Genetic Characteristics of Polycythemia Vera and Essential Thrombocythemia in Korean Patients

Eun-Hyung Yoo,^{1†} Kyung-Jin Park,^{2,3†} Hong-Hee Won,^{2,3} Jun-Hee Park,^{2,3} Jong-Ho Park,^{2,3} Seung-Tae Lee,⁴ Hee-Jin Kim,⁴ Soo-Mee Bang,⁵ Hyun-Sook Chi,⁶ Chul Won Jung,⁷ Sun-Hee Kim,⁴ Hongseok Yun,⁸ Choong-Hyun Sun,⁸ Inho Park,⁸ Seungmook Lee,⁸ Clarence Lee,⁹ Barry Merriman,⁹ Raymond Luo,¹⁰ Eileen Hwee Hong Tan,¹⁰ Keun-Joon Park,¹¹ Na-Kyung Yoo,¹¹ Jason J. Kang,⁹ and Jong-Won Kim^{4*}

¹Department of Laboratory Medicine, Konyang University Hospital, Myunggok Medical Research Institute, College of Medicine, Konyang University, Daejeon, Korea

² Department of Health Sciences and Technology, Samsung Advanced Institute for Health Science and Technology (SAIHST), Sungkyunkwan University, Seoul, Korea

³Samsung Biomedical Research Institute, Samsung Medical Center, Seoul, Korea

⁴Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

⁵Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

⁶Department of Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea

⁷Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

⁸Samsung SDS, Seoul, Korea

⁹Thermo Fisher Scientific, Carlsbad, California

¹⁰Thermo Fisher Scientific, Singapore, Singapore

¹¹Thermo Fisher Scientific, Seoul, Korea

Background: Despite recent advances in the investigation of myeloproliferative neoplasms (MPN), the impact of genetic heterogeneity on its molecular pathogenesis has not been fully elucidated. Thus, in this study, we aim to characterize the genetic complexity in Korean patients with polycythemia vera (PV) and essential thrombocythemia (ET). **Methods:** We conducted association studies using 84 single-nucleotide polymorphisms (SNPs) in 229 patients (96 with PV and 133 with ET) and 170 controls. Further, whole-genome sequencing was performed in six patients (two with *JAK2* V617F and four with wild-

type *JAK2*), and putative somatic mutations were validated in a further 69 ET patients. Clinical and laboratory characteristics were also analyzed. **Results:** Several germline SNPs and the 46 haplotype were significantly associated with PV and ET. Three somatic mutations in *MPDZ*, *IQCH*, and *CALR* genes were selected and validated. The frequency of the *CALR* mutation was 58.0% (40/69) in ET patients, who did not carry *JAK2*/*MPL* mutations. Moreover, compared with *JAK2* V617F-positive patients, those with *CALR* mutations showed lower hemoglobin and hematocrit levels ($P = 0.004$ and $P = 0.002$, respectively),

†These authors contributed equally to this work.

*Correspondence to: Jong-Won Kim, MD, PhD, Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea. E-mail: kimjw@skku.edu

Received 17 March 2015; Accepted 9 March 2016

DOI 10.1002/jcla.21981

Published online in Wiley Online Library (wileyonlinelibrary.com).

higher platelet counts ($P=0.008$), and a lower frequency of cytoreductive therapy ($P=0.014$). **Conclusion:** This study was the first comprehensive investigation of the genetic characteristics of Korean patients

Key words: calreticulin; essential thrombocythemia; haplotype; polycythemia vera; whole-genome sequencing

with PV and ET. We found that somatic mutations and the 46 haplotype contribute to PV and ET pathogenesis in Korean patients. *J. Clin. Lab. Anal.* **30**:1061–1070, 2016. © 2016 Wiley Periodicals, Inc.

INTRODUCTION

The discovery of the V617F mutation in the Janus Kinase 2 (*JAK2*) gene has been monumental in our understanding of the pathogenesis of myeloproliferative neoplasms (MPN). However, the *JAK2* V617F mutation is not disease specific. Indeed, it was identified in more than 95% of patients with polycythemia vera (PV) (1, 2) and in 50–60% of patients with essential thrombocythemia (ET) and primary myelofibrosis (PMF) (3–6).

Additional somatic mutations in the *JAK2* exon 12 (5, 7), *MPL* (4, 6), negative regulators (such as *SH2B3*) (3), and epigenetic regulators of DNA methylation [such as *TET2* (8), *DNMT3A* (9), *IDH1*, and *IDH2* (10)], and of chromatin structure [such as *EZH2* (11) and *ASXL1* (12)], have been identified in a minority of MPN patients. However, most of the mutations identified in MPN are low in frequency and lack disease specificity. In fact, there is a subset of patients with ET that is entirely negative for *JAK2* V617F and *MPL* mutations. The resultant dearth of molecular markers makes it difficult to diagnose, manage, and discover therapeutic targets in *JAK2/MPL* mutation-negative ET. In this regard, a novel driver mutation, considered responsible for *JAK2/MPL* mutation-negative ET, has been extensively investigated. Recent advances in whole-genome sequencing (WGS) and massively parallel sequencing technologies have allowed in-depth study of somatic and germline alterations in cancer (13–15). Despite various hurdles in the analysis of cancer WGS data, this technique promises to provide an accurate picture of cancer as a genetic disease and facilitates the discovery of novel mutations.

Interestingly, the discovery of a specific inherited *JAK2* 46/1 haplotype has presented evidence for a strong and complex interaction between somatic and germline genetics in MPN pathogenesis (16–18). The specific *JAK2* haplotype preferentially acquires the V617F mutation in *cis* with the predisposition allele. However, the underlying mechanism(s) remains obscure. The 46/1 haplotype promotes predisposition to the *JAK2* exon 12-mutated and *MPL*-mutated MPNs (19, 20). Moreover, the frequencies of this haplotype were comparable in different MPN subtypes,

and thus, it fails to explain the phenotypic diversity of the disease (16).

To date, none of the somatic and germline alterations discovered have been able to explain either the differential phenotypes or the molecular pathogenesis of MPN. Additional molecular alterations and/or mutant dosage modification may contribute toward both. In particular, no comprehensive investigation on germline and somatic alterations in Korean MPN patients have been reported. To address this, we performed a germline *JAK2* haplotype analysis and a WGS study for somatic alterations in Korean patients with PV and ET.

MATERIALS AND METHODS

JAK2 SNPs and Association Analysis

Patients and samples

The institutional review board of Samsung Medical Center approved this study. Additionally, verbal and written informed consent was obtained from all the patients. Two hundred twenty-nine patients from the Samsung Medical Center ($n=127$), Asan Medical Center ($n=63$), and the Bundang Hospital of Seoul National University ($n=39$) in Korea were enrolled. Among them, 96 had PV [male:female (M:F), 46:50; median age, 63 years (range, 26–84 years)] and 133 had ET [M:F, 57:76; median age, 54 years (range, 21–85 years)], in accordance with the 2008 World Health Organization (WHO) criteria. In addition, we also enrolled 170 healthy Korean individuals as controls [M:F, 97:73; median age, 53 years (range, 32–76 years)], attending the Health Promotion Center at our institution for a routine checkup. Total mononuclear cell and/or granulocyte DNA was isolated from the peripheral blood samples using the Wizard® Genomic DNA Purification Kit (Promega, Madison, WI) according to the manufacturer's instructions.

Selection of tagging SNPs and SNP genotyping

We selected 84 linkage disequilibrium (LD) tagSNPs using the Carlson method (21), with a minimum minor

allele frequency (MAF) of 5% and an r^2 value of 0.9 in *JAK2* genes, using the HapMap database. SNPs were genotyped using the SequenomTM MassARRAY system (San Diego, CA). The i-PLEX assay was performed according to the manufacturer's instructions (www.sequenom.com) using 5 ng of genomic DNA. The Sequenom SpectroDESIGNER software (Sequenom, San Diego, CA) was used for SNP assay design, and the SpectroTyper 4.0 (Sequenom) was used to call genotypes automatically, followed by manual review.

Association analysis

The frequency distribution of genotypes was examined for both the patients and the controls. The Hardy–Weinberg equilibrium was evaluated separately in the cases and the controls, using the chi-square test. Association analysis was performed with Fisher's exact test (genotype mode), and the Cochran–Armitage trend test was used to assess differences in genotype frequencies between the patients and controls. *P*-values were corrected by the false discovery rate. Haplotypes were constructed using the PHASE v 2.1 software, and LD was measured with the Haploview v 4.1 software.

Whole-genome Sequencing

Patients and samples

Six Korean ET patients (two with *JAK2* V617F and four with wild-type *JAK2* and *MPL*), diagnosed according to the 2008 WHO criteria, were included in the WGS analysis. *JAK2* V617F and *MPL* exon 10 were assessed by direct sequencing and/or allele-specific PCR. After being approved by the ethics committee at the Samsung Medical Center, granulocyte (tumor) and skin biopsy (control) samples were obtained, along with the patients' written informed consent and their clinical information (Table 1).

TABLE 1. Clinical and laboratory characteristics of six index patients included in the WGS analysis

Patients	ET1	ET2	ET3	ET4	ET5	ET6
Gender	F	M	F	F	F	M
Age (years)	51	52	44	48	42	37
<i>JAK2</i> V617F	d	n.d.	n.d.	d	n.d.	n.d.
<i>MPL</i> M515	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Hb ^a (g/dl)	12.6	16.5	13.3	12.9	12.8	15.6
Hct ^a (%)	38.2	47.5	40.1	39.4	39.3	46.5
WBC count ^a ($\times 10^3/\mu\text{l}$)	18.95	6.6	5.89	11.46	9.67	12.88
Platelet count ^a ($\times 10^3/\mu\text{l}$)	765	699	453	922	1,575	1,010

F, female; M, male; d, detected; n.d., not detected; Hb, hemoglobin; Hct, hematocrit; WBC, white blood cell.

^aCBC data at first referral to Samsung Medical Center.

Overall, 69 DNA samples of Korean ET patients with wild-type *JAK2* (57 from the Samsung Medical Center and 12 from the Asan Medical Center), including four index patients, were enrolled in our validation study for candidate somatic variants. Available clinical and laboratory data, including age at first referral, gender, complete blood counts, the presence or absence of splenomegaly, thrombosis, disease progression to acute leukemia or myelofibrosis, and cytoreductive therapy at follow-up, were collected. Clinical and laboratory data of 52 ET patients carrying *JAK2* V617F were also amassed to compare clinical phenotypes of ET patients with wild-type *JAK2*. We extracted 10 μg DNA each using the Wizard[®] Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. DNA quality and quantity were assessed by gel electrophoresis and Pico-Green dsDNA assay [Quanti-iTTM PicoGreen[®] dsDNA kit (Life Technologies, Grand Island, NY)], respectively.

Sequence analysis

Tumor-control paired samples from three patients (ET1, ET2, and ET3) were sequenced on the SOLiD[®] Next-Generation Sequencing platform (Life Technologies, Carlsbad, CA). The sequenced reads were mapped to the hg19 reference genome using the Burrow–Wheeler Aligner (BWA 0.6.2-r126). Duplication removal and recalibrations were performed using the Picard Tools 1.64 and GATK-Lite, respectively. Further, the genomic DNA libraries of the other three samples (ET4, ET5, and ET6) were introduced to the Complete Genomics sequencing platform (Mountain View, CA). The sequenced reads from these patients were mapped and analyzed using the Complete Genomics Analysis platform (CGA pipeline_2.0.3.6).

Prioritization of putative somatic mutations

Toward selection of putative mutations from the WGS data, we applied the following five filtering criteria: quality filtering, germline filtering, pathogenicity, population frequency, and variant location. First, we excluded the variants with segmental duplication, genotype quality <20, and depth of coverage <30-fold, as obtained from the SOLiD sequencing data. Moreover, we included only those variants that matched the quality criteria specified by the CGA Tools v.2.0.3.6, in case of the Complete Genomics sequencing data. Second, somatic variations were selected by comparing the patient's tumor and normal genomes. We further eliminated all MPN-specific polymorphisms identified in the six patients. Third, to predict the pathogenic effects of the mutations, variants were annotated using

the ANNOVAR tool including SIFT, Polyphen-2 LRT, MutationTaster, MutationAssesor, FATHMM, MetaSVM, MetaLR, and GERP scores (22). Mutations predicted as “deleterious” by at least one of the nine tools, and mutations that were not predicted using those tools were selected. Fourth, we assessed the population frequency using the 1,000 Genome database and ESP 6,500 and selected the mutations with maximum MAF <0.1%. Finally, missense and nonsense mutations in the coding sequence, as well as splicing mutations, were selected.

Mutation validation by Sanger sequencing

The putative somatic mutations identified by WGS in the six index patients were validated using Sanger sequencing. The relevant exons and their flanking intronic regions were amplified by PCR using primers designed by the authors. Cycle sequencing and subsequent analysis were performed using the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA) and the ABI 3100 Genetic Analyzer (Applied Biosystems), respectively. The validated mutations were screened in 69 *JAK2* V617F-negative ET patients using targeted exon sequence analysis.

Statistical Analyses

Statistical analyses were carried out using the PASW 22 software (IBM Corporation, Somers, NY). Differences in the distribution of continuous variables between categories were analyzed by either the Mann–Whitney test or the Kruskal–Wallis test, based on the mutation status. Patient groups with normal variables were compared by chi-square test.

P-values < 0.05 were considered to indicate statistical significance.

RESULTS

JAK2 SNPs and Association Analysis

Allele frequencies and association analysis of *JAK2* SNPs

Among the 84 SNPs, nine were excluded because they fell below the required 80% SNP genotyping success rate and three additional SNPs deviated from the Hardy–Weinberg equilibrium (at the 5% significance threshold) in the control samples. Therefore, 72 SNPs cleared the quality control. The call rates for each SNP were 92.7–100%.

Several *JAK2* tagging SNPs, including rs12342421 and rs10758677, were significantly associated with the PV and ET cases compared with the controls. They had a strong association with PV and the V617F mutation. Although *JAK2* SNPs were more significantly associated with V617F-positive ET, the distribution of significant SNP associations was similar in V617F-positive and V617F-negative ET. The association of *JAK2* SNPs according to disease phenotype and V617F mutation status are shown in Table 2 and Figure 1.

Haplotype analysis using genotyped SNPs

Haplotypes were constructed using six genotyped SNPs among the 14 SNPs previously analyzed in the WTCCC group. These were rs7864782, rs10758669, rs12342421, rs10974944, rs10974947, and rs2031904. Sixteen haplotypes were inferred in this study, and

TABLE 2. Association of *JAK2* SNP genotypes according to disease phenotype and V617F mutation status

Category	No. of associated <i>JAK2</i> SNPs (%)	Most significant SNP ^a		
		rs ID	<i>P</i> -value ^b	<i>P</i> -value ^c
All patients (<i>n</i> = 229) vs. control (<i>n</i> = 170)	37 (51.4)	rs12342421	1.35E-16	3.21E-13
V617F-positive PV and ET (<i>n</i> = 174) vs. control (<i>n</i> = 170)	32 (44.4)	rs12342421	8.56E-15	3.11E-19
PV (<i>n</i> = 96) vs. control (<i>n</i> = 170)	33 (45.8)	rs12342421	6.21E-27	9.67E-20
V617F-positive ET (<i>n</i> = 78) vs. control (<i>n</i> = 170)	21 (29.2)	rs10758677	3.34E-05	7.28E-05
V617F-negative ET (<i>n</i> = 55) vs. control (<i>n</i> = 170)	16 (22.2)	rs10758677	0.0003348	0.0002829
PV (<i>n</i> = 96) vs. ET (<i>n</i> = 133)	27 (37.5)	rs12342421	1.04E-11	3.60E-09
PV (<i>n</i> = 96) vs. V617F-positive ET (<i>n</i> = 78)	20 (27.8)	rs12342421	5.09E-09	6.43E-07
PV (<i>n</i> = 96) vs. V617F-negative ET (<i>n</i> = 55)	25 (34.7)	rs12342421	2.64E-08	1.88E-06
V617F-positive ET (<i>n</i> = 78) vs. V617F-negative ET (<i>n</i> = 55)	2 (2.8)	rs7023498	0.01526	0.0249

^aThe standard for the most significant SNP is the Cochran–Armitage trend *P*-value.

^bFisher’s exact test.

^cCochran–Armitage trend test.

Fig. 1. Associated *JAK2* SNPs according to disease phenotype and V617F mutation status.

eight were identified as being present in more than 5% of any group (Table 3). Homozygous ht1/ht1 haplotypes (referred to as 46/46 in previous studies) were identified more frequently in PV ($n = 34$; 36.2%) compared with ET ($n = 10$; 7.8%). Sixty-three PV patients (67.0%) and 71 ET patients (55.5%) had either the ht1 (46) or the ht11 (previously referred to as 1 haplotype) haplotype. However, the ht11 (1) haplotype was infrequent in the Korean MPN patients (0.045) and controls (0.01) compared with the WTCCC control frequency (0.124). The ht1/ht11 (46/1) haplotype had a significant association with PV and ET and the V617F mutation in Korean patients. Moreover, it might act as a risk factor for ET, irrespective of the V617F mutational status (Table 4).

Whole-genome Sequencing

Basic WGS statistics

We obtained >300 Gb per sample gross mapping yield from the Complete Genomics pipeline. Ninety-nine percent of the sequence had a depth of coverage of ≥ 5 fold, and more than 96% had a depth of ≥ 20 fold in all samples. The total SNP count was >3,400,000 per sample, with a novel SNP fraction of ~4%. Furthermore, the total number of insertions and deletions was >230,000 and 250,000 per sample, respectively. Data from the SOLiD platform showed performance comparable to that of the Complete Genomics pipeline. The depth of coverage was 38–111 fold in

TABLE 3. Haplotype Frequencies in Korean PV and ET Patients and Controls

Haplotypes		PV		ET		V617F negative	Control
		Homozygous V617F	Heterozygous V617F	Homozygous V617F	Heterozygous V617F		
ht1 (46)	ACCGGC	0.54	0.42	0.5	0.28	0.29	0.17
ht4	ACGCGC	0.06	0.1	0.13	0.09	0.12	0.15
ht7	AACGGC	0.08	0.07	0.19	0.02	0.05	0.02
ht9 (71/76)	AAGCGC	0.08	0.28	0.13	0.41	0.34	0.45
ht10	AAGCAT	0.01	0.02	0	0.08	0.01	0.02
ht11 (1)	GCCGGC	0.05	0	0	0.02	0.02	0.01
ht12	GACGGC	0.08	0.02	0	0.02	0.02	0.01
ht16 (8)	GAGCAT	0.03	0.05	0.06	0.09	0.11	0.13

TABLE 4. Association between 46/1 haplotype and susceptibility to PV and ET in Korean patients

Category	No. of haplotype	No. of 46/1 haplotype	Frequency of 46/1 haplotype	P-value (vs. controls)	OR (95% CI)
All ($n = 222$)	444	181	0.408	7.41e-13	3.212 (2.317–4.451)
PV ($n = 94$)	188	100	0.532	1.03e-17	5.303 (3.588–7.839)
ET ($n = 128$)	256	81	0.316	0.0001	2.160 (1.486–3.140)
V617F-positive ET ($n = 73$)	146	47	0.322	0.0004	2.216 (1.431–3.431)
V617F-negative ET ($n = 55$)	110	34	0.309	0.0033	2.088 (1.287–3.387)

OR, odds ratio.

tumor samples, and 23–27 fold in matched germline samples in the data acquired from the SOLiD platform. Other statistics can be provided on request.

Identification of candidate somatic mutations in *JAK2* V617F-negative ET

Through a five-step prioritization, most of the variations were filtered out, leaving a total of 42 somatic variants including *JAK2* V617F (Table 5). Each patient had 1–23 somatic variants, and 18 unique non-synonymous variants were identified in four V617F-negative ET patients. All the variants, except V617F, were non-recurrent in the six index patients. Among the above 18 variants, 5 were finally validated in the corresponding genes including multiple PDZ domain (*MPDZ*; ET2), *C19orf3* (ET2), *C4orf47* (ET3), IQ motif containing H (*IQCH1*; ET3), and calreticulin (*CALR*; ET5), by direct sequence analysis in each index patient. All of the five variants were heterozygous. In addition, there were one nonsense (*MPDZ*, p.R1044X), two missense (*C4orf47*, p.P64S; *IQCH*, p.H411P), one non-frameshift insertion (*C19orf33*, c. c.268_269insAAGAAGGAGAAGGGC), and one frameshift insertion (*CALR*, c.1154_1155 insTTGTC) mutations. We performed a further screening analysis of three mutations: *MPDZ*, *IQCH*, and *CALR*, in which only the *CALR* mutation was found to be recurrent.

CALR mutations and clinical phenotypes

The frequency of the *CALR* mutation was 58.0% (40/69) in Korean ET patients not carrying *JAK2* and

MPL mutations. The distribution of the *CALR* mutant types is shown in Table 6. We found all eight types of *CALR* mutation. The most common *CALR* mutation was type 1, c.1092_1143del (p. L367 fs*46; *n* = 21; 52.5%) followed by type 2, c.1154_1155insTTGTC (p.K385 fs*47; *n* = 13; 32.5%) among the *CALR*-mutated cases. Furthermore, other than ET5 (who was detected as *CALR*-mutated upon WGS analysis), the other three index patients with wild-type *JAK2* (ET2, ET3, and ET6) also carried a *CALR* mutation (type 1).

The clinical and laboratory characteristics of 121 Korean ET patients, based on their *JAK2* and *CALR* mutation status [*JAK2* V617F(+) (*n* = 52), *JAK2* V617F(-)/*CALR*(+) (*n* = 40), and *JAK2* V617F(-)/*CALR*(-) (*n* = 29)] are shown in Table 7. Compared with the attributes of the V617F-positive ET patients, the *CALR*-mutated patients showed lower hemoglobin (Hb) and hematocrit (Hct) levels (*P* = 0.004 and *P* = 0.002, respectively), higher platelet counts (*P* = 0.008), and lower frequencies of cytoreductive therapy (*P* = 0.014). Additionally, there was no difference in clinical and laboratory characteristics between the *JAK2* V617F(-)/*CALR*(+) and *JAK2* V617F(-)/*CALR*(-) patients. However, ET patients with the *CALR* type 2 mutation showed higher platelet counts than those with the type 1 mutation (*P* = 0.039).

DISCUSSION

In this study, we demonstrated that several *JAK2* SNPs confer susceptibility to PV and ET. Rs12342421 and rs10758677 were the SNPs most significantly associated with PV and ET, respectively. The association

TABLE 5. Number of candidate somatic variants in each prioritization strategy step, and list of final candidate genes in the six index patients

Steps	No. annotated variants					
	SOLiD			CGA		
	ET1	ET2	ET3	ET4	ET5	ET6
Quality filtering	13,678	26,043	27,379	–	–	–
Germline filtering	470	1,114	822	36	20	17
Pathogenicity	449	1,098	795	33	18	15
Population frequency	146	604	327	30	17	11
Classification	1	2	3	23	8	5
Candidate genes	<i>JAK2</i>	<i>MPDZ</i> <i>C19orf33</i>	<i>CCDC81</i> <i>IQCH1</i> <i>C4orf47</i>	<i>JAK2</i> Other 22 candidate genes ^a	<i>CALR</i> <i>SDK1</i> <i>ZNF169</i> <i>SLC15A3</i> <i>SYTL2</i> ^b <i>KLHL14</i> <i>BRD4</i>	<i>MAP10</i> <i>MUC17</i> <i>CIZ1</i> <i>CPNE6</i> <i>INSM1</i>

^aZRANB2, SMCP, KANSL3, IFT57, KLHL6, TBC1D9, OXRI, DNAH8, ASS1, KIAA1462, TENM4, ADAMTS20, NKX2-8, CCDC88C, PATL2, ZBTB7C, SIPR4, ASXL1, ARVCF, MKL1, LICAM, and PLXNA.
^bTwo coding-sequence variations in the SYTL2 gene were identified.

TABLE 6. Types and frequencies of *CALR* mutations in Korean *JAK2* V617F-negative ET patients

No.	<i>CALR</i> mutation		Frequency, <i>N</i> (%)
1	c.1092_1143del	p.L367 fs*46	21 (52.5)
2	c.1154_1155insTTGTC	p.K385 fs*47	13 (32.5)
3	c.1093_1126del	p.Q363 fs*54	1 (2.5)
4	c.1129_1154delinsTGTC	p.K377 fs*46	1 (2.5)
5	c.1101_1152del	p.K368 fs*45	1 (2.5)
6	c.1148_1163del	p.E383 fs*42	1 (2.5)
7	c.1131_1146delinsGTGCCTCCTCCT	p.E388 fs*41	1 (2.5)
8	c.1154delinsCTTGTC	p.K385 fs*47	1 (2.5)
Total			40

TABLE 7. Clinical and laboratory features of 122 Korean ET patients, stratified by the presence or absence of *JAK2* and *CALR* mutations

Variables	<i>JAK2</i> V617F (+) ^a	<i>JAK2</i> V617F (–)			<i>P</i> -value	
		Total	<i>CALR</i> (+) ^b	<i>CALR</i> (–) ^c	a vs. b	b vs. c
Male:female	21:31	27:42	18:22	9:20	0.655	0.241
Median age, years (range)	55 (25–77)	47 (12–81)	49 (19–79)	45 (12–81)	0.065	0.287
Median Hb, g/dl (range)	13.5 (9.4–16.9)	12.9 (7.4–16.8)	12.1 (7.7–16.5)	13.5 (7.4–16.8)	0.004	0.104
Median Hct, % (range)	41.5 (30.3–52.3)	39.1 (22.2–50.2)	37.7 (24.1–47.5)	40.7 (22.2–50.2)	0.002	0.096
White blood cells, ×10 ³ /μl (range)	9.45 (2.8–22.7)	8.01 (1.37–43.4)	8.2 (1.4–17.0)	7.6 (2.7–43.4)	0.153	0.675
Platelets, ×10 ³ /μl (range)	621 (199–1,271)	812 (8–2,107)	843 (8–2,107)	804 (60–1,887)	0.008	0.827
Splenomegaly, <i>N</i> (%)	14/52 (26.9)	13/57 (22.8)	8/33 (24.2)	5/25 (20.0)	0.806	1.000
Thrombosis, <i>N</i> (%)	10/52 (19.2)	9/57 (15.8)	4/33 (12.1)	5/25 (20.0)	0.551	0.470
Progression to myelofibrosis, <i>N</i> (%)	5/52 (9.6)	2/57 (3.5)	2/33 (6.1)	0/25 (0)	0.699	0.501
Progression to acute leukemia, <i>N</i> (%)	1/52 (1.9)	1/57 (1.8)	1/33 (3.0)	0/25 (0)	1.000	1.000
Hydroxyurea, <i>N</i> (%)	48/52 (92.3)	39/57 (68.4)	23/33 (69.7)	17/25 (68.0)	0.014	1.000

of both SNPs with MPNs has been reported in previous studies (18, 23). Haplotypes were constructed using six SNPs, which were previously analyzed by Jones et al. (16), to precisely compare the characteristics of Korean patients to those of Western patients with previously reported MPN risk haplotypes. Among 14 SNPs in the WTCCC group, eight SNPs were filtered out during LD tagSNP selection or quality control. The frequencies of the 46/1 haplotype in this study were 0.532, 0.322, and 0.309 in PV, V617F-positive ET, and V617F-negative ET, respectively. These results are in agreement with findings in Western MPN patients: Jones et al. reported the frequencies of the 46/1 haplotype in PV (0.54), V617F-positive ET (0.43–0.51), and V617F-negative ET (0.33) (16). These common characteristics in different ethnic groups suggest that germline variations are important for MPN pathogenesis and predispose individuals to these diseases irrespective of their ethnic background. Although the *JAK2* 46/1 haplotype was not associated with *JAK2* V617F-negative MPNs in previous Asian studies (24, 25), frequencies and odds ratios of the 46/1 haplotype in Korean patients did not differ between V617F-positive ET and V617F-negative ET. Pardanani et al.

also showed that the *JAK2* haplotype confers susceptibility to ET development independently of the V617F mutation (23). In earlier Asian studies, only a single tagging SNP was analyzed—either rs10974944 (24) or rs12340895 (25). In contrast, we performed a haplotype analysis of six *JAK2* SNP genotypes. Hence, our results might be of greater empirical value. Further, the frequency of the 1 haplotype in Koreans (control frequency 0.01) was significantly lower than that in the Western control group (WTCCC frequency 0.124). Therefore, the 46 haplotype may be more significantly associated with PV and ET, and the V617F mutation, in Korean patients.

We performed WGS analysis using tumor-control paired samples in six Korean ET patients. Five candidate genes—*CALR*, *MPDZ*, *IQCH*, *C19orf33*, and *C4orf47*—were confirmed by direct sequence analysis in each index patient. These variants included a *CALR* exon 9 mutation, which was recently identified as specific to V617F-negative MPN. During the mutation-screening process, we discovered 69 additional ET patients (including four index patients), who carried the wild-type *JAK2*. Surprisingly, all of the four index patients also carried *CALR* mutations. An innate low

sensitivity to insertion and deletion may be responsible for this. All three patients lacking the *CALR* mutation had a 52-bp deletion, c.1092_1143del (p. L367 fs*46), which was overlooked by both WGS platforms.

Recently, a somatic *CALR* mutation has been reported in ET and PMF patients, not carrying *JAK2* or *MPL* mutations (26–30). *CALR* is a highly conserved, multicompartamental, and functionally complex calcium-binding chaperone, in the lumen of the endoplasmic reticulum (ER). It forms a key component of the ER quality control machinery, ensuring proper glycoprotein folding and calcium homeostasis. Klampfl et al. identified novel exon 9 mutations of *CALR* in 67% and 88% of patients (with wild-type *JAK2* and *MPL*) with ET and PMF, respectively (26). Moreover, Nangalia et al. reported somatic *CALR* mutations in 70–84% of samples of MPN with wild-type *JAK2* (28). Somatic mutations in *CALR* are not found in PV, and they are mutually exclusive of *JAK2* and *MPL* mutations. In addition, Rumi et al. reported that the mutant allele burden of *CALR*-mutated patients was significantly higher than that of *JAK2*-mutated patients (32% vs. 18%; $P < 0.01$) (31). All *CALR* mutations are insertions or deletions resulting in frameshifts, and they cluster in the last exon. To date, >50 different types of mutations in *CALR* have been reported. A 52-bp deletion (type 1) and a 5-bp insertion (type 2) are the most prevalent and are found in >80% of patients with mutant *CALR* (32). The predicted C-terminal domain of mutant *CALR* is basic, lacks multiple calcium-binding sites, and has no KDEL sequence, which raises the possibility of cellular mislocalization.

Among the 69 Korean ET patients with wild-type *JAK2* and *MPL*, 58.0% carried *CALR* mutations. Although the types and distribution of the *CALR* mutations were similar to previous reports, their frequencies were slightly lower than that observed in a Caucasian group. However, the prevalence of the *CALR* mutations in Chinese patients was 52.7% (99/188) in *JAK2/MPL* mutation-negative ET (33), which was similar to that in the Korean ET patients. Thus, different races or ethnicities may influence the *CALR* mutation frequency, and direct sequence analysis might overlook low levels of mutants. Although complex insertion and/or deletions make it difficult to use sensitive assays, other precise detection methods can be considered.

Additionally, Klampfl et al. (26) found that transfection of mutant *CALR* into a Ba/F3 murine cell line induced IL-3-independent growth, which was associated with increased STAT5 phosphorylation. The most common *CALR* mutant (type 1) increased STAT5 phosphorylation and induced cytokine-independent

growth that could be blocked by a JAK inhibitor. Furthermore, *CALR*-mutant MPN patients demonstrated upregulation of JAK2, STAT1, EZH2, SUZ12, and KRAS expression (34). Collectively, these results suggest that *CALR* and *JAK2* mutations may have similar consequences, and the JAK-STAT pathway is central to the pathogenesis of MPN. However, the mechanism by which mutant *CALR* activates JAK-STAT signaling remains unclear (26).

In ET, *CALR* mutations have been correlated with male gender, younger age, lower leukocyte count, lower Hb level, higher platelet count, and thrombosis-free survival (29, 31). Moreover, *CALR*-mutated ET patients showed less frequent polycythemic, myelofibrotic, or leukemic transformation than those with *JAK2*-mutated ET (31). In this study, the clinical characteristics of *CALR*-mutated ET patients were different from those of *JAK2*-mutated ET patients. The former group had lower Hb/Hct levels, higher platelet counts, and a lower frequency of use of hydroxyurea. There was no male predominance in Korean *CALR*-mutated ET patients, which is in agreement with a report of Chinese patients (33). In the V617F-negative patients, comparison of *CALR*-mutated vs. wild-type *CALR* did not reveal any difference in clinical and laboratory parameters. ET patients carrying wild-type *JAK2*, with or without a *CALR* mutation, might display other common factors for disease phenotypes including an additional somatic mutation, or a germline modifier. In a collaborative study of 1027 ET patients, Tefferi et al. reported that a type 1, 52-bp deletion was associated with male predominance, while a type 2, 5-bp insertion correlated with a younger age (30). The platelet count, in particular, was significantly higher in type 2 vs. type 1 mutated patients. Additionally, type 2 patients in this study showed markedly higher platelet counts than type 1 patients.

In this study, four other candidate mutations existed concurrently with a *CALR* mutation. ET2 and ET3 patients showed three mutations in the *CALR/MPDZ/C19orf33* and *CALR/IQCH/C4orf47* genes, respectively. We excluded two open reading frame mutations during screening, due to difficulty in interpreting their roles in cancer pathogenesis (35). The candidate variants in *MPDZ* and *IQCH* were screened in 69 V617F-negative ET patients, and no recurrence was observed. However, we advise caution in drawing such conclusions, since our screened set of *CALR*-mutated patients was not large enough for validation.

MPDZ is a protein-coding gene and associated with congenital communicating hydrocephalus and leber congenital amaurosis. The *CALR/MPDZ/C19orf33*-mutated patients had normocellular bone marrow sustained thrombocytosis in peripheral blood; increased

number of large, mature megakaryocytes in bone marrow; and no evidence of *BCR-ABL* rearrangement, which is compatible with the diagnosis of ET. Considering the previously reported nonsense mutation in *MPDZ* p.R1044X in melanoma samples (35), the importance of this mutation should not be overlooked. Next, *IQCH* is a protein-coding gene located in chromosome 15q23. High expression of *IQCH* in platelets was reported (36). The *CALR/IQCH/C4orf47*-mutated patients showed a marked increase in platelet count after discontinuing the use of hydroxyurea and no splenomegaly. However, the relation between these somatic mutations and specific clinical manifestations remains unclear.

In summary, this study is the first comprehensive investigation of the genetic characteristics of Korean patients with PV and ET. We found that somatic mutations as well as the 46 haplotype might contribute to the pathogenesis of PV and ET in Korean patients. *CALR* mutations in V617F-negative ET patients have been reported by other groups, but we also identified this mutation in our WGS data. The *CALR* mutation types and frequencies were not significantly different from those of previous studies in Caucasian patients and were similar to those of Chinese patients. ET patients carrying the *CALR* mutation showed higher platelet counts, lower Hb/Hct levels and less frequent use of hydroxyurea compared with ET patients carrying *JAK2* V617F.

REFERENCES

1. Tefferi A, Lasho TL, Schwager SM, et al. The clinical phenotype of wild-type, heterozygous, and homozygous *JAK2*V617F in polycythemia vera. *Cancer* 2006;106:631–635.
2. Verstovsek S, Silver RT, Cross NC, Tefferi A. *JAK2*V617F mutational frequency in polycythemia vera: 100%, >90%, less? *Leukemia* 2006;20:2067.
3. Oh ST, Simonds EF, Jones C, et al. Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. *Blood* 2010;116:988–992.
4. Pardanani AD, Levine RL, Lasho T, et al. *MPL*S151 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood* 2006;108:3472–3476.
5. Pietra D, Li S, Brisci A, et al. Somatic mutations of *JAK2* exon 12 in patients with *JAK2* (V617F)-negative myeloproliferative disorders. *Blood* 2008;111:1686–1689.
6. Pikman Y, Lee BH, Mercher T, et al. *MPL*W515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med* 2006;3:e270.
7. Scott LM. The *JAK2* exon 12 mutations: a comprehensive review. *Am J Hematol* 2011;86:668–676.
8. Delhommeau F, Dupont S, Della Valle V, et al. Mutation in *TET2* in myeloid cancers. *N Engl J Med* 2009;360:2289–2301.
9. Stegelmann F, Bullinger L, Schlenk RF, et al. *DNMT3A* mutations in myeloproliferative neoplasms. *Leukemia* 2011;25:1217–1219.
10. Pardanani A, Lasho TL, Finke CM, Mai M, McClure RF, Tefferi A. *IDH1* and *IDH2* mutation analysis in chronic- and blast-phase myeloproliferative neoplasms. *Leukemia* 2010;24:1146–1151.
11. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene *EZH2* in myeloid disorders. *Nat Genet* 2010;42:722–726.
12. Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer* 2012;12:599–612.
13. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. *Nature* 2011;470:214–220.
14. Gui Y, Guo G, Huang Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* 2011;43:875–878.
15. Guo G, Gui Y, Gao S, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat Genet* 2012;44:17–19.
16. Jones AV, Chase A, Silver RT, et al. *JAK2* haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet* 2009;41:446–449.
17. 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;41:455–459.
18. Olcaydu D, Harutyunyan A, Jager R, et al. A common *JAK2* haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet* 2009;41:450–454.
19. Jones AV, Campbell PJ, Beer PA, et al. The *JAK2* 46/1 haplotype predisposes to *MPL*-mutated myeloproliferative neoplasms. *Blood* 2010;115:4517–4523.
20. Olcaydu D, Skoda RC, Looser R, et al. The ‘GGCC’ haplotype of *JAK2* confers susceptibility to *JAK2* exon 12 mutation-positive polycythemia vera. *Leukemia* 2009;23:1924–1926.
21. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–120.
22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
23. 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.
24. Tanaka M, Yujiri T, Ito S, et al. *JAK2* 46/1 haplotype is associated with *JAK2* V617F-positive myeloproliferative neoplasms in Japanese patients. *Int J Hematol* 2013;97:409–413.
25. Zhang X, Hu T, Wu Z, Kang Z, Liu W, Guan M. The *JAK2* 46/1 haplotype is a risk factor for myeloproliferative neoplasms in Chinese patients. *Int J Hematol* 2012;96:611–616.
26. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med* 2013;369:2379–2390.
27. Lundberg P, Karow A, Nienhold R, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* 2014;123:2220–2228.
28. Nangalia J, Massie CE, Baxter EJ, et al. Somatic *CALR* mutations in myeloproliferative neoplasms with nonmutated *JAK2*. *N Engl J Med* 2013;369:2391–2405.
29. Rotunno G, Mannarelli C, Guglielmelli P, et al. Impact of calreticulin mutations on clinical and hematological phenotype and outcome in essential thrombocythemia. *Blood* 2014;123:1552–1555.

30. Tefferi A, Pardanani A. Genetics: CALR mutations and a new diagnostic algorithm for MPN. *Nat Rev Clin Oncol* 2014; 11:125–126.
31. Rumi E, Pietra D, Ferretti V, et al. JAK2 or CALR mutation status defines subtypes of essential thrombocythemia with substantially different clinical course and outcomes. *Blood* 2014;123:1544–1551.
32. Cazzola M, Kralovics R. From Janus kinase 2 to calreticulin: the clinically relevant genomic landscape of myeloproliferative neoplasms. *Blood* 2014;123:3714–3719.
33. Fu R, Xuan M, Zhou Y, et al. Analysis of calreticulin mutations in Chinese patients with essential thrombocythemia: clinical implications in diagnosis, prognosis and treatment. *Leukemia* 2014;28: 1912–1914.
34. Rampal R, Al-Shahrour F, Abdel-Wahab O, et al. Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. *Blood* 2014;123:e123–e133.
35. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015;43:D805–D811. Available at: <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic>. Accessed on May 15, 2014.
36. Wang M, Weiss M, Simonovic M, et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics* 2012;11:492–500. Available at: <http://www.pax-db.org>. Accessed on May 15, 2014.