Myeloproliferative Neoplasms and Somatic Mosaicism in the 23andMe Participant Community

D.A. Hinds[1], K.E. Barnholt[1], J.L. Zehnder[2], A.K. Kiefer[1], C.B. Do[1], N. Eriksson[1], J.L. Mountain[1], U. Francke[1], J.Y. Tung[1], R.L. Levine[3], R.A. Mesa[4], J. Gotlib[5]

[1]23andMe, Inc, Mountain View, CA; [2]Department of Pathology and Department of Medicine/Hematology, Stanford University School of Medicine, Stanford, CA; [3]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY; [4]Division of Hematology & Medical Oncology, Mayo Clinic, Scottsdale, AZ; [5]Department of Medicine/Hematology, Stanford University School of Medicine, Stanford, CA.

Introduction

Myeloproliferative neoplasms (MPNs) are disorders that result in unregulated overproduction of one or more myeloid blood cell types by the bone marrow. Polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF) comprise the three classic MPNs. A somatic *JAK2* mutation, V617F, is present in 95% of PV and 50-60% of ET and PMF patients. Past work has identified a germline haplotype of *JAK2* associated with risk of developing a V617F-positive MPN[1-3]. This does not fully explain familial aggregation of MPNs.

Methods

We have recruited a web-based participatory cohort of patients with MPNs to better understand the genetic basis of these conditions. We have enrolled and collected saliva samples from more than 800 participants. Subjects have been genotyped using a derivative of the Illumina Human OmniExpress with additional custom content, including probes for *JAK2* V617F. We selected 447 unrelated individuals with self-reported diagnoses of classic MPNs, including ET (n=150), PV (n=149), PMF (n=60), and 88 with related multiple diagnoses ('MD'), i.e. PV+PMF. We used 65,051 additional unrelated 23andMe research participants as population controls. We imputed genotypes against the August 2010 release of 1000 Genomes haplotypes, using BEAGLE and minimac. We performed a GWAS of classic MPNs, adjusting for age, gender, and ancestry.

Results

In addition to replicating the known germline association at the *JAK2* locus, we see a strong association in the *TERT* gene, telomerase reverse transcriptase (Fig. 1, Table 1, Fig. 2). Other variants in *TERT* have been associated with a variety of solid tumors and with red blood cell count. While not genome-wide significant, our third ranking association is in the *CUL5* gene, at an imputed SNP in strong linkage disequilibrium (r^2=0.9) with rs1800056, a non-synonymous variant, F858L, in *ATM*, or ataxia telangiectasia mutated. This variant has previously been associated with chronic lymphocytic leukemia and breast cancer.

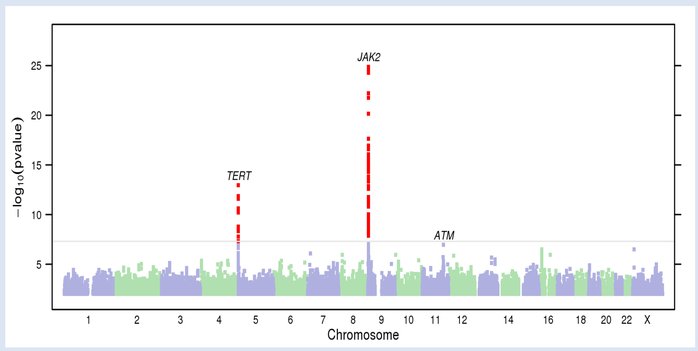

Figure 1. Manhattan plot of GWAS results.

SNP	Region	Alleles	Freq	P	OR	95% CI	Gene
rs1327494	9p24.1	A/G	0.268	8.6e-42	2.60	[2.27,2.98]	*JAK2*
rs2853677	5p15.33	A/G	0.435	1.1e-13	1.67	[1.46,1.92]	*TERT*
rs36034326	11q22.3	A/G	0.013	1.1e-07	3.14	[2.18,4.54]	*CUL5*
rs1800056	11q22.3	T/C	0.014	4.9e-06	2.65	[1.84,3.82]	*ATM*

Table 1. Association test results for principal findings. Frequencies and odds ratios are given for the second listed allele.

Consistent with a previous report[2], *JAK2* rs1327494 is more strongly associated with PV than with ET or MF. In contrast, we see consistent effects across diagnoses for *TERT* rs2853677 and *ATM* rs1800056 (Fig. 3).

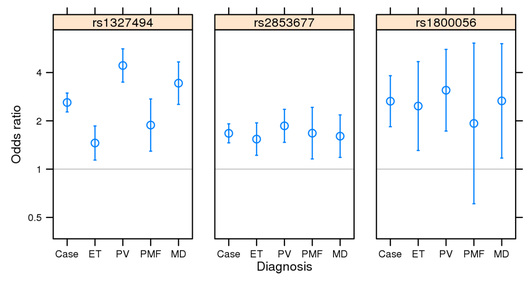

Figure 3. Effect sizes for lead SNPs versus MPN diagnoses.

Somewhat unexpectedly, we found that using probes for the V617F mutation on our genotyping arrays, we were able to detect the mutation in MPN study participants, with good agreement with self-reported mutation status (Fig. 4).

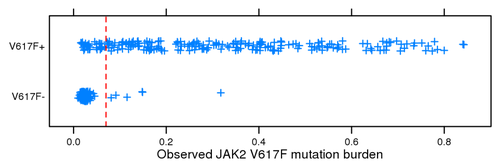

Figure 4. Self-reported V617F status versus observed V617F mutation burden, in 310 MPN cases. Burden is estimated as the ratio of probe intensity for the alternate allele to total intensity across both alleles.

We tested the three MPN risk variants for association with V617F+ and V617F− disease (Fig. 5). *JAK2* rs1327494 is much more strongly associated with V617F+ disease, but *TERT* rs2853677 is not; results for *ATM* rs1800056 are inconclusive.

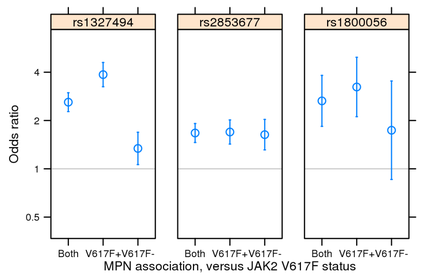

Figure 5. SNP effect sizes for V617F+ and V617F− MPN cases. Analyses shared the same set of GWAS controls.

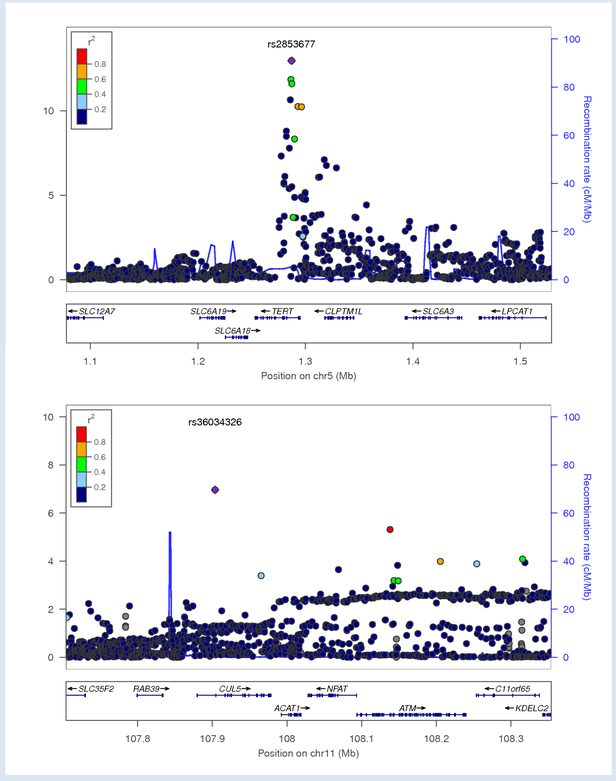

Figure 2. Regional association plots for *TERT* and *ATM* loci.

We found that we could also detect the V617F mutation in a subset of controls. The prevalence and extent of mutation burden increases with participant age (Fig. 6).

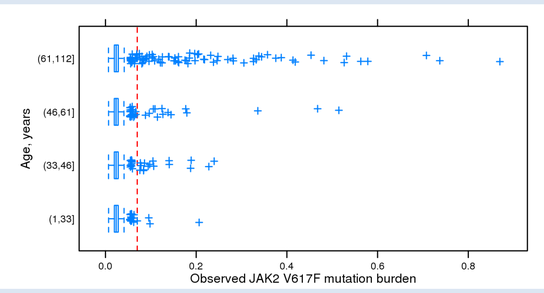

Figure 6. Observed somatic V617F mutation burden across quartiles of age, among 65,051 GWAS controls.

We tested whether each MPN risk variant was associated with V617F mutation status, among MPN cases, and among controls (Fig. 7). In cases, *JAK2* rs1327494 is strongly associated with V617F status, while *TERT* rs2853677 is not. The three variants are all associated with somatic V617F status among the controls.

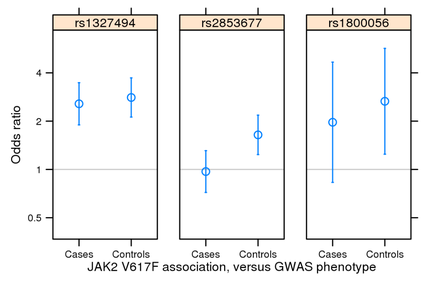

Figure 7. Associations with observed V617F status, in 447 GWAS cases, and 65,051 controls.

Discussion

Our study represents a significant addition to our understanding of the genetic predispositions for classic myeloproliferative neoplasms. The identification of risk alleles in *TERT* and *ATM* helps to link the genetics of MPNs with the genetics of other blood neoplasms and solid tumors. The work also demonstrates the power of web-based recruitment for studying uncommon diseases.

The rate of V617F positivity we see in our controls is higher than the prevalence of MPNs. It seems likely that this group includes individuals with undiagnosed or indolent disease, as well as some who will never develop an active hematological neoplasm. We are currently validating the V617F assay and are developing a strategy for returning these test results to our participants.

Acknowledgments

We thank the participants in the 23andMe Myeloproliferative Research Initiative, and 23andMe customers who have consented to participate in research for enabling this study. We also thank the employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

References

1. Jones AV et al. Nat Genet 2009;41:446-449.
2. Kilpivaara O et al. Nat Genet 2009;41:455-459.
3. Olcaydu D et al. Nat Genet 2009;41:450-454.