Leukemia (2014) 28, 1052–1059

www.nature.com/leu

OPEN

ORIGINAL ARTICLE

Targeted cancer exome sequencing reveals recurrent mutations in myeloproliferative neoplasms

E Tenedini[1,2,3], I Bernardis[1,2], V Artusi[1,2], L Artuso[1,2], E Roncaglia[1,2], P Guglielmelli[4], L Pieri[4], C Bogani[4], F Biamonte[4], G Rotunno[4], C Mannarelli[4], E Bianchi[2,3], A Pancrazzi[4], T Fanelli[4], G Malagoli Tagliazucchi[1,2], S Ferrari[1,2], R Manfredini[2,3], AM Vannucchi[4] and E Tagliafico[1,2] on behalf of AGIMM investigators

With the intent of dissecting the molecular complexity of Philadelphia-negative myeloproliferative neoplasms (MPN), we designed a target enrichment panel to explore, using next-generation sequencing (NGS), the mutational status of an extensive list of 2000 cancer-associated genes and microRNAs. The genomic DNA of granulocytes and *in vitro*-expanded CD3 + T-lymphocytes, as a germline control, was target-enriched and sequenced in a learning cohort of 20 MPN patients using Roche 454 technology. We identified 141 genuine somatic mutations, most of which were not previously described. To test the frequency of the identified variants, a larger validation cohort of 189 MPN patients was additionally screened for these mutations using Ion Torrent AmpliSeq NGS. Excluding the genes already described in MPN, for 8 genes (*SCRIB, MIR662, BARD1, TCF12, FAT4, DAP3, POLG* and *NRAS*), we demonstrated a mutation frequency between 3 and 8%. We also found that mutations at codon 12 of *NRAS* (*NRAS*G12V and *NRAS*G12D) were significantly associated, for primary myelofibrosis (PMF), with highest dynamic international prognostic scoring system (DIPSS)-plus score categories. This association was then confirmed in 66 additional PMF patients composing a final dataset of 168 PMF showing a *NRAS* mutation frequency of 4.7%, which was associated with a worse outcome, as defined by the DIPSS plus score.

Leukemia (2014) **28,** 1052–1059; doi:10.1038/leu.2013.302

Keywords: myeloproliferative neoplasms; mutational analysis; next-generation sequencing; *NRAS*; primary myelofibrosis; polycythemia vera

INTRODUCTION

The discovery of the *JAK2*V617F mutation in 2005[1] represented a major breakthrough in the understanding of the molecular pathogenesis of Philadelphia chromosome-negative chronic myeloproliferative neoplasms (MPN).[2] The *JAK2*V617F mutation is harbored by nearly all polycythemia vera (PV) patients and 50–60% of patients with essential thrombocythemia (ET) and primary myelofibrosis (PMF).[3] The resulting constitutively activated JAK/STAT signaling is considered central to the pathogenesis[4] and phenotype of MPN and therefore serves as a rational drug target for therapy.[5] Additional mutations were described at *MPL* codon 10 (W515L/K/A) in 5–8% of ET and PMF patients.[6–8] However, the significant proportion of *JAK2*V617F- and *MPL*W515L-negative MPN cases required additional effort to identify novel genetic lesions contributing to disease pathogenesis. Recent studies based on single-nucleotide polymorphism (SNP) array-based karyotyping resulted in the detection of several copy number alterations such as a loss of heterozygosity and a copy neutral loss of heterozygosity in genomic regions containing multiple members of the polycomb repressive complex 2[9–12] and other genes previously implicated in different hematological malignancies.[13] Moreover, via candidate gene sequencing, several novel *bona fide* somatic mutations[14] were detected at frequencies ranging from 1% to 20–30% in genes frequently mutated in other myeloid neoplasms, as well as MDS and acute myeloid leukemia, and some of these mutations have been positively correlated with clinical outcome.[15–18] Nevertheless, because a significant proportion of MPN cases are negative for molecular aberrations, a complete portrait of MPN genetic abnormalities remains to be depicted. The two major theoretical and technical drawbacks to the identification of new somatic mutations are represented, respectively, by the huge number of genes potentially involved in MPN tumorigenesis and by the availability of 'pure' germline control DNA. Buccal swabs and saliva have generally been considered as readily available sources of non-hematopoietic DNA, but detection of the *JAK2*V617F mutation in at least some of these samples suggested the presence of myeloid cell contamination. In addition, no evidence for germline transmission of the *JAK2*V617F mutation has been elucidated until now.

Given these previous results and with the goal of further exploring the molecular complexity of MPN, we investigated the incidence of mutations in genes already known to be implicated in cancer pathogenesis. Therefore, we designed a two-tiered next-generation sequencing (NGS) study. We first evaluated the somatic mutational status of a mostly inclusive list of known cancer-related genes in a 25 MPN sample learning set. We then tested the recurrence of the truly somatic variants in a broader validation set of 189 patients via an amplicon-sequencing NGS approach.

MATERIALS AND METHODS

Patients and samples

Patients were diagnosed as having PV, PMF and post PV-MF according to the World Health Organization (WHO)[19] and the International Working

[1]Center for Genome Research, University of Modena and Reggio Emilia, Modena, Italy; [2]Life Sciences Department, University of Modena and Reggio Emilia, Modena, Italy; [3]Center for Regenerative Medicine 'Stefano Ferrari', University of Modena and Reggio Emilia, Modena, Italy and [4]Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy. Correspondence: Professor E Tagliafico, Center of Genome Research, University of Modena and Reggio Emilia, Via Campi 287, 41125, Modena, Italy.
E-mail: enrico.tagliafico@unimore.it
Received 15 May 2013; revised 28 September 2013; accepted 7 October 2013; accepted article preview online 22 October 2013; advance online publication, 12 November 2013

Group for Myelofibrosis Research and Treatment (IWG-MRT) criteria.[20] All subjects provided informed written consent, and the study was performed under the Florence Institutional Review Board's approved protocol. The study was conducted in accordance with the Declaration of Helsinki. The presence of the *JAK2*V617F and *MPL*W515L mutations and the mutated allele burden were determined via quantitative real-time PCR (QRT–PCR), as previously described.[21]

The mutational status of *ASXL1, EZH2, IDH1*, as well as *IDH2, SRSF2, TET2, DNMT3A* and *CBL* was assessed using Sanger sequencing, as previously described.[18] Cytogenetic analysis was performed on Giemsa-stained slides. All patients were examined within 1 year of diagnosis.

Granulocyte and CD3+ T cell purification and genomic DNA extraction

Granulocytes were obtained via the density gradient centrifugation of peripheral blood samples, and CD3+ cells were immunomagnetically selected (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) from the peripheral blood mononuclear cell fraction recovered from the density gradient. After sorting, CD3+ cells were expanded *in vitro*. The purity of the CD3+ cells was determined using flow cytometry. The culture conditions as well the DNA extraction quality control procedures are fully described in the Materials and Methods section of the Supplementary Information.

Target enrichment and 454 sequencing

A solution-based capture custom panel was designed for target enrichment according to the NimbleGen (Roche NimbleGen, Inc., Madison, WI, USA) guidelines. The final list of genes and miRNAs were obtained by combining the complete list of mutations present in the latest release of the Sanger Institute Cancer Gene Census Database (www.sanger.ac.uk/genetics/CGP/Census) with the most inclusive list of DNA repair genes present in the OMIM Database (www.ncbi.nlm.nih.gov/omim) as well as a manually curated literature screening. The custom design was inserted into a 5-Mb SeqCap EZ Choice Library (Roche NimbleGen, Inc., Madison, WI, USA) containing the exonic portions of 1400 genes and 600 miRNA coding sequences (Supplementary Table 1).

A sample library preparation was performed using 500 ng of DNA from the granulocyte and CD3+T-cell samples. Each step of the working procedure to perform sequencing runs on the Roche 454 GS FLX platform is fully described in the Materials and Methods section of the Supplementary Information.

Variant detection, filtering and classification

The processing of the samples, evaluation of genuine somatic mutations and their classification are entirely described in the Materials and Methods section of the Supplementary Information.

Recurring variant validation test

Recurrence testing for genuine somatic variants was performed using Ion AmpliSeq technology with an Ion Torrent Personal Genome Machine (PGM) platform. The Ion AmpliSeq panel design, sample processing, barcoding and sequencing are fully described in the Materials and Methods section of the Supplementary Information.

N-RAS c.35 G>A mutation analysis

An independent set of 139 patients with PMF were recruited from the archive in Florence and used for *NRAS* and *KRAS* mutation analysis (see Materials and Methods section of the Supplementary Information for details).

Statistical analysis

The χ^2, Fisher's exact test (2×2 table) or χ^2 test for trend (larger contingency table) were used as appropriate to compare variables from different patient groups that had been categorized according to mutational status. The analysis of continuous variables among the groups was performed using the Mann-Whitney *U* test (two groups) or the Kruskal–Wallis test with the Dunn method for multiple comparisons. $P<0.05$ was considered to indicate statistical significance; all tests were two-tailed. Data were processed using SPSS Version 19.0 software (StatSoft, Tulsa, OK, USA).

RESULTS

A total of 25 tumor granulocyte and paired germline samples comprised the learning cohort. The NGS samples had been collected at the time of diagnosis in 9 PV subjects and 11 PMF subjects, while the additional 5 DNA samples were obtained from 5 of the 9 PV patients at the time that they evolved to post-PV myelofibrosis.

The learning cohort for PMF was deliberately selected as being predominantly *JAK2*V617F-negative (9/11 subjects), with only 2 *JAK2*V617F-positive patients. The stratification of patients according to the dynamic international prognostic scoring system (DIPSS)[22] and other clinical features of the cohort are summarized in Supplementary Table 2.

Identification of genuine somatic mutations in coding sequences and microRNAs

According to the NimbleGen procedure, the tumor granulocytic, germline salivary and CD3+ lymphocytic gDNA samples were sheared, barcoded and mixed in an appropriate number of template libraries that were subsequently captured. At the end of the gDNA sequence selection procedure, all the DNA libraries were checked using QRT–PCR to verify the relative fold enrichment of the panel; all libraries produced a fully satisfactory result and confirmed successful captures (data not shown). Next, each library was sequenced with 454 Titanium technology using the Roche GS FLX platform. To exclude unbalanced libraries, that is, libraries composed of a nonequimolar quantity of the samples, we checked for any possible unequal sequencing depth in the barcoded samples. As exemplified for the library shown in Supplementary Figure 1, balanced quantities of samples were pooled, captured and then processed.

DNA libraries were sequenced until the median 30-fold coverage was reached for each tumor or control sample. A very high capture specificity was observed (94% of unique reads in the target region) with a similar uniformity throughout the chromosomes (average standard deviation assessed to 1.6) (Figure 1).

Genuine germline sample selection

To identify somatic mutations in the MPN learning dataset, we sequenced paired DNA samples from granulocytes and germline cells. The candidate mutations detected in the granulocytes were then screened by subtracting those occurring in the germline to enable the identification of the variants that could be reliably considered truly acquired somatic variants.

In the first set of experiments, we employed DNA obtained from paired saliva samples, but this DNA consistently presented variants belonging to the neoplastic clone, notably the *JAK2*V617F mutation, with a comparable allele burden. These results prompted us to consider that saliva samples could not be considered genuine germline sources due to contamination by myeloid cells. Thus, we replicated the experiments using expanded CD3+ T cell DNA for control samples, and a very low level of somatic contamination was found in just 1 DNA sample from CD3+ T cells. Table 1 displays the mutational burden comparison for *JAK2, MPL* and *IDH2* in libraries prepared from different sources of DNA (granulocytes, saliva and CD3+ T cells). As a result, we discarded the data obtained from the salivary samples and selected for further analyses only those obtained from CD3+ T cells, which were definitively considered germline control samples.

Somatic mutations identification in genes and microRNAs in the learning cohort

The tumor and paired germline sample data were both mapped against the human reference genome (hg19). A total of 11006 and 9691 unique variants in 1057 and 1039 genes for the germline and

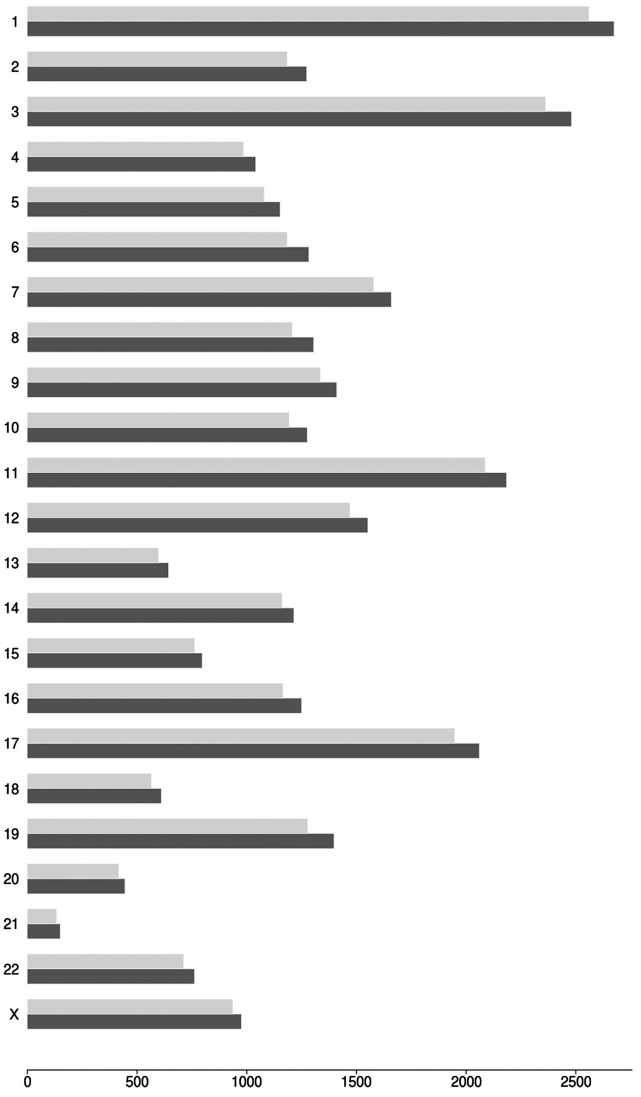

Figure 1. Enrichment uniformity landscape. The X-axis graphs the number of target regions included in the NimbleGen capture 'cancer exome' panel (approximately 29600 target regions in total). The light bars represent the number of target regions within the design, whereas the dark bars correspond to the effective number of enriched target regions for each chromosome. The Y-axis displays the chromosomes.

Table 1. Comparison of allele burden for selected mutations in libraries prepared from different DNA sources

Patient	*Granulocytes*	*CD3+ T cells*	*Salivary samples*
JAK2 V617F			
PV_4	64%	0%	67%
PMF_2	76%	0%	58%
PV_9–PPV_5	96–83%	0%	87%
PV_3	62%	0%	78%
PV_6–PPV_2	67–100%	14%	93%
IDH2 R140Q			
PMF_9	55%	0%	37%
MPL W515L			
PMF_9	80%	0%	38%

somatic samples, respectively, were detected, as shown in Supplementary Table 3.

The somatic variant identification procedure was intended to minimize the false positive somatic mutation rate. To this end, we used a two-step stringent approach. First, using a 'somatic' filter, we selected only DNA variants in tumor samples with no mutated reads either in the paired germline counterpart or in any other germline samples of the cohort. A 'functional' filter was then applied to eliminate the synonymous variants and all variants annotated in the 1000 Genomes database as having a frequency higher than 1% (see Materials and Methods for details). Supplementary Table 3 summarizes the narrowing of the unique detected variants after each of the analysis steps described above. Then, for evaluating for possible sequencing errors and the false positive rate, we validated these variants using a different type of NGS technology; specifically, we designed an Ion AmpliSeq panel containing all detected variants that was employed to re-sequence the same patient cohort via Ion Torrent PGM (1000-fold coverage).

Using this strategy, we estimated a very low sequencing error rate ($<1\%$), and we finally confirmed 136 genuine somatic, non-synonymous mutations affecting 121 genes. Twenty-five percent of these mutations are indexed in the dbSNP archive, and 2% of these specific variants are listed in COSMIC catalogue. The majority of mutations (89%) were estimated to be 'damaging' by at least 1 of the 5 algorithms that we used to investigate disease-causing potential (PolyPhen2, SIFT, Provean, Mutation Taster, LTR) (Supplementary Table 4 and Figure 2).

The vast majority of the identified somatic mutations were missense (92%), whereas the minority (8%) were indels (small insertions and deletions). Despite patients harboring different numbers of somatic mutations spanning from 1 to 21 variants (Table 2), only 14 genes appeared to be recurrently mutated (Figure 3) in at least two patients. It should be noted that the acquisition of additional mutations and/or the occurrence of loss of some mutations at the time of disease evolution from PV to post-PV myelofibrosis in patients for whom samples were available at both disease phases suggested the occurrence of sub-clone selection during disease evolution (Supplementary Table 5).

Five missense variants were identified in the 600 microRNAs coding sequences tested. These missense variants are summarized in Table 3. The *MIR662*, *MIR663* and *MIR542* sequences harbored missense mutations in their stem-loop coding region, and the *MIR17* mutation was shown to affect the miR-17-5p mature miRNA sequence 5 bases downstream of its seed region.

Somatic mutation recurrence in the validation cohort

To distinguish between the identified somatic variants and possible novel clonal drivers from clonal passenger mutations, we tested the recurrence of the above-described variants in a broader cohort of 189 patients diagnosed with PMF (91 samples, 48.2%), PV (50 patients, 26.4%) or post-PV myelofibrosis (48 samples, 25.4%). We utilized the Ion AmpliSeq panel and PGM sequencing as previously described to obtain an ultra-deep amplicon sequencing (see Supplementary Information for technical details) of the 141 variants, achieving a sample median of 1000-fold coverage. The clinical parameters of the patients comprising the validation set are summarized in Supplementary Table 2.

Excluding the *JAK2*, *MPL*, *IDH2, ASXL1, TET2, CBL* and *DNMT3A* known variants, 80 patients (42% of total) harbored at least 1 of the 141 somatic mutations tested for recurrence. Thirty somatic mutations (18.9% of the total) were displayed in at least 1 of the 189 patients; these were all missense mutations with the exception of a single frameshift mutation occurring in the *BRD4* gene (Table 4).

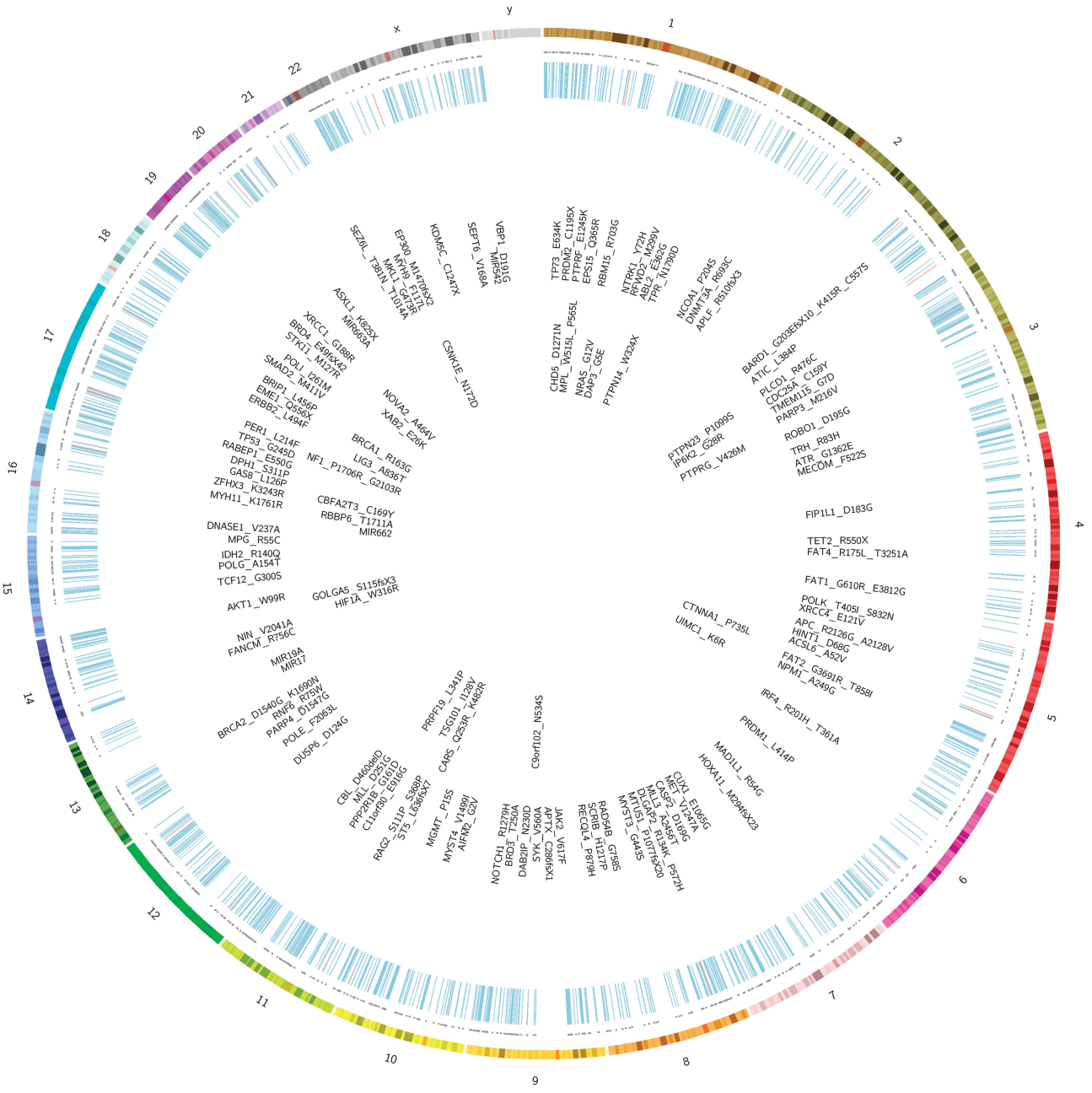

Figure 2. Circular diagram of mutations found in MPN. Chromosomes are illustrated in the outer perimeter. Grey dots show the 'cancer exome' regions of the NimbleGen panel, whereas the histograms show the captured (blue) and failed (red) target regions. MicroRNA or Gene Symbol with amino acidic change refers to the variants found in our cohort.

In addition, 8 genes (*SCRIB, MIR662, BARD1, TCF12, FAT4, DAP3, POLG* and *NRAS*) appeared as recurrently mutated in the cohort, some of which (*SCRIB* 7.9%, *MIR662* 7.4% and *BARD1* 5.3%) were more frequently mutated than previously identified, well-known mutational hotspots[18] (Table 4).

Correlations between recurrent mutations and clinical features: *NRAS* c.35 G>A mutation analysis

The groupwise associations between recurrent mutations and clinical and biological features were assessed using the χ^2 test or Fisher's exact test. Possibly because of the small number of subjects harboring each unique mutation abnormality, no significant association with clinical features was found, with the exception of mutations at codon 12 of *NRAS* (*NRAS*G12V and *NRAS*G12D) that occurred in 5 out 102 PMF patients included in the learning and validation cohorts. This association resulted in *P*-values <0.05 for the highest DIPSS-plus score categories. DIPPS-plus[23] effectively combines prognostic information from the DIPSS with karyotype, platelet count, and transfusion status to predict overall survival in PMF. This evidence prompted us to screen the *NRAS* gene for mutations in codon 12 in an independent cohort of 66 PMF patients via high-resolution melting analysis followed by Sanger sequencing validation, finding an additional 3 mutated subjects. Moreover, because *NRAS* and *KRAS* mutations have been described as possibly mutually exclusive, we also tested this cohort for *KRAS* mutations. As a whole, we found 8 of 168 MF patients (4.7%) harboring a heterozygous *NRAS* mutation in codon

Table 2. Number of genuine somatic, non-synonymous mutations harboured by each patient. Variants in MPN known mutated genes are shown in bold

Disease_# patient number	*Number of somatic mutations*	
PMF_1	4	FIP1L1D183G, **JAK2V617F**, PRPF19L341P, TPRN1790D
PMF_2	6	APCR2126G, BRCA1R163G, BRCA2K1690N, CTNNA1P735L, GAS8L126P, **JAK2V617F**
PMF_3	7	APLFR510fsX3, BRIP1L456P, DLGAP2R134K, ERBB2L494F, FANCMR756C, MPGR55C, RECQL4P879H
PMF_4	5	**DNMT3AR693C**, IRF4T361A, PLCD1R476C, PTPRFE1245K, ZFHX3K3243R
PMF_5	3	EME1Q556X, RABEP1E550G, SEPT6V168A
PMF_6	5	BARD1C557S, PER1L214F, SEZ6LT1014A, SEZ6LT381N, SYKV560A
PMF_7	10	ACSL6A52V, AIFM2G2V, AKT1W99R, ATRG1362E, CARSQ253R, IRF4R201H, RNF6R75W, TP53G245D, TP73E634K, TRHR83H
PMF_8	7	CDC25AC159Y, EPS15Q365R, FAT1E3812G, **MPLW515L**, NRASG12V, RFWD2M299V, TCF12G300S
PMF_9	10	**ASXL1K825X**, DAB2IPN230D, FAT2T858I, HIF1AW316R, **IDH2R140Q**, KDM5CC1247X, **MPLP565L**, **MPLW515L**, NPM1A249G, NRASG12V
PMF_10	3	CASP2D169G, HINT1D68G, SMAD2M411V
PMF_11	4	ATICL384P, C9orf102N534S, NOVA2A464V, RBBP6T1711A
PPV_1	2	**JAK2V617F**, STK11M127R
PPV_2	10	BARD1G203EfsX10, CHD5D1271N, CUX1E1065G, DNASE1V237A, DUSP6D124G, **JAK2V617F**, NF1G2103R, NF1P1706R, NOTCH1R1279H, PRDM2C1195X
PPV_3	3	**JAK2V617F**, **TET2R550X**, TMEM115G7D
PPV_4	7	**CBLD460delD**, DAP3G5E, GOLGA5S115fsX3, **JAK2V617F**, MTUS1P1077fsX20, NCOA1P204S, SCRIBH1217P
PPV_5	7	BRCA2D1540G, IP6K2G28R, **JAK2V617F**, MGMTP15S, NTRK1Y72H, POLII261M, XRCC4E121V
PV_1	4	FAT4R175L**, JAK2V617F**, RAG2S111P, RBM15R703G
PV_2	1	**JAK2V617F**
PV_3	21	ABL2E362G, BARD1K415R, C11orf30E916G, DLGAP2P572fsX72, FAT1G610R, FAT4T3251A, **JAK2V617F**, MAD1L1R54G, MECOMF522S, MLLD251G, MYH9F117L, NINV2041A, PARP4D1547G, POLEF2063L, POLKS832N, PRDM1L414P, RAD54BG758S, RAG2S368P, TSG101I128V, UIMC1K6R, VBP1D191G
PV_4	6	BRD4E49fsX42, JAK2V617F, LIG3A836T, POLKT405I, PTPRGV426M, XAB2E26K
PV_5	7	APCA2128V, CBFA2T3C169Y, **JAK2V617F**, PPP2R1BG161D, PTPN23P1099S, ST5L636fsX7, STK11M127R
PV_6	5	BRD3T250A, **JAK2V617F**, MKL1G473R, NF1G2103R, NOTCH1R1279H
PV_7	7	DPH1S311P, HOXA11M294fsX23, **JAK2V617F**, PARP3M216V, ROBO1D195G, TET2R550X, TMEM115G7D
PV_8	6	CARSK482R, CSNK1EN172D**, JAK2V617F**, PTPN14W324X, SCRIBH1217P, XRCC1G188R
PV_9	11	APTXC286fsX1, EP300M1470fsX2, FAT2G3691R, **JAK2V617F**, METV1247A, MLL3A2456T, MYH11K1761R, MYST3G443S, POLGA154T, POLII261M, XRCC4E121V

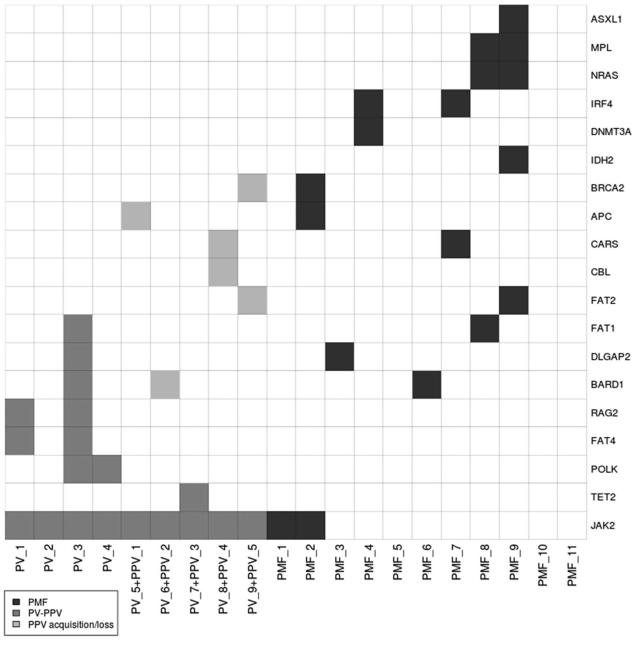

Figure 3. Heatmap of the found known variants and of genes presenting one or more variants in two or more patients in the data set. The horizontal axis presents the sequenced complete dataset of patients, with the PV samples grouped on the left, the evolution to post-PV MF patients (PPV) in the center and the PMF on the right. The vertical axis illustrates the recurrently mutated gene as exemplified in the legend.

12 (5 harbored the *NRAS*G12V mutation and 3 harbored the *NRAS*G12D mutation). In addition, 3 patients of 168 evaluated (1.6%) harbored heterozygous *KRAS* mutations (G12R, G12S, and G13D); the patient carrying the *KRAS*G13D mutation also harbored the *NRAS*G12V variant. Of note, *NRAS* variants preferentially clustered among *JAK2* wild-type subjects since only 1 of 8 mutated NRAS patients also harbored the *JAK2*V617F.

Finally, we confirmed a significant association of *NRAS* variants with DIPSS-plus scoring ($P = 0.022$) since all the 8 mutated patients were included in the highest (intermediate-2 and high) risk category, as shown in Table 5.

DISCUSSION

The discovery of the *JAK2*V617F mutation represents the single most significant contribution to the characterization of the pathophysiology of MPN thus far and has major implications also for the treatment of these malignancies. Mutations characterizing genes other than *JAK2* involve less than the 20% of MPN patients and often are co-expressed with the *JAK2* mutation; thus, even though the impact of the mutational status of a specific set of genes (*ASXL1, EZH2, SRSF2* and *IDH*) on disease outcome has been demonstrated in patients with PMF,[18] a comprehensive molecular landscape of MPN has not yet been completely depicted. Here, we performed the first large targeted NGS analysis aimed at exploring the mutational status of the broadest panel of known cancer-associated genes in MPN. This is the first report describing a robust NGS study design and an accurate data analysis pipeline aimed at minimizing the somatic mutation false positive rate. We also demonstrated that saliva samples are often heavily contaminated by myeloid cells and that expanded CD3+ T cells in culture therefore serve as the most

Table 3. Somatic mutations in microRNAs coding regions

miRNA ID	*Chromosome*	*SNP*	*Variant_effect*	*miRNA location*	*Region*
MIR662	chr16:820215	rs74656628	T>A p.T924S	MSLNL gene	Pre-miRNA
MIR17	chr13:92002884	unknown	A>G	intergenic	Mature sequence hsa-miR-17-5p
MIR19A	chr13:92003195	unknown	T>C	intergenic	Pre-miRNA
MIR542	chrX:133675465	unknown	C>T	intergenic	Pre-miRNA
MIR663A	chr20:26188912	rs7266947	A>C	intergenic	Pre-miRNA

Table 4. Recurrent mutations in the validation dataset

Chromosome	*Reference nucleotide*	*Variant*	*Gene*	*AA Change*	*Pathology (PV|PPV|PMF)*	*Freq (PV|PPV|PMF)*	*Freq tot*	*PV thrombotic risk (Low|High)*	*PPV-PMF IPSS (Low|Int-1|Int-2|High|Unknown)*
chr9	G	T	JAK2	V617F	50|48|72	100.00|100.00|79.12	90.0	14|36	17|33|41|21|8
chr8	T	G	SCRIB	H1217P	5|5|5	10.00|10.42|5.49	7.9	2|3	1|1|4|4|0
chr16	T	A	MIR662		2|6|6	4.00|12.50|6.59	7.4	1|1	2|1|6|3|0
chr2	C	G	BARD1	C557S	1|3|6	2.00|6.25|6.59	5.3	1|0	2|3|1|3|0
chr4	G	T	FAT4	R175L	2|2|3	4.00|4.17|3.30	3.7	0|2	0|2|1|1|1
chr15	G	A	TCF12	G300S	1|2|4	2.00|4.17|4.40	3.7	1|0	0|1|2|3|0
chr1	G	A	DAP3	G5E	3|1|2	6.00|2.08|2.20	3.2	0|3	1|0|1|1|0
chr1	C	A	NRAS	G12V	0|0|3	0.00|0.00|3.30	1.6	0|0	0|0|1|2|0
chr15	C	T	POLG	A154T	1|2|0	2.00|4.17|0.00	1.6	0|1	1|1|0|0|0
chr8	G	T	RECQL4	P879H	0|1|1	0.00|2.08|1.10	1.1	0|0	0|0|1|1|0
chr10	C	A	AIFM2	G2V	1|1|0	2.00|2.08|0.00	1.1	1|0	0|0|0|1|0
chr3	C	T	PTPN23	P1099S	1|0|1	2.00|0.00|1.10	1.1	0|1	0|1|0|0|0
chr2	G	A	DNMT3A	R693C	0|0|2	0.00|0.00|2.20	1.1	0|0	0|2|0|0|0
chr3	C	T	TMEM115	G7D	0|1|0	0.00|2.08|0.00	0.5	0|0	1|0|0|0|0
chr5	A	T	XRCC4	E121V	0|1|0	0.00|2.08|0.00	0.5	0|0	0|1|0|0|0
chr17	G	A	NF1	G2103R	0|1|0	0.00|2.08|0.00	0.5	0|0	0|1|0|0|0
chr4	C	T	TET2	R550X	0|1|0	0.00|2.08|0.00	0.5	0|0	1|0|0|0|0
chr17	C	G	NF1	P1706R	0|1|0	0.00|2.08|0.00	0.5	0|0	0|1|0|0|0
chr16	C	T	MPG	R55C	0|0|1	0.00|0.00|1.10	0.5	0|0	0|1|0|0|0
chr8	C	T	MYST3	G443S	0|0|1	0.00|0.00|1.10	0.5	0|0	0|0|1|0|0
chr15	C	T	IDH2	R140Q	0|1|0	0.00|2.08|0.00	0.5	0|0	1|0|0|0|0
chr1	T	C	RFWD2	M299V	0|0|1	0.00|0.00|1.10	0.5	0|0	0|1|0|0|0
chr19	C	CG	BRD4	E49fsX42	0|0|1	0.00|0.00|1.10	0.5	0|0	0|0|1|0|0
chr5	G	A	POLK	S832N	0|0|1	0.00|0.00|1.10	0.5	0|0	0|0|1|0|0
chr7	T	C	MAD1L1	R54G	1|0|0	2.00|0.00|0.00	0.5	1|0	0|0|0|0|0
chr22	C	T	MKL1	G473R	0|1|0	0.00|2.08|0.00	0.5	0|0	0|1|0|0|0
chr1	G	T	MPL	W515L	0|0|1	0.00|0.00|1.10	0.5	0|0	0|0|1|0|0
chr1	T	G	PDE4DIP	I303L	0|0|2	0.00|0.00|2.20	1.06	0|0	0|1|0|1|0
chr1	G	T	NTRK1	G613V	7|4|7	14.00|8.33|7.69	9.52	2|5	2|2|5|1|1
chr16	T	G	ZFHX3	T428P	1|0|1	2.00|0.00|1.10	1.06	1|0	0|0|1|0|0

reliable germline control for identifying true somatic mutations in MPN. We set up an analysis pipeline using the most stringent procedure to avoid false-positive calls of somatic mutations. In particular, paired germline and somatic DNA samples of the learning dataset were sequenced for reaching the same fold coverage. Moreover, we called somatic only those variants with no reads both in the paired germline DNA and in any other germline sample of the cohort. To these stringent 'somatic' filters, two additional controls were added to discard any possible polymorphisms (only variants with a frequency $<1\%$ in the 1000 genomes database were retained) as well as possible benign mutations (only non-synonymous variants were retained). Finally, all filtered variants were annotated with the functional effect prediction of five different algorithms.

Using this multistep bioinformatics pipeline, we finally identified 141 'genuine' somatic non-synonymous mutations affecting 121 genes and 5 miRNAs that were then tested for recurrence in a larger cohort of 189 patients. The variants found in the *SCRIB, MIR662, BARD1, TCF12, FAT4, DAP3, POLG* and *NRAS* genes were recurrent with a frequency higher than 3%. In particular, *SCRIB*, *MIR662* and *BARD1* showed frequencies of 7.9, 7.4, and 5.3%, respectively, which were higher than those described for some well-known mutational hotspots.[18]

Some findings appear to join some of these genes, suggesting the potential role of these genes in the pathogenesis of MPN. SCRIB and FAT4 are two proteins that regulate planar cell polarity differentiation. *SCRIB* encodes a cytoplasmic scaffolding protein consisting of leucine-rich repeats and PDZ domains that regulates protein–protein interactions,[24] while FAT4 belongs to the E-cadherin family and may control noncanonical Wnt/planar cell polarity signaling;[25] both pathways play a crucial role in the regulation of polarity and tissue homeostasis. Interestingly, TCF12 protein, also known as HEB, could be linked with this pathway. TCF12 forms heterodimers with other bHLH E-proteins and with chimeric protein AML1-ETO[26–28] and works as a transcriptional repressor of E-cadherin, thus playing an important role in cancer cell progression by enhancing the epithelial–mesenchymal transition process.[29] The endothelial-mesenchymal transition is a form of the more widely known epithelial–mesenchymal transition; similarly to epithelial–mesenchymal transition, endothelial-mesenchymal transition can be induced by transforming growth factor-β and allows a polarized cell, which normally interacts with

Table 5. Clinical features of PMF patients screened for NRAS mutations

Variables	*NRAS WT* (n = *160)*	*NRAS mutated* (n = *8)*	P
Follow-up, months; median (range)	35.6 (3.01–145.72)	28.0 (4.03–275.57)	ns
Age in years; median (range)	66 (19–90)	64 (48–84)	ns
Males (%)	109 (68.1%)	6 (75%)	ns
Hemoglobin, g/dL; median (range)	12.0 (5.0–16.0)	11.5 (10.0–13.0)	ns
Leukocytes, $\times 10^9$/L; median (range)	10.5 (3.0–20.0)	9.8 (1.4–99.6)	ns
Platelets, $\times 10^9$/L; median (range)	274 (19–1563)	287 (90–738)	ns
Constitutional symptoms; *n* (%)	85 (53.1%)	7 (87.5%)	0.044
Circulating blasts $\geqslant$1%; *n* (%)	35 (21.8%)	4 (57.1%)	ns
Cytogenetic categories; *n* (%)[a]			ns
Abnormal	48 (48%)	4 (50%)	
Unfavorable karyotype	15 (15%)	2 (25%)	
DIPSS-plus risk group; *n* (%)			0.022
Low	16 (10%)	0	
Intermediate- 1	49 (30.6%)	0	
Intermediate- 2	65 (40.6%)	6 (75%)	
High	30 (18.8%)	2 (25%)	
Palpable spleen; *n* (%)	98 (61.2%)	4 (50%)	ns
JAK2V617F; *n* (%)	96 (60%)	1 (12.5%)	0.001
Progression to acute leukemia; *n* (%)	12 (7.5%)	0	ns
Dead for disease progression; *n* (%)	49 (30.6%)	5 (62.5%)	0.043

[a]Evaluated on available data (*n* = 100/160 for RAS wild-type and 8/8 for RAS mutated. Abbreviations: DIPSS-plus, dynamic international prognostic scoring system-plus; IPSS, international prognostic scoring system.

the basement membrane via its basal surface, to undergo multiple changes that enable it to assume a mesenchymal cell phenotype, which includes enhanced migratory capacity, invasiveness, elevated resistance to apoptosis and an increased ability to induce fibrosis. Interestingly, endothelial-mesenchymal transition and the resulting endothelial cell fate have been recently implicated in the pathogenesis of PMF.[30–32]

The *SCRIB*H1217P variant is a missense mutation predicted to be damaging via SIFT and PolyPhen2. In particular, missense mutations in the same SCRIB c-terminal domain region significantly disrupt the membrane subcellular localization of the protein, and this was suggested to be one possible pathogenic mechanisms for planar cell polarity alterations in mammals.[33,34] Similarly, the *FAT4*R175L mutation is located in the extracellular cadherin domain, and Polyphen2 together with MutationTaster were used to reveal that this *FAT4* mutation is predicted to adversely affect protein function and alter the planar cell polarity environment. Furthermore, studies in *Drosophila* and mammalian cell lines have shown that SCRIB loss and RAS activation cooperate (interclonally or intraclonally) to promote invasion.[35–37] In Drosophila, interclonal cooperation in RasV12 and *Scrib*-minus tumor clones revealed a two-level mechanism in which *Scrib*-minus cells promote the neoplastic development of RasV12 cells. Specifically, this mechanism involves (1) the spread of stress-induced JNK activity from scrib-minus cells to RasV12-activated cells followed by (2) the expression of JAK/STAT-activating cytokines downstream of JNK.[38]

Further insights into the molecular and pathogenetic complexity of MPN emerged from the discovery of *BARD1*, *POLG* and *DAP3* mutations. All three of these variants are predicted to be damaging, and while DAP3 and POLG are mitochondrial proteins involved in DNA repair and apoptosis pathways,[39,40] BARD1 is a nuclear BRCA1-independent mediator between genotoxic stress and p53-dependent apoptosis[41] (see the Discussion section of the Supplementary Information for details).

In addition to above considerations, we were intrigued by the fact that the two PMF patients in the learning cohort showing the *NRASG12V* mutation presented a rapid progression into an accelerated form of the disease. Therefore, we attempted to validate this association in an additional 66 PMF cases, composing a final cohort of 168 PMF. Moreover, because *NRAS* and *KRAS* mutations have been described as potentially mutually exclusive,[42,43] we tested this additional cohort for *KRAS* mutations to verify mutual exclusivity. We found that 4.7% of the PMF patients harbored a heterozygous *NRAS* mutation in codon 12. Conceivably, considering all the 8 NRAS mutated patients all together, we found that this mutation was associated with a poorer prognosis as supported by the fact that the 8 patients clustered in the intermediate-2 and high-risk category of the DIPSS-plus score. Only 2% of all PMF patients harbored heterozygous *KRAS* mutations (G12R, G12S and G13D), while the co-occurrence of *NRAS*G12V and *KRAS*G13D or *JAK2*V617F mutations was observed only for a single patient (we were unable to determine whether the two mutations arose from the same or different clone since frozen cell samples were not available for this patient). The low number of patients harboring *KRAS* mutations alone, precluded any meaningful analysis of the association with disease progression. Overall, these data suggest that *NRAS* mutations specifically associate with a poorer outcome, although the molecular mechanism remains to be investigated.

Mutations in microRNAs warrant a separate discussion. Even if additionally specific studies are necessary to support a functional role of the *MIR662* variant rs74656628, some interesting features are worth mentioning. The human *MIR662* is an intragenic microRNA that resides in a non-coding exon sequence of the mesothelin-like gene. This is a heterozygous missense mutation that occurs in the precursor sequence of the micro-RNA.[38] A bioinformatics prediction analysis run with the *In-Silico*-Dicer and RNAFold suggested that this mutation could modify the RNA secondary structure and lead to the production of a different mature miRNA (see Supplementary Information for details).

In summary, this NGS study presents new data that contribute to elucidating the very high genomic complexity in MPN disorders and identifies new variants in cancer-related genes that are potentially involved in the pathogenesis of the disease and may deserve further studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Fondazione Cassa di Risparmio di Carpi (http://www.fondazionecrcarpi.it/), Associazione Italiana per la Ricerca sul Cancro (AIRC, Milano, http://www.airc.it/english/

obiettivi-risultati-english.asp), project number 10005 'Special Program Molecular Clinical Oncology 5 × 1000' to AGIMM (AIRC-Gruppo Italiano Malattie Mieloproliferative, http://www.progettoagimm.it), AIRC project number 12055, Italian Ministry of University & Research (FIRB Project 2011, project number RBAP11CZLK_002 and PRIN 2010–11, project number 2010NYKNS7, http://www.istruzione.it/web/ricerca) and Project Tecnopolo (Regione Emilia Romagna http://www.aster.it/tiki-index.php?page=TecnopoloMo), AIL Modena ONLUS (http://www.ailmodena.org/).

REFERENCES

1 Kralovics R, Teo SS, Buser AS, Brutsche M, Tiedt R, Tichelli A *et al.* Altered gene expression in myeloproliferative disorders correlates with activation of signaling by the V617F mutation of Jak2. *Blood* 2005; **106**: 3374–3376.
2 Campbell PJ, Green AR. The myeloproliferative disorders. *N Engl J Med* 2006; **355**: 2452–2466.
3 Milosevic JD, Kralovics R. Genetic and epigenetic alterations of myeloproliferative disorders. *Int J Hematol* 2013; **97**: 183–197.
4 Levine RL, Pardanani A, Tefferi A, Gilliland DG. Role of JAK2 in the pathogenesis and therapy of myeloproliferative disorders. *Nat Rev Cancer* 2007; **7**: 673–683.
5 Pardanani A, Laborde RR, Lasho TL, Finke C, Begna K, Al-Kali A *et al.* Safety and efficacy of CYT387, a JAK1 and JAK2 inhibitor, in myelofibrosis. *Leukemia* 2013; **27**: 1322–1327.
6 Kralovics R. Genetic complexity of myeloproliferative neoplasms. *Leukemia* 2008; **22**: 1841–1848.
7 Kilpivaara O, Levine RL. JAK2 and MPL mutations in myeloproliferative neoplasms: discovery and science. *Leukemia* 2008; **22**: 1813–1817.
8 Pikman Y, Lee BH, Mercher T, McDowell E, Ebert BL, Gozo M *et al.* MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS medicine* 2006; **3**: e270.
9 Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* 2010; **42**: 722–726.
10 Score J, Hidalgo-Curtis C, Jones AV, Winkelmann N, Skinner A, Ward D *et al.* Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. *Blood* 2012; **119**: 1208–1213.
11 Puda A, Milosevic JD, Berg T, Klampfl T, Harutyunyan AS, Gisslinger B *et al.* Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *Am J Hematol* 2012; **87**: 245–250.
12 Brecqueville M, Cervera N, Adelaide J, Rey J, Carbuccia N, Chaffanet M *et al.* Mutations and deletions of the SUZ12 polycomb gene in myeloproliferative neoplasms. *Blood CancerJ* 2011; **1**: e33.
13 Klampfl T, Harutyunyan A, Berg T, Gisslinger B, Schalling M, Bagienski K *et al.* Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood* 2011; **118**: 167–176.
14 Brecqueville M, Rey J, Bertucci F, Coppin E, Finetti P, Carbuccia N *et al.* Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. *Genes Chromosome Canc* 2012; **51**: 743–755.
15 Lasho TL, Jimma T, Finke CM, Patnaik M, Hanson CA, Ketterling RP *et al.* SRSF2 mutations in primary myelofibrosis: significant clustering with IDH mutations and independent association with inferior overall and leukemia-free survival. *Blood* 2012; **120**: 4168–4171.
16 Tefferi A. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. *Leukemia* 2010; **24**: 1128–1138.
17 Tefferi A, Jimma T, Sulai NH, Lasho TL, Finke CM, Knudson RA *et al.* IDH mutations in primary myelofibrosis predict leukemic transformation and shortened survival: clinical evidence for leukemogenic collaboration with JAK2V617F. *Leukemia* 2012; **26**: 475–480.
18 Vannucchi AM, Lasho TL, Guglielmelli P, Biamonte F, Pardanani A, Pereira A *et al.* Mutations and prognosis in primary myelofibrosis. *Leukemia* 2013; **27**: 1861–1869.
19 Tefferi A, Thiele J, Orazi A, Kvasnicka HM, Barbui T, Hanson CA *et al.* Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an *ad hoc* international expert panel. *Blood* 2007; **110**: 1092–1097.
20 Mesa RA, Verstovsek S, Cervantes F, Barosi G, Reilly JT, Dupriez B *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.
21 Guglielmelli P, Barosi G, Specchia G, Rambaldi A, Lo Coco F, Antonioli E *et al.* Identification of patients with poorer survival in primary myelofibrosis based on the burden of JAK2V617F mutated allele. *Blood* 2009; **114**: 1477–1483.
22 Scott BL, Gooley TA, Sorror ML, Rezvani AR, Linenberger ML, Grim J *et al.* The Dynamic International Prognostic Scoring System for myelofibrosis predicts outcomes after hematopoietic cell transplantation. *Blood* 2012; **119**: 2657–2664.
23 Gangat N, Caramazza D, Vaidya R, George G, Begna K, Schwager S *et al.* DIPSS plus: a refined Dynamic International Prognostic Scoring System for primary myelofibrosis that incorporates prognostic information from karyotype, platelet count, and transfusion status. *J Clin Oncol* 2011; **29**: 392–397.
24 Humbert PO, Grzeschik NA, Brumby AM, Galea R, Elsum I, Richardson HE. Control of tumourigenesis by the Scribble/Dlg/Lgl polarity module. *Oncogene* 2008; **27**: 6888–6907.
25 Wang Y. Wnt/Planar cell polarity signaling: a new paradigm for cancer therapy. *Mol Cancer Ther* 2009; **8**: 2103–2109.
26 Parker MH, Perry RL, Fauteux MC, Berkes CA, Rudnicki MA. MyoD synergizes with the E-protein HEB beta to induce myogenic differentiation. *Mol Cell Biol* 2006; **26**: 5771–5783.
27 Park S, Chen W, Cierpicki T, Tonelli M, Cai X, Speck NA *et al.* Structure of the AML1-ETO eTAFH domain-HEB peptide complex and its contribution to AML1-ETO activity. *Blood* 2009; **113**: 3558–3567.
28 Goardon N, Lambert JA, Rodriguez P, Nissaire P, Herblot S, Thibault P *et al.* ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *The EMBO journal* 2006; **25**: 357–366.
29 Lee CC, Chen WS, Chen CC, Chen LL, Lin YS, Fan CS *et al.* TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. *J Biol Chem* 2012; **287**: 2798–2809.
30 Wynn TA, Ramalingam TR. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nat Med* 2012; **18**: 1028–1040.
31 van Meeteren LA, ten Dijke P. Regulation of endothelial cell plasticity by TGF-beta. *Cell Tissue Res* 2012; **347**: 177–186.
32 Rosti V, Villani L, Riboni R, Poletto V, Bonetti E, Tozzi L *et al.* Spleen endothelial cells from patients with myelofibrosis harbor the JAK2V617F mutation. *Blood* 2013; **121**: 360–368.
33 Lei Y, Zhu H, Duhon C, Yang W, Ross ME, Shaw GM *et al.* Mutations in planar cell polarity gene scrib are associated with spina bifida. *PloS One* 2013; **8**: e69262.
34 Robinson A, Escuin S, Doudney K, Vekemans M, Stevenson RE, Greene ND *et al.* Mutations in the planar cell polarity genes CELSR1 and SCRIB are associated with the severe neural tube defect craniorachischisis. *Hum Mutat* 2012; **33**: 440–447.
35 Brumby AM, Richardson HE. Scribble mutants cooperate with oncogenic Ras or Notch to cause neoplastic overgrowth in Drosophila. *The EMBO journal* 2003; **22**: 5769–5779.
36 Dow LE, Elsum IA, King CL, Kinross KM, Richardson HE, Humbert PO. Loss of human Scribble cooperates with H-Ras to promote cell invasion through deregulation of MAPK signalling. *Oncogene* 2008; **27**: 5988–6001.
37 Wu M, Pastor-Pareja JC, Xu T. Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion. *Nature* 2010; **463**: 545–548.
38 Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P. Copy number variation of microRNA genes in the human genome. *BMC Genomics* 2011; **12**: 183.
39 Miller JL, Koc H, Koc EC. Identification of phosphorylation sites in mammalian mitochondrial ribosomal protein DAP3. *Protein Sci* 2008; **17**: 251–260.
40 Mukamel Z, Kimchi A. Death-associated protein 3 localizes to the mitochondria and is involved in the process of mitochondrial fragmentation during cell death. *J Biol Chem* 2004; **279**: 36732–36738.
41 Irminger-Finger I, Leung WC, Li J, Dubois-Dauphin M, Harb J, Feki A *et al.* Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis. *Molecular cell* 2001; **8**: 1255–1266.
42 Fernandez-Medarde A, Santos E. Ras in cancer and developmental diseases. *Genes & cancer* 2011; **2**: 344–358.
43 Paulsson K, Horvat A, Strombeck B, Nilsson F, Heldrup J, Behrendtz M *et al.* Mutations of FLT3, NRAS, KRAS, and PTPN11 are frequent and possibly mutually exclusive in high hyperdiploid childhood acute lymphoblastic leukemia. *Genes Chromosome Canc* 2008; **47**: 26–33.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)