Published in final edited form as:
Nat Genet. 2009 April ; 41(4): 455–459. doi:10.1038/ng.342.

A germline *JAK2* SNP is associated with predisposition to the development of *JAK2*^{V617F}-positive myeloproliferative neoplasms

Outi Kilpivaara[1,12], **Semanti Mukherjee**[2,3,12], **Alison M Schram**[1], **Martha Wadleigh**[4], **Ann Mullally**[4,5], **Benjamin L Ebert**[5,6], **Adam Bass**[4,6], **Sachie Marubayashi**[1], **Adriana Heguy**[1], **Guillermo Garcia-Manero**[7], **Hagop Kantarjian**[7], **Kenneth Offit**[8], **Richard M Stone**[4], **D Gary Gilliland**[4,5,6,9,10], **Robert J Klein**[2], and **Ross L Levine**[1,11]

[1]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York, USA

[2]Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York, USA

[3]Gerstner Sloan-Kettering Graduate School of Biomedical Sciences, New York, New York, USA

[4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

[5]Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA

[6]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

[7]Department of Leukemia, M.D. Anderson Cancer Center, Houston, Texas, USA

[8]Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA

[9]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA

[10]Harvard Stem Cell Institute, Boston, Massachusetts, USA

[11]Leukemia Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA

Abstract

Polycythemia vera, essential thrombocythemia and primary myelofibrosis are myeloproliferative neoplasms (MPN) characterized by multilineage clonal hematopoiesis[1–5]. Given that the identical somatic activating mutation in the *JAK2* tyrosine kinase gene (*JAK2*^{V617F}) is observed in most individuals with polycythemia vera, essential thrombocythemia and primary myelofibrosis[6–10], there likely are additional genetic events that contribute to the pathogenesis of these phenotypically distinct disorders. Moreover, family members of individuals with MPN are at higher risk for the development of MPN, consistent with the existence of MPN predisposition

Correspondence should be addressed to R.J.K. (kleinr@mskcc.org) or R.L.L. (leviner@mskcc.org).

[12]These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

The study was designed by O.K., S. Mukherjee, R.J.K. and R.L.L. with advice from K.O. SNP arrays were performed and analyzed by A.B., B.L.E. and R.L.L, and analysis of SNP array data for modifier and predisposition loci was performed by S. Mukherjee and R.J.K. Genotyping, sequence analysis and realtime PCR assays were performed by O.K., A.M.S., S. Marubayashi, A.H. and R.L.L. Principal component analysis was done by S. Mukherjee and R.J.K. Identification of subjects, sample collection and phenotypic assessment were done by M.W., A.M., G.G.-M., H.K., R.M.S, D.G.G. and R.L.L. The paper was written by O.K., S. Mukherjee, K.O., D.G.G., R.J.K. and R.L.L. All authors discussed the results and commented on the manuscript.

loci[11]. We hypothesized that germline variation contributes to MPN predisposition and phenotypic pleiotropy. Genome-wide analysis identified an allele in the *JAK2* locus (rs10974944) that predisposes to the development of *JAK2*^{V617F}-positive MPN, as well as three previously unknown MPN modifier loci. We found that *JAK2*^{V617F} is preferentially acquired in cis with the predisposition allele. These data suggest that germline variation is an important contributor to MPN phenotype and predisposition.

The presence of the identical *JAK2*^{V617F} allele in polycythemia vera, essential thrombocythemia and primary myelofibrosis suggests that there are additional genetic and epigenetic events that contribute to MPN pathogenesis. To date, most studies have focused on the identification of additional somatic events that are stochastically acquired by the MPN clone, whereas few studies have addressed the role of germline genetic variation in MPN pathogenesis. A recent analysis of 32 candidate SNPs in MPN tumor samples identified three SNPs in *JAK2* that were enriched in polycythemia vera[12]; although these results suggest there are host genetic variants that influence MPN phenotype, the findings are likely influenced by the high rate of somatic isodisomy at the *JAK2* locus in polycythemia vera. It has also been observed that there is familial clustering in MPN cases[13–15], suggesting that there are inherited MPN predisposition loci. We therefore used genome-wide SNP array data to identify MPN modifier and predisposition loci.

We first analyzed Affymetrix *StyI* SNP array data derived from granulocyte DNA from 181 subjects with polycythemia vera or essential thrombocythemia to identify loci significantly enriched in these conditions. We identified four loci with P values < 10^{-5}, including SNPs at three loci not previously implicated in either disease (Table 1). We confirmed that the minor allele at rs12500918 is more common in polycythemia vera than in essential thrombocythemia in a larger set of MPN samples (P = 0.01). Unbiased genome-wide analysis also suggested that germline variation at the *JAK2* locus (rs10974944) varied according to MPN phenotype. However, given acquired isodisomy leading to homozygosity for *JAK2*^{V617F} is more common in polycythemia vera than in essential thrombocythemia, we were concerned that individuals with poly-cythemia vera might have a spuriously high prevalence of the minor allele (G) owing to somatic alterations at the *JAK2* locus[6–8,10,16]. To exclude the effects of somatic variation at the *JAK2* locus, we analyzed germline DNA samples from 284 subjects with polycythe-mia vera or essential thrombocythemia for *JAK2* SNP rs10974944 and confirmed that the minor allele (G) was significantly more common in polycythemia vera than in essential thrombocythemia (P = 0.01).

We also noted that the frequency of the GG/CG genotypes at rs10974944 was much higher in MPN cases than in control populations, suggesting that rs10974944 is a MPN predisposition allele. Analysis of germline DNA from 324 subjects with polycythemia vera, essential thrombocythemia and primary myelofibrosis found that the GG/CG genotypes at rs10974944 were more common in MPN cases compared to WTCCC controls (OR = 3.1, P = 4.1×10^{-20}; Table 2), consistent with the G allele functioning as a dominant MPN predisposition allele. These data indicate that germline *JAK2* variation more strongly influences MPN predisposition than MPN phenotype. In contrast, allelic variation at rs12500918 is associated with MPN phenotype (P = 0.01) but not with MPN predisposition (P = 0.24) (Supplementary Table 1 online).

We then carried out principal component analysis of SNP array data derived from MPN samples and WTCCC controls to assess whether population substructure or ancestry might explain differences in rs10974944 allele frequency. We selected case and control individuals who cluster on the first two principal components consistent with ancestry from Northern and Western Europe (Fig. 1). We observed a similar relationship between genotypic variation at *JAK2* and MPN predisposition in these matched cases and controls, suggesting

that differences in the frequency of rs10974944 genotypes are not due to differences in population substructure. Moreover, the distribution of rs10974944 genotypes did not vary in eight different ancestry populations (Supplementary Fig. 1a online), or in myelodysplasia (P = 0.23) or acute myeloid leukemia (P = 0.10) samples compared to control samples (Supplementary Fig. 1b).

Given the high frequency of somatic mutations at *JAK2* in MPN[6–10,17], we hypothesized that rs10974944 germline variation might specifically predispose to *JAK2*-mutated MPN. We assessed rs10974944 genotype in 321 MPN cases that had been genotyped for $JAK2^{V617F}$ and for *JAK2* exon 12 mutations[17] and found that germline allelic variation at rs10974944 was strongly associated with $JAK2^{V617F}$-positive MPN (OR = 4.0, P = 7.7×10^{-22}) and much less strongly associated with $JAK2^{V617F}$-negative MPN (OR = 1.6, P = 0.06) (Table 3). Allelic variation at rs10974944 was strongly associated with predisposition to polycythemia vera (OR = 4.3, P = 1.0×10^{-16}) and essential thrombocythemia (OR = 2.1, P = 6.7×10^{-5}) (Supplementary Table 2 online). The higher odds ratio associated with allelic variation at rs10974944 in polycythemia vera is in part due to the higher incidence of *JAK2* mutations in this condition (95%) compared to essential thrombocythemia (65%), as shown by the higher association between rs10974944 genotype and $JAK2^{V617F}$-positive essential thrombocythemia (OR = 2.8, P = 3.0×10^{-5}).

Analysis of the haplotype structure of the *JAK2* locus in CEPH founders (Fig. 2a) shows that rs10974944 and $JAK2^{V617}$ are contained in a common haplotype block distinct from the promoter and 5′ exons of *JAK2.* This led us to hypothesize that the G allele at rs10974944 might predispose to somatic $JAK2^{V617F}$ mutations on the same strand. We investigated 42 subjects heterozygous for rs10974944 in their germline and a somatic homozygous $JAK2^{V617F}$ mutant clone ($JAK2^{V617F}$ allele burden > 50%) and found in 38 of 42 cases somatic conversion to a homozygous GG genotype at rs10974944 (Supplementary Fig. 2 online). Using allele-specific PCR on 45 subjects heterozygous for rs10974944 in their germline and hetero-zygous for $JAK2^{V617F}$ in their granulocyte DNA, we found that in 38 cases $JAK2^{V617F}$ was acquired in *cis* with the G allele at rs10974944 (P = 2.8×10^{-14}) (Fig. 2b). These data suggest that the G allele at rs10974944 favors the in *cis* acquisition of $JAK2^{V617F}$.

Given these findings, we then wished to see whether *JAK2* would be clearly detected as an MPN susceptibility locus in the context of testing SNPs genome-wide for association with MPN risk. To do so, we combined all unambiguous SNPs genotyped in MPN samples and WTCCC controls matched by principal component analysis with our data on rs10974944 and asked whether allele frequencies differed significantly between the two groups at any of the SNPs (Fig. 3). Four SNPs with *P* values < 10^{-7} were significantly associated with MPN risk after correcting for residual population stratification and multiple testing. One of these four SNPs is rs10974944, further supporting the hypothesis that this represents a true MPN risk SNP and that our findings are not due to population stratification.

The observation that a *JAK2* germline haplotype is markedly enriched in MPN cases compared to controls suggests that germline variation at the *JAK2* locus is an important contributor to MPN predisposition. Although genome-wide association studies have identified predisposition loci for a spectrum of human diseases, the loci identified in these studies have a modest effect on disease risk for the individual despite a larger population attributable risk. For example, a recent genome-wide association study identified six chronic lympho-cytic leukemia (CLL) predisposition loci, each of which had an odds ratio less than 1.6, but which accounted for between 12% and 39% of CLL attributable to a hereditary predisposition[18]. The GG/CC genotype at *JAK2* rs10974944 contributes significantly to the excess familial risk of MPN (OR = 2.80) with a population attributable risk of 46.0%. The

high population attributable risks for these predisposi-tion alleles reflect gene–gene and gene–environment interactions, and the familial relative risks due to these variants are much smaller, for example, less than 3% for CLL[19]. Nonetheless, the range of genotypic relative risks reported here for *JAK2*-positive MPN are among the highest described to date using a genome-wide association approach[20].

We found that somatic *JAK2* mutations were most commonly acquired in cis with the *JAK2* predisposition haplotype, suggesting a direct interaction between haplotype-specific genetic variation in the *JAK2* locus and secondary acquisition of somatic mutations on the same strand. We did not observe genotype-specific differences in *JAK2* expression (Supplementary Fig. 3 online), nor did we observe genotype-specific nonsynonymous sequence alterations or alterations in the 3′ UTR. These data suggest that rs10974944 favors the acquisition of *JAK2* mutations in cis by an unidentified mechanism. There may be haplotype-specific variation in a regulatory motif or genotype-specific splicing that increases the selective advantage of the $JAK2^{V617F}$ allele when it is acquired on the predisposition haplotype. Alternatively, it is possible that genotype-specific genomic variation in the *JAK2* haplotype block increases the somatic mutation rate at this locus, as has been observed for germline variants in the *APC* gene present in the Ashkenazi Jewish population that increase the rate of somatic *APC* mutations[21]. Although alternate activating *JAK2* muta-tions have been identified in acute leukemia samples and cell lines[22,23], the $JAK2^{V617F}$ allele overwhelmingly predominates in polycythemia vera, essential thrombocythema and primary myelofibrosis. It is thus possible that germline variation in the *JAK2* locus may be specifically associated with an increase in the rate of the guanine-to-thymidine substitution at *JAK2* codon 617.

It is likely there are additional germline loci important in MPN predisposition and pathogenesis. Our data suggests that germline variation at the *JAK2* locus has a minimal contribution to $JAK2^{V617F}$-negative MPN. Moreover, analysis of affected and unaffected members from 25 MPN kindreds did not reveal an association between *JAK2* rs10974944 genotype and MPN in these pedigrees (O.K., A.M., M.W., D.G.G. and R.L.L., unpublished data). Using genome-wide analysis of ancestry-matched cases and controls, we demonstrate that there are additional candidate loci that contribute to MPN predisposition. These data should be interpreted with the caveats that the MPN SNP data are generated from diseased tissue and that there is not complete coverage of the genome owing to strand ambiguous SNPs. In addition, this study was done on a relatively small number of samples, and it is likely that larger genome-wide association studies will identify additional germline alleles relevant to MPN pathogenesis. Taken together, these data indicate that germline variation is an important contributor to MPN pathogenesis. Moreover, this approach can be used to identify cancer predisposition loci in cancer SNP array data that is being generated by The Cancer Genome Atlas Project[24] and other large-scale cancer genomics studies.

METHODS

Study samples and SNP array analysis of MPN samples

Samples from subjects with MPN were obtained from the Harvard MPD Study case cohort[10], all cases provided informed consent. Acute myeloid leukemia, myelodysplasia, and familial MPN samples were collected using protocols approved by the Dana Farber Cancer Institute institutional review board; all subjects provided informed consent. DNA was extracted from granulocytes and buccal swabs as previously described[10], and RNA was extracted from subject cells stored in Trizol. We chose 217 granulocyte DNA samples, including 113 samples from subjects with polycythemia vera and 68 samples from subjects with essential thrombocythemia, for SNP array analysis on the basis of clonality studies and $JAK2^{V617F}$ mutational burden[25] in order to limit analysis to samples with > 80% MPN cells.

DNA samples were genotyped using Affymetrix 250K *StyI* arrays. We scanned arrays with the GeneChip Scanner 3000, and used the Affymetrix Genotyping Tools Version 2.0 to ascertain genotypes.

Genetic analysis for modifier loci

Analysis of genome-wide SNP array data was done using PLINK[26] unless otherwise described. To identify modifier loci, we tested all SNPs for frequency differences between 113 individuals with polycythemia vera and 68 individuals with essential thrombocythemia. Five different genetic models (genotypic, allelic, trend, dominant and recessive) were tested for each SNP, and significance at each SNP was assessed using adaptive permutation testing with a maximum of 10^8 permutations. Odds ratios and confidence intervals for Table 1 were computed based on logistic regression (–logistic option in PLINK).

Genotyping and expression analysis

Granulocyte and buccal DNA samples were genotyped using TaqMan SNP genotyping assays for rs10974944 and rs12500918 (Applied Biosystems) assays. DNA samples from CEU HapMap founders were used as controls. We measured expression of *JAK2* and *HPRT1* using TaqMan Gene Expression Assays (Applied Biosystems).

JAK2 rs10974944 mutation clonal analysis

A 3-kb PCR product containing rs10974944 and exon 14 of *JAK2,* where 1849G > T (V617F) resides, was amplified from $JAK2^{V617F}$ -positive subjects heterozygous for rs10974944 in the germline using allele-specific forward primers and a common reverse primer followed by sequencing the 1849G > T site with the reverse primer (Supplementary Table 3 online). To validate the allele-specific PCR assay we cloned the 3-kb fragment from 11 cases using the TOPO TA Cloning Kit (Invitrogen), and sequenced sufficient individual colonies from each subject to ascertain which genotype at rs10974944 was in *cis* with the 1849G > T (V617F) allele in granulocyte DNA from each informative subject.

Principal component analysis of MPN cases and controls

For principal component analysis we used genome-wide data from the 217 MPN cases and from 3,000 controls from the Wellcome Trust Case Control consortium[27], which were genotyped with the Affymetrix GeneChip 500k Mapping Array Set, of which the 250 K Sty chip is a subset. Before analysis, we carried out quality control filtering of both samples and SNP separately for cases and controls and then merged the dataset using the common set of SNPs present in the two cohorts. To do so, we first filtered out the ambiguous SNPs (A/Tor G/C alleles) to ensure we unambiguously know strand when we merge the two datasets. We removed 35,218 ambiguous markers (out of 231,786) from the MPN genotype dataset and 77,934 ambiguous markers (out of 486,661) from the WTCCC control cohort. The quality control filters and quality assessment removed subjects with low genotype completion rates (< 90%). Further data cleaning of the autosomal SNPs typed in both datasets retained SNPs that have a minor allele frequency (MAF) > 5%, a rate of missing genotype < 1%, and are in Hardy-Weinberg equilibrium in the WTCCC controls (exact test $P > 10^{-7}$). In total, 62,775 markers were identified for analysis and used in the merged case and control dataset.

To investigate potential population stratification biases that could be introduced by the shared controls, we carried out principal component analysis using EIGENSTRAT[28] . To reduce the linkage disequilibrium between markers, we first used PLINK to filter markers such that all remaining markers are in low LD ($r^2 < 0.1$, calculated in sliding windows 50 SNPs wide, shifted and recalculated every 5 SNPs). We applied the EIGENSTRAT program with default parameters and no outlier removal to infer axes of variation in the combined

dataset. The case and controls that clustered together on the eigenvector plot (with the first two axes of variation) were used for the association analysis.

The main SNP of interest in *JAK2,* rs10974944, has G and C alleles and was therefore eliminated by our filtering for ambiguous SNPs. To see at what rank it would appear in a GWAS for MPN risk alleles, we included it in our genome-wide association analysis. Specifically, we included the germline genotypes generated using TaqMan for the cases with the genotypes provided by the WTCCC data for the controls. A test of allelic association was done using –assoc in PLINK.

Statistical analysis

The frequencies of the genotypes between cases and controls were compared using Pearson's χ^2 test and, when required, Fisher's exact test. The ANOVA test was used for comparison of $JAK2^{V617F}$ allele burden between different genotypes. SPSS version 16.0 for Windows (SPSS) was used for all statistical tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge the subjects who have contributed to our understanding of these disorders. We thank S. Thomas, I. Dolgalev and T. Landers for assistance with high-throughput resequencing, A. Viale for assistance with JAK2 expression analysis, and T. Kirchhoff for advice and suggestions. This study makes use of data generated by the Wellcome Trust Case-Control Consortium; a full list of the investigators who contributed to the generation of the data are available from http://www.wtccc.org.uk and funding was provided by the Wellcome Trust under award 076113. This work was supported by grants from the National Institutes of Health, the Starr Cancer Consortium, the Myeloproliferative Disorders Foundation, the Howard Hughes Medical Institute, the Doris Duke Charitable Foundation and the Kristen Amico Sesselman Leukemia Research Fund. O.K. is supported by a grant from the Academy of Finland. D.G.G. is an Investigator of the Howard Hughes Medical Institute and is a Doris Duke Charitable Foundation Distinguished Clinical Scientist. Work in the laboratory of R.J.K. is supported by Memorial Sloan Kettering Cancer Center through US National Institutes of Health grant P30 CA008748. R.L.L. is an Early Career Award recipient of the Howard Hughes Medical Institute and a Clinical Scientist Development Award recipient of the Doris Duke Charitable Foundation and is the Geoffrey Beene Junior Chair at Memorial Sloan Kettering Cancer Center.

References

1. Adamson JW, Fialkow PJ, Murphy S, Prchal JF, Steinmann L. Polycythemia vera: stem-cell and probable clonal origin of the disease. N Engl J Med. 1976; 295:913–916. [PubMed: 967201]
2. Gilliland DG, Blanchard KL, Levy J, Perrin S, Bunn HF. Clonality in myeloproliferative disorders: analysis by means of the polymerase chain reaction. Proc Natl Acad Sci USA. 1991; 88:6848–6852. [PubMed: 1862109]
3. El Kassar N, Hetet G, Li Y, Briere J, Grandchamp B. Clonal analysis of haemopoietic cells in essential thrombocythaemia. Br J Haem. 1995; 90:131–137.
4. Tsukamoto N, et al. Clonality in chronic myeloproliferative disorders defined by X-chromosome linked probes: demonstration of heterogeneity in lineage involvement. Br J Haematol. 1994; 86:253–258. [PubMed: 7911034]
5. Jamieson CH, et al. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation. Proc Natl Acad Sci USA. 2006; 103:6224–6229. [PubMed: 16603627]
6. James C, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature. 2005; 434:1144–1148. [PubMed: 15793561]
7. Baxter EJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet. 2005; 365:1054–1061. [PubMed: 15781101]

8. Kralovics R, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med. 2005; 352:1779–1790. [PubMed: 15858187]
9. Zhao R, et al. Identification of an acquired JAK2 mutation in polycythemia vera. J Biol Chem. 2005; 280:22788–22792. [PubMed: 15863514]
10. Levine RL, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell. 2005; 7:387–397. [PubMed: 15837627]
11. Landgren O, et al. Increased risks of polycythemia vera, essential thrombocythemia, and myelofibrosis among 24577 first-degree relatives of 11039 patients with myelo-proliferative neoplasms in Sweden. Blood. 2008; 112:2199–2204. [PubMed: 18451307]
12. Pardanani A, Fridley BL, Lasho TL, Gilliland DG, Tefferi A. Host genetic variation contributes to phenotypic diversity in myeloproliferative disorders. Blood. 2008; 111:2785–2789. [PubMed: 18006699]
13. Cario H, Goerttler PS, Steimle C, Levine RL, Pahl HL. The JAK2V617F mutation is acquired secondary to the predisposing alteration in familial polycythaemia vera. Br J Haematol. 2005; 130:800–801. [PubMed: 16115144]
14. Bellanne-Chantelot C, et al. Genetic and clinical implications of the Val617Phe JAK2 mutation in 72 families with myeloproliferative disorders. Blood. 2006; 108:346–352. [PubMed: 16537803]
15. Pietra D, et al. Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders. Blood. 2008; 111:1686–1689. [PubMed: 17984312]
16. Scott LM, Scott MA, Campbell PJ, Green AR. Progenitors homozygous for the V617F mutation occur in most patients with polycythemia vera, but not essential thrombocythemia. Blood. 2006; 108:2435–2437. [PubMed: 16772604]
17. Scott LM, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. N Engl J Med. 2007; 356:459–468. [PubMed: 17267906]
18. Di Bernardo MC, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet. 2008; 40:1204–1210. [PubMed: 18758461]
19. Hemminki K, Forsti A, Bermejo JL. The 'common disease-common variant' hypothesis and familial risks. PLoS ONE. 2008; 3:e2504. [PubMed: 18560565]
20. Hemminki K, Forsti A, Lorenzo Bermejo J. New cancer susceptibility loci: population and familial risks. Int J Cancer. 2008; 123:1726–1729. [PubMed: 18623115]
21. Laken SJ, et al. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. Nat Genet. 1997; 17:79–83. [PubMed: 9288102]
22. Bercovich D, et al. Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. Lancet. 2008; 372:1484–1492. [PubMed: 18805579]
23. Mercher T, et al. JAK2T875N is a novel activating mutation that results in myelo-proliferative disease with features of megakaryoblastic leukemia in a murine bone marrow transplantation model. Blood. 2006; 108:2770–2779. [PubMed: 16804112]
24. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. [PubMed: 18772890]
25. Levine RL, et al. X-inactivation-based clonality analysis and quantitative JAK2V617F assessment reveal a strong association between clonality and JAK2V617F in PV but not ET/MMM, and identifies a subset of JAK2V617F-negative ET and MMM patients with clonal hematopoiesis. Blood. 2006; 107:4139–4141. [PubMed: 16434490]
26. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
27. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]
28. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]

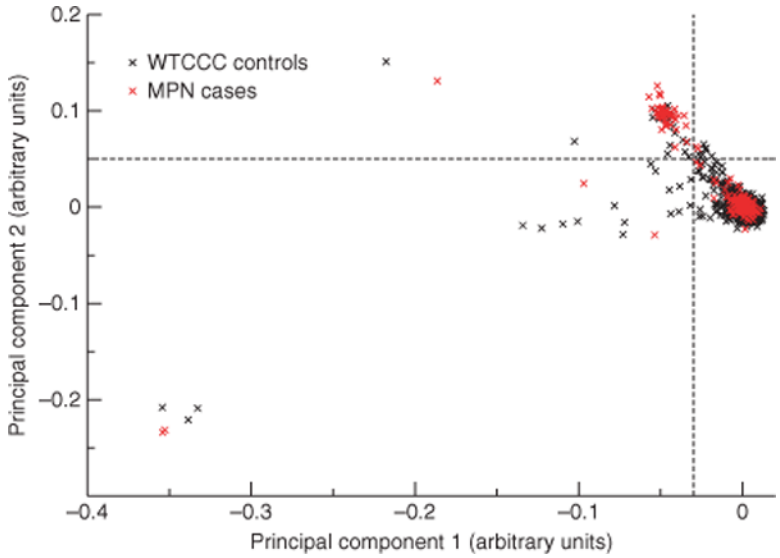

Figure 1.
Principal component analysis of MPN cases and WTCCC controls. The two axes represent the first two principal components. The lower right quadrant contains the cluster of individuals believed to be of northern and western European ancestry, who were selected for further analysis.

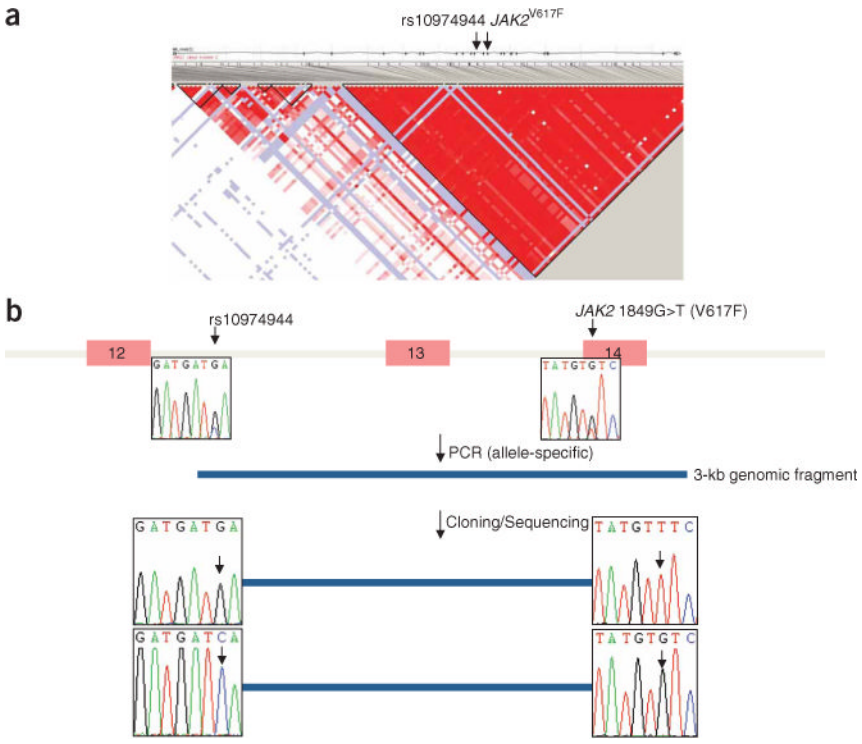

Figure 2.

JAK2^{V617F} is acquired in cis with *JAK2* SNP rs10974944. (**a**) Haplotype structure of the *JAK2* locus in CEPH HapMap founders, showing that rs10974944 and exon 14 of *JAK2* are in a shared haplotype approximately 3 kb apart. (**b**) Precise location of rs10974944 in relation to exons 12, 13 and 14, and the result of long-range PCR and clonal sequence analysis of this 3-kb fragment in a subject who was heterozygous for rs10974944 in his germline and heterozygous for *JAK2*^{V617F} in affected tissue. Analysis shows that in this subject the G allele at rs10974944 is in cis with the mutant T allele at *JAK2*^{V617}, whereas the C allele is in cis with the wild-type G allele at *JAK2*^{V617}. The G allele was found to be in cis with the mutant T allele in 38 of 45 *JAK2*^{V617F}-positive MPN cases who were heterozygous for rs10974944, suggesting that *JAK2*^{V617F} is almost always acquired in cis with the risk allele at rs10974944.

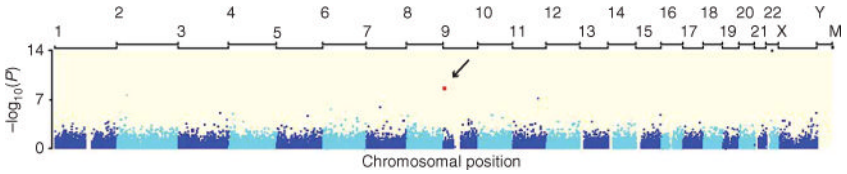

Figure 3.
Genome-wide SNP analysis of MPN cases and WTCCC controls. The arrow marks the position of rs10974944.

Table 1

Candidate modifier loci identified in genome-wide SNP array analysis

SNP	Chromosomal locus (gene)	Genetic model	P value (unadjusted)	Permutation P value	Odds ratio (95% CI)
rs12500918	4q31	Allelic	1.0×10^{-7}	1.0×10^{-7}	3.6 (2.0–6.3)
rs1524395	7p11	Allelic	4.4×10^{-6}	8.3×10^{-6}	3.4 (2.0–6.0)
rs10974944	9p24 (JAK2)	Genotypic	2.5×10^{-7}	9.7×10^{-6}	1.9 (1.2–2.7)
rs2279784	3q21 (KALRN)	Trend	1.5×10^{5}	1.0×10^{-5}	7.4 (2.8–20)
rs12500918 genotype			PV		ET
AA			50 (38.8%)		29 (24.4%)
AC			64 (49.6%)		63 (52.9%)
CC			15 (11.6%)		27 (22.7%)
					$P = 0.0134$
Germline rs10974944 genotype			PV		ET
GG			44 (26.7%)		19 (16.0%)
CG			84 (50.9%)		56 (47.1%)
CC			37 (22.4%)		44 (37.0%)
					$P = 0.011$

PV, polycythemia vera; ET, essential thrombocythemia.

Table 2

Germline genotype for *JAK2* SNP rs10974944 in MPN cases and WTCCC controls

(a) Germline genotype for *JAK2* SNP rs10974944 and MPN predisposition		
rs10974944 genotype	**MPN**	**WTCCC**
GG	70 (21.6%)	195 (6.5%)
CG	161 (49.7%)	1,139 (38.0%)
CC	93 (28.7%)	1,665 (55.5%)
	P	OR (95% CI)
GG vs. CG/CC	1.5×10^{-21}	4.0 (2.9–5.4)
GG/CG vs. CC	4.1×10^{-20}	3.1 (2.4–4.0)
GG vs. CC	5.1×10^{-32}	6.4 (4.6–9.1)
CG vs. CC	2.1×10^{-12}	2.5 (1.9–3.3)
(b) Germline genotype for *JAK2* SNP rs10974944 in MPN cases and matched WTCCC controls according to principal component analysis		
rs10974944 genotype	MPN	WTCCC
GG	18 (20.5%)	192 (6.5%)
CG	44 (50.0%)	1,121 (37.9%)
CC	26 (29.5%)	1,646 (55.6%)
	P	OR (95% CI)
GG vs. CG/CC	3.5×10^{-07}	3.7 (2.2–6.3)
GG/CG vs. CC	1.3×10^{-06}	3.0 (1.9–4.8)
GG vs. CC	2.3×10^{-10}	5.9 (3.2–11.0)
CG vs. CC	0.00017	2.5 (1.5–4.2)

Table 3

Germline genotype for *JAK2* SNP rs10974944 in *JAK2*^{V617F}-positive MPN cases, *JAK2*^{V617F}-negative MPN cases and WTCCC controls

	***JAK2*^{V617F}-positive MPN**	***JAK2*^{V617F}-negative MPN**	**WTCCC**
GG	60 (24.5%)	10 (13.5%)	195 (6.5%)
CG	127 (51.8%)	31 (41.9%)	1,139 (38.0%)
CC	58 (23.7%)	33 (44.6%)	1,665 (55.5%)
Total	245	74	2,999
JAK2^{V617F}-positive MPN	*P*	OR (95% CI)	
GG vs. CG/CC	8.4×10^{-24}	4.7 (3.4–6.5)	
GG/CG vs. CC	77×10^{-22}	4.0 (3.0–5.5)	
GG vs. CC	6.9×10^{-37}	8.8 (6.0–13.1)	
CG vs. CC	7.9×10^{-14}	3.2 (2.3–4.4)	
JAK2^{V617F}-negative MPN	*P*	OR (95% CI)	
GG vs. CG/CC	0.017	2.2 (1.1–4.4)	
GG/CG vs. CC	0.06	1.6 (1.0–2.5)	
GG vs. CC	0.008	2.6 (1.3–5.3)	
CG vs. CC	0.21	1.4 (0.8–2.3)	