ORIGINAL ARTICLE

Targeted cancer exome sequencing reveals recurrent mutations in myeloproliferative neoplasms

E Tenedini1,2,3, I Bernardi1,2, V Artusi1,2, L Artuso1,2, E Roncaglia1,2, P Guglielmelli1,2, L Pieri1, C Bogani1, F Biamonte1, G Rotunno1, C Mannarelli1, E Bianchi2,3, A Pancrazzi4, T Fanelli4, G Malagoli Tagliazucchi1,2, S Ferrari1,2, R Manfredini2,3, AM Vannucchi4 and E Tagliafico1,2 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 (SCRB, 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 (NRASG12V and NRASG12D) 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 JAK2V617F mutation in 20051 represented a major breakthrough in the understanding of the molecular pathogenesis of Philadelphia chromosome-negative chronic myeloproliferative neoplasms (MPN).2 The JAK2V617F 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 and phenotype of MPN and therefore serves as a rational drug target for therapy.5 Additional mutations were described at codon 12 of MPL (WS15L/K/A) in 5–8% of ET and PMF patients.5–8 However, the significant proportion of JAK2V617F- and MPLW515L-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 complex23–12 and other genes previously implicated in different hematological malignancies.13 Moreover, via candidate gene sequencing, several novel bona fide somatic mutations14 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 JAK2V617F mutation in at least some of these samples suggested the presence of myeloid cell contamination. In addition, no evidence for germline transmission of the JAK2V617F 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 Group.20 MPN cases included 189 patients with well-documented clinical, hematological and molecular characteristics.21–23

1Center for Genome Research, University of Modena and Reggio Emilia, Modena, Italy; 2Life Sciences Department, University of Modena and Reggio Emilia, Modena, Italy; 3Center for Regenerative Medicine ‘Stefano Ferrari’, University of Modena and Reggio Emilia, Modena, Italy and 4Department 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

Copyright © 2014 Macmillan Publishers Limited All rights reserved
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.

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.

Recurrent variant validation test

Recurrent 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 χ², Fisher’s exact test (2×2 table) or χ² 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 JAK2V617F-negative (9/11 subjects), with only 2 JAK2V617F-positive patients. The stratification of patients according to the dynamic international prognostic scoring system (DIPSS)17 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 qDNA 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 JAK2V617F 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 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 BDRD4 gene (Table 4).
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 hotspots18 (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 \( \chi^2 \) 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 (NRASG12V and NRASG12D) 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

---

**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.
The discovery of the JAK2V617F 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, 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

**DISCUSSION**

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

| Disease_# | Number of somatic mutations |
|-----------|-----------------------------|
| PMF_1     | 4 FIP1L1D183G, JAK2V617F, PRPF19L341P, TPRN1790D |
| PMF_2     | 6 APCR2126G, BRCAlR163G, BRCAlK1690N, CTNN1P735L, GAS8L126G, JAK2V617F |
| PMF_3     | 7 APLFR8105X3, BRRP1L456R, DLRP2R134K, ERBB2L494E, FANCMR756C, MPRGR55C, RECOl4P879H |
| PMF_4     | 5 EM1Q556X, RAPEB1550G, SEPT6V168A |
| PMF_5     | 3 BARD1C557S, PER1L214F, SEZILT1014A, SEZ6LT381N, SYK560A |
| PMF_6     | 5 ABL2E362G, BARD1K415R, C11orf30E916G, DLGAP2P572fsX72, FAT1G610R, FAT4T3251A |
| PMF_7     | 10 ABL2E362G, BARD1K415R, C11orf30E916G, DLGAP2P572fsX72, FAT1G610R, FAT4T3251A |
| PMF_8     | 7 CDC25AC159Y, EPS15Q3635, FAT1E3812G, MPLW515L, NRASG12V, RFWD2M299V, TCF12G300S |
| PMF_9     | 10 AXSL1K282X, DAB2IPN230D, FAT2B851I, HIF1A1W366R, IDH2R140Q, KDM5CC124X, MPLP565L, MPLW515L, NPM1A249G, NRASG12V |
| PMF_10    | 3 CASP2D169G, HINTT1D68G, MAD2M411V |
| PMF_11    | 4 APCR2126G, BRCAlR163G, BRCAlK1690N, CTNN1P735L, GAS8L126G, JAK2V617F |
| PMF_12    | 4 ATICL384P, C9orf102N534S, NOVA2A464V, RBPP6T1711A |
| PPV_2     | 1 JAK2V617F, STK11M27R |
| PPV_3     | 10 BARD1D2033E, CHD5D1271N, CUX1E1065G, DUSP6D124G, JAK2V617F, NFI1P1706R, NOTCH1R1279H, PRDM2C1195X |
| PPV_4     | 3 JAK2V617F, TET2R550X, TME1M15G7D |
| PPV_5     | 7 BRCA2D1540G, IP6K2G28R, JAK2V617F, MGMTTP155, NTRK1Y72H, POLI1261M, XRC4E1217V |
| PV_1      | 4 FATR175L, JAK2V617F, RAG2S111P, RBM15R703G |
| PV_2      | 1 JAK2V617F |
| PV_3      | 21 ABL2E362G, BARD1D415R, C11orf30E916G, DLGAP2P572fsX72, FAT1G610R, FAT4T3251A, JAK2V617F, MAD2M411V, MECONM522S, MLLD251G, MYH9F117L, NINV1024I, PARPD5417G, PLEF2063L, POLKS32N, PRMD1L141P, RAD54BG75S, RAG2S368P, TSG11011L28, UIMCIK16R, VBPI1D191G |
| PV_4      | 6 BRD4E494X8, JAK2V617F, LIG3AS843, POLK4T405I, PTPG4V263, XAB2E268 |
| PV_5      | 5 APCA2128V, CBAF2T3C169Y, JAK2V617F, PPPR1B161D, PTN23P1099S, ST5L366fsX7, STK1IM27R |
| PV_6      | 5 BRT3T250A, JAK2V617F, MKLI4735R, NFI1G2103R, NOTCH1R1279H |
| PV_7      | 7 DPH1S311P, HOXA11M294fsX23, JAK2V617F, PAR3M216V, ROB01D195G, TET2R550X, TME1M15G7D |
| PV_8      | 6 CARSK482R, CSNK1E172D, JAK2V617F, PTTPN4W324X, SCRIBH1217P, XRC1C180R |
| PV_9      | 11 APTX2C865X1, E3PO3M14706fsX2, FAT2G3691R, JAK2V617F, METVI1247A, MLL32456T, MYH11K1761R, MYST3G443S, POLGA154T, POLI1261M, XRC4E1217V |

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.
and MIR662 recurrent with a frequency higher than 3%. In particular, SCRIB, larger cohort of 189 patients. The variants found in the fied 141 'genuine' somatic non-synonymous mutations affecting prediction of five different algorithms. all filtered variants were annotated with the functional effect mutations (only non-synonymous variants were retained). Finally, 1000 genomes database were retained) as well as possible benign polymorphisms (only variants with a frequency
additional controls were added to discard any possible sample of the cohort. To these stringent 'somatic' filters, two reads both in the paired germline DNA and in any other germline coverage. Moreover, we called somatic only those variants with no learning dataset were sequenced for reaching the same fold procedure to avoid false-positive calls of somatic mutations. MPN. We set up an analysis pipeline using the most stringent reliable germline control for identifying true somatic mutations in MPN. We set up an analysis pipeline using the most stringent method to avoid false-positive calls of somatic mutations. SCRIB and FAT4 are two proteins that regulate planar cell polarity signaling;25 both pathways play a crucial role in the transition process.29 The endothelial-mesenchymal transition is a cell polarity signaling;25 both pathways play a crucial role in the transition process.29 The endothelial-mesenchymal transition is a cell progression by enhancing the epithelial–mesenchymal transition process.29 The endothelial-mesenchymal transition can be induced by transforming growth factor-β and allows a polarized cell, which normally interacts with

### 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 | Pre-miRNA |
| MIR19A   | chr13:92000195 | unknown | T>C | intergenic | Pre-miRNA |
| MIR542   | chrX:133657465 | unknown | C>T | intergenic | Pre-miRNA |
| MIR662A  | 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 | 90.00 | 14|36 | 17|33|41|21|18 |
| chr8       | T                     | G       | SCRIB| H1217P   | 5|5 |5 | 10.00|10.42|5.49 | 7.90 | 2|3 | 1|14|4|0 |
| chr7       | T                     | A       | AML2 | unknown  | 2|6 |6 | 4.00|12.06|5.69 | 7.40 | 1|1 | 2|16|3|0 |
| chr7       | C                     | G       | BARD1| C557S    | 1|3 |6 | 2.00|6.25|6.59 | 5.30 | 1|10 | 2|3|1|30 |
| chr4       | G                     | T       | FAT4 | R175L    | 2|2 |3 | 4.00|17.33|3.70 | 3.70 | 0|2 | 0|1 | 2|1 |1 |
| chr15      | G                     | A       | TCF12 | G300S    | 1|2 |4 | 2.00|17.44|3.70 | 3.70 | 1|0 | 0|12|3|0 |
| chr1       | G                     | A       | DAP3 | G5E      | 3|1 |2 | 6.00|2.08|2.20 | 3.20 | 0|3 | 0|1 | 1|0 |1 |
| chr1       | A                     | NNAS    | G12V | unknown  | 0|3 |3 | 0.00|0.00|3.30 | 1.60 | 0|0 | 0|1 | 2|0 |
| chr15      | C                     | T       | POLG | A154T    | 1|2 |0 | 2.00|17.00|1.00 | 1.00 | 0|1 | 1|0 | 1|0 |
| chr8       | G                     | T       | RECLQ4| P8794    | 0|1 |1 | 0.00|2.08|1.10 | 1.10 | 0|0 | 0|0 | 0|1 |
| chr10      | C                     | A       | AIFM2 | G2V      | 1|1 |1 | 2.00|2.00|1.00 | 1.00 | 0|1 | 0|0 | 0|1 |
| chr3       | C                     | T       | PTPN23| P10995   | 1|0 |1 | 2.00|0.00|1.10 | 1.10 | 0|0 | 0|1 | 1|0 |
| chr2       | G                     | A       | DNM3TA| R693C    | 0|2 |2 | 0.00|0.00|2.20 | 1.00 | 0|2 | 0|0 | 0|1 |
| chr3       | T                     | T       | TMEM115| G7D      | 0|1 |0 | 0.00|0.00|2.00 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr5       | A                     | T       | XRC4 | E121V    | 0|1 |0 | 0.00|0.00|2.00 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr17      | G                     | A       | NFI    | G2103R   | 0|1 |0 | 0.00|0.00|2.00 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr4       | C                     | T       | TET2 | R550X    | 0|1 |0 | 0.00|0.00|2.00 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr17      | C                     | G       | NFI    | P1706R   | 0|1 |0 | 0.00|0.00|2.00 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr16      | C                     | T       | MPG   | R55C     | 0|1 |0 | 0.00|0.00|1.10 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr8       | C                     | T       | MYST3 | G443S    | 0|1 |0 | 0.00|0.00|1.10 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr15      | C                     | T       | IDH2 | R140Q    | 0|1 |0 | 0.00|0.00|2.08 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr1       | T                     | C       | RFWD2 | M299V    | 0|1 |1 | 0.00|0.00|1.10 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr19      | C                     | CG      | BRD4 | E495fsX42| 0|1 |0 | 0.00|0.00|1.10 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr5       | G                     | A       | POLK | S832N    | 0|1 |0 | 0.00|0.00|1.10 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr7       | T                     | C       | MAD1L1| R54G     | 1|0 |0 | 2.00|0.00|0.00 | 0.50 | 0|10 | 0|0 | 0|0 |
| chr22      | C                     | T       | MKL1 | G473R    | 0|1 |0 | 0.00|0.00|2.00 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr1       | G                     | T       | MPL   | W515L    | 0|1 |0 | 0.00|0.00|1.10 | 0.50 | 0|0 | 0|1 | 0|0 |
| chr1       | T                     | G       | POED4D| I303L    | 0|2 |2 | 0.00|0.00|2.00 | 1.06 | 0|0 | 0|1 | 1|0 |
| chr1       | G                     | T       | NTRK1 | G613V    | 7|4 |7 | 14.00|8.33|7.69 | 9.52 | 2|5 | 2|2|5|11 |
| chr16      | T                     | G       | ZFHX3 | T428P    | 1|0 |1 | 2.00|0.00|1.10 | 1.06 | 0|1 | 0|0 | 0|1 |
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 SCRIBH1217P 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 FAT4R175L 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 apoptosis41 (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 NRASG12V and KRASG13D or JAK2V617F 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.fondazionecarpi.it/), Associazione Italiana per la Ricerca sul Cancro (AIRC, Milano, http://www.airc.it/english/
Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)