Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Ortmann CA, Kent DG, Nangalia J, et al. Effect of mutation order on myeloproliferative neoplasms. N Engl J Med 2015;372:601-12. DOI: 10.1056/NEJMoa1412098

Supplementary Appendix

Supplement to: Mutation order affects tumor evolution and clinical course of disease

*Christina A. Ortmann, *David G. Kent, Jyoti Nangalia, Yvonne Silber, David C. Wedge, Jacob Grinfeld, E. Joanna Baxter, Charles E. Massie, Elli Papaemmanuil, Suraj Menon, Anna L. Godfrey, Danai Dimitropoulou, Paola Guglielmelli, Beatriz Bellosillo, Carles Besses, Konstanze Döhner, Claire N. Harrison, George S. Vassiliou, Alessandro Vannucchi, Peter J. Campbell and Anthony R. Green
*denotes equal contribution

Table of Contents

1. **Supplementary Methods** 2

Patients and samples 2

Peripheral blood Isolation 2

Colony assays 3

HSPC Isolation 3

In vitro cultures and secondary colony assays 4

Gene expression analysis 4

Screen for additional mutations in the initial cohort of patients 5

Identification of an extended cohort of JAK-first and TET-first patients. 5

Analysis of clinical correlates 7

2. **Supplementary Figures** 8

Figure S1: Clonal composition of 24 TET2/JAK2V617F-mutant MPN patients. 8

Figure S2: Clonal composition at different time points of 12 TET2/JAK2V617F-mutant MPN patients. 9

Figure S3: Age of presentation and progenitor analysis for TET2-first and JAK2-first patients. 10

Figure S4: HSPC isolation and characterization 12

Figure S5: Functional clustering of gene expression changes 13

Figure S6: Occurrence of thrombosis in patients where order could not be established. 14

3. Supplementary Tables 15

Table S1: Clinical and laboratory characteristics of 24 TET2/JAK2V617F-mutant MPN patients. 15

Table S2: Tabular view of clonal evolution data from Figure S1 and S2. 18

Table S3: Targeted gene sequencing and exome sequencing for mutations in known cancer genes. 20

Appendix 1: Cluster enrichment scores for individual gene expression comparisons as determined by DAVID using Gene Ontology, KEGG, and OMIM databases. 21

4. Supplementary References 21

1. Supplementary Methods

Patients and samples

246 patients with a JAK2V617F mutation were screened for mutations in TET2 (*TET2* primers previously described[1]). Diagnoses were based on the British Committee for Standards in Haematology (BCSH) guidelines.

Peripheral blood Isolation

Fresh venous blood samples (40-60ml, Lithium-Heparin tubes) were collected from patients with TET2 mutations and mononuclear cells (MNCs) were isolated using Lymphoprep (Axis Shield PLC) according to the manufacturer's instructions. In some patients, frozen MNCs were available. In addition to our Cambridge cohort patients, we obtained peripheral blood from 2 additional TET2 and JAK2V617F mutant patients from collaborators. The study was approved by the Cambridge and

Eastern Region Ethics Committee, and was carried out in accordance with the principles of the Declaration of Helsinki. Informed written consent was obtained from all patients before participating.

Colony assays

MNCs were plated at a density of 3-30x10^4/ml in Methocult (H4531, STEMCELL Technologies (STEMCELL)) supplemented with 1unit/ml erythropoietin alpha and incubated at 37°C for 10 -14 days. Individual BFU-E were plucked into 50 µl Buffer RLT (QIAGEN) and a portion was used for genomic DNA extraction by isopropanol precipitation including seeDNA (GE Healthcare). Each colony was sequenced by Sanger sequencing for both JAK2V617F and each individual *TET2* mutation(s). Reliable sequencing results from colony DNA were achieved by design of individual primers for each *TET2* mutation covering short amplicons of 150-200bp (available upon request). Mutational order was only assigned when at least 2 concordant colonies were observed.

For the Ruxolitinib studies, 3x10^6 -1.3x10^7 MNCs were thawed from frozen stores of 4 TET2-first patients and 4 JAK2-first patients. Each sample was divided into two fractions (one with 100nM Ruxolitinib and one without). 5mL of each fraction was divided across 2 wells each and colonies were grown and genotyped as described above.

HSPC Isolation

MNCs were washed with and stored at 4°C overnight in PBS supplemented with 10% fetal bovine serum and 5mM EDTA. In the morning, cells were spun at 300 x g for 7 minutes and the resulting pellets were treated with 50uL of 1mg/mL DNAse (STEMCELL) to avoid clumping. Cells were then depleted of differentiated hematopoietic cells using the EasySep Human Hematopoietic Progenitor enrichment kit (#19056 STEMCELL Technologies) with the following modifications: All samples were processed in 500uL of recommended medium for the first step, the progenitor enrichment cocktail was used at 5uL/mL cells, and the sample was only taken through one round of separation. Cells were then stained with antibodies to isolate the various progenitor compartments: anti-CD34 Per-CpCy5.5 (clone 581 Biolegend), anti-CD38 FITC (clone HIT2, BD Biosciences), anti-CD90 APC (clone 5E10, BD Biosciences), anti-FLK2 PE (clone BV10A4H2, Biolegend), anti-CD10 APC-Cy7 (clone HI10a, Biolegend), and anti-CD45RA Horizon V450 (HI100, BD Biosciences). HSPCs were isolated as $lin^-CD34^+CD38^-CD45RA^-CD90^+$, CMPs as $lin^-CD34^+CD38^+CD90^-CD10^-FLK2^+CD45RA^-$, GMPs as $lin^-CD34^+CD38^+CD90^-CD10^-$

FLK2^{+}CD45RA^{+} and MEPs as lin^{-}CD34^{+}CD38^{+}CD90^{-}CD10^{-}FLK2^{-}CD45RA^{-} (Supplementary Figure 1B). Cells were sorted using a BD Influx sorter equipped with 355nm, 405nm, 488nm, 561nm, and 640nm lasers into 1.5mL tubes containing 2% PBS or sorted directly into individual wells of a 96-well U-bottomed plate. The CMP, MEP and GMP fractions were placed directly into a CFC assay using Methocult 04435 (STEMCELL) as described by the manufacturer (Supplementary Figure 1C).

In vitro cultures and secondary colony assays

HSPCs were sorted as single cells into STEMSPAN medium (STEMCELL) supplemented with the cc100 cytokine cocktail (STEMCELL), which was used to maintain the stem and progenitor cell fraction in culture before harvesting and placing into a secondary CFC assay using Methocult 04435 as above. For proliferation studies, HSPC-derived clones were counted on day 10 (<50 cells were counted directly and larger clones were estimated based on diameter and binned into small (50-500) and medium (500+) categories). For secondary colony counts, each 10-day single HSPC-derived culture was placed into one CFC assay and all colonies were scored. Colony genotyping was either performed on 10-day clones or on pools of secondary colonies.

Gene expression analysis

Transcriptional profiling was performed on individual erythroid colonies from 7 patients (4 TET2-first, 3 JAK2-first) from whom sufficient numbers of WT, single mutant and double mutant colonies could be collected from the same sample (ArrayExpress accession E-MTAB-3086). For each patient, BFU-E colonies in RLT buffer with the same genotype were pooled and total RNA was extracted (Qiagen Rneasy kit) and assessed by nanoelectrophoresis (Bioanalyzer, Agilent Technologies) as previously described[2]. If the RNA yield of one or more genotypes of a patient was below 100ng, amplification with the MessageAmp II aRNA amplification kit (Ambion) was performed for all RNA samples of this patient assay. Biotinylated RNA was prepared with the Illumina TotalPrep RNA amplification kit (Ambion) and used for hybridization onto Illumina HU12 v4 Expression BeadChips. Quantile normalization was performed using *Lumi* and log2 transformed data were subjected to differential gene expression analysis using the Bioconductor package *limma*[3]. Patient identity and sample amplification were accounted for in the design matrix. Correlation between technical replicates was also controlled for. Genes were first stratified based

on p-value (0.1) and then a fold change cutoff of >1.25 was applied across the patient cohorts for genes to be considered up- or down-regulated. Enrichment within these lists of genes from different functional categories was determined using the online DAVID functional clustering tool using Gene Ontology, KEGG, and OMIM databases and the top enrichment scores are presented in Figure 4 for each comparison (see Appendix A for complete analysis and individual category p values and false discovery rates).

Screen for additional mutations in the initial cohort of patients

Previously published exome sequencing data[4] for 9 patients was screened for recurrent MPN driver mutations (Supplementary Table 2). 13 of 14 patients in whom the order of acquisition could be determined were further screened using a targeted sequencing approach for 111 genes or genetic regions implicated in myeloid malignancies; as previously described[1] which included a broad selection of recurrent MPN-associated mutations. CALR mutational status was determined to be negative in all patients.

Identification of an extended cohort of JAK-first and TET-first patients.

In order to identify an extended cohort of JAK-first and TET-first patients, we undertook a targeted sequencing study using genomic DNA from whole blood or peripheral blood granulocytes from >1000 MPN samples collected from UK centres and Italy. Written informed consent was obtained from all patients with local ethics approval. Samples underwent whole genome amplification and targeted sequencing was performed using an Agilent SureSelect custom RNA bait set for 111 genes and genetic regions implicated in myeloid malignancies as described previously[4], as well as a further a panel of 65 genes selected based on results of previously published MPN exome sequencing data[3]. A total of 918 samples were successfully sequenced. Mutation calling for TET2 and JAK2V617F utilised previously described algorithms.[1]

90 patients were identified with JAK2V617F and mutated TET2. To determine the order of acquisition of JAK2 and TET2 mutations in double-mutant patients, we first calculated the fraction of cells bearing each mutation. This calculation is affected by the presence of copy number changes in the region of either mutation. Therefore, allele frequencies of reference and alternative alleles of all SNPs within the 1000 genomes dataset[5] lying on chromosomes 4 and 9 were obtained. Visual inspection of b-allele frequency (BAF) and normalised coverage (logR) plots identified samples subject to loss-of-heterozygosity (LOH) in the region of either TET2 or JAK2.

For samples with LOH, informative heterozygous SNPs were identified as those loci with BAF between 0.01 and 0.99 (inclusive) and a minimum of 5 reads. For each informative heterozygous SNP, bootstrap re-sampling of reference and alternative reads was performed (n=10, 000) and the fraction of cells bearing a copy number change, *f*, was calculated under 2 different assumptions: loss of DNA leading to a copy number state 1+0; or copy neutral LOH, resulting in a copy number state of 2+0. The respective equations used are:

$$f = 2 - 1 / b$$

$$f = 2 * b - 1$$

where *b* is the b-allele frequency.

Using the same bootstrap re-sampling technique, the number of reads reporting either a TET2 or JAK2 mutation was re-sampled from the observed data. The fraction of cells bearing a mutation, *M*, was calculated for each sample assuming either copy number loss or copy neutral LOH, respectively, using the following equations:

$$M = m * [f + 2*(1-f)]$$

$$M = 2 * m$$

where *M* is the fraction of cells bearing a mutation and *m* is the fraction of reads reporting a mutation.

If a mutation has occurred prior to copy neutral LOH, it will be present on 2 chromosomal copies. In this case, the following equation is used:

$$M = 2 * \mathrm{m} - f$$

When comparing the fraction of cells bearing each mutation, copy number changes in JAK2 were assumed to be copy neutral LOH. Copy number changes in TET2 were assumed to be copy number losses unless high allele frequencies indicated the presence of copy neutral LOH. Differences in fraction of cells bearing each mutation were assessed at a significance level of 0.01, i.e. they were significant only if 99% of the bootstrap samples reported the same ordering. A

comparison of the fraction of cells bearing TET2 and JAK2 mutations in each sample has 4 possible outcomes:

- There is no significant difference in the fraction of cells bearing each mutation
- The fraction of cells bearing each mutation is significantly different, but the sum of the 2 values is less than or equal to 1. In this case, the 2 mutations may be present in separate subclonal populations and temporal ordering cannot be determined with confidence.
- JAK2 is present in a significantly larger proportion of cells and the sum of the proportions is greater than 1. JAK2 has occurred first in these samples.
- TET2 is present in a significantly larger proportion of cells and the sum of the proportions is greater than 1. TET2 has occurred first in these samples.

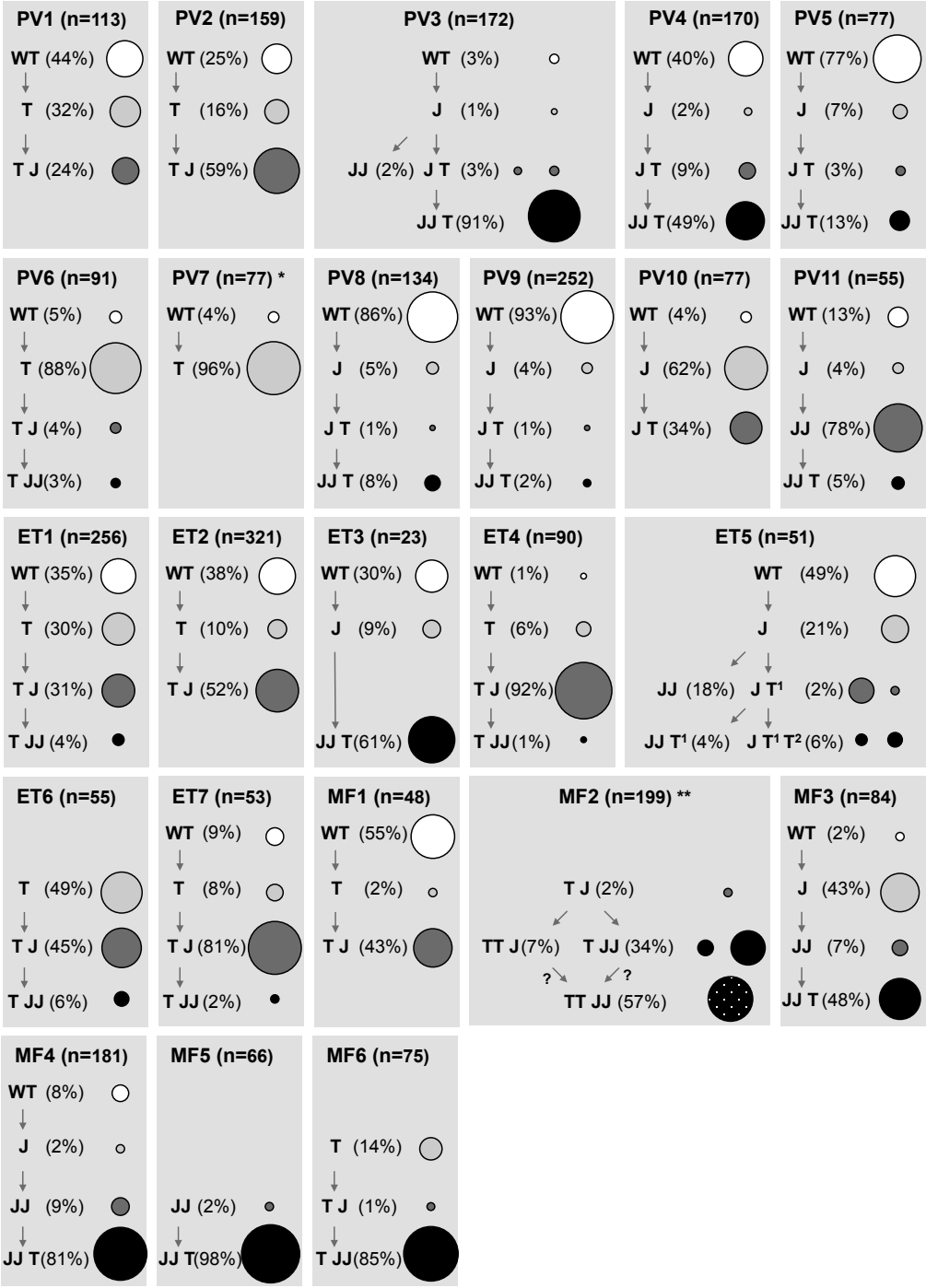
In the first 2 cases, order is undetermined. In this manner, an additional 18 JAK2-first and 6 TET2-first patients were identified for the follow-up cohort.

Analysis of clinical correlates

Patient characteristics at diagnosis, such as age, gender, prior thrombotic history, diagnostic hematological parameters and MPN subtype, as well as adverse outcome events, such as thrombosis, death or disease progression, were collected for all patients. These parameters were analysed for associations with order of acquisition. Baseline patient and laboratory features were compared using Students t-test and chi-squared (Fisher's exact test) analysis for continuous and categorical variables respectively. Kaplan-Meier and log-rank analyses were used to compare rates of thrombosis-free survival, defined as the time between diagnosis and the development of the patient's first arterial (cerebrovascular or coronary) or venous (portal vein, splanchnic vein, calf vein) events, or to last follow-up (for censored patients). To determine whether order of acquisition was an independent risk factor for the development of thrombosis, progression or death, multivariate analysis was undertaken using cox proportional hazards modelling and encorporated the following variables: acquisition order, gender, age, white blood cell count, MPN subtype, cytoreductive therapy and history of prior thrombosis.

2. Supplementary Figures

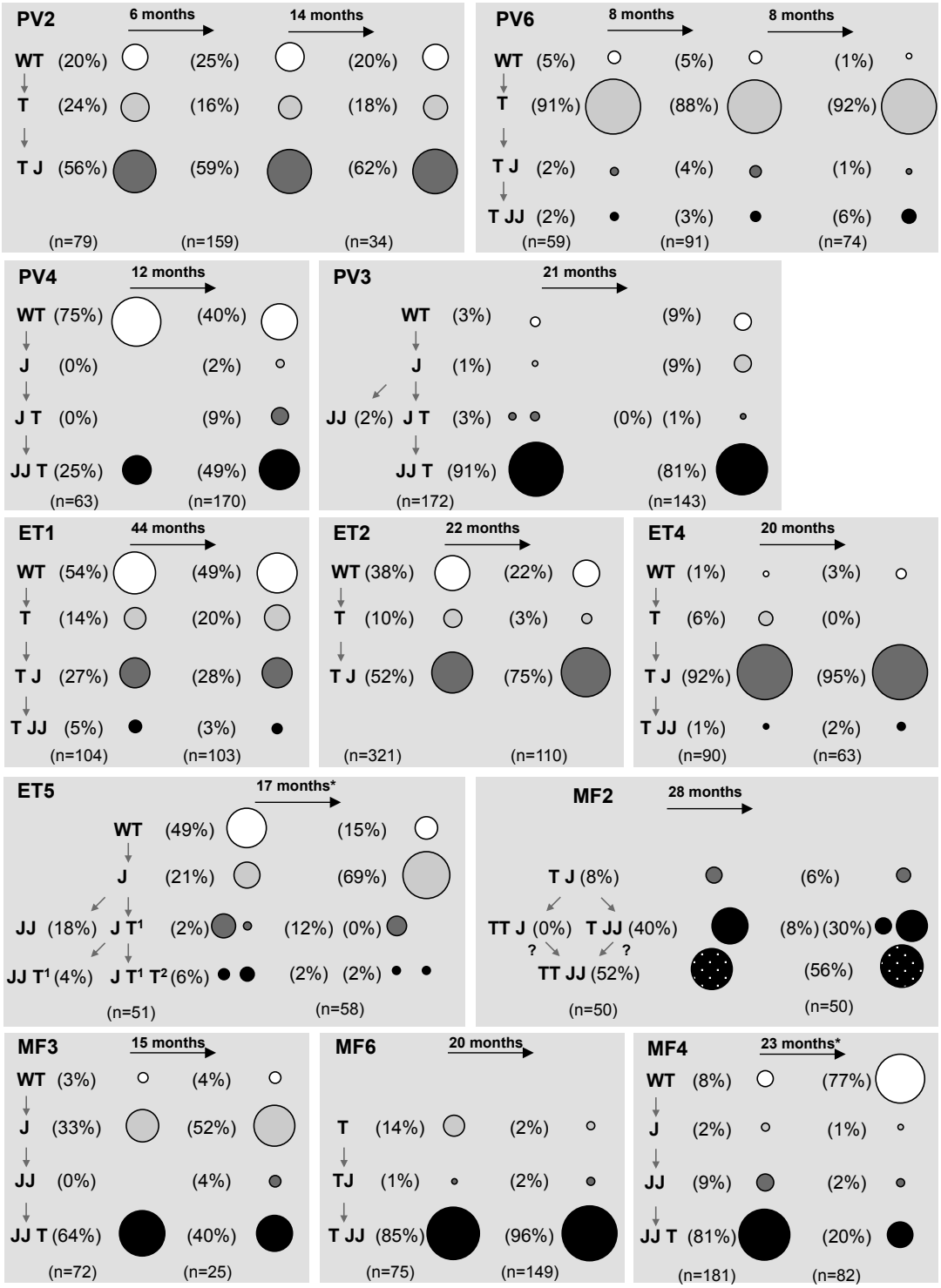
Figure S1: Clonal composition of 24 TET2/JAK2V617F-mutant MPN patients.



Each panel represents the clonal pattern of BFU-E grown from peripheral blood mononuclear cells and representative assays are shown. The number of colonies picked and genotyped is shown (n).

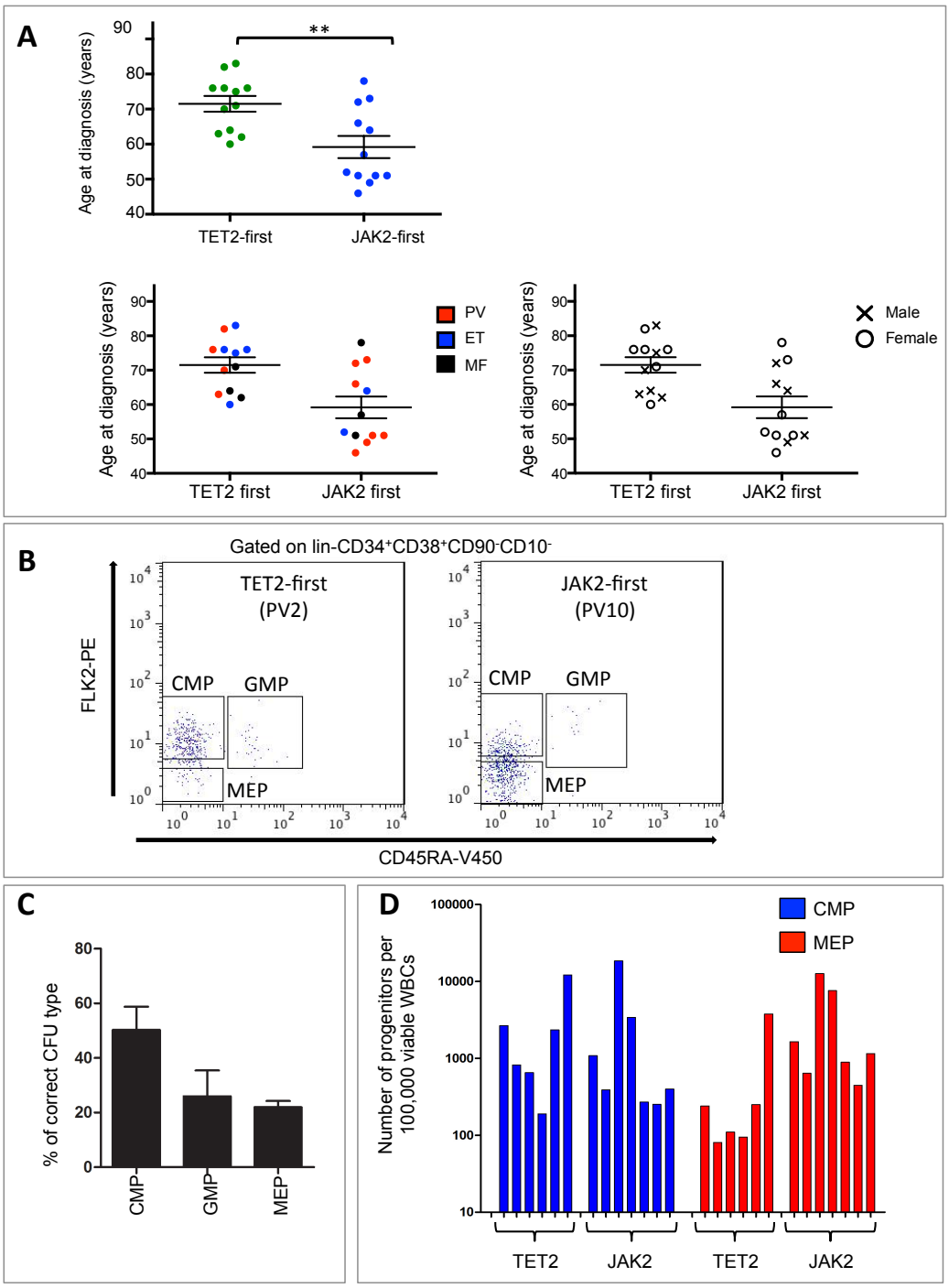
*In PV7, the additional JAK2V617F mutation was only detected in erythropoietin-independent erythroid colonies (EEC) and was reported previously[6]. **In MF2, the mutational order was established from analysis of CFU-GM colonies.

Figure S2: Clonal composition at different time points of 12 TET2/JAK2V617F-mutant MPN patients.



Each patient presented in this figure was re-analyzed to determine if the clonal composition of the BFU-E compartment changed over time. The amount of time between assay is indicated for each patient and the number of colonies analysed is presented at the bottom of each time point. *Treatment with cytoreductive agents was changed in 2 patients in between assays: ET5: anagrelide had been discontinued at follow up; MF4: ruxolitinib treatment was introduced 11 months before follow-up sample. Blood counts were stable between assays.

Figure S3: Age of presentation and progenitor analysis for TET2-first and JAK2-first patients.



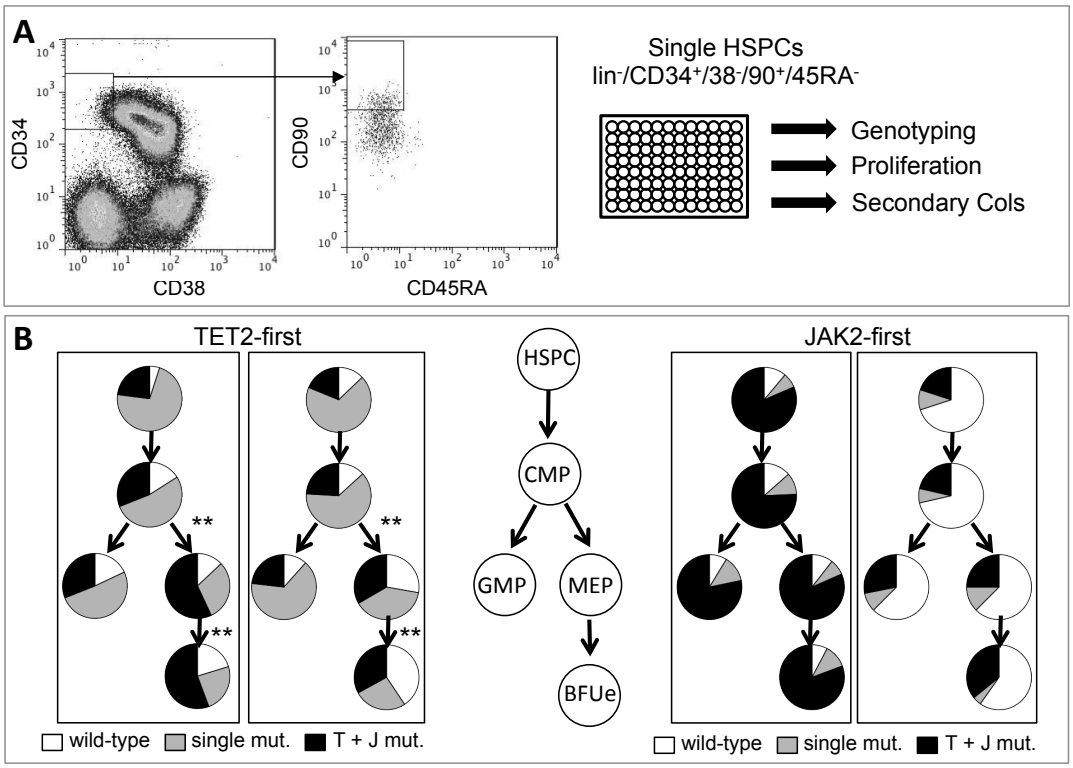
A) The mean age of presentation is shown for 24 MPN patients in whom mutational order of TET2 and JAK2 was determined by colony genotyping (upper panel). On average, TET2-first patients presented 12.3 years later than JAK2-first patients (p=0.0043 by t-test). The lower panels show a two-way ANOVA analyses with order and MPN phenotype (p=0.0195 for effect of order, left panel) or order and sex (p=0.0068 for effect of order, right panel) with no significant effect of disease phenotype or age.

B) Isolation strategy for CMPs, MEPs, and GMPs – cells were first gated as $CD34^{+}CD38^{+}CD10^{-}CD90^{-}$ and then sub-fractionated according to the method described in Doulatov et al.[7] with CMPs as $FLK2^{+}CD45RA^{-}$, GMPs as $FLK2^{+}CD45RA^{+}$, and MEPs as $FLK2^{-}CD45RA^{-}$. Plots are shown for one TET2-first patient and one JAK2-first patient to exemplify the differences observed in Figure 2C.

C) Progenitors were isolated and seeded in a methylcellulose-based colony assay. Colonies were morphologically scored as mixed, erythroid, or granulocytic/monocytic and the percentage of the correct colony type expected from the progenitor is depicted.

D) Absolute numbers of progenitor fractions for Figure 2C are displayed in this bar graph. CMPs are in blue, MEPs in red and GMPs in green.

Figure S4: HSPC isolation and characterization



A) The FACS isolation strategy for HSPCs is shown in the top panel, first selecting lin$^-$ CD34$^+$CD38$^-$ cells and then gating on CD90$^+$CD45RA$^-$ cells. Single HSPCs were deposited into individual wells of a 96-well plate and cultured for 10 days to determine proliferation capacity and clones were placed into secondary CFC assays to assess progenitor expansion and to be genotyped.

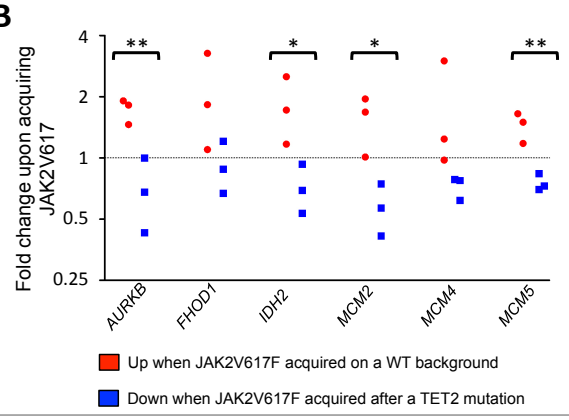
B) Individual stem and progenitor cell fractions of four patients were purified, grown for 10-14 days as colonies in methylcellulose and genotyped to determine the frequency of each genotype (wild-type, single mutant, double mutant) in the HSPC, common myeloid progenitor (CMP), granulocyte/macrophage progenitor (GMP), megakaryocyte/erythrocyte progenitor (MEP) and BFU-E compartments. Each pie shows the ratio of wild-type (white), single mutant (grey), or double mutant (black, T+J) cells. While the JAK2-first patients show similar patterns in all compartments, the TET2-first patients diverge at the MEP level, showing an over-representation of the double mutant clone which bears the JAK2V617F mutation. Statistical differences are determined by comparing the single mutant to double mutant ratio in each compartment to the HSPC fraction using a Fisher's Exact test (** indicates $p<0.01$)

Figure S5: Functional clustering of gene expression changes

A

	Up-regulated	Down-regulated
Single mut. vs. WT		
JAK2-first (JAK2 effect on WT)	**Set 1** **747 genes** translation/ribosomes (17.15) mitochondrial proteins (6.08) ribosome biogenesis (6.11)	**Set 2** **815 genes** cell cycle regulation (4.65) metabolism / proteolysis (4.42) postranscriptional regulation (3.44)
TET2-first (TET2 effect on WT)	**Set 3** **27 genes** no specific enrichment	**Set 4** **44 genes** ribonuclease activity (2.36) host defense (1.74)
Double vs. single mut.		
JAK2-first (TET2 effect on JAK2)	**Set 5** **96 genes** nucleotide binding (1.50) development/patterning (1.49) protein kinase signaling (1.42)	**Set 6** **97 genes** oxidative phosphorylation (2.13) ribosome biogenesis (1.60)
TET2-first (JAK2 effect on TET2)	**Set 7** **241 genes** inflammatory response (2.94) ROS metabolism (2.62) cellular response to stress (2.50)	**Set 8** **298 genes** cell cycle progression (20.52) DNA Replication (11.11) chromosome regulation (6.23)

B

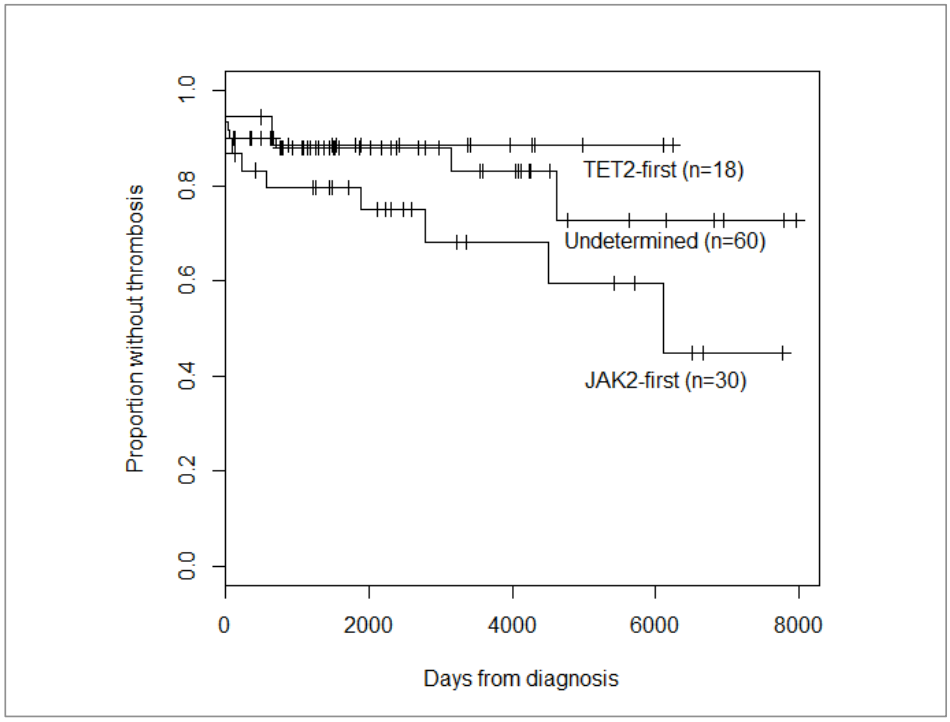


A) Gene expression arrays were performed on pooled BFU-E colonies of the same genotype within the same patient. Differentially expressed genes were analyzed using the online DAVID functional clustering tool using Gene Ontology, KEGG, and OMIM databases and the top enrichment scores are presented for each comparison (see Appendix 1 for complete analysis and individual category p-values and false discovery rates). The most enriched clusters across all comparisons were up-regulation of translational machinery when JAK2V617F was acquired on a wild-type background (Set 1) and downregulation of cell cycle progression genes when JAK2V617F was acquired on a TET2-mutant background (Set 8).

B) Six genes from Figure 3F were selected for validation by quantitative RT-PCR of material isolated from pooled colonies. The red dots represent individual JAK2-first patients and the fold difference that each gene is up-regulated upon JAK2 acquisition on a wild-type background (wild-

type set to 1). The blue dots represent individual TET2-first patients and the fold difference that each gene is down-regulated upon JAK2V617F acquisition on a TET2-single mutant background (single mutant set to 1). *represents p<0.05, ** represents p<0.01 as determined by a one-tailed Student's t-test. WT= wild type

Figure S6: Occurrence of thrombosis in patients where order could not be established.



In order to ensure that our order-calling algorithm (which excluded patients where biclonal disease could not ruled out) did not bias our dataset, we also analysed thrombotic events in those patients where order could not be determined. Reassuringly, these data show an intermediate risk for thrombotic-free survival, suggesting that the algorithm was not significantly biased.

3. Supplementary Tables

Table S1: Clinical and laboratory characteristics of 24 TET2/JAK2V617F-mutant MPN patients.

Mutational order	Patient identifier	Diagnosis	Gender	Age at diagnosis	Disease duration at assay (months)	TET2 mutations	Karyotype	Therapy at colony assay (previous therapies before colony assay)	Blood count at presentation				
									Hb	WBC	Plt	Mono	Neutr
TET2-first	**ET1**	ET	M	75	84	c.3120_3121het_insA	na	hydroxyurea	12.9	11.7	681	0.7	1.9
	ET2	ET	F	76	93	c.3890G>GA; p.G1297GE	na	hydroxyurea	14.7	9.6	875	0.6	7.5
	ET4	ET	M	83	13	c.4076G>GT; p.R1359RL	na	hydroxyurea	15.8	12.9	856	1.1	10.0
	ET6	ET	F	76	109	c.1340het_delG	normal	hydroxyurea	14.8	10.6	1054	0.4	7.4
	ET7	ET	F	60	61	c.2656C>CT; p.Q886QX	na	hydroxyurea	12.9	7.1	695	0.4	4.9
	MF1	post-ET MF	M	62 *(74*)*	161 *(10*)*	c.650het_delC	na	hydroxyurea, (anagrelide)	*13.6 (9.7#)*	8.4 (15.3 *#)*	461 (337 *#)*	1.04 (1.57 *#)*	9.4 (11.3 *#)*
	MF2	pMF	F	71	22	c.5447T>A; p.L1816X	normal	nil	13.0	7.9	92	unk	6.4
	MF6	post-PV MF	M	64 *(77*)*	140 *(5*)*	c.3866het_delG	normal	hydroxyurea, anagrelide	18.0 (9.7 #)	10.7 (3.7 #)	426 (457 #)	0.5 (0.1 #)	7.8 (2.5 #)
	PV1	PV	M	70	31	c.2887het_delC	normal	hydroxyurea	18.2	10.4	977	0.7	6.7
	PV2	PV	M	63	6	c.817C>CT; p.Q273QX	na	venesection	18.5	9.6	336	1.0	7.3
	PV6	PV	F	76	54	c.5035_5036het_insT	na	hydroxyurea	18.8	8.8	534	0.9	5.8
	PV7	PV	F	82	108	c.1969_1972het_delTCAC	normal	pipobroman, (hydroxyurea)	17.9	10.3	505	0.7	6.5
JAK2-first	**ET3**	ET	F	52	190	c.2248_2252het_delATAAA**	na	hydroxyurea, (busulfan, anagrelide)	11.2	3.5	900	unk	unk
	ET5	ET	M	64	130	c.2850delTGCTC and c.1081C>CT; Q361QX	na	hydroxyurea, anagrelide; anagrelide discontinued at follow-up	15.0	9.0	1400	unk	6.8
	MF3	pMF	F	78	16	c.1954C>CT; p.Q652QX	na	nil	10.4	8.7	103	0.2	6.8
	MF4	pMF	F	57	17	c.3356T>TG; p.L1119LX	na	hydroxyurea; ruxolitinib for 11 months at follow-up	10.2	20.0	257	0.9	16.9
	MF5	post-PV MF	F	51 *(60*)*	155 *(38*)*	c.3728A>AG; p.K1243KR	normal	hydroxyurea	20.1 (10.3 #)	17.1 (17.0 #)	345 (239 #)	0.3 (0.4 #)	13.7 (15.2 #)
	PV3	PV	F	46	20	c.4896het_delA	na	venesection	17.5	8.0	346	0.3	5.9
	PV4	PV	M	49	67	c.5326het_delT	normal	hydroxyurea	20.1	12.1	292	unk	9.6
	PV5	PV	M	66	159	c.4926C>CA; p.C1642CX	normal	hydroxyurea	19.2	10.4	451	0.2	7.7
	PV8	PV	M	51	113	c.2419G>GT; p.E807EX**	normal	imatinib, venesection	19.7	9.2	250	unk	unk
	PV9	PV	F	51	200	c.686het_delC**		venesection	18.4	7.2	155	unk	unk
	PV10	PV	M	72	10	c.3356T>TG; p.L1119LX	na	hydroxyurea	15.9	13.8	704	0.9	10.8
	PV11	PV	F	73	108	c.4894C>CT; p.Q1632QX	na	nil, (hydroxyurea)	15.1	8.9	634	0.5	7.0

Table S1: Disease duration is stated as the time from diagnosis to first colony assay, blood counts are at the time of presentation with chronic phase disease; *, time from diagnosis to, or age at, myelofibrotic transformation; **, genotype determined by absence of mutation in a proportion of colonies; # blood counts at transformation to MF; Hb, hemoglobin (g/dl); WBC, white

blood cell count (10^9/l); Plt, platelets (10^9/l); Mono, monocytes (10^9/l); Neutr, neutrophils (10^9/l); na, not assessed; unk unknown.

Table S2: Tabular view of clonal evolution data from Figure S1 and S2.

Order	ID	time (months)	Clones	Col #
TET2-first	ET1	0	WT (54%) => T (14%) => TJ (27%) =>TJJ (5%)	104
		3	WT (35%) => T (30%) => TJ (31%) =>TJJ (4%)	256
		14	WT (50%) => T (22%) => TJ (27%) =>TJJ (1%)	198
		44	WT (49%) => T (20%) => TJ (28%) =>TJJ (3%)	103
	ET2	0	WT (38%) => T (10%) => TJ (52%)	321
		3	WT (22%) => T (8%) => TJ (70%)	245
		22	WT (22%)=> T (3%) => TJ (75%)	110
	ET4	0	WT (1%) => T (6%) => TJ (92%) =>TJJ (1%)	90
		20	WT (3%) => T (0%) => TJ (95%) =>TJJ (2%)	63
	ET6		T (49%) => TJ (45%) =>TJJ (6%)	55
	ET7		WT (9%) => T (8%) => TJ (81%) =>TJJ (2%)	53
	MF1		WT (55%) => T (2%) => TJ (43%)	48
	MF2	0	TJ (8%)=>TTJ (0%) and TJJ (40%) => TTJJ (52%)	50
		5	TJ (2%) =>TTJ (7%) and TJJ (34%) => TTJJ (57%)	199
		28	TJ (6%) =>TTJ (8%) and TJJ (30%) => TTJJ (56%)	50
	MF6	0	T (14%) => TJ (1%) =>TJJ (85%)	75
		20	T (2%) => TJ (2%) =>TJJ (96%)	149
	PV1		WT (44%) => T (32%) => TJ (24%)	113
	PV2	0	WT (20%) => T (24%) => TJ (56%)	79
		6	WT (25%) => T (16%) => TJ (59%)	159
		20	WT (20%) => T (18%) => TJ (62%)	34
	PV6	0	WT (5%) => T (91%) => TJ (2%) =>TJJ (2%)	59
		8	WT (5%) => T (88%) => TJ (4%) =>TJJ (3%)	91
		16	WT (1%) => T (92%) => TJ (1%) =>TJJ (6%)	74
	PV7		WT (4%) => T (96%) => TJ (0%) (TJ in EEC only)	77
JAK2-first	ET3		WT (30%) => J (9%) => JJT (61%)	23
	ET5	0	WT (49%) => J (21%) => JJ (18%) and JT1 (2%) => JJT1 (4%) and JT1T2 (6%)	51
		2	WT (40%) => J (27%) => JJ (29%) and JT1 (0%) => JJT1 (2%) and JT1T2 (2%)	55
		17*	WT (15%) => J (69%) => JJ (12%) and JT1 (0%) => JJT1 (2%) and JT1T2 (2%)	58
	MF3	0	WT (3%) => J (33%) => JJ (0%)=> JJT (64%)	72
		10	WT (2%) => J (43%) => JJ (7%) => JJT (48%)	84
		15	WT (4%) => J (52%) => JJ (4%) => JJT (40%)	25
	MF4	0	WT (8%) => J (2%) => JJ (9%) => JJT (81%)	181
		3	WT (3%) => J (1%) => JJ (20%) => JJT (76%)	110
		23 *	WT (77%) => J (1%) => JJ (2%) => JJT (20%)	82
	MF5		JJ (2%) => JJT (98%)	66
	PV3	0	WT (3%) => J (1%) => JJ (2%) and JT (3%) => JJT (91%)	172
		21	WT (9%) => J (9%) => JJ (0%) and JT (1%) => JJT (81%)	143
	PV4	0	WT (75%) => J (0%) => JT (0%) => JJT (25%)	63
		12	WT (40%) => J (2%) => JT (9%) => JJT (49%)	170
	PV5		WT (77%) => J (7%) => JT (3%) => JJT (13%)	77
	PV8		WT (86%) => J (5%) => JT (1%) => JJT (8%)	134
	PV9		WT (93%) => J (4%) => JT (1%) => JJT (2%)	252
	PV10		WT (4%) => J (62%) => JT (34%)	77
	PV11		WT (13%) => J (4%) => JJ (78%) => JJT (5%)	55

Each patient is listed alongside the clonal hierarchy from the first time analysed and then sequentially (by number of months following first analysis). The percentage of each colony genotype is listed in their phylogenetic order for each patient. WT, no mutation in JAK2 or TET2; T, TET2 mutation; J, JAK2V617F mutation; JJ, JAK2V617F homozygosity; TT, compound heterozygote for TET2.

Table S3: Targeted gene sequencing and exome sequencing for mutations in known cancer genes.

Mut. Order	Patient ID	Mutations present in known cancer gene	Somatic	Oncogenic	Seq. Method
TET2-first	ET1 (PD6339)				Targeted
	ET2 (PD19239)				Targeted
	ET4 (PD19242)				Targeted
	ET6 (PD6638)	ASXL1 c.2422_2423delCC, p.P808fs*13 (24%) TET2 c.4393C>T, p.R1465* (24%) FLT3 c.2647T>G, p.S883A (35%)	unconfirmed unconfirmed unconfirmed	YES YES possible	Targeted
	ET7 (PD6548)				Exome
	MF1(PD10053)	DNMT3a p.R882H (c.2645G>A) 56.9% SF3B1 p.K700E (c.2098A>G) 49.4%	YES YES	YES YES	Exome
	MF2 (PD5008)	TET2 p.Y1245fs*22 (c.3730_3731delCT) 3.7%	YES	YES	Exome
	MF6 (PD4775)				Exome
	PV1 (PD5124)				Targeted
	PV2 (PD4940)				Exome
	PV6 (PD5158)				Exome
	PV7 (PD5147)				Targeted
JAK2-first	ET3 (PD11406)				Targeted
	ET5 (PD5637)				Targeted
	MF3 (PD4991)	TET2 c.5730delA; p.K1911fs*39 (16%)	YES	likely	Targeted
	MF4 (PD4178)				Exome
	MF5 (PD4060)	SH2B3 p.L458P (c.1373T>C) 49.5%	YES	possible	Exome
	PV3 (PD5177)				Targeted
	PV4 (PD19240)				Targeted
	PV5 (PD4779)				Targeted
	PV8 (PD19241)				Targeted
	PV9 (ND)				
	PV10 (PD19243)				Targeted
	PV11 (PD4781)				Exome

Each patient is listed alongside any additional mutations that were found in either the exome sequencing studies or the targeted gene sequencing screen, the latter of which screened for the presence of mutations in over 100 commonly recurrent mutations in hematological malignancies as previously described[1] as well as CALR status[4]. Where a germline DNA sample was unavailable for comparison, mutations were filtered for previously reported SNPs as described previously[4]. Mutations in black are predicted to be oncogenic. Mutations in grey are of unknown significance as these mutations that have not been reported previously in the 'Catalogue of Somatic Mutations in Cancer (COSMIC)'[8]. The known JAK2 and TET2 mutations were detected following targeted capture, sequencing and application of analysis algorithms in 85% and 92% of samples respectively. In two patients (PV1 and ET6) JAK2V617F was not identified by the analysis algorithms due to low coverage in the region. Only ET6, MF1, MF2, and MF3 showed additional mutations in known oncogenes that were confirmed as somatically acquired (DNMT3A, SF3B1, ASXL1, and a second TET2 mutation). Notably, the additional TET2 mutation in MF3 (a JAK2-first patient) was subclonal to the highly dominant JAK2V617F patient and did not change the “JAK2-first” categorization of the patient.

Appendix 1: Cluster enrichment scores for individual gene expression comparisons as determined by DAVID using Gene Ontology, KEGG, and OMIM databases.

Differentially expressed genes were analyzed using the online DAVID functional clustering tool using Gene Ontology, KEGG, and OMIM databases and the top enrichment scores are presented for each comparison (see Appendix 1 for complete analysis and individual category p-values and false discovery rates). The most enriched clusters across all comparisons were up-regulation of translational machinery when JAK2V617F was acquired on a wild-type background (Set 1) and downregulation of cell cycle progression genes when JAK2V617F was acquired on a TET2-mutant background (Set 8). An enriched cluster was defined as having an enrichment score greater than 1.

4. Supplementary References

1. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 2013;122:3616-27; quiz 99.

2. Chen E, Beer PA, Godfrey AL, et al. Distinct clinical phenotypes associated with JAK2V617F reflect differential STAT1 signaling. Cancer cell 2010;18:524-35.

3. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology 2004;3:Article3.

4. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. The New England journal of medicine 2013;369:2391-405.

5. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.

6. Scott LM, Beer PA, Bench AJ, Erber WN, Green AR. Prevalance of JAK2 V617F and exon 12 mutations in polycythaemia vera. British journal of haematology 2007;139:511-2.

7. Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, Dick JE. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nature immunology 2010;11:585-93.

8. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic acids research 2011;39:D945-50.