-value (corrected for PCA stratification), and Odds Ratio (with 95% confidence intervals). Genomic positions refer to Human Assembly GRCh37/hg19.

rsID (position)	Minor Allele / Major Allele	MAF Cases / MAF Controls	Logistic Regression p-value	Odds Ratio (95% CI)
rs9884090 (ch3:119116150)	A/G	0.10/0.18	P = 0.001	0.52 (0.34-0.81)
rs28383233 (ch6:32584153)	G/A	0.68/0.40	$P = 3.87x10^{-9}$	2.57 (1.87-3.53)
rs1064994 (ch6:32611195)	C/T	0.42/0.11	$P = 5.015x10^{-9}$	2.86 (2.06-3.99)

Table 2 – French cohort replication of lead SNPs identified in the UK discovery cohort. Displayed are Minor/Major Alleles, Minor Allele Frequencies (MAF), logistic regression p-value (corrected for PCA stratification), and Odds Ratio (with 95% confidence intervals). Genomic positions refer to Human Assembly GRCh37/hg19.

rsID (position)	Minor Allele / Major Allele	MAF Cases / MAF Controls	Logistic Regression p-value	Odds Ratio (95% CI)
rs9884090 (ch3:119116150)	A/G	0.08/0.19	$P = 1.60x10^{-10}$	0.47 (0.36-0.60)
rs28383233 (ch6:32584153)	G/A	0.64/0.41	$P = 1.22x10^{-42}$	3.70 (2.81-4.03)
rs1064994 (ch6:32611195)	C/T	0.22/0.11	$P = 5.03x10^{-25}$	2.89 (2.39-3.49)

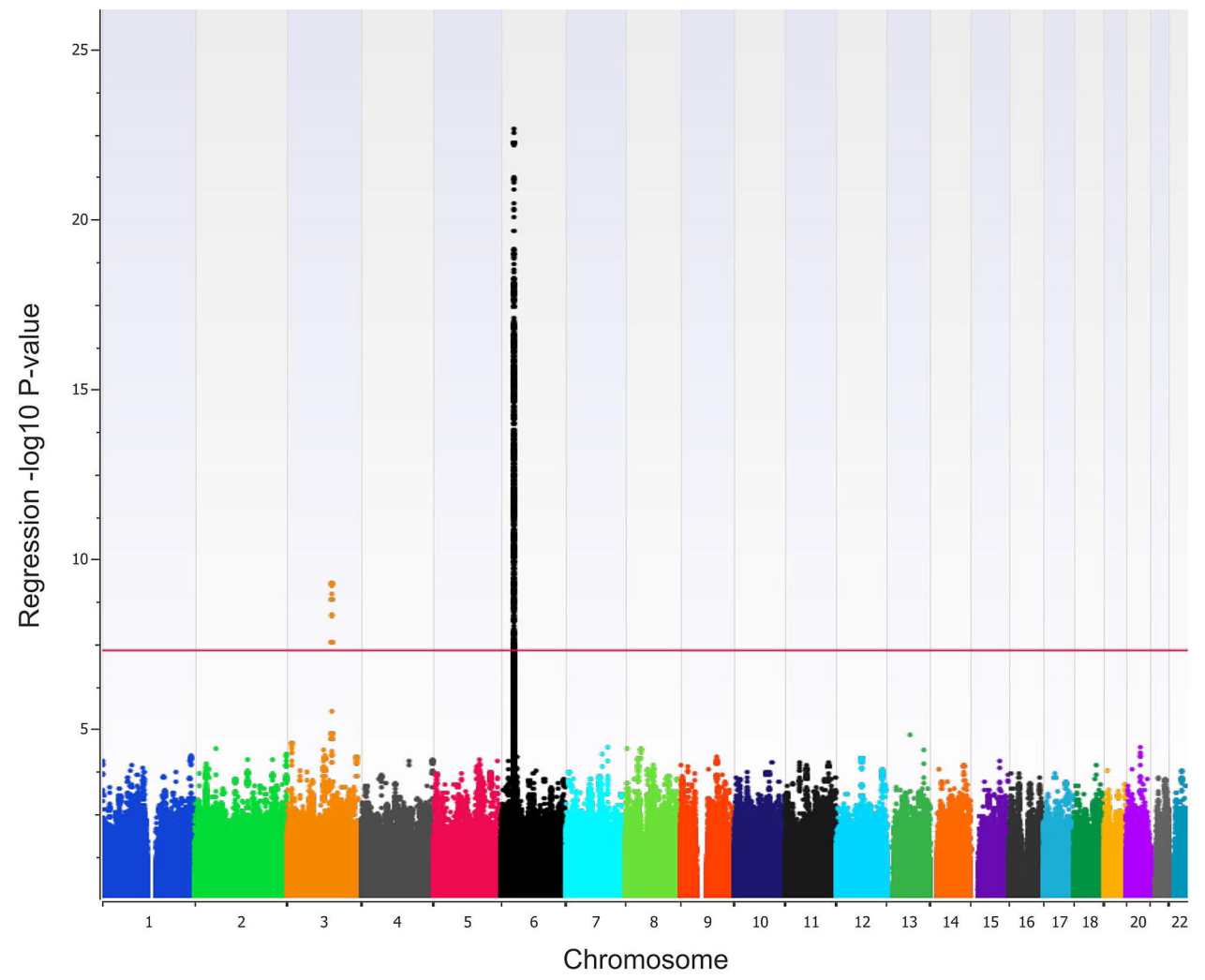
Table 3 – Meta-analysis combining UK and French Cohorts, showing lead SNPs identified in the UK discover cohort. Displayed are Minor/Major Alleles, Minor Allele Frequencies (MAF), logistic regression p-value (corrected for PCA stratification), and Odds Ratio (with 95% confidence intervals). Genomic positions refer to Human Assembly GRCh37/hg19.

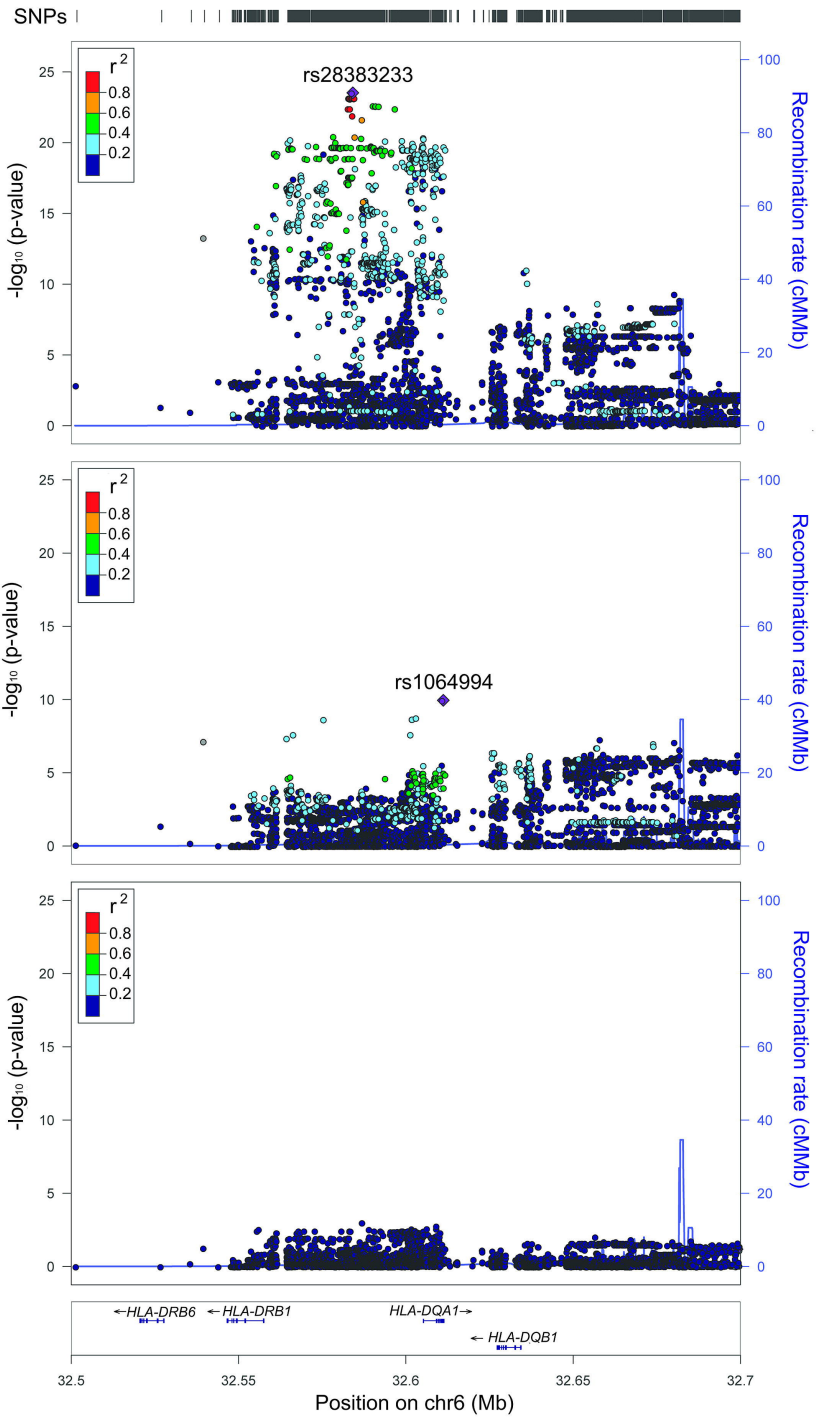
FIGURE LEGENDS

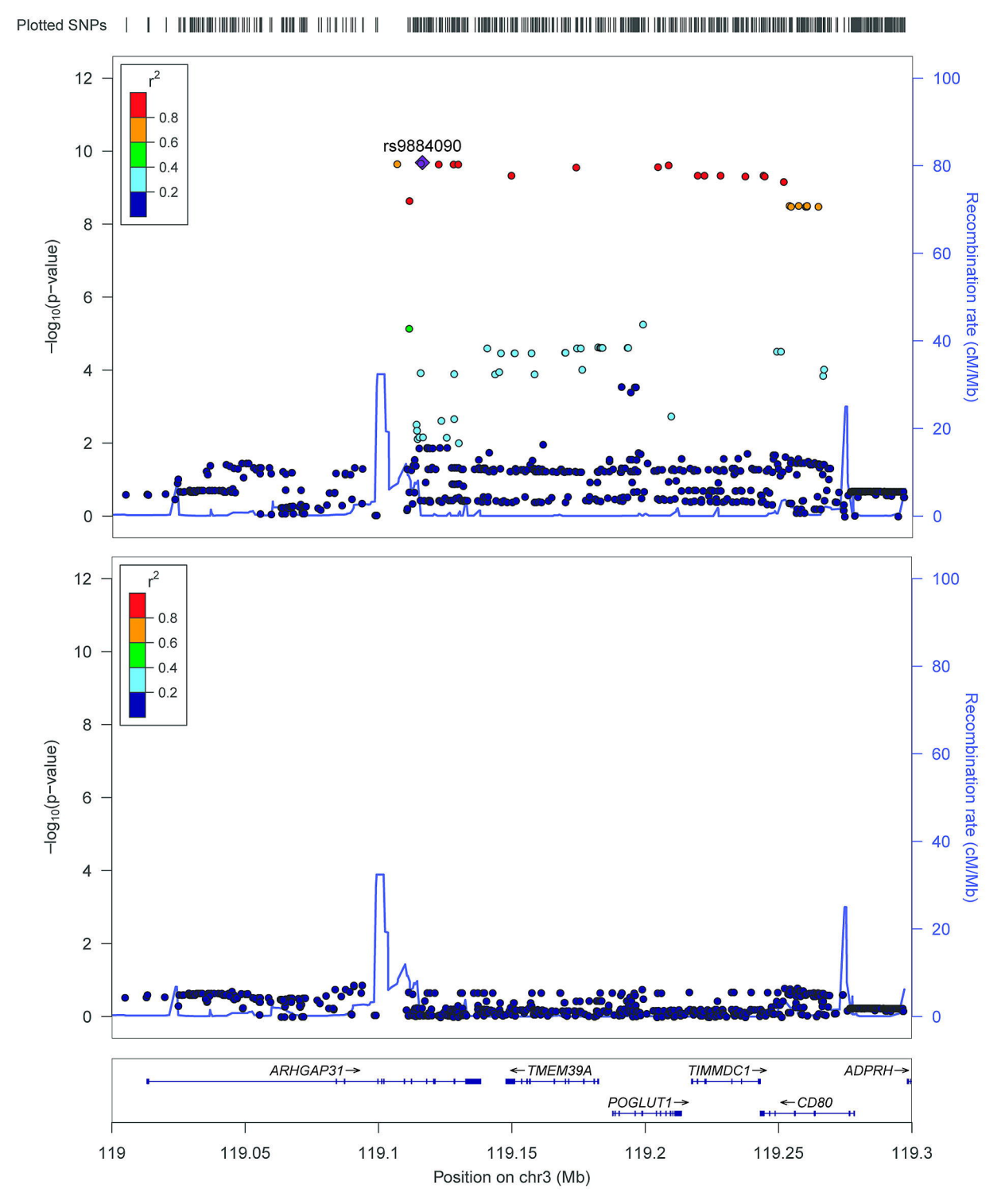
Figure 1 – Manhattan plot of genome wide association analysis comparing UK iTTP discovery cohort compared with controls. The X axis shows chromosome location, and the Y axis shows negative logarithmic p-values. Standardised genome wide significant $5x10^{-8}$ is depicted by the red line. The HLA peak is visualised on chromosome 6 (black), in addition to the novel chromosome 3 association (orange).

Figure 2 – Locus zoom plots of the chromosome 6 peak in the UK discovery cohort. The upper plot (a) shows the unconditioned analysis with the lead SNP rs28383233, and the middle plot (b) shows analysis conditioned on the lead SNP rs28383233, revealing independent association with rs1064994. The lower plot (c) shows analysis conditioned on both rs28323233 and rs1064994. Genomic positions refer to Human Assembly GRCh37/hg19.

Figure 3 – Locus zoom plots of the chromosome 3 peak in the UK discovery cohort. The upper plot (a) shows the unconditioned analysis and the lower plot (b) shows associations of the same markers when conditioned on the lead SNP, rs9884090. Genomic positions refer to Human Assembly GRCh37/hg19.


Regression -log10 P-value
25
20
15
10
5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22
Chromosome


SNPs
rs28383233
rs1064994
r2
0.8
0.6
0.4
0.2
-log10 (p-value)
Recombination rate (cMMb)
HLA-DRB6
HLA-DRB1
HLA-DQA1
HLA-DQB1
32.5
32.55
32.6
32.65
32.7
Position on chr6 (Mb)


Plotted SNPs
r2
0.8
0.6
0.4
0.2
rs9884090
−log10(p−value)
Recombination rate (cM/Mb)
ARHGAP31
TMEM39A
TIMMDC1
ADPRH
POGLUT1
CD80
119
119.05
119.1
119.15
119.2
119.25
119.3
Position on chr3 (Mb)

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

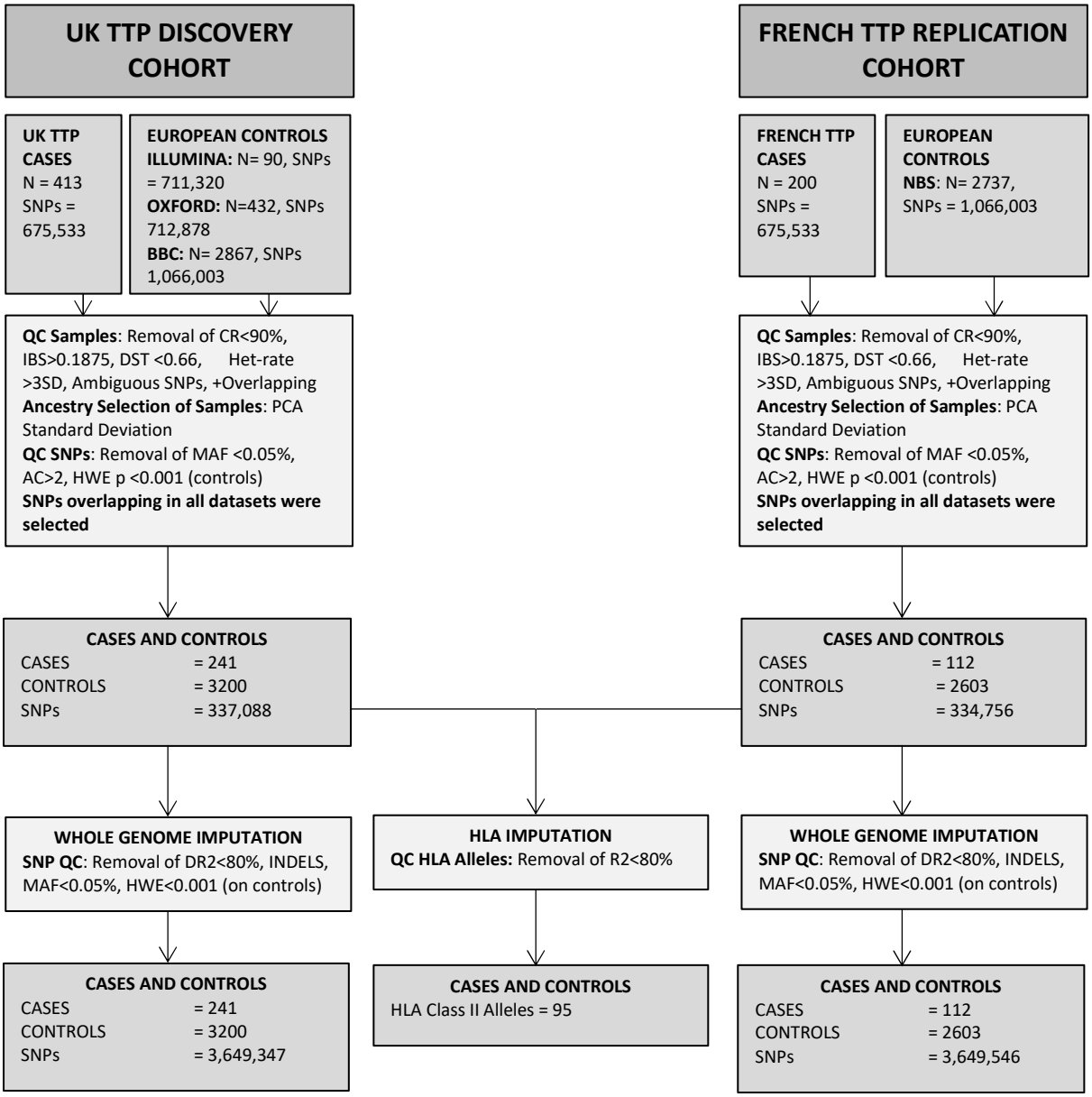
Binding sites of transcription factors that were identified through functional annotation in the region of interest, with potential functional importance were obtained from FactorBook.[24] Searching for specific cells lines (HEPG2) the position weight matrix (PWM) binding motifs of transcription factors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

MAST/MEME

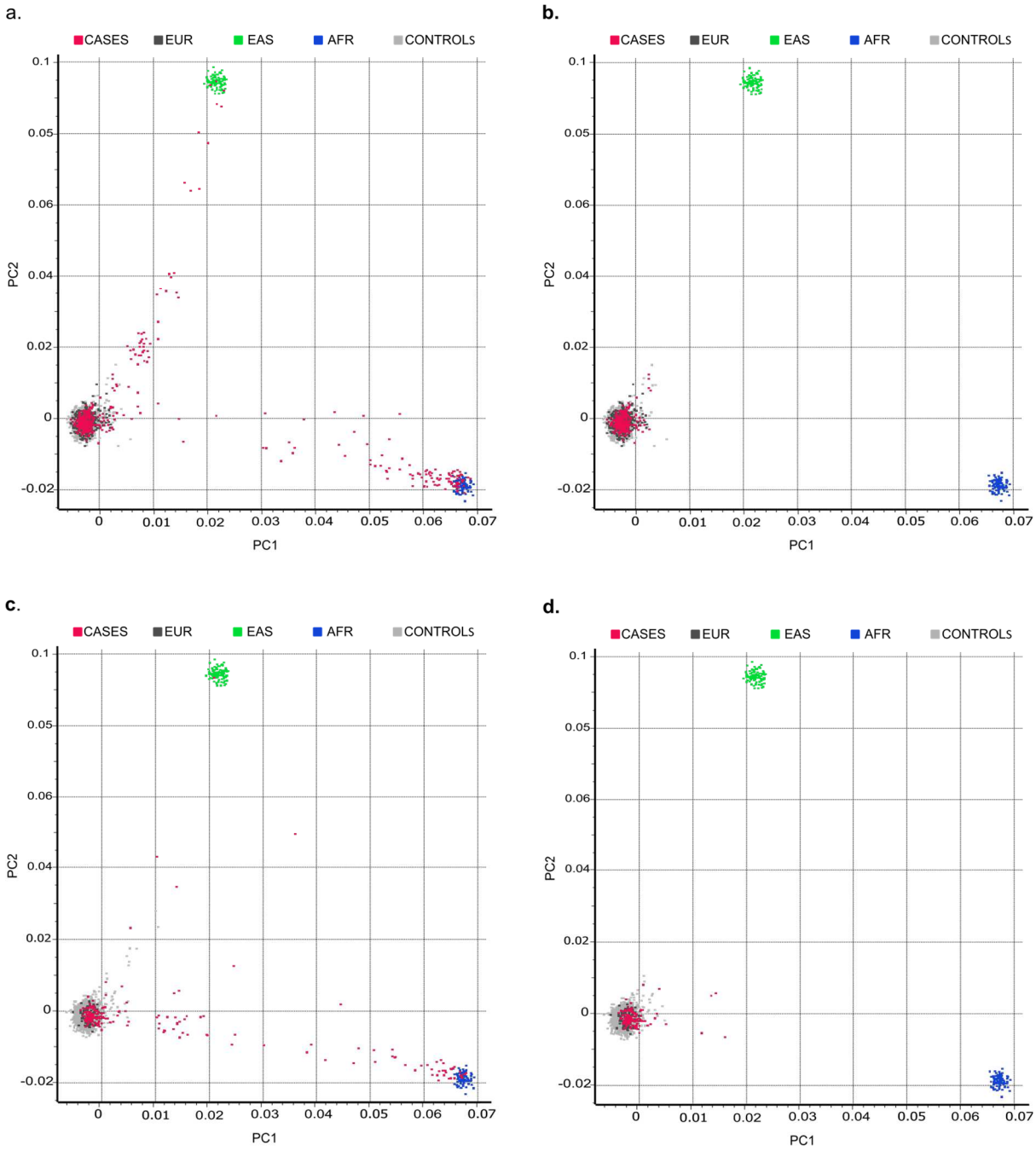
PWM binding motifs that were obtained from factor book were analysed along-side haploblock genetic variants, obtained from UCSC. A 80bp DNA sequence (40bp flanking, listed below) were analysed for potential DNA-transcription factor binding).[25]

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

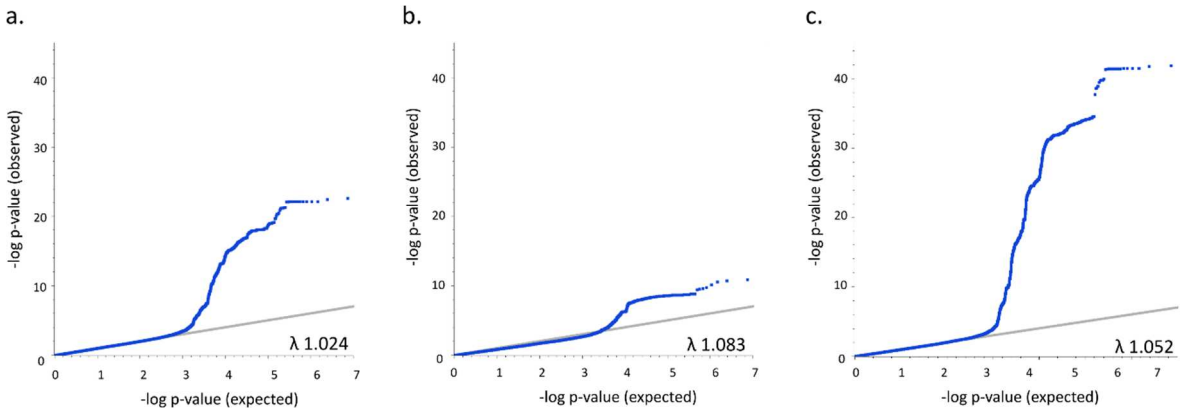
SUPPLEMENTARY FIGURES / RESULTS



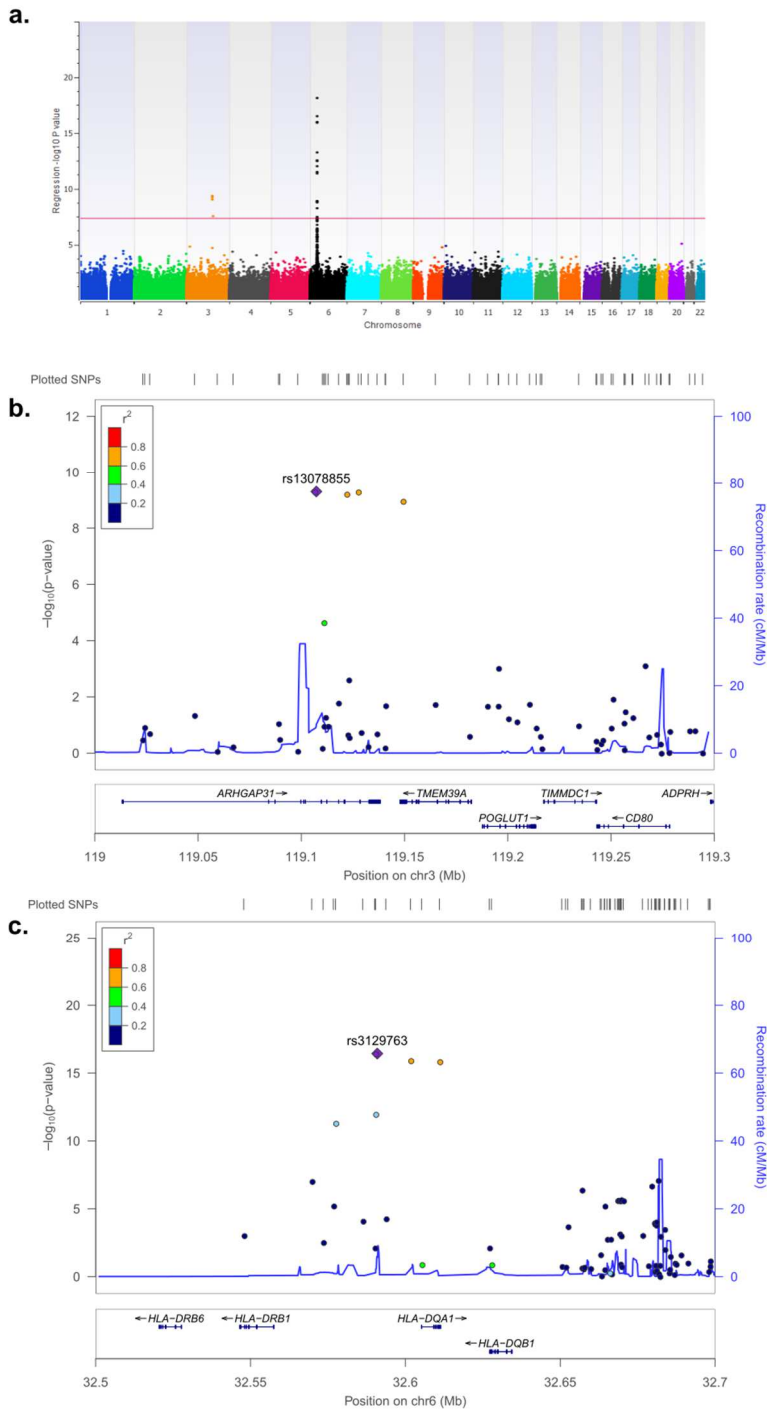
Supplemental Figure 1 – Summary of Quality Control in UK Discovery Cohort and French Replication Cohort.



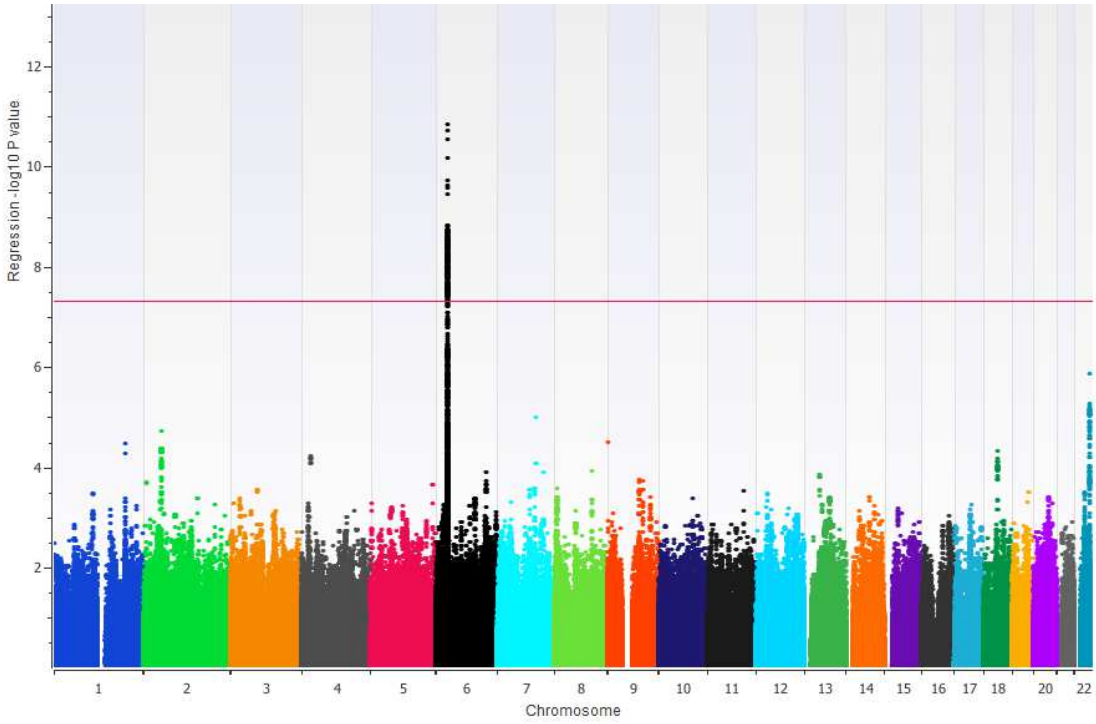
Supplemental Figure 2 – Principal Component Analysis in UK and French Cohorts. Cases are shown in red, and control genotypes in grey, and controls with known genetic ancestry are shown in black (EUR, European), blue (AFR, African) and green (EAS, East Asian). a. UK discovery cohort (without ethnicity 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.



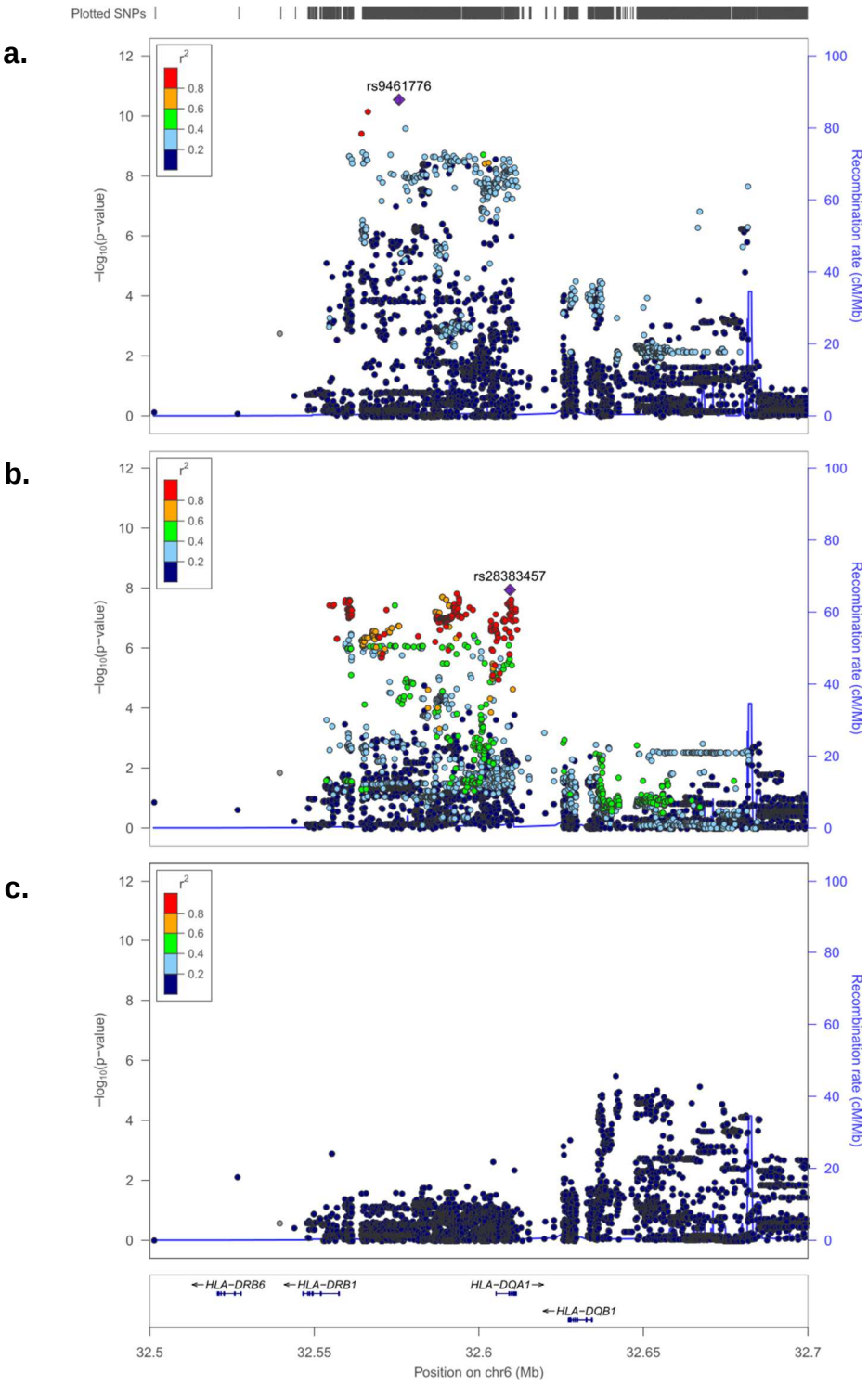
Supplemental Figure 3 – QQ plots, observed against expected p-values, for a. UK discovery population, b. French replication cohort, and c. Combined Analysis.



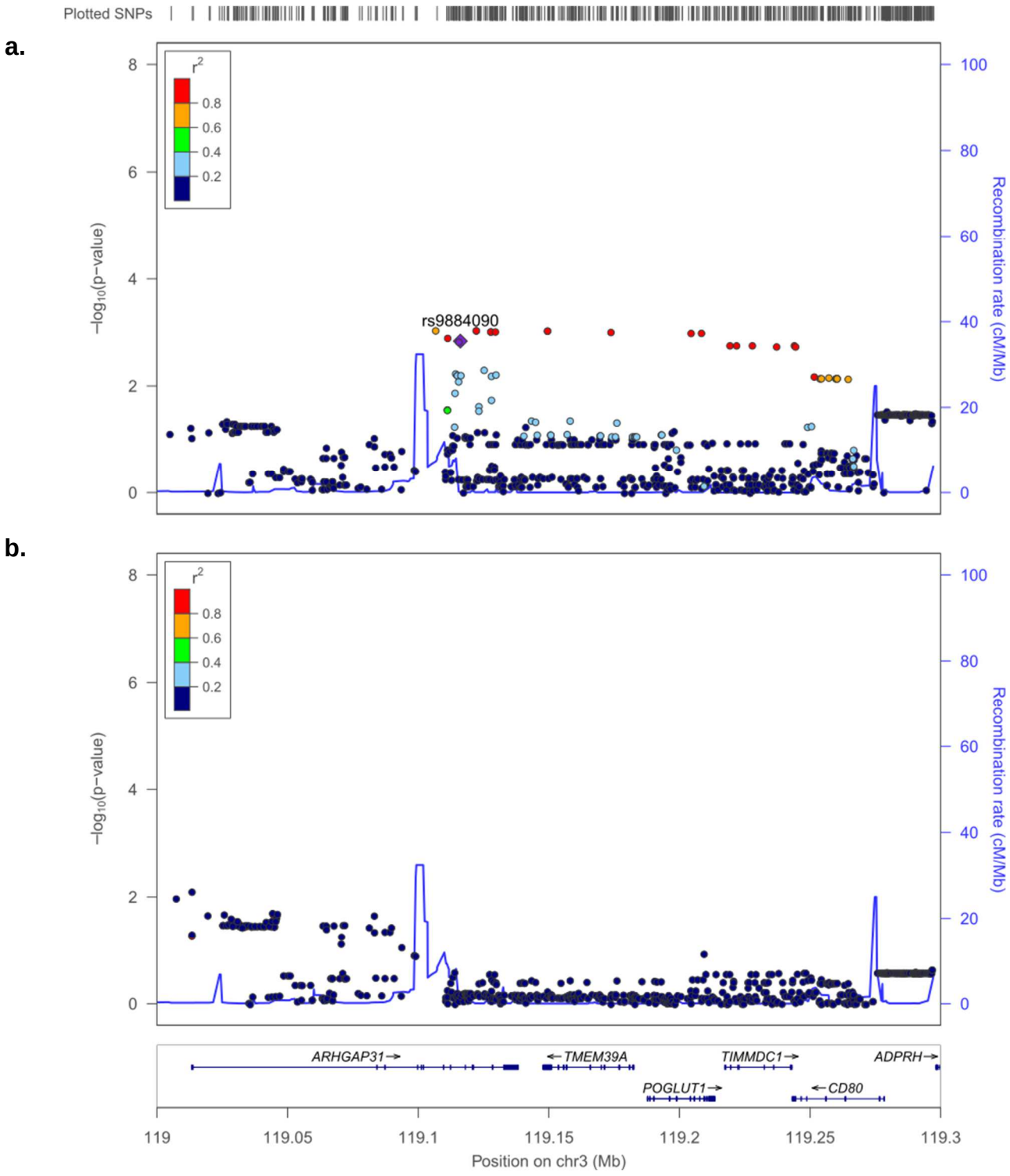
Supplemental Figure 4 – UK Cohort Directly Genotyped GWAS - (a) Manhattan plot of genome wide association test for directly genotyped SNPs only, comparing UK cases against controls, utilising the logistic regression method, corrected for top 10 principal components for stratification. The X axis shows 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).



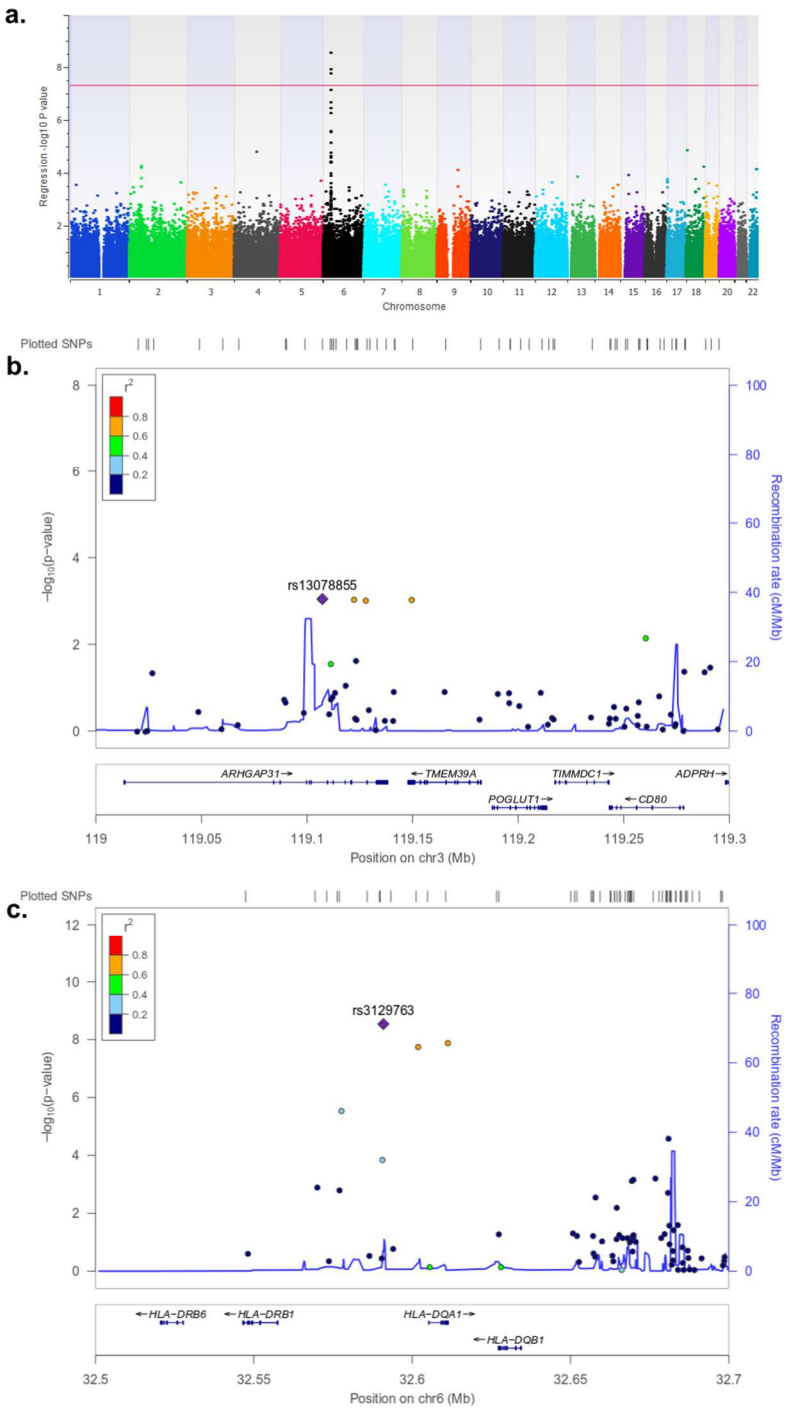
Supplemental Figure 5 – French Cohort Imputed GWAS – Manhattan plot of genome wide association test comparing French Replication cohort compared against controls, utilising the logistic regression method, corrected for top 10 principal components for stratification, in all imputed SNPs. The X axis shows 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).



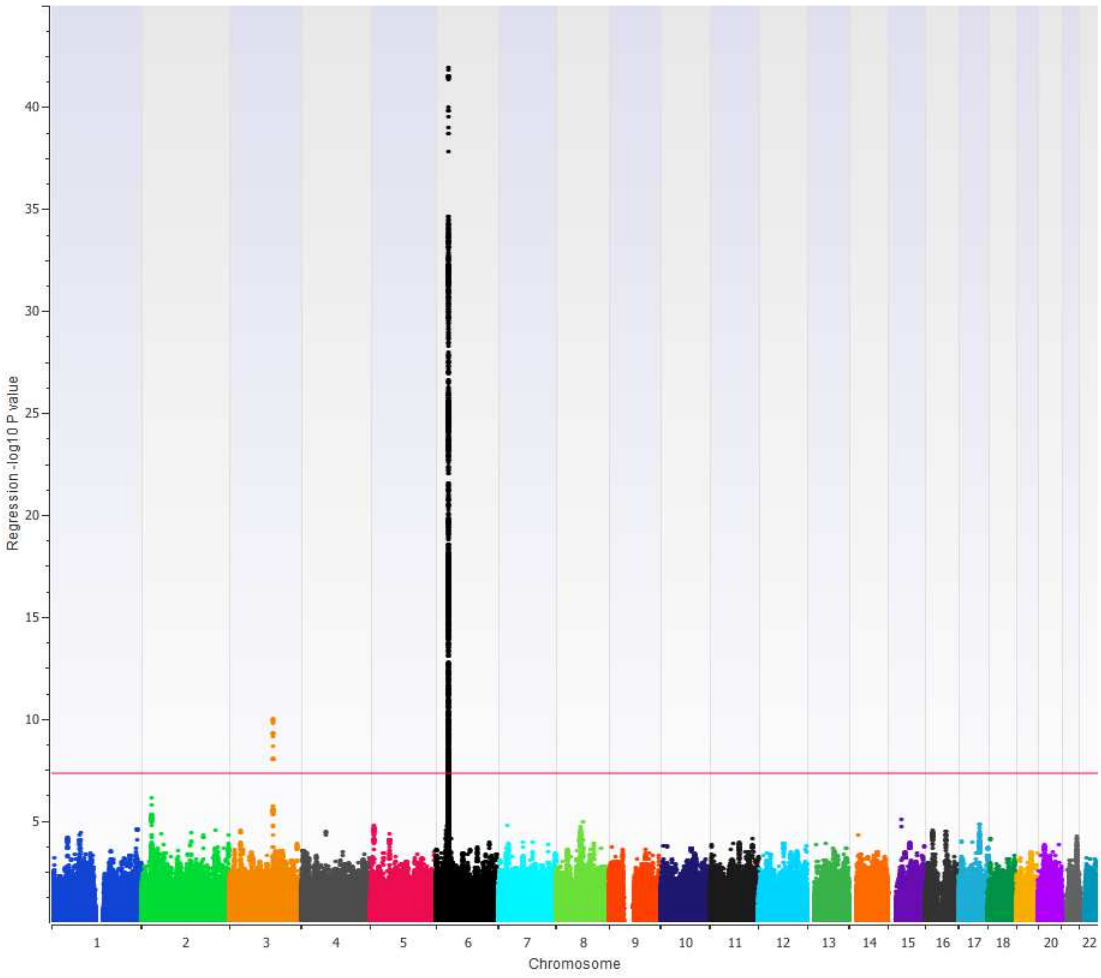
Supplemental Figure 6 – Locus zoom plot of the chromosome 6 peak in the French replication cohort. Genomic area displayed is 32.5 Mb to 32.7 Mb on chromosome 3 (Human Assembly GRCh37/hg19). The X axis shows chromosome location, and the left Y axis shows logarithmic p-values (logistic regression), and 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.