blood

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Genome-wide analysis defines genetic determinants of MPN subtypes and identifies a sex-specific association at CDH22/CD40

Tracking no: BLD-2025-028489R2

William Tapper (Southampton University, United Kingdom) Ahmed Dawoud (University of Southampton, United Kingdom) Joannah Score (University of Southampton,) Andrew Chase (University of Southampton, United Kingdom) E. Joanna Baxter (University of Cambridge, United Kingdom) Joanne Ewing (Heart of England NHS Foundation Trust, United Kingdom) Louise Wallis (Royal Bournemouth Hospital, United Kingdom) Paola Guglielmelli (University of Florence, Italy) Dolors Colomer (Hospital Clínic. IDIBAPS-CIBERONC, Spain) Beatriz Bellosillo (Hospital del Mar, Spain) Monste Gomez (Hematology Department, Hospital Clínico Universitario, Spain) Juan Carlos Hernandez-Boluda (Hospital Clínico Valencia, Spain) Carlos Besses (Hospital del Mar, Spain) Francisco Cervantes (IDIBAPS, Hospital Clinic, Spain) Steffen Koschmieder (RWTH Aachen University, Germany) Anthony Green (University of Cambridge, United Kingdom) Andreas Reiter (University Hospital Mannheim, Heidelberg University, Germany) Alessandro Vannucchi (Department of Experimental and Clinical Medicine, Centro di Ricerca e Innovazione Malattie Mieloproliferative (CRIMM), AOU Careggi, University of Florence, Italy) Claire Harrison (Guy's and St Thomas' NHS Foundation Trust, United Kingdom) Nicholas Cross (University of Southampton, United Kingdom)

Abstract:
To identify genetic variants that influence myeloproliferative neoplasm (MPN) phenotype, we undertook a two-stage case-only genome-wide association study using cohorts from the UK (including UK Biobank), Spain, Germany and Italy. MPN subtype [essential thrombocythemia (ET); polycythemia vera (PV)] were compared to each other, to healthy controls and stratified analyses was performed based on chromosome 9p aberrations, JAK2 V617F mutation burden and sex. The ET versus PV analysis identified known associations: (i) at HBS1L-MYB that increased ET risk (PMETA=7.93x10-6, OR=1.28) and reduced PV risk (PMETA=9.43x10-5, OR=0.81) and (ii) at GFI1B-GTF3C5 that predisposed to PV only (PMETA=1.43x10-9, OR=1.38). Two further linked intronic SNPs, rs2425786 and rs2425788, at CDH22/CD40 were significant in females only (PMETA=2.67x10-8) with predisposition to PV (PMETA=0.0006, OR=1.3) and reduction of ET risk (PMETA=7.82x10-5, OR=0.75). Associations with JAK2, TERT, ATM, TET2, PINT, GFI1B and SH2B3 were confirmed (PMETA<5x10-8) and nine further loci were replicated (PMETA<0.05). A polygenic risk score consisting of 48 SNPs from 31 loci demonstrated moderate discriminative performance for ET and PV (AUC=0.718) and was improved by optimization for disease subtype (AUCET=0.724 and AUCPV=0.755). Overall, our results reveal that multiple germline variants influence MPN phenotype with HBS1L-MYB and a novel sex-specific association with CDH22/CD40 being the strongest determinants.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: The study was designed and overseen by WJT and NCPC. Data analysis was performed by WJT and AAZD. JS and AJC prepared samples for genotyping. All other authors provided samples and/or clinical or laboratory data from their respective centers. The manuscript was drafted by WJT and all authors contributed to the final version.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: The array data generated in this study have been deposited at BioStudies (www.ebi.ac.uk/biostudies) under accession number S-BSST1772.

Clinical trial registration information (if any):

GENETIC DETERMINANTS OF MPN SUBTYPE

Context of Research

Some patients who acquire *JAK2* V617F develop polycythemia vera but others develop essential thrombocythemia. The reason for this phenotypic difference is incompletely understood

Aim of This Study

To identify genetic variants that influence myeloproliferative neoplasm (MPN) phenotype, including variants that have gender-specific effects

Findings

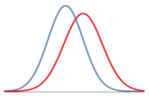
Genetic characterisation

- SNP array genotyping
- Imputation

Genome-wide association analyses

- ET versus PV
- ET versus controls
- PV versus controls

Polygenic Risk Score (PRS)



- Optimised for ET and PV

Female-specific association for variants within:

- *CDH22* (ET versus PV $P_{meta} = 2.67 \times 10^{-8}$)
  - Reduced ET risk ($P_{meta} = 7.82 \times 10^{-5}$, OR = 0.75)
  - Elevated PV risk ($P_{meta} = 0.0006$, OR = 1.30)
- eQTL for increased expression of *CD40* ($P = 3.80 \times 10^{-7}$)

PRS constructed from 48 SNPs in 31 independent loci

- Optmised PRS show stronger associations with ET and PV versus PRS based on platelets and red blood cell traits

Conclusion: Multiple germline variants influence MPN phenotype, including a novel female-specific association with *CDH22/CD40*.

*Tapper et al. DOI: 10.xxxx/**blood**.2025xxxxxx*

Genome-wide analysis defines genetic determinants of MPN subtypes and identifies a sex-specific association at *CDH22/CD40*

William J Tapper[1*], Ahmed A Z Dawoud[1], Joannah Score[1], Andrew J Chase[1], E Joanna Baxter[2], Joanne Ewing[3], Louise Wallis[4], Paola Guglielmelli[5], Dolors Colomer[6], Beatriz Bellosillo[7], Montse Gomez[8], Juan Carlos Hernández-Boluda[8], Carlos Besses[7], Francisco Cervantes[6], Steffen Koschmieder[9], Anthony R Green[2], Andreas Reiter[10], Alessandro Vannucchi[5], Claire Harrison[11], Nicholas C P Cross[1*]

[1] Faculty of Medicine, University of Southampton, Southampton, UK
[2] Department of Haematology, University of Cambridge, Cambridge, UK
[3] Department of Haematology, Birmingham Heartlands Hospital, Birmingham, UK
[4] Department of Haematology, Royal Bournemouth Hospital, Bournemouth, UK
[5] Center Research and Innovation of Myeloproliferative Neoplasms, DMSC, University of Florence, AOU Careggi, Florence, Italy
[6] Hospital Clinic, Institut Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Universitat Barcelona, Barcelona, Spain
[7] Hospital del Mar, Hospital del Mar Research Institute, Barcelona, Spain
[8] Hospital Clínico Universitario, INCLIVA, University of Valencia, Valencia, Spain,
[9] Department of Hematology, Oncology, Hemostaseology, and Stem Cell Transplantation, Faculty of Medicine, RWTH Aachen University, Aachen, Germany
[10] University Hospital Mannheim, Heidelberg University, Mannheim, Germany
[11] Department of Haematology, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

* Corresponding authors
William J Tapper: University of Southampton, Duthie Building (808), Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK.
wjt@soton.ac.uk

Nicholas C P Cross: Wessex Genomics Laboratory Service, Salisbury District Hospital, Salisbury SP2 8BJ, UK.
ncpc@soton.ac.uk

Data Sharing statement

The array data generated in this study have been deposited at BioStudies (www.ebi.ac.uk/biostudies) under accession number S-BSST1772.

Key Points

Genetic variation at HBS1L-MYB and CDH22/CD40 are the strongest determinant of MPN phenotype, but the latter is only seen in females

Polygenic risk scores for MPN are improved when optimised for disease subtype

Abstract

To identify genetic variants that influence myeloproliferative neoplasm (MPN) phenotype, we undertook a two-stage case-only genome-wide association study using cohorts from the UK (including UK Biobank), Spain, Germany and Italy. MPN subtype [essential thrombocythemia (ET); polycythemia vera (PV)] were compared to each other, to healthy controls and stratified analyses was performed based on chromosome 9p aberrations, *JAK2* V617F mutation burden and sex. The ET versus PV analysis identified known associations: (i) at *HBS1L-MYB* that increased ET risk (P_{META}=7.93x10^{-6}, OR=1.28) and reduced PV risk (P_{META}=9.43x10^{-5}, OR=0.81) and (ii) at *GFI1B-GTF3C5* that predisposed to PV only (P_{META}=1.43x10^{-9}, OR=1.38). Two further linked intronic SNPs, rs2425786 and rs2425788, at *CDH22/CD40* were significant in females only (P_{META}=2.67x10^{-8}) with predisposition to PV (P_{META}=0.0006, OR=1.3) and reduction of ET risk (P_{META}=7.82x10^{-5}, OR=0.75). Associations with *JAK2*, *TERT*, *ATM*, *TET2*, *PINT*, *GFI1B* and *SH2B3* were confirmed (P_{META}<5x10^{-8}) and nine further loci were replicated (P_{META}<0.05). A polygenic risk score consisting of 48 SNPs from 31 loci demonstrated moderate discriminative performance for ET and PV (AUC=0.718) and was improved by optimization for disease subtype (AUC_{ET}=0.724 and AUC_{PV}=0.755). Overall, our results reveal that multiple germline variants influence MPN phenotype with *HBS1L-MYB* and a novel sex-specific association with *CDH22/CD40* being the strongest determinants.

Introduction

Common, low penetrance genetic variants contribute to the risk of developing MPN and also phenotypic pleiotropy in these disorders[1-10]. In a prior genome-wide association study (GWAS), we found that genetic variation at *MECOM*, *TERT*, *JAK2* and *HBS1L-MYB* predisposes to *JAK2*-unmutated MPN[11]. Targeted analysis of these four variants demonstrated that rs9376092 at *HBS1L-MYB* and the *JAK2* 46/1 haplotype specifically influence whether *JAK2* V617F mutated cases present with PV or ET. It is likely that variation at other loci influence MPN phenotype and the primary aim of this study was to identify inherited genetic factors on a genome-wide basis that influence whether *JAK2* V617F positive MPN patients present with polycythemia vera (PV) or essential thrombocythemia (ET). Secondary aims were to explore gender effects and the efficacy of phenotype-specific polygenic risk scores.

Methods

We performed a two-stage case-only GWAS with 556 ET and 556 PV patients at stage 1, all *JAK2* V617F positive. Selected SNPs were tested for replication in four independent *JAK2* V617F positive stage 2 cohorts (ET, n=703; PV, n=715) plus MPN cases from UK Biobank (ET, n=322; PV, n=506) (Supplementary Table 1). ET or PV cases were compared to healthy controls and stratified analyses was performed based on chromosome 9p aberrations, *JAK2* V617F variant allele frequencies (VAF) and sex. Final effect sizes and significance levels were estimated by meta-analysis. Detailed methods and expanded results are in the Supplementary Material.

Results and Discussion

After quality control, a total of 7,267,872 SNPs (658,066 observed, 6,609,806 imputed) and 1069 patients (535 ET and 534 PV) remained for analysis at stage 1 (Supplementary Figure 1, Supplementary Table 1). ET and PV cases were compared using logistic regression and the first five principal components from multidimensional scaling to correct for population stratification (Supplementary Figure 2). Twenty nine genome-wide significant SNPs were identified ($P<5x10^{-8}$), however all but two were linked to the 46/1 *JAK2* haplotype[8] (Supplementary Figure 3).

We selected 93 SNPs for replication in a case only analysis using binary logistic regression to compare ET and PV; final significance levels and effect sizes were determined by a fixed effects inverse variance-weighted meta-analysis which combined evidence from the two stages. Two linked SNPs ($r^2=0.91$) with genome-wide significance were identified in the *HBS1L-MYB* intergenic region, rs9399137 ($P_{meta}=2.28x10^{-10}$) and rs9376092 ($P_{meta}=4.35x10^{-9}$). SNPs at four additional loci (*ZBTB7C*-

CTIF, *ADORA1*, *GFI1B-GTF3C5*, *LINC02398*) were identified with suggestive levels of significance (Table 1, Supplementary Table 2).

To determine if these six SNPs associate with MPN subtype, we compared ET or PV cases from stage 1 and UK Biobank against healthy controls from the WTCCC2 (n=5,195) and UK Biobank (n=326,027) and combined the evidence using a fixed effects meta-analysis. As summarised in Table 1, the two *HBS1L-MYB* SNPs and *ADORA1* SNP were associated with an increased risk of ET and reduced risk of PV. In contrast, variation at *GFI1B-GTF3C5* was only associated with an elevated risk of PV and, consistent with this finding, was significantly associated with 9p chromosome aberrations and *JAK2* V617F VAF (see Supplementary Material). Finally, variation at *LINC02398* and *ZBTB7C-CTIF* was associated with an increased risk of PV, with the latter also associated with a reduced risk of ET. These findings indicate a multifactorial genetic influence of constitutional genotype on MPN phenotype. The most significant association for each SNP is summarised in Figure 1.

To investigate the possibility of sex differences in SNP-disease associations, ET and PV cases from stage 1 and UK Biobank were stratified by gender and analysed against each other and controls. Two linked SNPs (r^2=1.0) within *CDH22*, rs2425786 in intron 5 and rs2425788 in intron 4, were identified with genome-wide significance (rs2425786 $P_{meta}=2.67x10^{-8}$, rs2425788 $P_{meta}=3.45x10^{-8}$) (Table 1, Supplementary Figure 4). In comparison with healthy female controls, these SNPs were associated with a reduced risk of ET (rs2425786 $P_{meta}=7.82x10^{-5}$, OR=0.75; rs2425788 P_{meta}=0.0001, OR=0.75) and an elevated risk of PV (rs2425786 P_{meta}=0.0006, OR=1.30; rs2425788 P_{meta}=0.0006, OR=1.29). While sex-related differences have previously been reported in MPN[12,13] this represents the first instance of a sex-specific genetic association with phenotypic predisposition.

CDH22 encodes cadherin 22, which is essential for maintaining the structure and function of several tissues, including the hematopoietic microenvironment[14]. However, *CDH22* does not appear to be expressed in hematopoietic cells and eQTL analysis indicates that rs2425786 is associated with increased expression of the neighbouring gene *CD40* ($P=3.80x10^{-7}$; Supplementary Material and Supplementary Table 3). CD40 is expressed in hematopoietic cells and encodes a cell surface receptor belonging to the tumour necrosis factor receptor superfamily. Consequently, it is a potential candidate that merits further investigation.

The mechanism underlying the female-specific effect of rs2425786 is unclear, but it may involve hormonal influences, differential gene regulation, or sex-specific immune modulation. We used data

from UK Biobank to evaluate whether the effects of the *CDH22/CD40* SNPs were mediated by or interacted with hormonal biomarkers (sex hormone binding globulin [SHBG] and testosterone [TT]) or the inflammatory biomarker C-reactive protein (CRP). The SNPs were associated with a reduced risk of ET (rs2425786 P_{CRP}=0.0016, OR=0.69; rs2425786 P_{TT}=0.0051, OR=0.68) and an increased risk of PV (rs2425786 P_{CRP}=0.0264, OR=1.32; rs2425786 P_{TT}=0.0459, OR=1.31), independently of CRP (Supplementary Table 4) and testosterone (Supplementary Table 5), with no evidence of significant interactions. Adjustment for SHBG did not attenuate the SNPs associations for ET versus PV (rs2425786 P_{SHBG}=0.0017, OR=0.56) and ET versus controls (rs2425786 P_{SHBG}=0.0065, OR=0.71), and no significant interactions were observed (Supplementary Table 6). A similar trend towards increased risk of PV was shown, although it did not reach nominal significance (rs2425786 P_{SHBG}=0.0656, OR=1.27).

To further investigate potential sex-linked biological pathways, we reviewed phenome-wide association study results which revealed a significant association between rs2425786 and complications of labour and delivery (OR=0.95, P=1.48x10^{-4})[15], suggesting a possible link to female-specific physiological processes. Some genes are differently regulated in males and females due to differences in the epigenetic landscape. Interestingly, aberrant demethylation of the promoter region of *CD40LG*, which encodes the CD40 ligand, on the inactive X chromosome can lead to biallelic expression in females. This abnormal expression pattern has been linked to a higher prevalence of immune-related diseases[16,17] and elevated levels of IgM in females[18]. This female-specific mechanism may be relevant to the observed association between *CDH22/CD40* SNPs and increased risk of PV in women, and we plan to explore this using bulk and single cell methylation/expression analysis in relation to rs2425786 genotype and MPN phenotype

To estimate an individual's genetic risk for developing MPN, and specifically ET or PV, we calculated three polygenic risk scores (PRS_{MPN}, PRS_{ET}, PRS_{PV}) using 48 SNPs (Supplementary Table 7). The PRS_{MPN} exhibited moderate performance in UK Biobank, achieving an AUC value of 0.635 which increased to 0.718 when covariates for age, sex and ancestry (first 10 principal components) were included (Figure 2). Individuals with scores in the highest decile were estimated to have a 4.88-fold increased risk of MPN versus those in the lowest decile. The PRS_{ET} and PRS_{PV} showed a slight improvement with an AUC of 0.724 for ET and 0.755 for PV, respectively, when adjusting for covariates. The relative risk of disease associated with scores in the top versus bottom decile were 5.78 for ET and 4.66 for PV.

In a recent study, Guo et al 2024[19] showed that a PRS for platelet traits in healthy individuals (pct and plt) were associated with ET and that a PRS for red blood cells (hgb, hct, rbc and mchc) were risk factors for PV. An additional PRS consisting of MPN-associated SNPs also increased the risk of ET and PV, but to a lesser extent. We computed PRS for the six blood cell traits using all available SNPs (Supplementary Table 8) and our tailored PRS (PRS_{ET} and PRS_{PV}) and assessed their relationship with ET and PV in UK Biobank. We confirmed the association of platelet traits with ET (PRS_{pct} $P_{fdr}=7.16x10^{-17}$, OR=1.63; PRS_{plt} $P_{fdr}=6.08x10^{-14}$, OR=1.54) and red blood cell traits with PV (PRS_{hgb} $P_{fdr}=5.09x10^{-16}$, OR=1.47; PRS_{hct} $P_{fdr}=1.58x10^{-13}$, OR=1.42; PRS_{rbc} $P_{fdr}=7.01x10^{-11}$, OR=1.38; PRS_{mchc} $P_{fdr}=1.17x10^{-3}$, OR=1.18) using univariable logistic regression (Supplementary Table 9). However, our tailored PRS had the strongest association with a diagnosis of ET (PRS_{ET} $P=1.92x10^{-16}$, OR=1.58) and PV (PRS_{PV} $P=7.62x10^{-18}$, OR=1.48) using multivariable logistic regression and correcting for either platelet traits with ET or red blood cell traits with PV along with age, sex, *JAK2* V617F VAF and 10 principal components (Supplementary Table 10).

According to the per allele odds ratio and minor allele frequency, rs2425786 (*CDH22/CD40)* is estimated to account for the largest proportion of the population attributable fraction (19.6%) followed by rs9399137 (*HBS1L-MYB*; 9.7%). The intergenic SNP between *GFI1B* and *GTF3C5*, rs3011271, accounts for a further 6.3% of the PAF. Based on a multiplicative model without interaction, these three genetic risk factors are estimated to have a combined PAF of 32% (Supplementary Table 11) indicating that they play a substantial role in influencing MPN phenotype.

Our findings highlight the importance of considering the possibility of gender-specific effects in studies that explore the connection between genetic variation and patient phenotype, and this may extend beyond presenting features to clinical management issues such as adverse events and outcomes following treatment.

Acknowledgements

Part of this research has been conducted using the UK Biobank Resource under Application Number 35273.

Authorship Contributions

The study was designed and overseen by WJT and NCPC. Data analysis was performed by WJT and AAZD. JS and AJC prepared samples for genotyping. All other authors provided samples and/or clinical or laboratory data from their respective centers. The manuscript was drafted by WJT and all authors contributed to the final version.

Disclosure of Conflicts of Interest

None of the authors declare any relevant conflicts of interest

References

1. Bao EL, Nandakumar SK, Liao X, et al. Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature*. 2020;586(7831):769-775.
2. Loh PR, Genovese G, McCarroll SA. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature*. 2020;584(7819):136-141.
3. Grinfeld J, Nangalia J, Baxter EJ, et al. Classification and Personalized Prognosis in Myeloproliferative Neoplasms. *N Engl J Med*. 2018;379(15):1416-1430.
4. Trifa AP, Bănescu C, Bojan AS, et al. MECOM, HBS1L-MYB, THRB-RARB, JAK2, and TERT polymorphisms defining the genetic predisposition to myeloproliferative neoplasms: A study on 939 patients. *Am J Hematol*. 2018;93(1):100-106.
5. Hinds DA, Barnholt KE, Mesa RA, et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood*. 2016;128(8):1121-1128.
6. Jones AV, Campbell PJ, Beer PA, et al. The JAK2 46/1 haplotype predisposes to MPL-mutated myeloproliferative neoplasms. *Blood*. 2010;115(22):4517-4523.
7. Olcaydu D, Harutyunyan A, Jäger R, et al. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet*. 2009;41(4):450-454.
8. Jones AV, Chase A, Silver RT, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet*. 2009;41(4):446-449.
9. Kilpivaara O, Mukherjee S, Schram AM, et al. A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nat Genet*. 2009;41(4):455-459.
10. Pardanani A, Fridley BL, Lasho TL, Gilliland DG, Tefferi A. Host genetic variation contributes to phenotypic diversity in myeloproliferative disorders. *Blood*. 2008;111(5):2785-2789.
11. Tapper W, Jones AV, Kralovics R, et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat Commun*. 2015;6:6691.
12. Moliterno AR, Braunstein EM. The roles of sex and genetics in the MPN. *Int Rev Cell Mol Biol*. 2022;366:1-24.
13. Karantanos T, Chaturvedi S, Braunstein EM, et al. Sex determines the presentation and outcomes in MPN and is related to sex-specific differences in the mutational burden. *Blood Adv*. 2020;4(12):2567-2576.
14. Cao ZQ, Wang Z, Leng P. Aberrant N-cadherin expression in cancer. *Biomed Pharmacother*. 2019;118:109320.
15. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. 2018;50(11):1593-1599.
16. Liao J, Liang G, Xie S, et al. CD40L demethylation in CD4(+) T cells from women with rheumatoid arthritis. *Clin Immunol*. 2012;145(1):13-18.
17. Lian X, Xiao R, Hu X, et al. DNA demethylation of CD40l in CD4+ T cells from women with systemic sclerosis: a possible explanation for female susceptibility. *Arthritis Rheum*. 2012;64(7):2338-2345.
18. Lleo A, Liao J, Invernizzi P, et al. Immunoglobulin M levels inversely correlate with CD40 ligand promoter methylation in patients with primary biliary cirrhosis. *Hepatology*. 2012;55(1):153-160.
19. Guo J, Walter K, Quiros PM, et al. Inherited polygenic effects on common hematological traits influence clonal selection on JAK2(V617F) and the development of myeloproliferative neoplasms. *Nat Genet*. 2024;56(2):273-280.

Locus	SNP	Fixed effect meta-analysis*																	
		ET vs PV (6)		ET vs controls (2)		PV vs controls (2)		ET/PV vs controls (2)		9p aUPD/CNG vs controls (1)		*JAK2* V617F VAF (3)		ET vs PV females (2)		ET vs control females (2)		PV vs control females (2)	
		P	OR	P	OR	P	OR	P	OR	P	OR	P	BETA	P	OR	P	OR	P	OR
HBS1L-MYB	rs9399137	**2.28×10^{-10}**	1.47	7.93×10^{-6}	**1.28**	9.43×10^{-5}	**0.81**	0.2967	1.04	0.2928	0.90	0.0025	-0.111	1.99×10^{-7}	1.78	0.0001	1.31	0.0002	0.71
	rs9376092	**4.35×10^{-9}**	1.41	2.27×10^{-7}	**1.32**	0.0049	**0.86**	0.1609	1.06	0.5493	0.95	0.0043	-0.102	1.75×10^{-6}	1.67	1.08×10^{-5}	1.36	0.0089	0.80
ZBTB7C-CTIF	rs8087061	†**1.67×10^{-6}**	0.54	0.0028	**0.74**	0.0005	**1.31**	0.6180	1.03	0.0431	1.32	†0.0658	0.151	0.0086	0.61	0.0137	0.72	0.3855	1.12
ADORA1	rs3766568	†**3.99×10^{-5}**	1.34	0.0030	**1.17**	0.0031	**0.86**	0.8722	0.99	0.9935	1.00	†0.02302	-0.056	0.0042	1.34	0.0203	1.17	0.0719	0.86
LINC02398	rs2244740	**7.06×10^{-5}**	0.61	0.1106	0.80	0.0013	**1.38**	0.2538	1.10	0.0715	1.38	0.0721	0.133	0.1017	0.66	0.1938	0.80	0.3226	1.18
GFI1B-GTF3C5	rs3011271	4.77×10^{-5}	0.78	0.7076	1.02	**1.43×10^{-9}**	**1.38**	3.57×10^{-6}	1.21	3.44×10^{-9}	1.71	2.35×10^{-8}	0.207	0.0086	0.73	0.8167	1.02	0.0004	1.35
	rs520812	0.0111	0.83	0.8614	0.99	**1.22×10^{-6}**	**1.34**	0.0007	1.18	1.28×10^{-6}	1.63	0.0002	0.159	0.0251	0.73	0.5725	0.95	0.0317	1.24
FAM135B	rs12550019	0.0419	0.90	0.9332	1.00	**2.48×10^{-5}**	**1.22**	0.0018	1.12	0.0009	1.31	0.2214	0.039	0.1703	0.87	0.7929	1.02	0.0494	1.16
CDH22	rs2425786	†3.93×10^{-5}	0.75	0.0364	0.89	0.0024	1.15	0.3591	1.03	0.0998	1.15	†0.3328	0.045	**2.67×10^{-8}**	0.56	7.82×10^{-5}	**0.75**	0.0006	**1.30**
	rs2425788	†4.60×10^{-5}	0.75	0.0333	0.89	0.0030	1.15	0.3968	1.03	0.1124	1.14	†0.3667	0.042	**3.45×10^{-8}**	0.56	0.0001	**0.75**	0.0006	**1.29**

Table 1. Summary of the most significant SNPs following meta-analysis.

Locus, HGNC gene symbol with flanking genes shown for intergenic SNPs; SNP, rs identifier from dbSNP; Fixed effect meta-analysis was used to generate significance levels (P) and effect sizes (OR or BETA) except for 9p aUPD/CNG which was only available in the stage 1 case control cohort. Comparative groups or trait investigated are shown by column titles and the number of independent cohorts used for meta-analysis is shown in parentheses. †rs8087061 and rs3766568 failed replication QC (HWE $P<1E^{-10}$ and call rate <90% respectively) while the *CDH22* SNPs (rs2425786 and rs2425786) were not selected for replication genotyping. As a result, these SNPs are only tested in two cohorts for the ET vs PV analysis and one cohort for association with *JAK2* V617F. The most significant P-value across all analyses is highlighted in bold. Odds ratios (OR) in bold highlight the most significant subtype-specific associations in comparisons of either ET or PV cases with controls. SNPs associated with both subtypes have two bolded ORs, while those associated with only one subtype have a single bolded OR. Numbers in brackets indicate the number of cohorts tested for each comparison.

Figure legends

Figure 1. Forest plot and meta-analysis for the most significant SNPs. Forest plots showing the odds ratios, 95% confidence intervals (CI), percentage weight contributed to the overall meta-analysis and p-value for each SNP with or approaching a genome-wide level of significance. The most significant association for each SNP with a genome-wide or suggestive level of significance is shown. Odds ratios greater than 1 for the ET vs PV comparison indicate an increased risk of PV while those less than 1 increase the risk of ET. The SNP subtotals show the OR and CI for a fixed-effects meta-analysis; Cochran's Q test and I^2 statistics showed that for each SNP there was no evidence of heterogeneity between cohorts. Each SNP is significant in at least one of the replication cohorts tested and has evidence for the same trend in the remaining populations. GWAs significant P-values are highlighted in bold.

Figure 2. Evaluation of PRS optimised for disease subtype. Panels represent PRS optimised for ET and PV cases (A) ET cases (B) and PV cases (C). Density plots compare the distribution of Z-scaled PRS in cases and controls. Receiver operating characteristic curves showing the predictiveness of the PRS alone or with covariates (age, sex and first 10 principal components). Decile plots of relative disease risk in each decile versus the lowest decile.

Figure 1

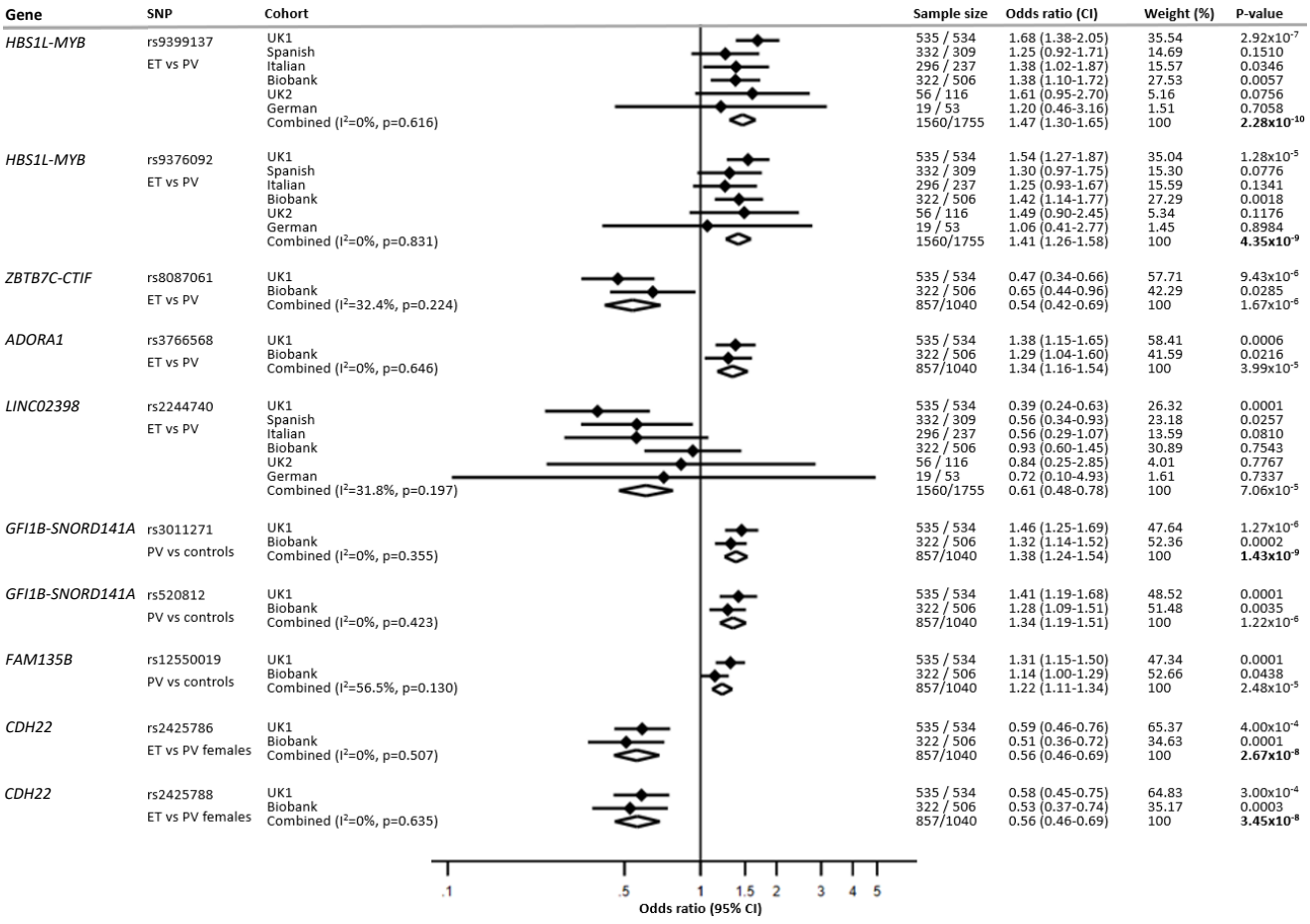
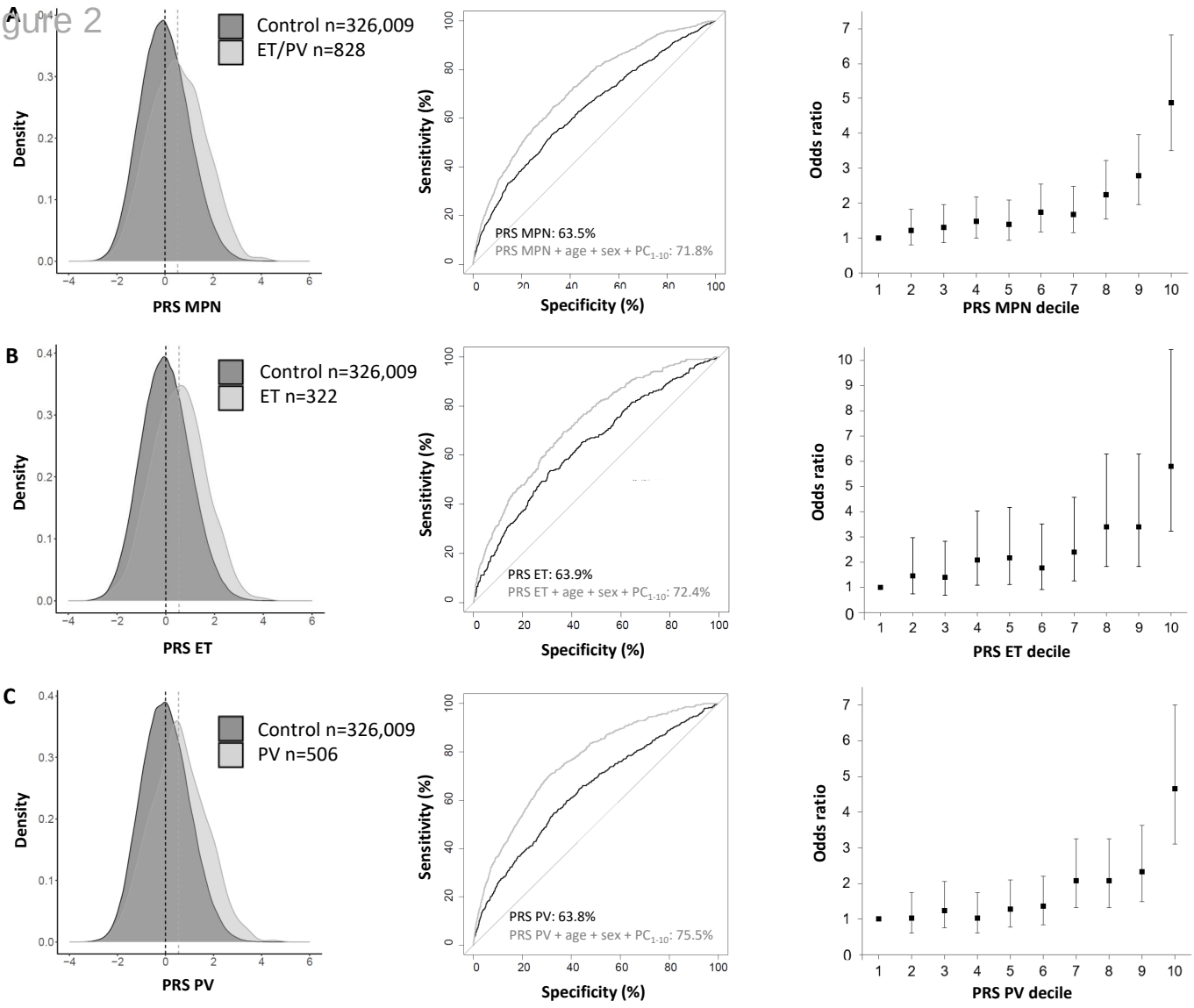


Figure 1

Figure 2



Figure