의학박사 학위논문

골수증식종양과 그 이차 이환의 역학 및 관련된 유전자변이의 임상적 의의

Epidemiology and implications of mutation profiling in myeloproliferative neoplasms and their secondary transformations

2019년 2월

서울대학교 대학원
의학과 중개의학전공
변 자 민

A thesis of the Doctor of Philosophy degree

Epidemiology and implications of mutation profiling in myeloproliferative neoplasms and their secondary transformations

골수증식종양과 그 이차 이환의 역학 및 관련된 유전자변이의 임상적 의의

February 2019

Department of Medicine
Translational Medicine Major
Seoul National University College of Medicine
Ja Min Byun

Abstract

Introduction: Myeloproliferative neoplasms (MPN) are a group of hematopoietic stem cell disorders characterized by aberrant clonal proliferation and a tendency towards secondary transformation. Polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF) comprise the three major subgroups of BCR-ABL1 negative MPN. The discovery of the JAK2 V617F mutation has allowed for the development of novel therapeutic agents and has encouraged the efforts in molecular diagnostics for MPN. Such efforts led to evidence of clonality and mutational events preceding the acquisition of JAK2 V617F, and identification of recurrent mutations in genes distinct from JAK2 including MPL and CALR. Despite the better understanding of the disease, however, a subset of MPN patients experience secondary transformation leading to dismal prognosis and the mechanisms that contribute to transformation is not as well delineated. For this study, we first delineated the epidemiology of secondary transformation of MPN. Then, using next-generation sequencing, we have defined the mutational profile of a cohort of 41 patients and the impact of the somatic mutations on clinical outcome.

Methods: For population study, we used the National Health Insurance and Korean Health Insurance Review and Assessment Service database. For genetic study, we analyzed the mutational status of 17 polycythemia vera, 16 essential thrombocythemia, 8 primary myelofibrosis who tested positive for JAK2 V617F by PCR.

Results: In a population study, the 5-year cumulative incidence of transformation to secondary myelofibrosis and acute myeloid leukemia

were 1.4% and 1.91% in ET patients, 0.69% and 0.97% in PV patients, and not applicable and 16.54% in PMF patients, respectively.

For genetic study, during a median follow-up of 39.2 months, 4 leukemic transformations (9.8%) and 3 secondary myelofibrotic transformations (7.3%) were documented. Interestingly, splicing genes were more frequently mutated in PMF, resembling myelodysplastic syndrome. On the other hand, mutations in epigenetic regulators were significantly more frequent in PV. When prognostic implications of mutations were analyzed, we found that TP53 mutation was associated with shorter OS (P=0.028), and patients with splicing gene mutations were associated with reduced overall survival (P=0.028) and increased risk of secondary transformation (P=0.035). ASXL1 mutation was also related to increased risk of secondary transformation (P=0.014). Notably, all of the patients with secondary leukemia had at least one mutation of acute myeloid leukemia drivers while all of the patients who transformed into secondary MF harbored at least one mutation in tumor suppressor genes, suggesting different genes are involved in each scenario. Lastly, we provide evidence that most putative driver somatic mutations are already present at the time of MPN diagnosis in small clones.

Conclusions: Our findings provide a better understanding of the mutational landscape of MPN in their chronic state and secondary transformed state.

Keywords: Myeloproliferative neoplasms; Secondary myelofibrosis; Secondary acute myeloid leukemia; Next-generation sequencing; Epidemiology; Mutation profiling

Student Number: 2015 – 30899

CONTENTS

LIST OF TABLES AND FIGURES

LIST OF ABBREVIATIONS

JAK2;	Janus kinase 2
MPL;	Myeloproliferative leukemia virus oncogene
CALR;	Calreticulin
PCR;	polymerase chain reaction
VAF;	variant allele frequency
TP53;	Tumor protein 53
ASXL1;	Additional sex combs-like 1

Introduction

Myeloproliferative neoplasms (MPN) represent a heterogeneous group of clonal hematopoietic disorders characterized by excessive terminally differentiated myeloid cells [1,2]. Polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF) comprise the three major subgroups of BCR-ABL1 negative MPN [3]. Although most MPN patients have a relatively indolent clinical course, some patients experience progression to secondary myelofibrosis (sMF) and transformation to acute myeloid leukemia (AML) [4]. Moreover, others suffer from thromboembolic events and/or various constitutional symptoms ranging from organomegaly to cytopenias, which can greatly compromise the quality of life [5,6]. All in all, with expected increment in number, MPNs impose a cumulative threat to public health, inducing substantial economic and social burdens.

To this end, we carried out this study in two parts to provide a better overall understanding of MPN. The first part of this study focused on descriptive epidemiology of MPN, as such information provides an understanding of the disease secular trends. We conducted a nationwide population-based study for comparative analyses of prevalence, incidence, crude medical costs and secondary transformation rates of MPN in Korea. The second part of this study focused on the genetic

landscapes of MPN. The main questions addressed were (1) why different subtypes of MPN show different clinical manifestations despite harboring similar genetic mutations; and (2) what are the risk factors of secondary transformations. In recent years, there were significant advances in understanding of the pathophysiology of MPN. Specifically, the discovery of the *JAK2 V617F* mutation, which is mutated in approximately 63% of MPN patients [7,8,9], was a huge breakthrough that allowed for the development of novel therapeutic agents and further efforts in molecular diagnostics for MPN. Such efforts led to evidence of clonality and mutational events preceding the acquisition of *JAK2 V617F*, and identification of recurrent mutations in genes distinct from *JAK2* including myeloproliferative leukemia virus oncogene (*MPL*) [10] and calreticulin (*CALR*) [11]. Unfortunately, despite the better understanding of the disease, certain issues remained unanswered. For example, we are yet to understand why some patients present as ET while others as PV or PMF despite harboring *JAK2 V617F* mutation as a common denominator. Also, even with the use of *JAK2* inhibitors a subset of MPN patients still experience secondary transformation leading to dismal prognosis, but the mechanisms that contribute to transformation is not as well delineated [12,13]. To address these questions, we focused on BCR-ABL1 negative, *JAK2 V617F* positive MPN patients and using next generation sequencing (NGS) we attempted to define the mutational profiles and the impact of additional somatic mutations in relation to *JAK2* on clinical

outcomes.

Material and Methods

Population study: data source

In Korea, medical insurance is in the form of universal public insurance. The National Health Insurance (NHI) program, operated by the Korean Ministry for Health and Welfare, is the sole and mandatory insurance system that covers approximately 98% of the overall Korean population [14,15]. The ranges of insurance coverage are comprehensive; it includes routine health checkups, diagnostic and therapeutic interventions, preventative care, hospitalization, and rehabilitation. On the other hand, the Korean Health Insurance Review and Assessment Service (HIRA) is a government affiliated organization responsible for insurance claims review and quality control of the NHI. Since the Korean population itself is fairly ethnically homogenous, both the NHI and HIRA database can be readily used for nationwide analyses. For the epidemiology part of this study, we utilized information from the HIRA and NHI database between January 2004 and December 2013.

Population study: case identification

Using HIRA and NHI claims database, patients with MPNs were retrospectively identified using the Korean Classification of Diseases (KCD). Although International Classification of Disease (ICD-10) is the

widely used reference, many countries use adapted versions according to their circumstances. KCD is essentially the same as ICD, except KCD codes are made up of five-digit codes for some diseases to clarify the conditions of patients. From 2004 to 2007, KCD 4th edition based on 2001 World Health Organization (WHO) classification was applied. For KCD-4, the codes for ET are D47.3 plus D75.2 and the code for PV is D45. From 2008 to 2010, KCD 5th edition based on 2008 WHO classification was used. For KCD-5, the codes remained the same as KCD-4. From 2011 to 2013, KCD 6th edition, also based on 2008 WHO classification, was employed. For KCD-6, the code for ET is D47.3, for PV D45, and for PMF D47.4. Since the code for PMF was established in 2011, data regarding PMF was available from 2011. If cases identified as PMF had the diagnosis of PV or ET in the previous years, they were counted as PV or ET cases, respectively. These codes were used in both the main diagnosis and sub-diagnosis for querying.

Population study: statistical analysis

The population size of each group was obtained from the annual reports by Statistics Korea (www.kostat.go.kr). The annual prevalence and incidence were standardized using the population structure data from the respective year, which was produced by a population census. The data with Hb <3g/dL or 30 >g/dL were eliminated regarding as clerical error. As for the costs associated with MPN, all claims made under the

diagnosis of PV, ET and PMF have been tallied to estimate the total amount and the patterns. The data on costs were accumulated in Korea Won, then converted into US dollars based on an exchange rate of Korean Won 1200 = US $1. All the analyses were performed using SAS v9.4 (SAS Institute Inc., Cary, NC, USA).

Genetic study: patient samples

This part of the study was carried out at Seoul National University Hospital during the period between May 2010 and April 2016. Patients diagnosed with MPN by cyto-morphological evaluation of their bone marrow samples according to 2008 WHO criteria, negative for BCR-ABL1 fusion as assessed by polymerase chain reaction (PCR) and positive for *JAK2 V617F* by PCR were deemed eligible for analysis. For sequencing, DNA extracted from formalin-fixed paraffin embedded bone marrow aspirates were quantified. The DNA collected at the time of MPN diagnosis was analyzed using whole genome sequencing (WGS) and target gene sequencing, while the DNA collected at the time of leukemia diagnosis was analyzed using WGS only. Saliva DNA collected at the time of remission was used as matched normal sample.

Genetic study: next generation sequencing

A total of 47 target genes providing diagnostic information in

myeloid malignancies were selected based on literature evidence: *CALR, MYC, ETV6, CEBPA, MLL, BRCA2, MPL, SETBP1, PTPN11, ARID1A, ARID2, ATM, BCOR, CTNNA1, EPHB1, FANCA, LRP1B, MEN1, KMT2D, MSH2, NF1, PHF6, PTCH1, RB1, SMARCA4, SUZ12, SH2B2, IKZF1, JAK2, NRAS, KRAS, FLT3, CBL, RUNX1, NPM1, TP53, TET2, ASXL1, EZH2, IDH1/2, UTX, DNMT3A, U2AF1, ZRSR2, SF3B1,* and *SRSF2* **(Table 1)**. Targeted sequencing was performed with a customized design: TruSeq® Custom Amplicon (Illumina, San Diego, CA) using the MiSeq® sequencing platform (Illumina). TruSeq Custom Amplicon is a fully integrated end-to-end amplicon sequencing solution, including online probe design, assay and sequencing. Online probe design was performed by entering into the Design Studio software (Illumina) the 47 genes. Once the design is completed, TruSeq Custom Amplicon kit produces the required targeted amplicons with the necessary adapters and indices for sequencing on the MiSeq® system without any additional processing. Library preparation and sequencing runs have been performed according to the manufacturer's procedure. Quantified libraries were sequenced using the 2 × 150 bp configuration (300 cycles) and run on V2 sequencing flow cell. Final targeted sequencing panel consisting 2228 amplicons covering 47 genes was produced. The mean coverage depth was 1056X.

As for WGS, paired-end sequencing was performed using the Illumina HiSeq® platform with 100bp read length (Illumina, San Diego,

CA). For the MPN and AML sample, the mean coverage depth was 60.2X and for the normal sample, 37.5X.

After sequencing reads produced, raw de-multiplexed reads from the MiSeq® sequencer were aligned to the reference human genome (UCSC build hg19) using the Burrows-Wheeler Aligner (BWA) [17], running in paired-end mode. To ensure a good call quality and to reduce the number of false positives, samples underwent Base Quality Score Recalibration (BQSR), using the Genome Analysis Toolkit (GATK) [18]. Putative somatic variant calls were detected with VarScan 2 [19] and subsequently filtered to exclude non-somatic calls, with an allelic fraction less than 1%, or with a read depth less than 10. Variant calls were annotated with biological information using ANNOVAR [19]. Mutations were annotated with the 1000 Genomes project, dbSNP (version 138) and Catalogue of Somatic Mutations in Cancer (COSMIC), version 68. We only focused on mutations with the exon-based region and filtered out all synonymous and unknown genes after annotation (**Fig. 1**).

The validity of the somatic mutations was checked against the publicly accessible COSMIC v68 database and functional interpretation was performed using SIFT 1.03, PolyPhen 2.0 and ClinVar. Single-nucleotide polymorphisms (SNP) were annotated according to the NCBI dbSNP database.

Genetic study: statistical analysis

Differences between groups were assessed using a Student's t-test or one-way analysis of variance for continuous variables, and Pearson chi-square test for categorical variables, as indicated. The overall survival (OS) and transformation free survival (TFS) curves were estimated using the Kaplan-Meier method. OS was defined as the time from MPN diagnosis to death or the last follow-up. TFS as the duration from the date of diagnosis to disease transformation or death. All data were analyzed using the Statistical Package for the Social Sciences software (IBM® SPSS®Statistics, version 22.0). P values of < 0.05 were considered statistically significant.

Ethics

This study was approved by the Institutional Review Board of Seoul National University Hospital (approval number: IRB No. H-1607-198-782) and was conducted in accordance with the Principles of the Declaration of Helsinki.

Table 1. Myeloproliferative neoplasm target sequencing information.

Gene	Chr	Start	End	Length
KRAS	chr12	25358179	25403854	45675
NRAS	chr1	115247084	115259515	12431
CEBPA	chr19	33790839	33793430	2591
IDH1	chr2	209100952	209119806	18854
ETV6	chr12	11802787	12048325	245538
SETBP1	chr18	42260137	42648475	388338
MYC	chr8	128748314	128753680	5366
ZRSR2	chrX	15808573	15841382	32809
EPHB1	chr3	134514098	134979307	465209
SRSF2	chr17	74730196	74733493	3297
SH2B2	chr7	101928404	101962178	33774
TP53	chr17	7571719	7590868	19149
NF1	chr17	29421944	29704695	282751
BRCA2	chr13	32889616	32973809	84193
NPM1	chr5	170814707	170837888	23181
FLT3	chr13	28577410	28674729	97319
TET2	chr4	106067031	106200960	133929
MPL	chr1	43803474	43820135	16661
ASXL1	chr20	30946146	31027122	80976
DNMT3A	chr2	25455829	25565459	109630
EZH2	chr7	148504463	148581441	76978
IDH2	chr15	90627211	90645708	18497
MLL	chr11	118307204	118397539	90335
PTPN11	chr12	112856535	112947717	91182
RUNX1	chr21	36160097	36421595	261498
PHF6	chrX	133507341	133562822	55481
MSH2	chr2	47630205	47710367	80162
CBL	chr11	119076985	119178859	101874
ARID2	chr12	46123619	46301819	178200
ARID1A	chr1	27022521	27108601	86080
SF3B1	chr2	198256697	198299771	43074
ATM	chr11	108093558	108239826	146268
FANCA	chr16	89803958	89883065	79107

PTCH1	chr9	98205263	98279247	73984
RB1	chr13	48877882	49056026	178144
KDM6A	chrX	44732422	44971845	239423
U2AF1	chr21	44513065	44527688	14623
MLL2	chr12	49412757	49449107	36350
MEN1	chr11	64570985	64578766	7781
SMARCA4	chr19	11071597	11172958	101361
BCOR	chrX	39910498	40036582	126084
CTNNA1	chr5	138089106	138270723	181617
IKZF1	chr7	50344377	50472798	128421
LRP1B	chr2	140988995	142889270	1900275
JAK2	chr9	4985244	5128183	142939
SUZ12	chr17	30264043	30328057	64014
CALR	chr19	13049413	13055304	5891
TP53	chr17	7576534	7577160	626
TP53	chr17	7578171	7578559	388
TP53	chr17	7579306	7579917	611
NF1	chr17	29559712	29560236	524
NF1	chr17	29663345	29663937	592
NF1	chr17	29683972	29684392	420
NF1	chr17	29562623	29563044	421
NF1	chr17	29664831	29665162	331
NF1	chr17	29548862	29549010	148
ARID1A	chr1	27098985	27099483	498
ARID1A	chr1	27099831	27100394	563
ARID1A	chr1	27092706	27093062	356
ARID1A	chr1	27100814	27101716	902
ARID2	chr12	46123614	46123925	311
ARID2	chr12	46231098	46231495	397
ARID2	chr12	46287197	46287509	312
ARID2	chr12	46230366	46230779	413
ARID2	chr12	46285557	46285884	327
SMARCA4	chr19	11113699	11114078	379
SMARCA4	chr19	11168925	11169044	119
SMARCA4	chr19	11106883	11107225	342
SMARCA4	chr19	11151968	11152241	273
SMARCA4	chr19	11170423	11170868	445
LRP1B	chr2	141459704	141460127	423

LRP1B	chr2	141201893	141202253	360
SF3B1	chr2	198264773	198265163	390
SF3B1	chr2	198267274	198267764	490
SF3B1	chr2	198269794	198270201	407
SF3B1	chr2	198266460	198266859	399
FLT3	chr13	28608018	28608549	531
FLT3	chr13	28623515	28623916	401
NPM1	chr5	170819708	170819987	279
EZH2	chr7	148506157	148506487	330
EZH2	chr7	148543556	148543695	139
EZH2	chr7	148523540	148523729	189
ATM	chr11	108143253	108143584	331
ATM	chr11	108098346	108098620	274
ATM	chr11	108235803	108236237	434
ATM	chr11	108141785	108142138	353
ATM	chr11	108186544	108186845	301
MLL2	chr12	49447755	49448814	1059
MLL2	chr12	49418355	49418734	379
MLL2	chr12	49435694	49436118	424
MLL2	chr12	49437977	49438310	333
MLL2	chr12	49421580	49421929	349
MLL2	chr12	49437412	49437786	374
MLL2	chr12	49422605	49423264	659
MLL2	chr12	49443459	49446497	3038
MLL2	chr12	49446692	49447429	737
MLL2	chr12	49415560	49416141	581
MLL2	chr12	49424057	49428723	4666
MLL2	chr12	49439697	49440578	881
MLL2	chr12	49432999	49435493	2494
MLL2	chr12	49436338	49437216	878
MLL	chr11	118361905	118362038	133
MLL	chr11	118390327	118390784	457
PTPN11	chr12	112915449	112915824	375
PTPN11	chr12	112924273	112924439	166
CBL	chr11	119155673	119156281	608
MPL	chr1	43803514	43803907	393
MPL	chr1	43812110	43812610	500
MEN1	chr11	64574477	64574696	219

MEN1	chr11	64577116	64577586	470
DNMT3A	chr2	25466761	25467212	451
DNMT3A	chr2	25505254	25505585	331
DNMT3A	chr2	25463165	25463604	439
DNMT3A	chr2	25468883	25469183	300
ASXL1	chr20	31015925	31016230	305
ASXL1	chr20	31019118	31019487	369
RUNX1	chr21	36259134	36259414	280
IDH2	chr15	90628042	90628624	582
IDH2	chr15	90630338	90630812	474
IDH2	chr15	90631585	90631984	399
RB1	chr13	49039128	49039509	381
RB1	chr13	48954183	48954382	199
FANCA	chr16	89805006	89805966	960
FANCA	chr16	89865568	89865645	77
FANCA	chr16	89812986	89813303	317
FANCA	chr16	89845203	89845416	213
FANCA	chr16	89849261	89849515	254
FANCA	chr16	89877109	89877484	375
FANCA	chr16	89836239	89836672	433
SUZ12	chr17	30267299	30267510	211
CTNNA1	chr5	138266156	138266629	473
PHF6	chrX	133549040	133549254	214
KDM6A	chrX	44941815	44942039	224
KDM6A	chrX	44918245	44918716	471
BRCA2	chr13	32900232	32900424	192
BRCA2	chr13	32953881	32954287	406
PTCH1	chr9	98244225	98244490	265
PTCH1	chr9	98278745	98279107	362
PTCH1	chr9	98247961	98248161	200
BCOR	chrX	39922855	39923210	355
U2AF1	chr21	44514575	44514903	328
PHF6	chrX	133547512	133548001	489
CALR	chr19	13049942	13050450	508
CALR	chr19	13050861	13051706	845
CALR	chr19	13054345	13054729	384
TET2	chr4	106155094	106158599	3505
JAK2	chr9	5080223	5080688	465

JAK2	chr9	5090440	5090916	476
MSH2	chr2	47630325	47630546	221
IKZF1	chr7	50459421	50459566	145
U2AF1	chr21	44515542	44515858	316

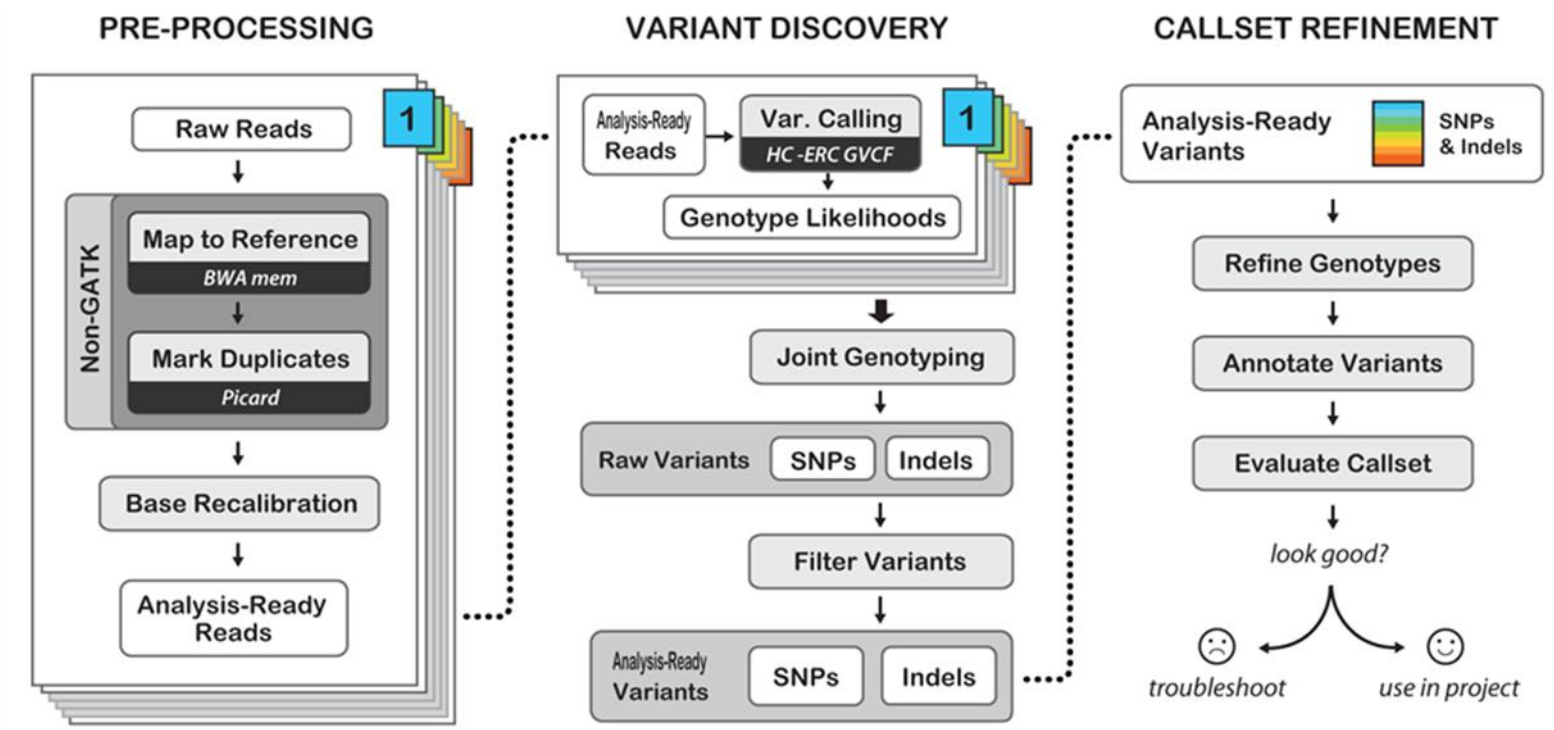


Figure 1. Process of data generation.

Results

Prevalence and incidence

Prevalence steadily increased over time for MPNs as a whole, as well as for all subgroups in MPNs. Prevalence was the highest for ET (range 4.1-9.0 per 100,000), followed by PV (range 2.8-5.4 per 100,000) and PMF (range 0.5 to 0.9 per 100,000) (**Table 2** and **Fig. 2**). The mean number of prevalent cases of ET was 3,442, with a female predominance. ET prevalence almost doubled during the study period, from 4.1 in 2004 to 9.0 in 2013. Similar pattern was evident for PV, with its prevalence increasing from 2.8 in 2004 to 5.4 in 2013. The mean number of prevalent cases of PV was 2 044, with male predominance. The annual prevalence of all MPNs, only available from the year 2011 and onwards due to reasons mentioned above, was 12.463 in 2011, 14.061 in 2012 and 15.291 in 2013.

Incidence rates were similar by year. The estimated incidence rates were between 2.0 to 3.0 per 100,000 per year for ET, and 1.0 to 1.5 per 100,000 per year for PV. The incidence of PMF ranged from 0.3 to 0.5 per 100,000 per year. More women were newly diagnosed (incident cases) with ET. On the other hand, fewer women were diagnosed with PV (**Table 2**)

Table 2. Annual incidence and prevalence of myeloproliferative neoplasms in the Korean population.

Cohort	Year	ET				PV				PMF			
		N(M,%)	M	F	Overall	*N*(M,%)	M	F	Overall	*N*(M,%)	M	F	Overall
Incidence													
	2004	975(47.0)	1.880	2.135	2.007	626 (73.5)	1.888	0.685	1.288				
	2005	1147 (46.7)	2.192	2.512	2.351	644 (74.5)	1.963	0.674	1.320				
	2006	1265 (46.3)	2.386	2.779	2.582	712 (68.1)	1.975	0.929	1.453				
	2007	1386 (45.3)	2.543	3.084	2.813	601 (72.9)	1.774	0.663	1.220				
	2008	1491 (44.8)	2.691	3.330	3.010	484 (70.7)	1.378	0.574	0.977				
	2009	1329 (42.6)	2.270	3.071	2.670	521 (68.7)	1.436	0.656	1.047				
	2010	1180 (43.6)	2.031	2.642	2.336	499 (71.7)	1.414	0.559	0.988				
	2011	1003 (45.2)	1.783	2.172	1.977	505 (68.7)	1.366	0.624	0.995	277 (59.6)	0.649	0.442	0.546
	2012	1151 (42.7)	1.929	2.590	2.259	579 (72.5)	1.647	0.625	1.136	192 (51.0)	0.384	0.369	0.377
	2013	1154 (41.6)	1.876	2.638	2.256	609 (71.1)	1.692	0.689	1.191	180 (62.2)	0.438	0.266	0.352
	Mean	1208 (44.5)	2.158	2.695	2.426	578 (71.3)	1.653	0.668	1.162	216 (57.9)	0.490	0.359	0.425
Prevalence													
	2004	1997 (45.1)	3.698	4.525	4.110	1379 (63.1)	3.571	2.102	2.838				
	2005	2427 (44.5)	4.420	5.533	4.975	1568 (64.3)	4.126	2.298	3.214				
	2006	2818 (44.4)	5.094	6.413	5.752	1839 (62.8)	4.699	2.803	3.754				
	2007	3239 (44.0)	5.775	7.377	6.574	1872 (64.0)	4.852	2.742	3.800				
	2008	3660 (43.7)	6.438	8.342	7.388	1897 (63.0)	4.818	2.836	3.829				
	2009	3812 (42.8)	6.546	8.775	7.659	2062 (62.3)	5.150	3.132	4.143				
	2010	3931 (43.1)	6.701	8.867	7.782	2152 (61.6)	5.239	3.277	4.260				

	2011	3720 (42.7)	6.254	8.414	7.332	2326 (61.2)	5.605	3.561	4.585	277 (59.6)	0.649	0.442	0.546
	2012	4217 (42.7)	7.054	9.503	8.277	2570 (61.8)	6.223	3.863	5.044	377 (54.1)	0.800	0.680	0.740
	2013	4597 (42.3)	7.593	10.386	8.989	2779 (61.7)	6.702	4.164	5.434	444 (56.1)	0.973	0.763	0.868
	Mean	3442 (43.3)	5.957	7.814	6.884	2 044 (62.4)	5.099	3.078	4.090	366 (56.3)	0.807	0.628	0.718

Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis, M, male; F, female.

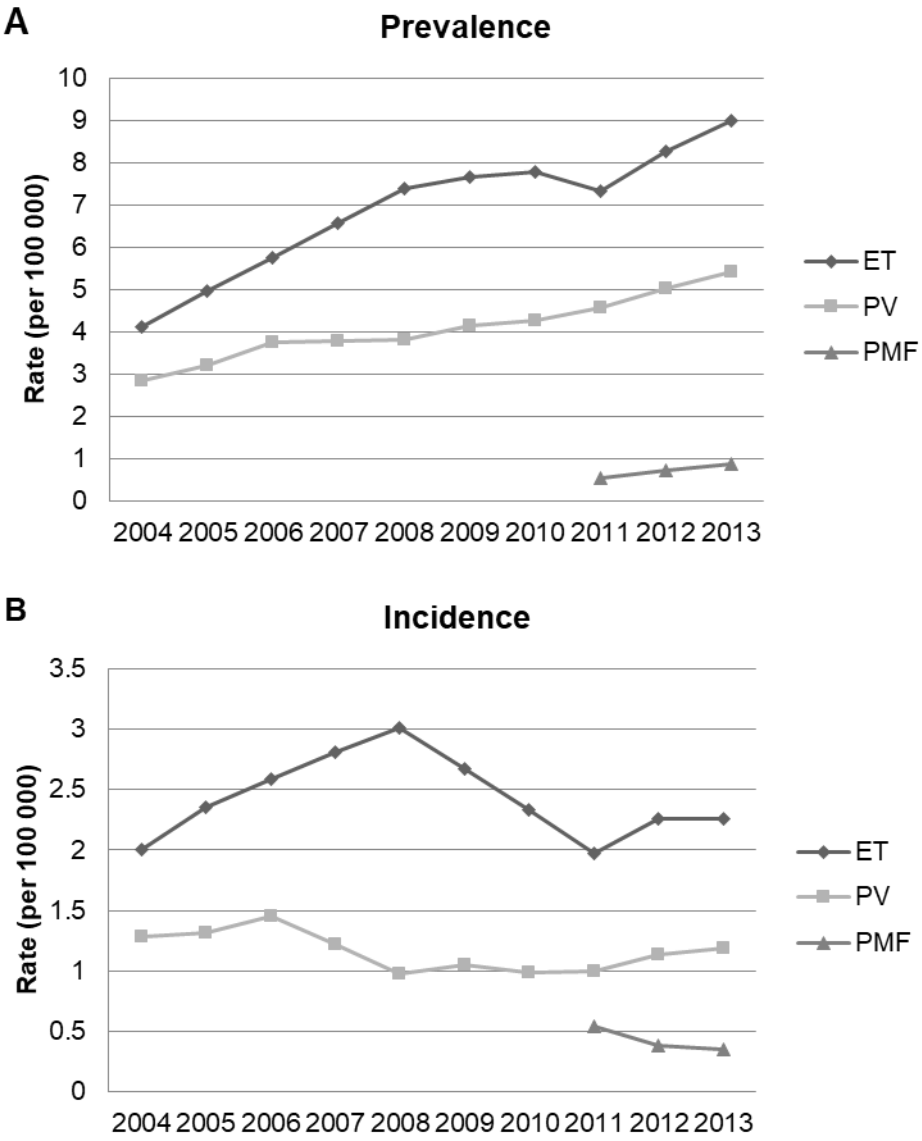


Figure 2. Pattern of myeloproliferative neoplasm (A) prevalence and (B) incidence. Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis.

Age at diagnosis

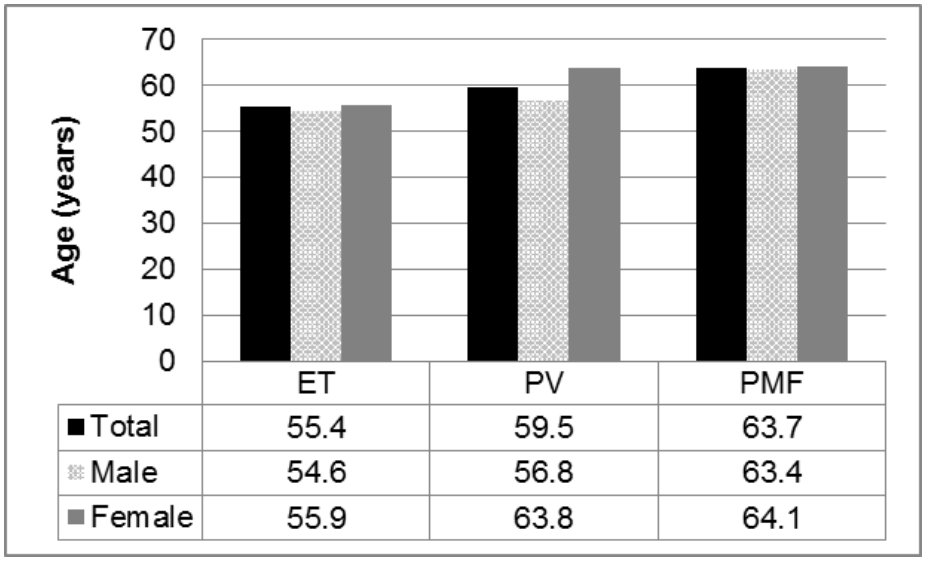
Mean age of ET was 55.4 years (54.6 years for males, 55.9 years for females) and that of PV was 59.5 years (56.8 years for males, 63.8 years for females). Mean age for PMF was the oldest at 63.7 years (63.4 years for males, 64.1 years for females) (**Fig. 3**).

As for newly diagnosed MPNs, mean onset age for ET was youngest at 46.6 years (45.4 years for males, 47.6 years for females), followed by PV at 53.7 years (51.5 years for males, 59.4 years for females) and PMF at 63.3 years (63.1 years for males, 63.6 years for females). Males were predisposed to earlier development of all MPNs compared to females (**Fig. 3**).

Medical cost

Annual cumulative medical cost for MPNs has increased by more than 5 times, from approximately US $ 2.1 million in 2004 to US $ 11.0 million in 2013. During the study period, ET incurred the highest total cost and most frequent hospital visits followed by PV then PMF. On the other hand, PMF incurred the highest cost per person, followed by ET then PV (**Table 3**). Analyses of the increase in the medical cost per patient in the MPN prevalent age group (40's, 50's and 60's) showed that the increase in medical costs related to MPN treatment was largely due to the increase in general healthcare expense. During the 10-year

period, the general healthcare expense per patient for those in their 40's, 50's and 60's increased by 1.51, 1.61, and 1.76 times, respectively. During the same period, disease specific medical cost for PV increased by 1.65, 1.54, and 1.49 times in each age group, and for ET by 1.26, 1.30, and 1.61 times, respectively. During the 3 years between 2011 and 2013 when costs data regarding MF was available, the general healthcare expenses per patient increased by 1.05, 1.02, and 1.01 times for each age group, while disease specific medical cost for MF increased by 3.49, 1.10, and 1.07 times, respectively.



	ET	PV	PMF
Total	55.4	59.5	63.7
Male	54.6	56.8	63.4
Female	55.9	63.8	64.1

Figure 3. Age of myeloproliferative neoplasms patients at diagnosis. Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis.

Table 3. Estimation of medical cost.

Classification	Year	Patients (*N*)	No of encounters (*N*)	Total annual medical cost (US$)*	Annual medical cost per person (US$/person)*	Average annual medical cost per person over the decade (US$/person)*
ET	2004	1997	11894	1408306	705	1035
	2005	2427	14199	1574374	649	
	2006	2818	17207	2406144	854	
	2007	3239	19673	2650972	818	
	2008	3660	24028	3680967	1006	
	2009	3812	26994	3983520	1045	
	2010	3931	28658	4556894	1159	
	2011	3720	29267	4779461	1285	
	2012	4217	36037	5280256	1252	
	2013	4597	39876	5940973	1 292	
	Total	34 418	247 833	35428532	10065	
PV	2004	1379	9297	671118	487	748
	2005	1568	10992	794049	506	
	2006	1839	12392	1020666	555	
	2007	1872	13557	1264484	675	
	2008	1897	15150	1388193	732	
	2009	2062	17195	1723047	836	
	2010	2151	18352	1911130	888	
	2011	2326	20682	2046252	880	
	2012	2570	24366	2268922	883	
	2013	2779	26743	2612665	940	
	Total	20 443	168 726	15700526	7382	
PMF	2011	277	2540	1156302	4174	5000
	2012	377	4688	2190818	5811	
	2013	444	5642	2 142521	4825	
	Total	1098	12 870	5489641	14811	

Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis;
*US $1 is approximately equivalent to Korean Won 1 200

Prevalence and incidence by race/ethnicity

All specified MPNs prevalence was lower among Koreans than North Americans (**Table 4**). Although there was a wide variation regarding European MPNs prevalence data, generally all specified MPNs prevalence was also lower among Koreans. ET was most common in Koreans in contrast to PV in Caucasian populations.

All specified incidence was lower among Koreans (or East Asians) compared to other ethnicities (**Table 5**). Whites showed the highest incidence of MPNs overall (ET 9.7, PV 12.3, and PMF 3.3), and East Asians represented in our study showed the lowest incidence (ET 2.2-2.7, PV 1.1-1.4, and PMF 0.4). East Asians and blacks showed the highest incidence of ET while whites and Hispanics showed highest incidence of PV.

Table 4. Comparative prevalence of myeloproliferative neoplasms.

		Korea (Current study)			Norway*	IHCIS (USA)			MarketScan (USA)			Orphanet (Europe)	RARECARE (Europe)
	Year	*'04-'07*	*'08-'10*	*'11-'13*	*'04-'12*	*'08*	*'09*	*'10*	*'08*	*'09*	*'10*	*Unknown aggregated period*	*'88-'02*
Prevalence	ET	5.4	7.6	8.2	8.6	54.1	56.3	57.0	38.7	43.0	43.7	24.0	4.0
	PV	3.4	4.1	5.0	9.2	56.4	56.3	57.2	45.0	46.7	48.2	30.0	5.0
	PMF	NA	NA	0.7	3.0	1.7	2.0	2.3	1.4	1.3	1.7	2.7†	0.5†

Abbreviation: IHCIS, Integrated Health Care Information Solutions; USA, United States of America; RARECARE, Surveillance of Rare Cancers in Europe; ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis; NA, not available.
*Prevalence is based on 2011 Norwegian population
†Any myelofibrosis (primary and secondary myelofibrosis)

Table 5. Comparative incidence of myeloproliferative neoplasms.

		East Asians(Current study)			Whites	Hispanics	Blacks	Asia/Pacific Islanders
	Year	*'04-'07*	*'08-'10*	*'11-'13*	*'01-'12*	*'01-'12*	*'01-'12*	*'01-'12*
Incidence	ET	2.3	2.7	2.2	9.7	6.4	11.5	7.4
	PV	1.4	1.1	1.1	12.3	7.2	7.5	7.5
	PMF	NA	NA	0.4	3.3	2.2	2.4	2.4

Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis; NA, not available.

Secondary transformation cumulative incidence

To calculate the cumulative incidence (CI) secondary transformation of MPN, cases with an episodic recording of diagnostic codes and those with ambiguous disease status and/or inappropriate data were excluded. At the end, there were a total of 7,471 patients (median age 60 years, range 11-106) who were diagnosed ET (N = 4,405), PV (N = 2,470) or PMF (N = 596), and with appropriate follow-up data to trace the disease course regarding transformation.

Among ET patients (median follow-up duration 46.7 months), 223 patients (5.06%) underwent any transformation during study period (**Table 6**). More specifically, 68 ET patients transformed to sMF, 6 ET patients transformed to sMF then subsequently AML, and 66 ET patients transformed to AML. Among PV patients (median follow-up duration 47.4 months), 33 patients (1.36%) underwent transformation to either sMF (N = 13) or AML (N = 20), respectively. Among PMF patients (median follow-up duration 31.9 months), 81 patients (13.59%) underwent transformation to AML.

Five-year CI of transformation to sMF and AML were 1.4% and 1.91% in ET patients, 0.69% and 0.97% in PV patients, and not applicable and 16.54% in PMF patients, respectively.

Table 6. Disease transformation to secondary myelofibrosis or acute myeloid leukemia.

	Non-transformed MPN			Secondary MF			Secondary AML			
	non-transformed ET	non-transformed PV	non-transformed PMF	Total	ET→sMF	PV→sMF	Total	ET→AML	PV→AML	MF→AML
N	4175	2437	514	82	67	13	166	60	20	80
Age (Years)										
Median (Range)	61(13-106)	58(11-100)	63(21-89)	61(20-90)	61(20-90)	59(48-73)	69(39-94)	72(39-94)	73.5(43-82)	66(33-83)
0-9	0	0	0	0	0	0	0	0	0	0
10~19	75	23	0	0	0	0	0	0	0	0
20-29	163	81	6	2	2	0	0	0	0	0
30-39	348	189	25	3	3	0	5	1	0	4
40-49	596	396	65	9	9	2	11	2	2	7
50-59	843	624	103	16	16	5	29	7	2	19
60-69	847	581	146	21	21	5	46	16	4	21
70-79	960	436	144	11	11	1	58	25	10	23
≥80	343	107	25	5	5	0	17	9	2	6
Sex										
Male	1835	1615	306	39	39	8	108	34	10	59
Female	2340	822	208	28	28	5	58	26	10	21
Year										

2008	425	207	66	15	15	3	20	8	2	10
2009	417	226	72	11	11	2	36	14	4	17
2010	378	240	46	12	12	2	16	5	1	7
2011	428	255	31	5	5	1	22	9	3	8
2012	478	262	37	5	5	1	22	8	4	10
2013	489	294	49	3	3	2	14	4	1	9
2014	490	315	62	4	4	0	15	4	4	7
2015	557	307	78	7	7	0	12	5	0	7
2016	513	331	73	5	5	2	9	3	1	5
FU duration, (months) median (range)	46.05 (6.02-114.61)	47.34 (6.02-113.68)	33.4 (6.02-113.78)	64.57 (8.65-113.65)	64.57 (8.65-113.65)	74.11 (9.01-112.86)	35.21 (6.15-110.20)	47.2 (6.15-110.20)	33.11 (6.78-103.22)	19.97 (6.78-103.91)

Abbreviation: MPN, myeloproliferative neoplasms; sMF, secondary myelofibrosis; AML, acute myeloid leukemia; ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis; FU, follow-up.

Frequency and distribution of mutations among ET, PV and PMF patients

For mutation profiling, we characterized a cohort of 41 MPN patients, consisting of 17 PV, 16 ET and 8 PMF patients. The baseline clinical parameters of the patients at diagnosis of MPN are summarized in **Table 7**. Interestingly, there were no patients in high risk group. The median follow-up was 39.2 months (range 1-118 months). During follow-up, 4 leukemic transformations (9.8%) and 3 secondary myelofibrotic transformations (7.3%) were documented.

Most patients harbored more than one mutation (**Fig. 4**), but 5 patients only had *JAK2 V617F* mutation (i.e. one mutation only). There were 4 PV patients and 4 ET patients harboring more than 5 mutations, while none of the PMF patients were associated with more than 4 mutations per patient.

Apart from *JAK2*, 8 genes *ASXL1* (N=8), *KMT2D* (N=7), *TET2* (N=6), *MSH2* (N=5), *CALR* (N=4), *NPM1* (N=4), *SETBP1* (N=4), *ARID2* (N=4) were mutated in more than 10% of the analyzed MPN patients (**Figs. 5 and 6**). The most frequently observed mutations affected genes implicated in epigenetic regulation. No mutations were found in the genes *ETV6, MYC, SMARCA4, UTY, NRAS* and *IKZF1*.

Table 7. Baseline characteristics of the 41 patients.

	Total	PV	ET	PMF
Total (N, %)	41	17	16	8
Male (N, %)	19 (46.3)	6 (35.3)	8 (50.0)	5 (62.5)
Age (median, range)	63 (35-79)	66 (45-76)	62 (35-79)	58 (46-66)
Laboratory findings at diagnosis (mean, ± SD)				
Hemoglobin (g/dL)	15.0 (3.1)	17.5 (2.2)	13.9 (2.0)	12.1 (3.0)
Platelet (10^3/μl)	657 (466)	489 (279)	943 (542)	439 (350)
WBC (10^3/μl)	16.8 (22.9)	15.8 (13.9)	20.2 (33.6)	12.5 (8.9)
IPSS risk group (N, %)				
Low	3 (37.5)	NA	NA	3 (37.5)
Intermediate-1	2 (25.0)	NA	NA	2 (25.0)
Intermediate-2	3 (37.5)	NA	NA	3 (37.5)
High	0	NA	NA	0
D-IPSS risk group (N, %)				
Low	3 (37.4)	NA	NA	3 (37.4)
Intermediate-1	4 (50.0)	NA	NA	4 (50.0)
Intermediate-2	1 (12.5)	NA	NA	1 (12.5)
High	0	NA	NA	0
Secondary transformation (N, %)				
Total	7 (17.1)	3 (17.6)	1 (6.3)	3 (37.5)
AML	4	1	0	3
sMF	3	2	1	NA

Abbreviation: PV, polycythemia vera; ET, essential thrombocythemia; PMF, primary myelofibrosis; SD, standard deviation; WBC, white blood count; IPSS, International Prognostic Scoring System; NA, not applicable; D-IPSS, Dynamic International Prognostic Scoring System; AML, acute myeloid leukemia; sMF, secondary myelofibrosis.

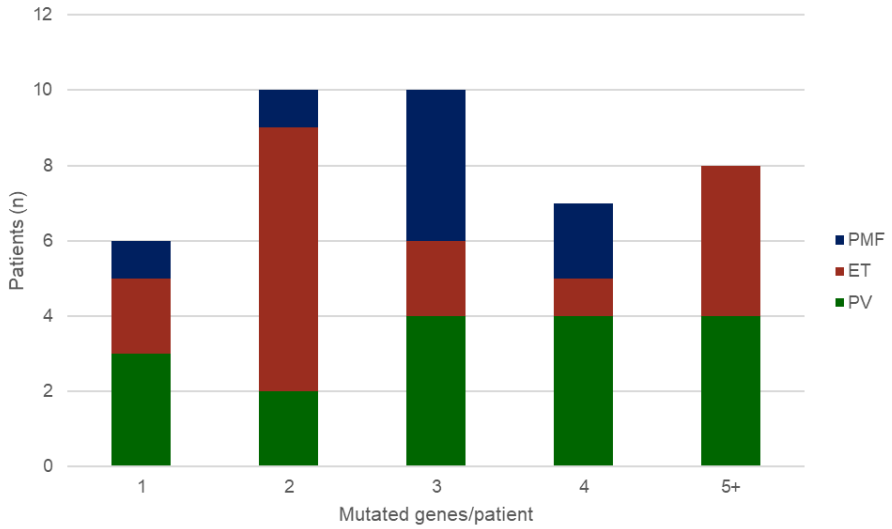


Figure 4. Number of mutations per patient. 41 myeloproliferative neoplasm (MPN) patients were investigated for mutations by the developed next generation sequencing-based MPN gene panel. Patients are grouped by number of mutated genes per person. For each group, the number of patients is shown, Color code represents MPN subgroups as indicated. Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis.

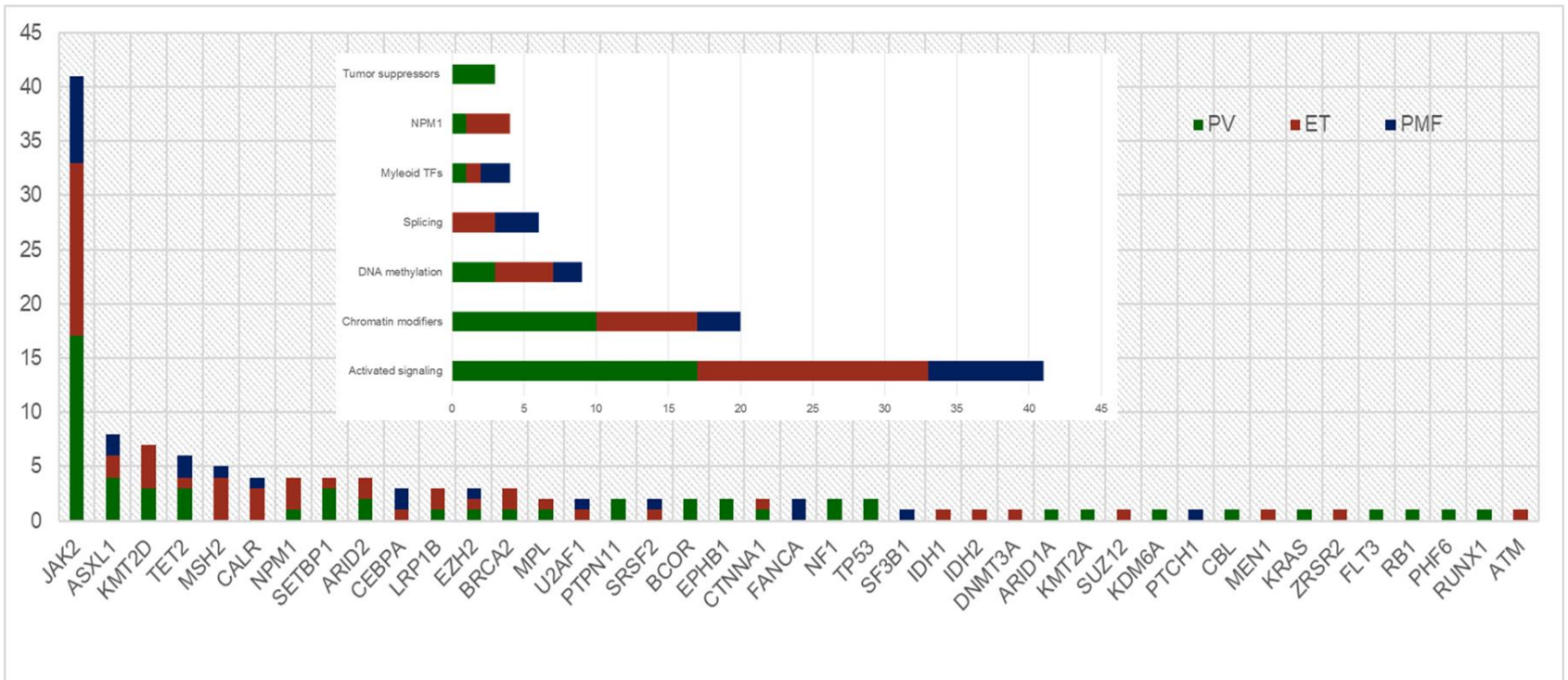


Figure 5. Mutation frequency. Number of patients carrying mutations in the respective genes are shown. Color code represents MPN subgroups as indicated. Boxed figure shows mutated genes grouped by functional pathways. Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis.

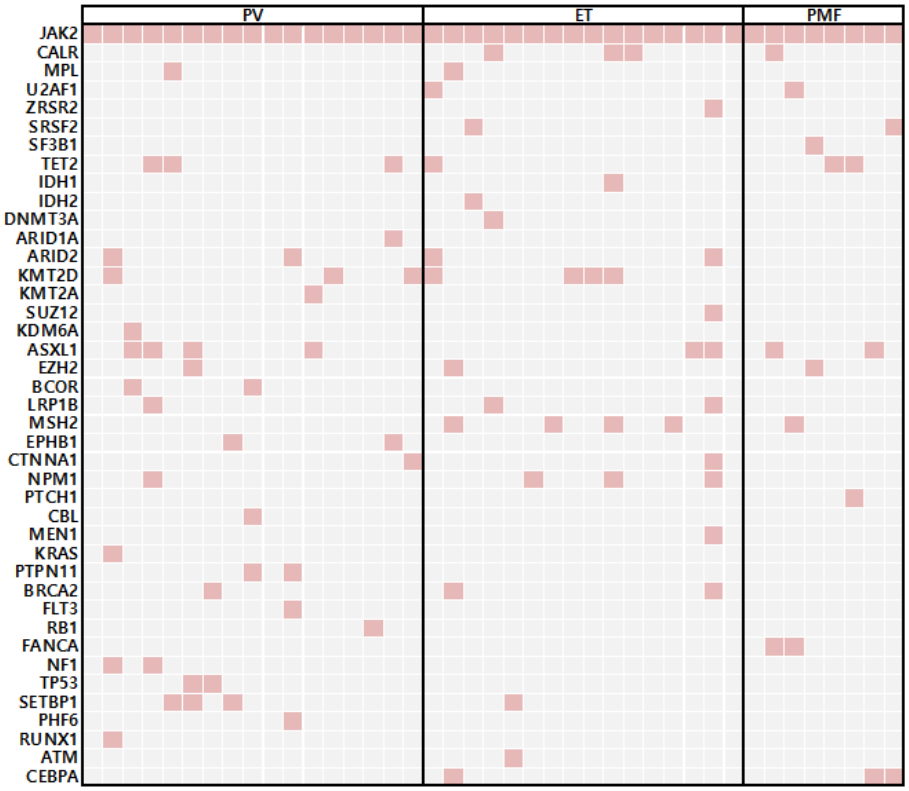


Figure 6. Distribution of mutations. Distribution of mutated genes is shown on a single sample level. Each column represents one patient sample. Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis.

Distinct mutation patterns between ET, PV and PMF patients

Overall, there were differences in mutation frequency per MPN subgroup (**Fig. 7A**). Interestingly ,however, PMF patients were associated with lower number of mutated genes per patient: in PV patients 3.35 mutated genes/patient were present and in ET patients 3.31 mutated genes/patient, while in PMF patients 2.88 mutated genes/patient were noted. These subgroup-specific mutation frequencies were caused by mutations in genes usually not analyzed in routine diagnosis of MPN.

The mutation frequencies of the splicing genes consisting of *SF3B1, SRSF2, ZRSR2* and *U2AF1* were significantly different between MPN subgroups (**Fig. 7B**). Splicing genes were more frequently mutated in PMF (0.38 mutated genes/patient) than in ET (0.19) and not mutated in PV patients. On the other hand, mutations in epigenetic regulators were significantly more frequent in PV (P=0.028) (**Fig. 7C**). By analyzing the epigenetic regulator's subgroups, namely chromatin modification and DNA methylation, two distinct patterns were observed. While there were no differences in DNA methylation gene mutations across MPN subgroups, PV patients harbored more mutations of genes involved in chromatin modification compared to ET (P=0.085) and PMF (P=0.008).

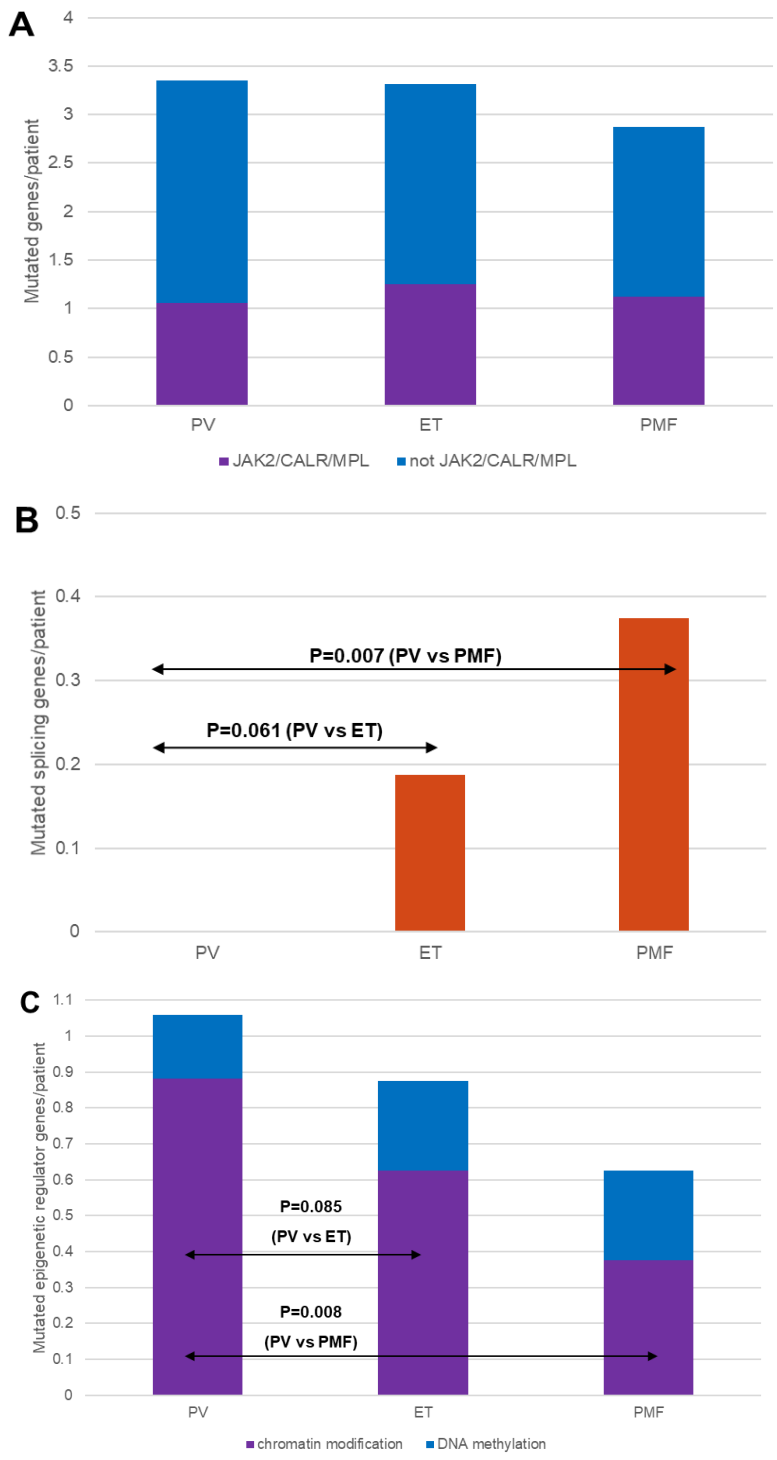

A
Mutated genes/patient
4
3.5
3
2.5
2
1.5
1
0.5
0
PV
ET
PMF
JAK2/CALR/MPL
not JAK2/CALR/MPL
B
Mutated splicing genes/patient
0.5
0.4
0.3
0.2
0.1
0
P=0.007 (PV vs PMF)
P=0.061 (PV vs ET)
PV
ET
PMF
C
Mutated epigenetic regulator genes/patient
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
P=0.085
(PV vs ET)
P=0.008
(PV vs PMF)
PV
ET
PMF
chromatin modification
DNA methylation

Figure 7. Subgroup specific mutation patterns. (A) mutation frequency per patient; **(B)** mutation frequency of splicing genes; **(C)** mutation frequency of epigenetic regulator genes. Epigenetic regulator genes are subdivided in chromatin modification genes and NDA methylation genes. Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis.

Clinical relevance of mutations and temporal sequence of mutations

In PMF patients with splicing gene mutation, these splicing mutations seem to occur after *JAK2 V617F* mutation. Also, these patients were associated with significantly higher hemoglobin (mean± SD, 14.5±1.2 vs. 10.5±2.6 g/dL, P=0.035) and platelet (mean± SD, 717.2±275.6 vs. 260.5±248.4 x10^3/μl, P=0.049) count at MPN diagnosis compared to those without splicing mutations (**Table 8**).

On the other hand, in PV when the temporal sequence of somatic mutations acquisition was considered, it appeared that PV patients acquired *ASXL1* (mean VAF 25.9%) or *EZH2* (VAF 31.5%) mutations after *JAK2 V617F* mutation (mean VAF 55.6% for *AXL1*, 50.8% for *EZH2*), while ET and PMF patients acquired *ASXL1* (mean VAF 44.1%) or *EZH2* (VAF 41.5%) prior to *JAK2 V617F* (mean VAF 26.1% for *ASXL1*, 28.4% for *EZH2*) (**Fig. 8**).

Among PV patients, those with *ASXL1* mutation were associated with significantly lower hemoglobin count at MPN diagnosis compared to those without (15.2±1.8 g/dL vs 18.2±1.8 g/dL, P=0.013, **Table 8**). These patients were associated with more thrombotic events compared to PV patients without *ASXL1* mutation (P<0.001). More specifically, among *ASXL1* mutated PV patients, 2 patients had cerebrovascular infarctions around the time of PV diagnosis, 1 had symptomatic portal

vein thrombosis 2 years after diagnosis and 1 had a past history of myocardial infarction. Overall, males harbored more *ASXL1* mutations (P=0.022).

Table 8. Specific mutations and their clinical implications

(A) Additional splicing gene mutation acquisition in *JAK2* positive primary myelofibrosis is associated with higher hemoglobin and platelet count at diagnosis; **(B)** Additional *ASXL1* mutation in *JAK2* positive polycythemia vera is associated with lower hemoglobin count at diagnosis. Data presented as mean (±standard deviation). Hb, hemoglobin; Plt, platelet; WBC, white blood count.

A

	Splicing mutation	**No splicing mutation**	**P**
Hb (g/dL)	14.5 (1.2)	10.5 (2.6)	0.035
Plt (10^3/µl)	717.2 (275.6)	260.5 (248.4)	0.049
WBC (10^3/µl)	14.7 (9.6)	10.2 (8.8)	0.513

B

	ASXL1 mutated	**No ASXL1**	**P**
Hb (g/dL)	15.2 (1.8)	18.2 (1.8)	0.013
Plt (10^3/µl)	376.5 (304.2)	523.6 (274.6)	0.374
WBC (10^3/µl)	25.6 (28.3)	12.7 (4.0)	0.431

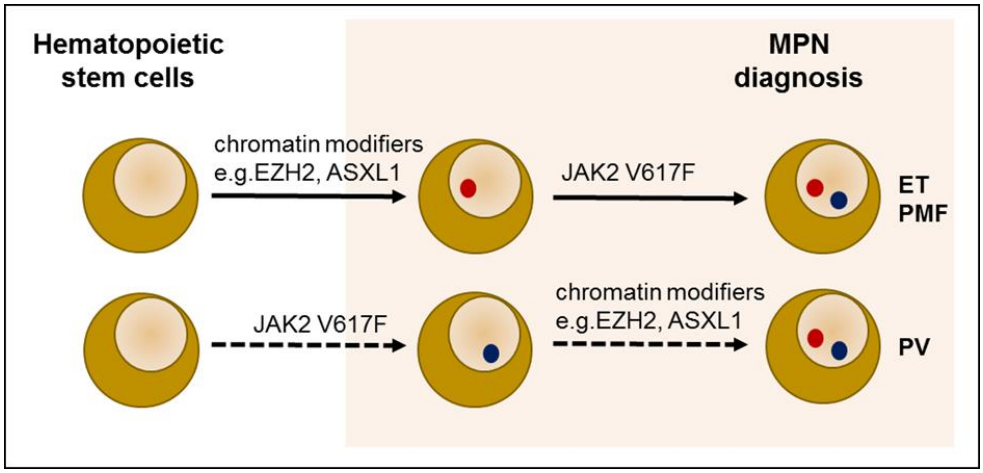


Figure 8. Temporal sequence of mutation acquisition and clinical phenotypes. Abbreviation: ET, essential thrombocythemia; PV, polycythemia vera; PMF, primary myelofibrosis.

Prognostic implications of mutations

We analyzed the impact of the number of mutations other than JAK2 V617F on survival and transformation. We found that the number of somatic mutations did not play a significant role in reduced overall survival (no additional mutations versus one or more additional mutations, P=0.440) or increased risk of secondary transformation (no additional mutations versus one or more additional mutations, P=0.296). When we analyzed the role of JAK2 V617F allelic burden on survival, those with higher allele burden (>50%) did not show a propensity towards secondary transformation (P=0.468).

When patients were analyzed per mutation (**Fig. 9**), we observed that patient with TP53 mutation were associated with shorter overall survival (P=0.039, **Fig. 9A**). There were no patients harboring TP53 mutation experiencing secondary transformation, thus transformation free survival was not analyzed. The number of each mutation of individual splicing gene was low, thus they were combined as a group for survival analyses. As shown in **Fig. 9B**, patients with splicing gene mutations were associated with reduced overall survival (P=0.028) and increased risk of secondary transformation (P=0.035). In addition, patients with ASXL1 mutation showed trends toward decreased overall survival (P=0.086, **Fig. 9C**) and significantly increased risk of secondary transformation (P=0.014, **Fig. 9C**).

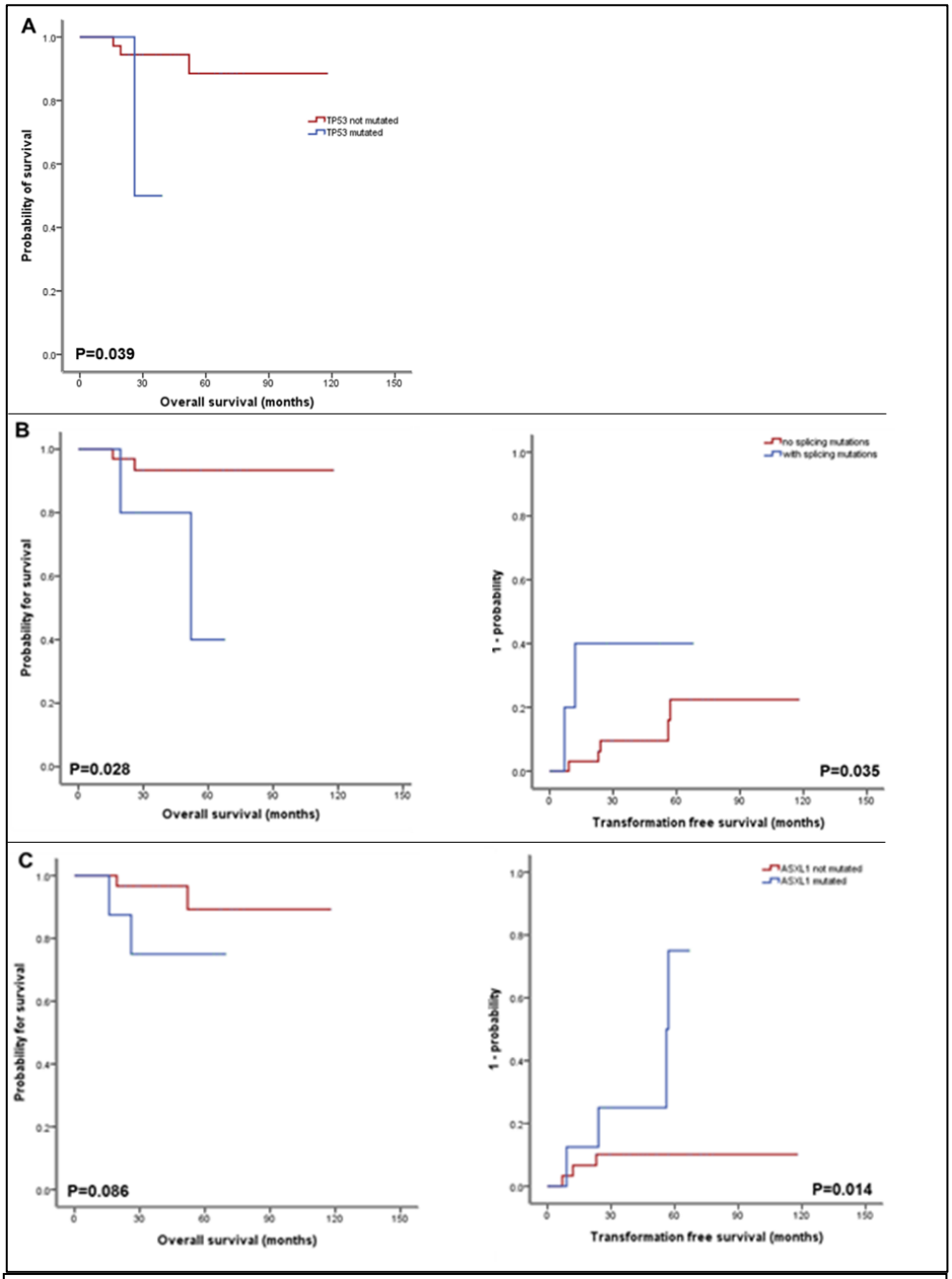


Figure 9. Survival curves stratified by their mutational status. (A) overall survival between *TP53* mutated patients versus *TP53* unmutated patients (P=0.039); **(B)** survival curves for patients with splicing mutations versus those without splicing mutations. Left, overall survival (P=0.028) and right, transformation free survival (P=0.035); **(C)** survival curves for *ASXL1* mutated patients versus *ASXL1* unmutated patients. Left, overall survival (P=0.086) and right, transformation free survival (P=0.014).

Focusing on secondary transformations

The mutational landscapes of 7 patients who experienced secondary transformation are presented as **Table 9**. All of the patients with secondary leukemia had at least one mutation of AML drivers. Notably, 2 of the PMF patients harbored ASXL1 mutation and one had SRSF2 mutation. On the other hand, sMF transformation showed a different pattern of mutations. All of the patients who transformed into sMF harbored at least one mutation in tumor suppressor genes.

For a better understanding of the disease transformation process, WGS was performed on the PMF and secondary AML bone marrow samples of a single patient experiencing secondary transformation. Going from PMF to AML, the total number of mutations increased (**Fig. 10**) but the majority of mutations were randomly distributed across chromosomes. There were no differences in mutation signatures between PMF and secondary AML (**Fig. 11**), indicating that disease progression was not accompanied by gain of a new mutation nor was it caused by a specific increase in putative driver mutations. In fact, when VAF density for putative driver mutations was compared (**Fig. 12**), an increment of allele burdens was not observed.

Table 9. Seven patients experiencing secondary transformation.

MPN	TFS (mo)	# mut	JAK2 allele burden	AML driver								Tumor suppressor			
				ASXL1	EZH2	KMT2D	TET2	SRSF2	U2AF1	TP53	NPM1	FANCA	LRP1B	ARID2	RB1
Secondary AML															
PMF	56	4	52	50								47			
PMF	9	3	9	49											
PMF	12	3	71					39							
PV	24	5	51	37	32					14					
Secondary MF															
ET	7	4	42			15	52		27					50	
PV	57	6	58	17			55				19		14		
PV	23	2	55												50

The numbers in each box represents allele burden in VAF.

Abbreviation: MPN, myeloproliferative neoplasm; TFS, transformation free survival; mo, months; mut, mutation; AML, acute myeloid leukemia; PMF, primary myelofibrosis; PV, polycythemia vera; MF, myelofibrosis; ET, essential thrombocythemia.

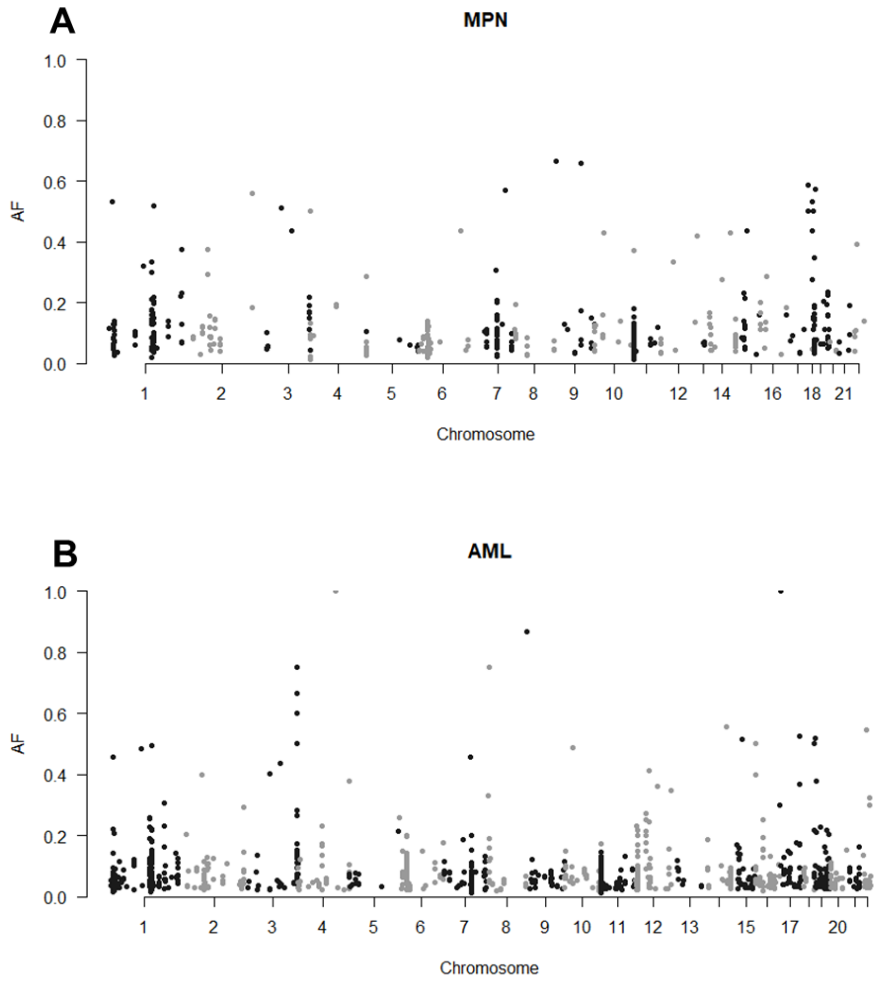


Figure 10. Number of mutations per chromosome in a single patient experiencing secondary transformation from primary myelofibrosis (PMF) to acute myeloid leukemia (AML) (A) in PMF state; **(B)** in secondary AML state.

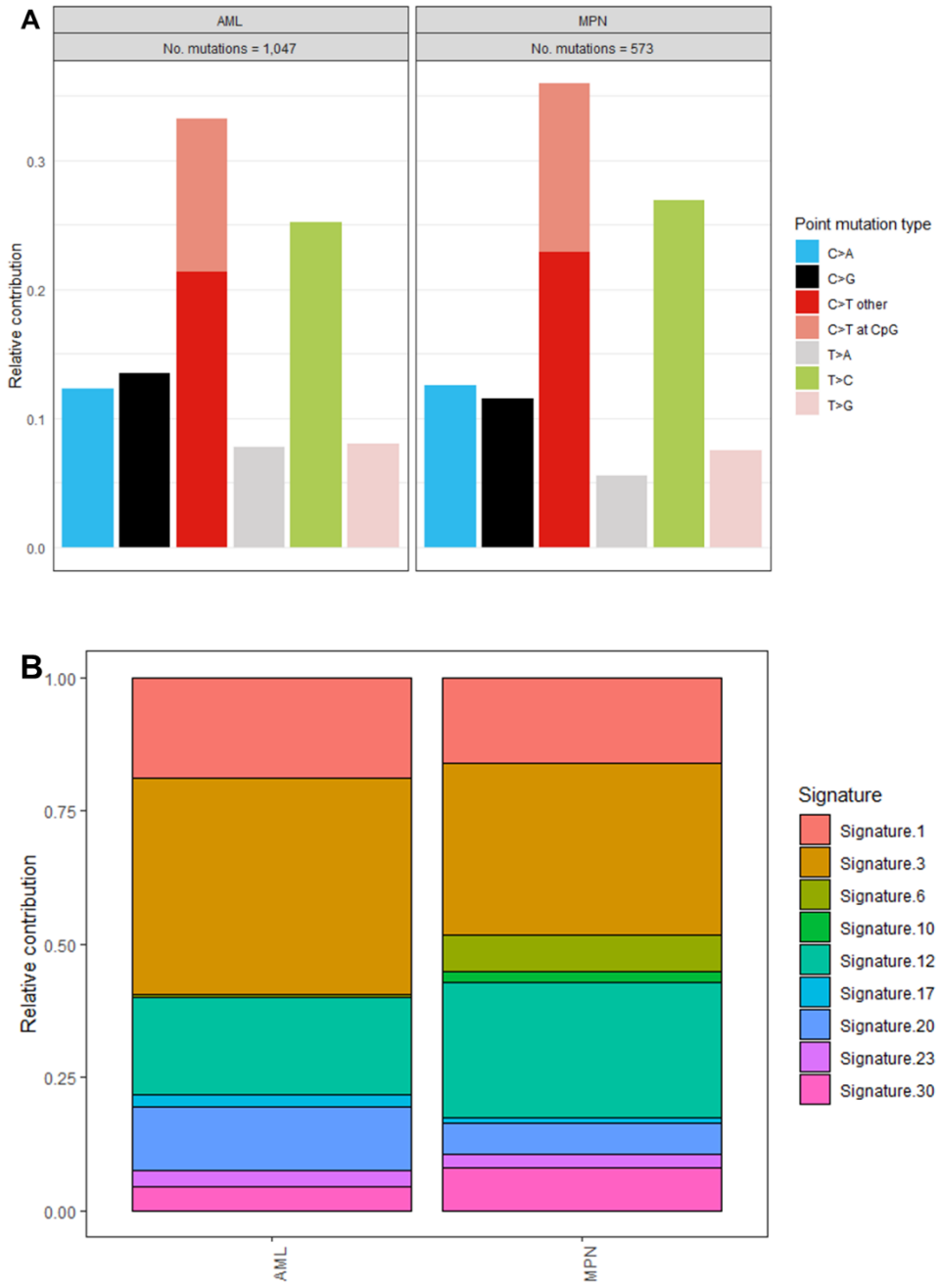


Figure 11. Mutation signature of a single patient experiencing secondary transformation (A) point mutation frequency; **(B)** mutational signature.

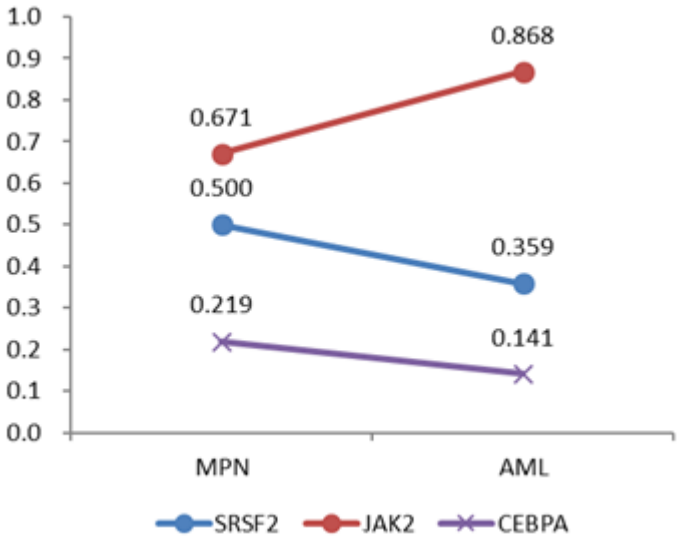


Figure 12. Representative analysis of VAF of mutations occurring at chronic myeloproliferative neoplasm (MPN) state and secondary acute myeloid leukemia (AML) state from a single patient.

Discussion

Firstly, based on population-based extensive analyses, we found ET to be the most common MPN etiology in Korea, followed by PV then PMF. There was no definite pattern regarding its incidence, but the prevalence of ET doubled during the past decade, from 4.110 in 2004 to 8.989 in 2013. The prevalence of PV also showed a rapid increase from 2.838 in 2004 to 5.434 in 2013, while its incidence (0.977-1.453) remained similar throughout the study period. Data on PMF was only available for recent 3 years, but the pattern of escalating prevalence was also evident for PMF. Furthermore, despite the fact that there were far fewer PMF patients compared to ET or PV, medical utilization costs were the greatest for patients with PMF than for those with ET or PV. The 5-year CI of transformation to sMF and AML were 1.4% and 1.91% in ET patients, 0.69% and 0.97% in PV patients, and not applicable and 16.54% in PMF patients, respectively.

A wide range of MPNs prevalence and incidence estimates have been reported across the world over time [10-27]. Such differences in epidemiological estimates reflect the evolution in diagnosis and classifications of MPN over time. Since our results are based on a single database source from a homogenous population, we feel that our data is suitable for evaluating the impact of the new classification on trends in

MPNs epidemiology. The incidence of both ET and PV decreased from 2008 to 2011, then increased in 2012 and plateaued in during the following year. The drop in the incidence rate in 2008 may reflect stricter but more accurate diagnosis for MPNs with the introduction of *JAK2* mutation analysis, whilst the gradual increment followed by stabilization may be associated with the solidified diagnostic techniques and matured clinicians' perception of the disease. Following this secular trend, we attempted to comparatively analyze the prevalence and incidence of MPNs across geographies using data published after 2008. Unfortunately, there were little data available, but from the data obtained, we gathered that PV is more prevalent than ET in both Europe and North America (**Tables 4 and 5**). In comparison, we found ET to be the most common MPNs subtype in Koreans. The presumption that ET is more prevalent than PV in Korea from previous reports [28] have been corroborated by our large-scaled study. In addition, we also found that although all regions of the world share similar patterns of increment in MPNs, the absolute prevalence rate in Korea is only 1/10 of that in North America. These seemingly ethnical differences should be interpreted with caution as many MPNs patients go under-detected and underdiagnosed.

Previous studies have reported that most Caucasian patients with MPNs were older than 60 years old at diagnosis [29,30]. The mean age for 381 ET patients in a Japanese cohort was 58.4 (±15.7) and that for 266 PV patients was 59.2 (±12.1) [31]. In our cohort, we found that the

mean age of ET was comparatively young at 55.4 years. The mean age for PV (59.5 years) and PMF (63.7) were similar to other cohorts. Additionally, we found that ET develops at a relatively young age (mean age at first diagnosis 46.6 years), and males were associated with earlier development compared to females. The mean age at first diagnosis of PV was also relatively young at 53.7 years. Only PMF patients seemed to be diagnosed in their sixties. The median age (interquartile range) at diagnosis was 55 years (43-67 years) for PV and 52 years (27-68 years) for ET, which was similar to previous studies from China reporting the median age at diagnosis of PV and ET to be 54 and 52 years, respectively 32,33].

Secondly, results from this study add insights to the implications of mutation profiling in MPN and suggest new findings regarding the secondary transformation of the disease. Several genes have been proposed as diagnostic or prognostic markers for MPN in the past few years. In this study we selected *JAK2 V617F* positive patients to investigate the roles of somatic mutations other than the so-called driver mutations in MPN. As expected, the 3 MPN subgroups showed different mutational spectrums at their chronic state. Notably, the group of analyzed splicing genes consisting of *U2AF1, ZRSR2, SRSF2,* and *SF3B1* was most frequently mutated in PMF (**Fig. 6**), less frequently in ET, and none in PV. Mutations in genes involved in splicing are best known from the MDS pathogenesis [9, 10]. MPN are known to share

many of the same genetic mutations as MDS and it is thought that the phenotypic differences between these 2 disease entities arise mainly from the individual composite of the same mutations, including the order of mutation acquisition and the interactions between clones and subclones [11]. Taken this into consideration, our findings suggest that PMF is more closely related to MDS than other MPN subgroups, which can also account for the higher risk of secondary leukemia in PMF compared to ET or/and PV. Secondary leukemic transformation occurs in 20%, 1% and 4% over a 10-year period in PMF, ET and PV, respectively [12]. In previous studies, *SRSF2* has shown to play a role in secondary leukemia development from both MDS [13] and PMF [34]. Likewise, one of the PMF patients experiencing secondary leukemia transformation in our cohort also harbored SRSF2 mutation. This again reinforces our suggestion that PMF is closely related to MDS, and the secondary leukemic transformation arising from PMF and MDS may share similar pathophysiology.

In the context of secondary leukemia transformation of PMF, another gene that deserves attention is *ASXL1*. The possible non-overlapping contributions of *ASXL1* and *JAK2* in myeloid transformation has been suggested by Abdel-Wahab et al. [35]. Similarly, Vannucchi et al. [34] reported leukemia-free survival was negatively impacted by *ASXL1* mutation in PMF patients, while Engle et al. [36] showed expansion of *ASXL1* clones with the transformation of PMF to secondary

leukemia. Also in our cohort, 2 of the 3 PMF patients experiencing secondary leukemia transformation harbored *ASXL1* mutation. Moreover, the patients harboring *ASXL1* mutation were associated with shorter transformation-free survival (P=0.014, **Fig. 9C**). All in all, our findings further support the role of *SRSF2* and *ASXL1* in secondary leukemia transformation of PMF in previous reports (**Table 9**) [37, 38].

Interestingly, there were 4 patients (3 ET and 1 PV) harboring NPM1 mutation but only one of them showed the secondary transformation. Upon inspection, this was because one patient was lost to follow-up thus information on transformation could not be gathered, while another one harbored mutation on exon 8 (c.G646C p.E216QNPM1). The last patient harbored mutation on exon 5, and considering the known AML driver is located on exon 3, the mutation found in our patients might not be deleterious, which in turn explains the long transformation free survivals.

Although secondary MF occurs as part of natural history of ET and PV, with rates of transformation ranging from 5-10% in ET and 10-15% in PV [39], this end of MPN secondary transformation is somewhat less understood. A recent study from AGIMM group [40] showed that the molecular landscape of secondary MF is different from PMF and suggested that unknown mutational events might contribute to the disease progression. There were 3 patients with secondary MF patients in our cohort and except for the fact that all of them had mutations of

tumor suppressor genes, there were no common grounds to establish a distinctive pattern.

The effect of mutation order in MPN has been shown to alter clinical features and clonal evolution [41,42]. Focusing on *ASXL1, EZH2* and *JAK2*, we found that *ASXL1* or *EZH2* mutation acquisition after *JAK2* leads to PV, while *ASXL1* mutation acquisition before *JAK2* leads to ET or PMF (**Fig. 8**). It is interesting to note that both *ASXL1* and *EZH2* are key regulators of PRC2, and are known to play a role in MPN initiation and disease progression. As such, the difference in the temporal sequence of epigenetic regulators mutation acquisition might decide the phenotype of *JAK2* positive MPN, and also explain the natural disease course of each MPN subtype. For example, as *ASXL1* deletion is associated with defective hematopoietic stem cells self-renewal properties leading to profound cytopenias and dysplasias, this sequence of mutations may explain why ET is more closely related to pre-fibrotic PMF.

Also among PV patients, *ASXL1* mutation was associated with lower hemoglobin, platelet and higher WBC count at PV diagnosis. This finding is in agreement with the recent report that *ASXL1* mutation is correlated with features of PMF, such as increased reticulin fibrosis and reduced myeloid erythroid ratio [43]. Meanwhile, splicing genes mutated PMF had higher hemoglobin, platelet and WBC count at PMF diagnosis. It would be interesting to see if this subpopulation of patients represent

masked PV and pre-fibrotic MF.

In conclusion, our findings provide a better understanding of MPN in Korea and we showed the sequence of somatic mutations in relation to *JAK2* influences the clinical behavior of the disease. We also provide evidence that each subtype of MPN harbors distinct patterns of somatic mutations and acquisition order, and that mutations in *TP53*, *ASXL1*, and splicing genes are associated with prognosis of MPN.

References

1. Sabattini E, Bacci F, Sagramoso C, Pileri SA: WHO classification of tumours of haematopoietic and lymphoid tissues in 2008: an overview. *Pathologica.* 2010;102:83-87.
2. Campbell PJ, Green AR: The myeloproliferative disorders. *N Engl J Med.* 2006;355:2452-2466.
3. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009;114:937-951.
4. Levine RL, Gilliland DG: Myeloproliferative disorders. *Blood.* 2008; 112:2190-2198.
5. Mesa RA, Schwager S, Radia D, Cheville A, Hussein K, Niblack J, et al. The Myelofibrosis Symptom Assessment Form (MFSAF): an evidence-based brief inventory to measure quality of life and symptomatic response to treatment in myelofibrosis. *Leuk Res*. 2009;33:1199-1203.
6. Scherber R, Dueck AC, Johansson P, Barbui T, Barosi G, Vannucchi AM, et al. The Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF): international prospective validation and reliability trial in 402 patients. *Blood*. 2011;118:401-408.
7. Wadleigh M, Tefferi A. Classification and diagnosis of myeloproliferative neoplasms according to the 2008 World Health Organization criteria. *Int J Hematol*. 2010;91:174-179.
8. Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR, Tichelli A, Cazzola M, Skoda RC: A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005;352:1779-1790.
9. Pellagatti A, Boultwood J: Splicing factor gene mutations in the myelodysplastic syndromes: impact on disease phenotype and therapeutic applications. *Adv Biol Regul.* 2017;63:59-70.
10. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M *et al*: Frequent pathway mutations

of splicing machinery in myelodysplasia. *Nature.* 2011;478:64-69.

11. Sperling AS, Gibson CJ, Ebert BL: The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer.* 2017;17:5-19.
12. Rampal R, Ahn J, Abdel-Wahab O, Nahas M, Wang K, Lipson D, Otto GA, Yelensky R, Hricik T, McKenney AS *et al*: Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. *Proc Natl Acad Sci U S A.* 2014;111:E5401-5410.
13. Kim T, Tyndel MS, Kim HJ, Ahn JS, Choi SH, Park HJ, Kim YK, Yang DH, Lee JJ, Jung SH *et al*: The clonal origins of leukemic progression of myelodysplasia. *Leukemia.* 2017;31:1928-1935.
14. Kim DS. Introduction: health of the health care system in Korea. *Soc Work Public Health*. 2010;25:127-141.
15. Kim DS. Special issue on the national health care system of South Korea. *Soc Work Public Health*. 2010;25:125-126.
16. Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26:589-595.
17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297-1303.
18. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK: VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;22:568-576.
19. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38:e164.
20. Moulard O, Mehta J, Fryzek J, Olivares R, Iqbal U, Mesa RA. Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union.*Eur J Haematol*. 2014;92:289-297.
21. Mehta J, Wang H, Iqbal SU, Mesa R. Epidemiology of myeloproliferative

neoplasms in the United States. *Leuk Lymphoma*. 2014;55:595-600.

22. Phekoo KJ, Richards MA, Moller H, Schey SA, South Thames Haematology Specialist C. The incidence and outcome of myeloid malignancies in 2,112 adult patients in southeast England. *Haematologica*. 2006;91:1400-1404.
23. Buhr T, Georgii A, Choritz H. Myelofibrosis in chronic myeloproliferative disorders. Incidence among subtypes according to the Hannover Classification. *Pathol Res Pract*. 1993;189:121-132.
24. Ania BJ, Suman VJ, Sobell JL, Codd MB, Silverstein MN, Melton LJ, 3rd. Trends in the incidence of polycythemia vera among Olmsted County, Minnesota residents, 1935-1989. *Am J Hematol*. 1994;47:89-93.
25. Christtina Roaldsnes AW, Mette Nogaard, Waleed Ghanima. Epidemiology of Myeloproliferative Neoplasms in Norway. *Blood*. 2014;124: S634 (abstrac 1858).
26. Johansson P. Epidemiology of the myeloproliferative disorders polycythemia vera and essential thrombocythemia. *Semin Thromb Hemost*. 2006; 32:171-173.
27. Srour SA, Devesa SS, Morton LM et al. Incidence and patient survival of myeloproliferative neoplasms and myelodysplastic/myeloproliferative neoplasms in the United States, 2001-12. *Br J Haematol.* Prepublished on April 7, 2016, as doi: 10.1111/bjh.14061
28. Choi CW, Bang SM, Jang S, Jung CW, Kim HJ, Ki, HY, et al. Guidelines for the management of myeloproliferative neoplasms. *Korean J Intern Med* 2015; 30: 771-788.
29. Mesa RA, Li CY, Ketterling RP, Schroeder GS, Knudson RA, Tefferi A. Leukemic transformation in myelofibrosis with myeloid metaplasia: a single-institution experience with 91 cases. *Blood*. 2005;105:973-977.
30. Rollison DE, Howlader N, Smith MT, Strom SS, Merritt WD, Ries LA, et al. Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. *Blood*. 2008;112:45-52.
31. Dan K, Yamada T, Kimura Y, Usui N, Okamoto S, Sugihara T, et al. Clinical features of polycythemia vera and essential thrombocythemia in

Japan: retrospective analysis of a nationwide survey by the Japanese Elderly Leukemia and Lymphoma Study Group. *Int J Hematol*. 2006;83:443-449.

32. Bai J, Ai L, Zhang L, Yang F, Zhou Y, Xue Y. Incidence and risk factors for myelofibrotic transformation among 272 chinese patients with JAK2-mutated polycythemia vera. *Am J Hematol.* 2015;90:1116-1121
33. Fu R, Xuan M, Lv C, Zhang L, Li H, Zhang X, et al. External validation and clinical evaluation of the International Prognostic Score of Thrombosis for Essential Thrombocythemia (IPSET-thrombosis) in a large cohort of Chinese patients. *Eur J Haematol.* 2014;92:502-509
34. Vannucchi AM, Lasho TL, Guglielmelli P, Biamonte F, Pardanani A, Pereira A, Finke C, Score J, Gangat N, Mannarelli C *et al*: Mutations and prognosis in primary myelofibrosis. *Leukemia.* 2013;27:1861-1869.
35. Abdel-Wahab O, Manshouri T, Patel J, Harris K, Yao J, Hedvat C, Heguy A, Bueso-Ramos C, Kantarjian H, Levine RL *et al*: Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res.* 2010;70:447-452.
36. Engle EK, Fisher DA, Miller CA, McLellan MD, Fulton RS, Moore DM, Wilson RK, Ley TJ, Oh ST: Clonal evolution revealed by whole genome sequencing in a case of primary myelofibrosis transformed to secondary acute myeloid leukemia. *Leukemia.* 2015;29:869-876.
37. Sokol K, Tremblay D, Bhalla S, Rampal R, Mascarenhas JO: Implications of Mutation Profiling in Myeloid Malignancies-PART 2: Myeloproliferative Neoplasms and Other Myeloid Malignancies. *Oncology (Williston Park).* 2018;32:e45-e51.
38. Guglielmelli P, Biamonte F, Rotunno G, Artusi V, Artuso L, Bernardis I, Tenedini E, Pieri L, Paoli C, Mannarelli C *et al*: Impact of mutational status on outcomes in myelofibrosis patients treated with ruxolitinib in the COMFORT-II study. *Blood.* 2014;123:2157-2160.
39. Mesa RA, Verstovsek S, Cervantes F, Barosi G, Reilly JT, Dupriez B, Levine R, Le Bousse-Kerdiles MC, Wadleigh M, Campbell PJ *et al*: Primary myelofibrosis (PMF), post polycythemia vera myelofibrosis (post-PV MF), post essential thrombocythemia myelofibrosis (post-ET

MF), blast phase PMF (PMF-BP): Consensus on terminology by the international working group for myelofibrosis research and treatment (IWG-MRT). *Leuk Res.* 2007;31:737-740.

40. Rotunno G, Pacilli A, Artusi V, Rumi E, Maffioli M, Delaini F, Brogi G, Fanelli T, Pancrazzi A, Pietra D *et al*: Epidemiology and clinical relevance of mutations in postpolycythemia vera and postessential thrombocythemia myelofibrosis: A study on 359 patients of the AGIMM group. *Am J Hematol.* 2016;91:681-686.
41. Lundberg P, Karow A, Nienhold R, Looser R, Hao-Shen H, Nissen I, Girsberger S, Lehmann T, Passweg J, Stern M *et al*: Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood.* 2014;123:2220-2228.
42. Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J, Baxter EJ, Massie CE, Papaemmanuil E, Menon S *et al*: Effect of mutation order on myeloproliferative neoplasms. *N Engl J Med.* 2015;372:601-612.
43. Wong WJ, Hasserjian RP, Pinkus GS, Breyfogle LJ, Mullally A, Pozdnyakova O. JAK2, CALR, MPL and ASXL1 mutational status correlates with distinct histological features in Philadelphia chromosome-negative myeloproliferative neoplasms. *Haematologica.* 2018;103:e63-e68.

초 록

서론: BCR-ABL1 음성 골수증식종양 (MPN)은 크게 진성 적혈구증가증, 본태성 혈소판증가증, 일차성 골수섬유증 3가지 주요 하위그룹으로 나뉜다. 최근 차세대염기서열분석 (NGS)의 발전과 함께 JAK2 V617F, MPL, CALR등의 반복되는 유전자 변이가 MPN 발병에 중요한 역할을 함이 규명되었다. 하지만 이러한 질병 이해의 발전에도 불구하고 일부의 고위험군 골수증식종양 환자들은 혈전색전증 등의 합병증을 경험하고, 또 일부의 환자들은 2차 이환와 함께 매우 불량한 예후를 보인다. 불행히도 아직까지 어떠한 이유로 이러한 합병증이 생기는지에 대한 명확한 근거는 제시되지 않고 있다. 또한, 비슷한 유전자 변이를 바탕으로 하고 있음에도 왜 어떠한 환자들은 적혈구증가증으로 나타나고 어떠한 환자들은 본태성 혈소판증가증으로 나타나는지에 대해서도 알려져 있지 않다. 이에, 연구자들은 본 연구를 통해 (1) 한국의 골수증식종양 현황과 골수증식정질환 이차이환 현황 파악과 (2) 각 골수증식종양 유전자변이 특성과 이들이 임상 경과에 미치는 영향에 대해 통합적으로 분석하고자 하였다.

방법: 연구의 첫번째 부분인 역학분석을 위해서는 국민건강보험공단과 건강보험심사평가원 데이터베이스를 사용하였다. 두번째 부분인 유전자 변이 분석을 위해서는 서울대학교병원 코호트를 바탕으로 하여 NGS 분석을 진행하였다.

결과: 한국에서는 골수증식성 질환중 본태성 혈소판증가증이

가장 많은 것으로 확인을 하였으며, 이차성 급성골수성백혈병으로의 5년 누적 이환률은 본태성 혈소판증가증에서 1.91%, 진성 적혈구증가증에서 0.97%, 일차성 골수섬유증에서 16.54%였다.

유전자 분석 결과 골수증식종양의 종류에 따라 서로 다른 유전자변이 발현양상을 보였는데, 일차성 골수섬유증의 경우 splicing gene의 변이를 주로 나타냈고, 진성 적혈구증가증의 경우 epigenetic regulator의 변이를 주로 나타내었다. 특히 일차성 골수섬유증에서 splicing gene변이가 있는 경우 진단시 헤모글로빈과 혈소판 수치가 변이가 없는 경우에 비하여 의미있게 높게 나타났다. 진성 적혈구증가증에서는 epigenetic regulator중 ASXL1 변이가 혈전색전증의 위험도를 높이는 것으로 나타났다.

생존에 영향을 미치는 유전자변이를 분석한 결과 TP53 변이가 있는 경우 전체 생존기간이 의미 있게 짧았다 (P=0.028). 이차 이환과 관련이 높은 유전자변이는 ASXL1, TP53, splicing gene 변이들 이었다.

마지막으로, ASXL1이나 EZH2의 변이가 JAK2 V617F 변이 이전에 나타나면 본태성 혈소판증가증 또는 일차성 골수섬유증으로 발현하는 반면, ASXL1이나 EZH2의 변이가 JAK2 V617F 변이 이후에 나타나면 진성 적혈구증가증의 표현형을 보이는 것을 관찰하였다.

결론: 본 연구 결과 골수증식종양이 만성적 상태에서 이차 이환하는 상태의 역학을 확인하였고, 그 유전적 배경에 대한 이해를 넓힐 수 있었다.

주요어:골수증식종양; 이차 이환; 차세대염기서열분석; 발현체분석; 역학; 유전자변이 발현양상

학번: 2015 - 30899