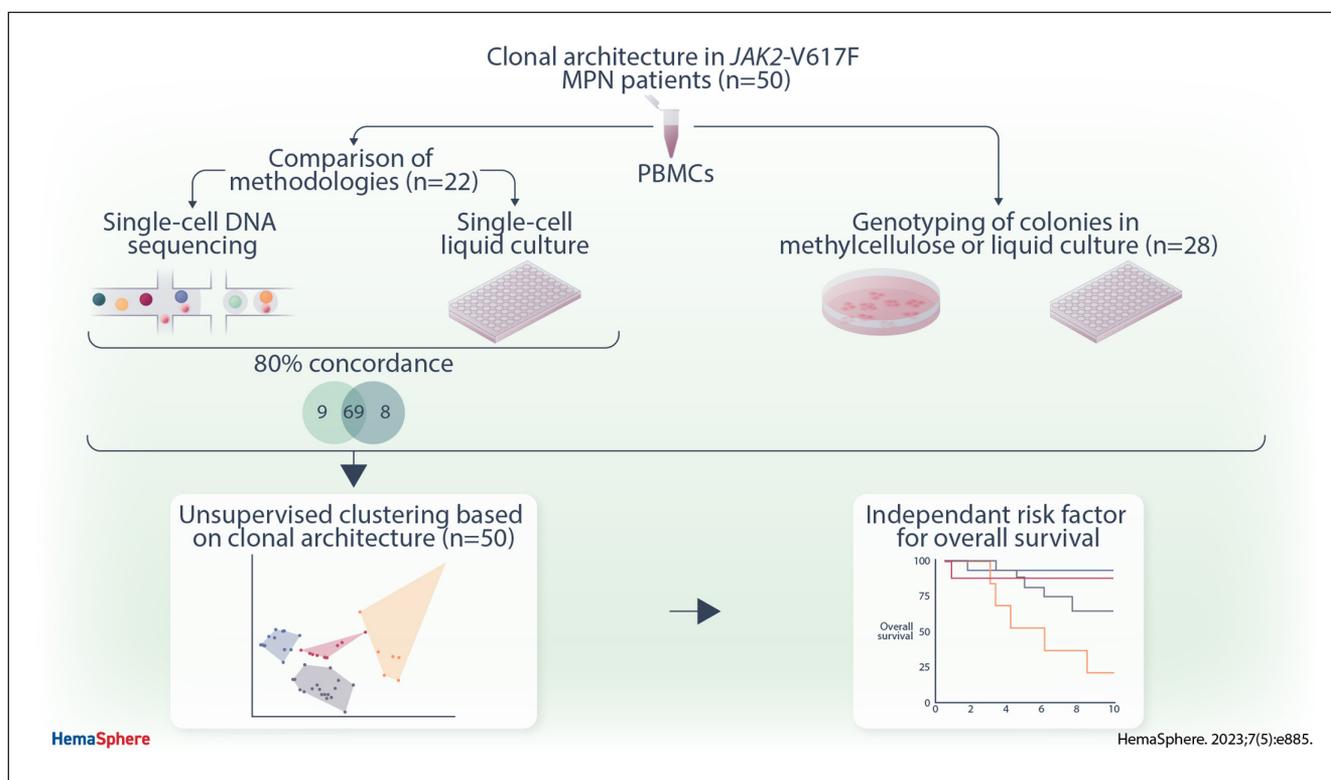


Article
Open Access

Impact of Clonal Architecture on Clinical Course and Prognosis in Patients With Myeloproliferative Neoplasms

Damien Luque Paz^{1,2}, Michael S. Bader³, Ronny Nienhold^{1,*}, Shivam Rai¹, Tiago Almeida Fonseca¹, Jan Stetka¹, Hui Hao-Shen¹, Gabriele Mild-Schneider¹, Jakob R. Passweg³, Radek C. Skoda¹

GRAPHICAL ABSTRACT



Article

Open Access

Impact of Clonal Architecture on Clinical Course and Prognosis in Patients With Myeloproliferative Neoplasms

Damien Luque Paz^{1,2}, Michael S. Bader³, Ronny Nienhold^{1,*}, Shivam Rai¹, Tiago Almeida Fonseca¹, Jan Stetka¹, Hui Hao-Shen¹, Gabriele Mild-Schneider¹, Jakob R. Passweg³, Radek C. Skoda¹

Correspondence: Radek C. Skoda (radek.skoda@unibas.ch).

ABSTRACT

Myeloproliferative neoplasms (MPNs) are caused by a somatic gain-of-function mutation in 1 of the 3 disease driver genes *JAK2*, *MPL*, or *CALR*. About half of the MPNs patients also carry additional somatic mutations that modify the clinical course. The order of acquisition of these gene mutations has been proposed to influence the phenotype and evolution of the disease. We studied 50 *JAK2*-V617F-positive MPN patients who carried at least 1 additional somatic mutation and determined the clonal architecture of their hematopoiesis by sequencing DNA from single-cell-derived colonies. In 22 of these patients, the same blood samples were also studied for comparison by Tapestry single-cell DNA sequencing (scDNAseq). The clonal architectures derived by the 2 methods showed good overall concordance. scDNAseq showed higher sensitivity for mutations with low variant allele fraction, but had more difficulties distinguishing between heterozygous and homozygous mutations. By unsupervised analysis of clonal architecture data from all 50 MPN patients, we defined 4 distinct clusters. Cluster 4, characterized by more complex subclonal structure correlated with reduced overall survival, independent of the MPN subtype, presence of high molecular risk mutations, or the age at diagnosis. Cluster 1 was characterized by additional mutations residing in clones separated from the *JAK2*-V617F clone. The correlation with overall survival improved when mutation in such separated clones were not counted. Our results show that scDNAseq can reliably decipher the clonal architecture and can be used to refine the molecular prognostic stratification that until now was primarily based on the clinical and laboratory parameters.

INTRODUCTION

BCR::ABL1-negative myeloproliferative neoplasms (MPNs) are clonal disorders of hematopoietic stem cells driven by a somatic gain-of-function mutation in either *JAK2*, *MPL*, or *CALR*.¹⁻⁹ MPN patients at diagnosis are assigned to 1 of the 3 subtypes: polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF).^{10,11} The prognosis of MPNs is mostly dependent on the presence or absence of complications such as thrombotic or hemorrhagic events, and on the progression to secondary myelofibrosis or acute myeloid

leukemia.¹² Next-generation sequencing (NGS) allowed efficiently detecting additional gene mutations, which are often associated with less favorable prognosis.^{13,14} Additional somatic mutations can be acquired before or after the driver gene mutation or can represent a separate clone,¹⁵ and the order of acquisition has been proposed to influence the phenotype and the evolution of the disease.¹⁶

Several methods have been developed to determine the subclonal structure in patients with >1 somatic mutation. Genotyping of colonies derived from single cells, grown either in semi-solid media (eg, methylcellulose) or in liquid culture, is considered the gold standard for determining clonal architecture, but this approach is labor-intensive and difficult to use for high-throughput applications for larger cohorts of patients. Methods for genotyping of single cells without expansion in culture have first been developed at the RNA level and more recently also on DNA level. The Tapestry single-cell DNA sequencing (scDNAseq) method was previously used to determine the clonal architecture in patients with acute myeloid leukemia (AML)^{17,18} and also in 8 MPN patients treated by MDM2 inhibitors,¹⁹ but a validation by direct comparison with genotyping of single-cell colonies has not yet been reported.

In the present study, we performed a side-by-side comparison of the clonal architectures derived by genotyping of single-cell colonies with results obtained by the Tapestry scDNAseq method. We used samples from MPN patients with *JAK2*-V617F and at least 1 additional somatic mutation. We also tested whether

¹Department of Biomedicine, Experimental Hematology, University Hospital Basel and University of Basel, Switzerland

²Univ Angers, Nantes Université, CHU Angers, Inserm, CNRS, CRCI2NA, Angers, France

³Division of Hematology, University Hospital Basel, Switzerland

*Current address for R. Nienhold: Institute of Pathology and Molecular Pathology, University Hospital Zurich, Zurich, Switzerland.

Supplemental digital content is available for this article.

Copyright © 2023 the Author(s). Published by Wolters Kluwer Health, Inc.

on behalf of the European Hematology Association. This is an open-access

article distributed under the terms of the Creative Commons Attribution-Non

Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible

to download and share the work provided it is properly cited. The work cannot be

changed in any way or used commercially without permission from the journal.

HemaSphere (2023) 7:5(e885).

<http://dx.doi.org/10.1097/HS9.0000000000000885>.

Received: December 7, 2022 / Accepted: March 31, 2023

the order of acquisition of mutations and the sequential versus bi-clonal patterns have an impact on the phenotype and overall survival (OS).

METHODS

Patients and samples

Collection of blood samples and clinical data of MPN patients at the University Hospital Basel, Switzerland was approved by the local Ethics Committee (Ethik Kommission Beider Basel), and written informed consent was obtained from all patients in accordance with the Declaration of Helsinki. The diagnosis of MPN was established according to the World Health Organization and International Consensus Classification criteria.^{10,11}

Next-generation sequencing

The mutational profiles of patients were derived by targeted NGS of granulocyte DNA as previously described,¹³ with minor modifications. After fragmentation of DNA by enzymatic digestion or sonication, the complete coding regions of the 52 genes listed in Suppl. Table S1 were captured using a custom set of probes (Agilent SureSelectXT). Libraries were generated and loaded on HiSeq2000 or NovaSeq6000 for 100-bp paired-end sequencing. DNA samples were processed and sequenced in duplicates, and only sequence alterations that were present in duplicate samples were retained. Mapping of reads and variant calling was performed using the CLC genomics workbench (v7.5), with a variant allele frequency (VAF) threshold of 5%. The variants were annotated with gnomAD2.1, Cosmicv89, ClinVar, and tools for predicting the functional impact of mutations (Polyphen2, SIFT, MutationTaster). Patients with *JAK2-V617F* and one or more additional somatic mutations were selected for the study.

Single-cell DNA sequencing

We selected 22 patients with *JAK2-V617F* and at least 1 additional somatic gene mutations for detailed analysis. A vial of frozen peripheral blood mononuclear cells (PBMCs) from each of these patients was thawed and then split for scDNAseq and single-cell liquid culture for genotyping of colonies. For Tapestry scDNAseq (Mission BioTM), a custom panel of 97 amplicons was designed, covering the 31 known mutated gene regions of the selected 22 patients. Additional 66 amplicons were designed to cover mutational hotspots and genes not previously detected to be mutated by NGS in the 22 selected patients, including *ASXL1*, *CALR*, *DNMT3A*, *EZH2*, *IDH1*, *IDH2*, *JAK2*, *KRAS*, *MPL*, *NRAS*, *RUNX1*, *SF3B1*, *SRSF2*, *TET2*, *TP53*, and *U2AF1*. Briefly, single-cell libraries were built from 120,000 PBMCs loaded into a microfluidic cartridge for a theoretical final number of 10,000 cells captured. Quality was accessed by Bioanalyser DNA1000 chip (Agilent). Finally, libraries were pooled for a 2 × 150 bp sequencing on an Illumina NovaSeq for a theoretical number of 52 million reads per sample. Data were processed by the Tapestry Insight v2.2 software and a custom R script. Briefly, variant calling was performed with GATKv3.7 in each cell and variants were filtered in order to remove low confidence mutations. Uncertainties concerning homozygous versus heterozygous mutations due to potential allele drop-out were manually curated by inspecting the coverage and the base quality score for each genomic position with a variant in all clones detected. Preparation of samples and libraries and bioinformatic analyses are detailed in Supplemental Methods.

Genotyping of colonies

For liquid cultures, CD34-positive and Lin-negative cells from PBMC were single-cell sorted into 96-well plates containing 100 μ L of free medium consisting of StemSpam SFEMII

(Stemcell Technologies) with human cytokines stem cell factor (SCF) (20 ng/mL), interleukin-3 (IL-3) (20 ng/mL), IL-6 (20 ng/mL), IL-9 (20 ng/mL), IL-11 (20 ng/mL), thrombopoietin (TPO) (20 ng/mL), granulocyte-colony stimulating factor (G-CSF) (20 ng/mL), granulocyte monocyte-colony stimulating factor (GM-CSF) (50 ng/mL), and erythropoietin (EPO) (3 U/mL) (Peprotech, Rocky Hill, NJ) and fresh media (50 μ L) was added on day 8 of culture. PBMCs were seeded at a density of 100,000 cells/mL in methylcellulose media containing cytokines (H4034, StemCell technologies) and cultivated for 14 days at 37°C and 5% CO₂. On day 14, DNA was extracted from colonies (either from liquid culture or from methylcellulose assay) using Chelex 100 Resin (Cat. No. 143-2832, Bio-Rad Laboratories, Hercules, CA). Known somatic mutations were genotyped either by allele-specific polymerase chain reaction (AS-PCR) for *JAK2-V617F*,²⁰ or by Sanger sequencing of amplicons covering the mutated gene regions (details in Supplemental Methods).

JAK2-V617F quantification by digital PCR

To determine the VAF of *JAK2-V617F* in the cell populations used for scDNAseq and liquid culture assays, aliquots of unfractionated PBMCs and CD34+ cells from PBMCs were taken for DNA extraction followed by digital PCR. Briefly, DNA was extracted using QIAmp DNA micro kit (Qiagen) and then 80 ng of DNA was loaded on a 20 k-chipv2 (ThermoFisher) with 20,000 partitions for allele-specific PCR using Taqman probes with FAM marker for wildtype *JAK2* and VIC marker for *JAK2-V617F* (assay Hs000000038_rm, ThermoFisher). After 39 cycles, fluorescence in each well was scored by a QuantStudio3D (ThermoFisher) analyzer, and the VAF was derived using the QuantStudio3D analysis suite. The sensitivity of digital PCR for *JAK2-V617F* was 0.034%.

Statistics

Accuracy of clonal architecture methodology to derive the *JAK2-V617F* VAF was accessed by a nonparametric Passing-Bablok regression with the digital PCR measurement as a reference. For unsupervised clustering, a multivariate dimensional reduction of 10 clonal architecture parameters was performed by a factor analysis of mixed data analysis (Supplemental Methods) and followed by a hierarchical classification based on the ward distance (R package FactoMineR). Groups' characteristics were compared using Fisher exact test for categorical variables and the Kruskal-Wallis test followed by Dunn tests for multiple comparisons for continuous variables. Survival estimates were obtained with the Kaplan-Meier method and statistical tests performed using Cox models. Statistical analysis was performed using R software (www.R-project.org, Vienna, Austria).

Additional methods are described in the online version of this article.

RESULTS

Comparison of methodologies for determining the clonal architecture of hematopoiesis in a cohort of MPN patients carrying a *JAK2-V617F* mutation

From our cohort of 225 MPN patients positive for the *JAK2-V617F* mutation, we selected 50 patients that carried at least 1 additional somatic gene mutations by targeted NGS (Figure 1). In these 50 patients, we characterized the clonal architecture by either single-cell liquid cultures or colony assays in methylcellulose. In 22 of these 50 patients, we performed scDNAseq and genotyping of single-cell liquid culture colonies starting from the same vial of frozen PBMCs (Figure 2A). Of the remaining 28 patients, 7 were studied by liquid cultures and 21 by methylcellulose colony assays. scDNAseq of PBMCs allowed scoring a median of 4367 (range, 459–8009) individual cells per patient

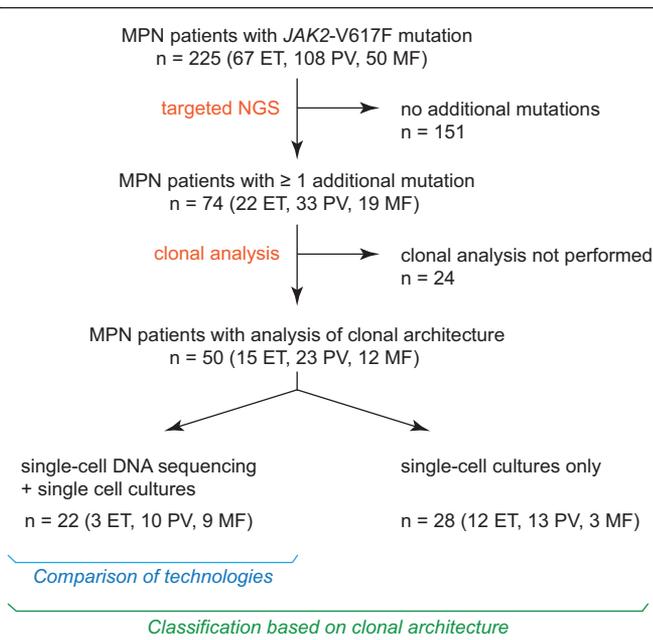


Figure 1. Overview of MPN patients included in the study. A flow chart of patients with MPN included in this study and analyses that were performed are shown. NGS, targeted next-generation sequencing with a panel of 68 genes. ET = essential thrombocythemia; MF = myelofibrosis; MPN = myeloproliferative neoplasms; PV = polycythemia vera.

sample and only cells were scored with a sufficient number of sequence reads in all gene regions previously defined as mutated by bulk NGS was obtained, that is, median of 2786 cells per sample (range, 277–6717). Lower coverage mostly corresponded to mutations located in GC-rich regions (eg, *SRSF2*-P95). Liquid cultures were performed with CD34-positive cells from PBMCs with a colony-forming efficiency of 60% (range, 14%–86%), which allowed us to obtain a median of 144 colonies per patient (range, 27–256). While the DNA from single colonies was genotyped by Sanger sequencing or AS-PCR, and for each individual patient only the gene mutations previously identified by bulk NGS were inspected, a larger but identical set of gene regions was sequenced by Tapestry scDNAseq in all patients, which comprised all gene regions found to be mutated by bulk NGS in any of the 22 patients. Because Tapestry allows to simultaneously probe multiple amplicons, we also added and covered a number of known mutational hotspots for some of the genes and analyzed a total of 97 amplicons in all 22 patients.

In addition to *JAK2*-V617F present in all 22 patients, 37 somatic mutations in 12 genes detected by bulk NGS in granulocyte DNA (Suppl. Table S2) were analyzed using the 2 methods to reconstruct the clonal architecture. A concordance of 97% for detecting these mutations was found between the 2 methods (Figure 2C). One mutation detected only by scDNAseq was a *EZH2* gene mutation in patient P499A that was missed by genotyping of a total of 78 colonies (Suppl. Figure S1D), and one *TET2* gene mutation was missed by scDNAseq due to a coverage failure of the amplicon located at this position. Using these 59 mutations, a total of 86 individual clones or subclones could be identified, with a concordance of 80% was achieved between the 2 methods (Figure 2C and Suppl. Table S3). The higher number of clones compared with mutations ($n = 27$) was due to the distinction between heterozygous and homozygous subclones that were each counted separately ($n = 24$), and to somatic mutations acquired twice in 2 separate clones in the same patient ($n = 3$). The 8 clones detected only by scDNAseq were small homozygous clones corresponding to late events (Suppl. Figure S1B). Conversely, genotyping of colonies detected

8 subclones in 5 patients that were not found by scDNAseq, of which 7 subclones represented intermediate ancestral subclones in the transition from heterozygosity to homozygosity (Figure 3A and Suppl. Figure S1A).

Thus, scDNAseq appears to be less reliable in distinguishing heterozygosity from homozygosity in some cases. This could be due to an allele drop-out artifact in heterozygous cells. We also used unfractionated PBMCs for scDNAseq, whereas in the single-cell liquid cultures, the progeny of CD34+ progenitors was scored. The difficulties in distinguishing heterozygosity from homozygosity is inherent to the scDNAseq methodology. To correctly identify cells heterozygous for a mutation, scDNAseq must amplify and sequence the relevant locus from both the chromosomes (Suppl. Figure S2A). If only one of the loci is amplified from a heterozygous cell, the genotype will be wrongly assigned to either wildtype or homozygous. Allele drop-out typically results in a reduced frequency of reads for the homozygous or wildtype false-positive genotypes (Suppl. Figure S2). Identification of allele drop-out artifacts was facilitated by detection of the 2 counterparts of the allele bias resulting in false-positive homozygous and wildtype cells occurring at a low but similar frequency (Suppl. Figure S3).

Since with scDNAseq in each of the 22 patients we simultaneously screened for mutations in 17 genes using 97 amplicons, we also detected 11 new mutations in addition to the 59 somatic mutations previously identified by NGS (Figure 3B and Suppl. Figure S1C). We found that these newly detected mutations were already present in the original NGS alignments from granulocyte DNA, but with a VAF below the cutoff threshold of 5% (data not shown). Thus, genotyping of colonies is very reliable for reconstructing the clonal architecture, but the sensitivity for detecting mutations with low VAF by genotyping of colonies is dependent on the number of colonies that are available for analysis.

A very good correlation was found between the *JAK2*-V617F VAF determined by scDNAseq and digital PCR in DNA from unfractionated PBMCs ($r = 0.993$) (Figure 4A, left panel). VAF calculated from genotyping of CD34+ cell-derived colonies in liquid culture compared with digital PCR of DNA from the original CD34+ cells showed more variability, with an overall trend toward lower values by genotyping of colonies (Figure 4A, right panel). The cloning efficiencies in liquid culture may differ for individual combinations of gene mutations, in some cases lowering and in other cases increasing the likelihood that a CD34+ progenitor cell will expand and initiate a colony.

To estimate the effects of clonal selection bias caused by the single-cell cultures, we performed a detailed analysis in 1 patient from whom bone marrow cells were available. We obtained sufficient numbers of CD34+ cells for scDNAseq and single-cell liquid cultures and compared the 2 methodologies using same starting cell population (Figure 4). Heterozygous *JAK2*-V617F mutation was found in 3.1% of CD34+ bone marrow cells by scDNAseq, whereas 10% of colonies were heterozygous for *JAK2*-V617F by genotyping of single-cell-derived colonies. The VAF for *JAK2*-V617F was 5% in peripheral blood granulocytes and 6% in total bone marrow cells. Thus, the higher estimate of *JAK2*-V617F-positive CD34+ progenitors in bone marrow from this patient is likely due to the increased cloning efficiency of *JAK2*-mutant progenitors compared with wildtype progenitors. Interestingly, the *SF3B1* subclone was slightly underrepresented by genotyping of colonies (15% versus 29%), suggesting that the *SF3B1* mutation decreased the cloning efficiency compared with wildtype progenitors. Interestingly, the presence of *JAK2*-V617F/*SF3B1* double mutant cells/colonies that carry an identical but rare *SF3B1* mutation indicates that *JAK2*-V617F was acquired twice in this patient (P368H). A similar pattern was also found in P099D (Suppl. Figure S1).

Overall, our data demonstrate that the 2 methodologies in most cases produced concordant results yielding the same clonal architectures. Genotyping of colonies may shift the frequencies of subclones when gene mutations increase or decrease cloning efficiencies,

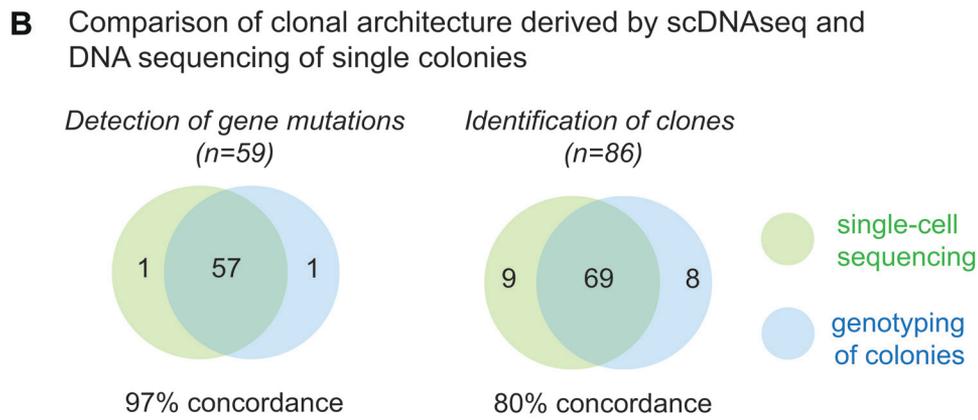
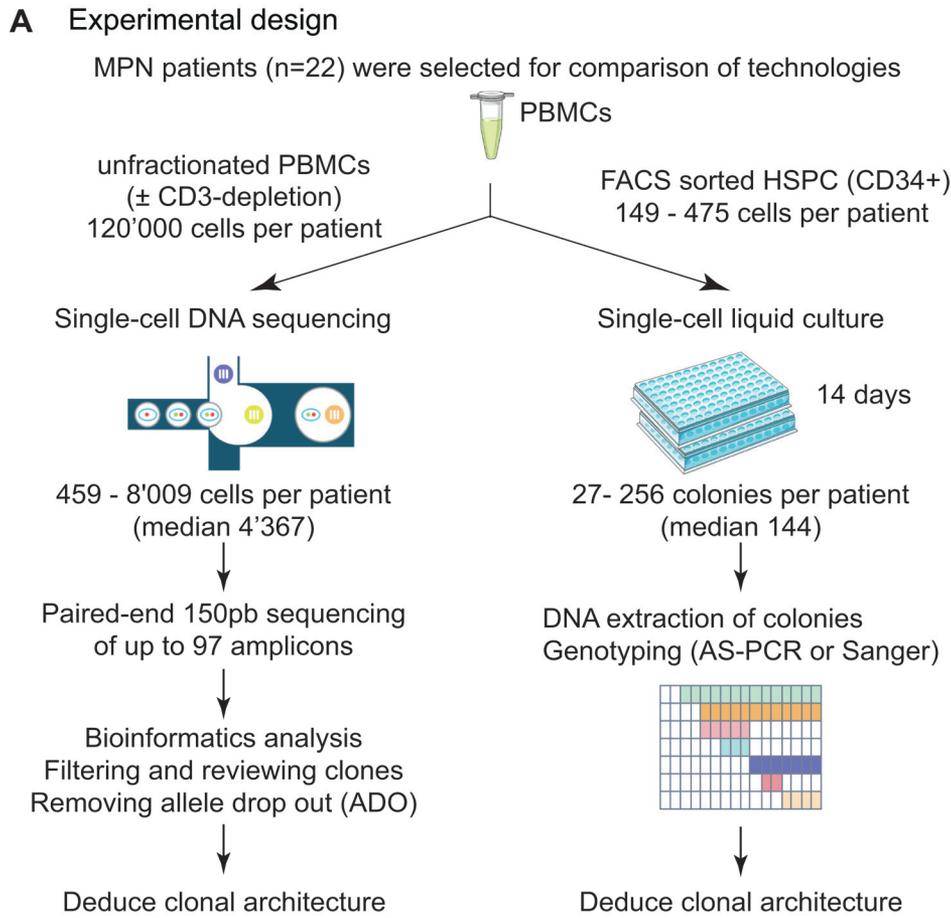


Figure 2. Comparison of methodologies to define the clonal architecture of hematopoiesis in MPN patients carrying >1 somatic gene mutation. (A) Experimental design. Starting material were frozen PBMCs. (B) Venn plots showing the concordance and differences in detecting known somatic mutations (left panel), or derived individual clones (right panel) using scDNAseq vs genotyping DNA from individual colonies. AS-PCR = allele-specific PCR; FACS = fluorescence activated cell sorting; HSPC = hematopoietic stem and progenitor cells; MPN = myeloproliferative neoplasms; PBMCs = peripheral blood mononuclear cells; scDNAseq = single-cell DNA sequencing.

whereas scDNAseq can detect small subclones and additional mutations, but has more difficulties distinguishing heterozygous from homozygous mutations due to allele drop-out artifacts.

Correlation between clonal architecture and clinical features of MPN patients with JAK2-V617F

To obtain a cohort with sufficient numbers of patients with fully characterized clonal architecture, we combined data from the 22 patients studied by both scDNAseq and genotyping of

colonies with data from 28 additional patients studied solely by genotyping of colonies (Figure 1). With these 50 patients, unsupervised clustering using only data on clonal architecture, including 10 parameters listed in Supplemental Methods, defined 4 well-separated clusters (Figure 5A). The clusters did not differ in the distribution of the individual mutated genes, but cluster 2 showed the least complex mutational landscape and cluster 4 had the most complex profile with the largest number of additional mutations per patient (Figure 5B). The

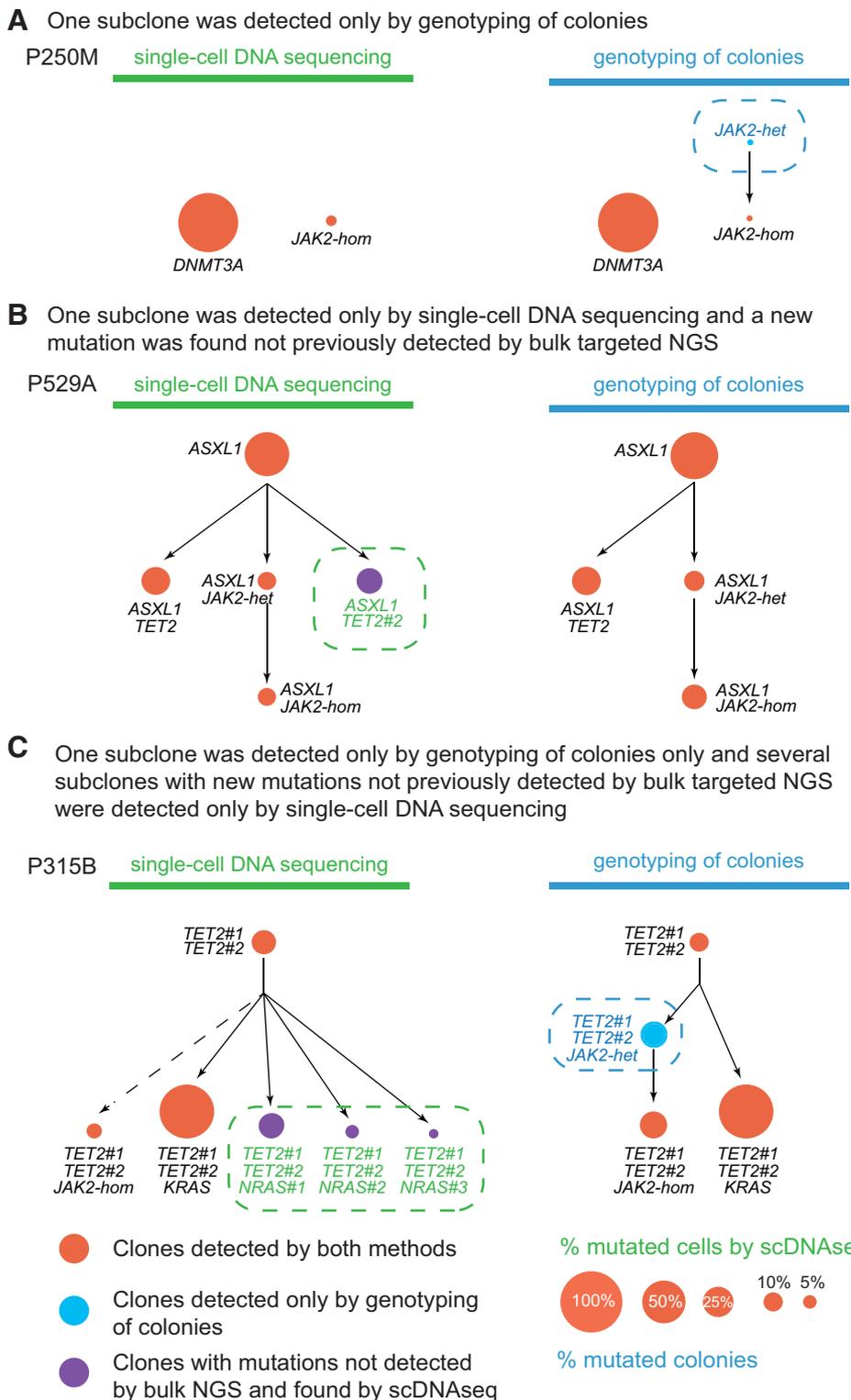
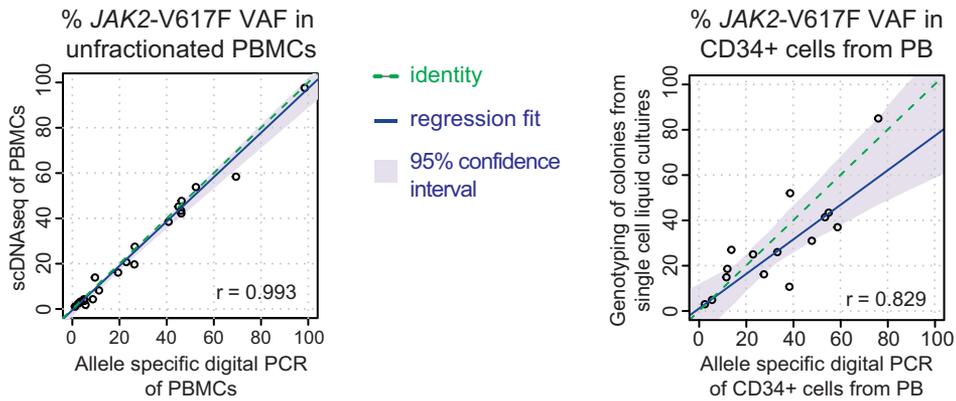


Figure 3. Examples illustrating concordance and discordance in detecting subclones by single-cell DNA sequencing vs genotyping of colonies. Examples of 3 patients in whom differences in the derived clonal architecture were noted when comparing the 2 methodologies. The deduced order of acquisition of mutations is drawn from top to bottom. Each subclone is represented by a circle and its size represents the percentage related to the sum of all mutated subclones, which was set to 100%. (A) In P250M, 1 subclone was detected only by genotyping of colonies (marked by dashed blue circle). (B) In P529A, 1 subclone was detected only by single-cell DNA sequencing and corresponded to a mutation not detected by bulk NGS (marked by dashed green circle). (C) In P315B, some subclones were detected only by single-cell DNA sequencing (marked by dashed green circle) and 1 subclone was detected only by genotyping of colonies (marked by dashed blue circle). NGS = next-generation sequencing; scDNAseq = single-cell DNA sequencing.

proportion of patients with at least 1 mutation in genes previously associated with adverse outcome in myelofibrosis (ie, *ASXL1*, *CBL*, *EZH2*, *IDH1*, *IDH2*, *KRAS*, *NRAS*, *SRSF2*,

U2AF1, and *TP53*) was 27% for cluster 1, 67% for cluster 2, 37% for cluster 3, and 100% for cluster 4. The order in which the additional mutations were acquired with respect

A JAK2-V617F VAF in 22 patients studied by scDNAseq and single cell liquid cultures



B Comparison of data derived by scDNAseq and genotyping of colonies from FACS-sorted CD34+ bone marrow cells and PBMCs

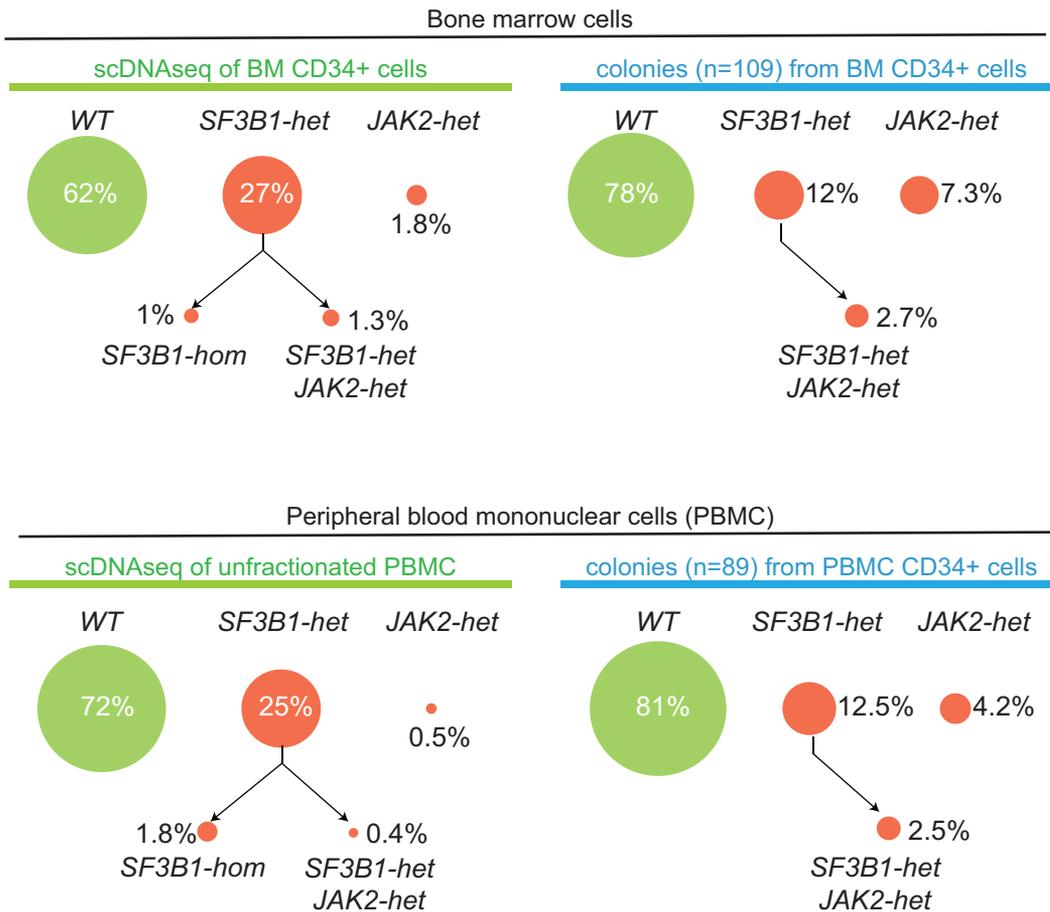
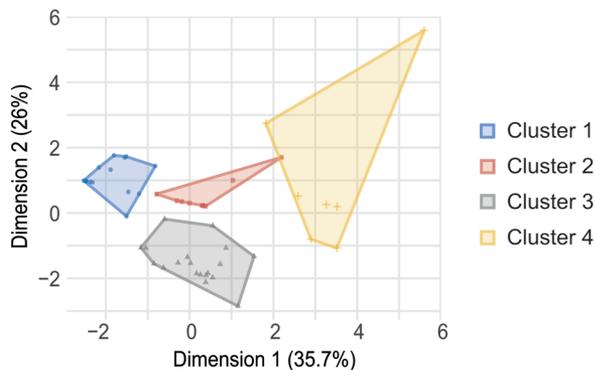


Figure 4. Comparison of scDNAseq and liquid culture for determination of VAF and size of clones. (A) Correlation between the VAF measured by digital AS-PCR in bulk cell populations (PBMC or CD34-positive cells) vs VAF deduced from scDNAseq (left panel) or from genotyping DNA from individual colonies (right panel). Nonparametric Passing-Bablok regression was performed and the correlation was estimated by Pearson r parameter. (B) Comparison of data derived by scDNAseq and genotyping of colonies from FACS-sorted CD34+ bone marrow cells of patient P368H (top panel). PBMC were also analyzed from the same sampling date (bottom panel) by scDNAseq on total PBMC and colonies derived from sorted CD34+ cells. Each circle represents the percentage related to the sum of all cells or colonies genotyped. Wildtype cells/colonies are represented by a green circle and mutated clones by red circles. AS-PCR = allele-specific PCR; FACS = fluorescence activated cell sorting; PBMCs = peripheral blood mononuclear cells; scDNAseq = single-cell DNA sequencing; VAF = variant allele frequencies.

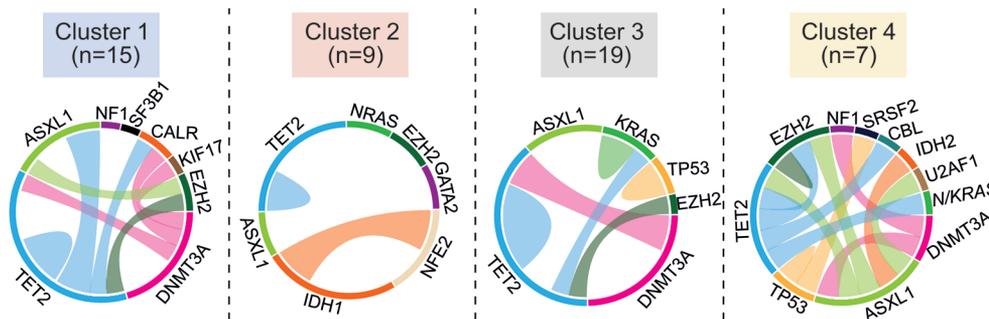
to JAK2-V617F also differed between clusters. In cluster 2, JAK2-V617F was always the first event, whereas in cluster 3 at least 1 additional mutation always preceded the acquisition

of JAK2-V617F (Figure 5C). In cluster 1, the majority of additional mutations represented a clone separate from JAK2-V617F. Cluster 4 was more complex, without a strict order

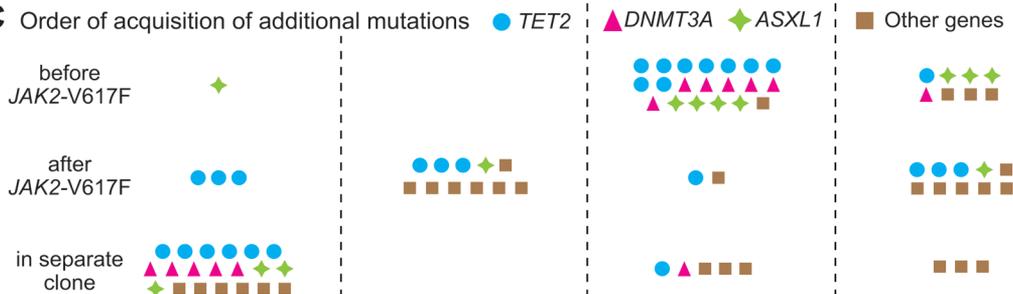
A Unsupervised clustering based on clonal architecture



B Co-occurrence of additional somatic mutations within the clusters



C Order of acquisition of additional mutations



D Number and percentages of patients with a subclone homozygous for *JAK2-V617F*

Cluster 1	Cluster 2	Cluster 3	Cluster 4
3/15 (20%)	8/9 (89%)	14/19 (74%)	7/7 (100%)

E Number and percentages of MPN subtypes within each cluster

ET	7 (47%)	0	8 (42%)	0
PV	6 (40%)	8 (89%)	7 (37%)	2 (29%)
MF	2 (13%)	1 (11%)	4 (21%)	5 (71%)
Total	15 (100%)	9 (100%)	19 (100%)	7 (100%)

Figure 5. Analysis of clonal architecture allowed defining clusters with specific molecular landscapes. (A) Map derived by Factor Analysis of Mixed Data methodology using clonal architecture parameters. The 4 clusters obtained after a clustering based on the ward distance are indicated by colored areas. (B) Circos plots illustrating co-occurrence of somatic mutations in the same individual. The length of the arc corresponds to the frequency of the mutation, whereas the width of the ribbon corresponds to the relative frequency of co-occurrence of 2 mutations in the same patient. (C) Order of acquisition of additional somatic mutations shown in relation to the *JAK2-V617F* driver mutation for each of the 4 clusters. Each symbol represents 1 gene mutation, as indicated at the bottom. (D) Proportion of patients with at least 1 subclone with homozygous *JAK2-V617F* mutation detectable in each cluster. (E) Proportions of MPN subtypes across clusters. MPN = myeloproliferative neoplasms.

of events. The order of events in all 50 patients studied is also depicted in fish-plots (Suppl. Figure S4). The clusters also differed in the presence or absence of subclones homozygous for

JAK2-V617F. In cluster 1, only 20% of patients had a homozygous subclone, whereas in cluster 4, a homozygous subclone was detectable in all patients (Figure 5D).

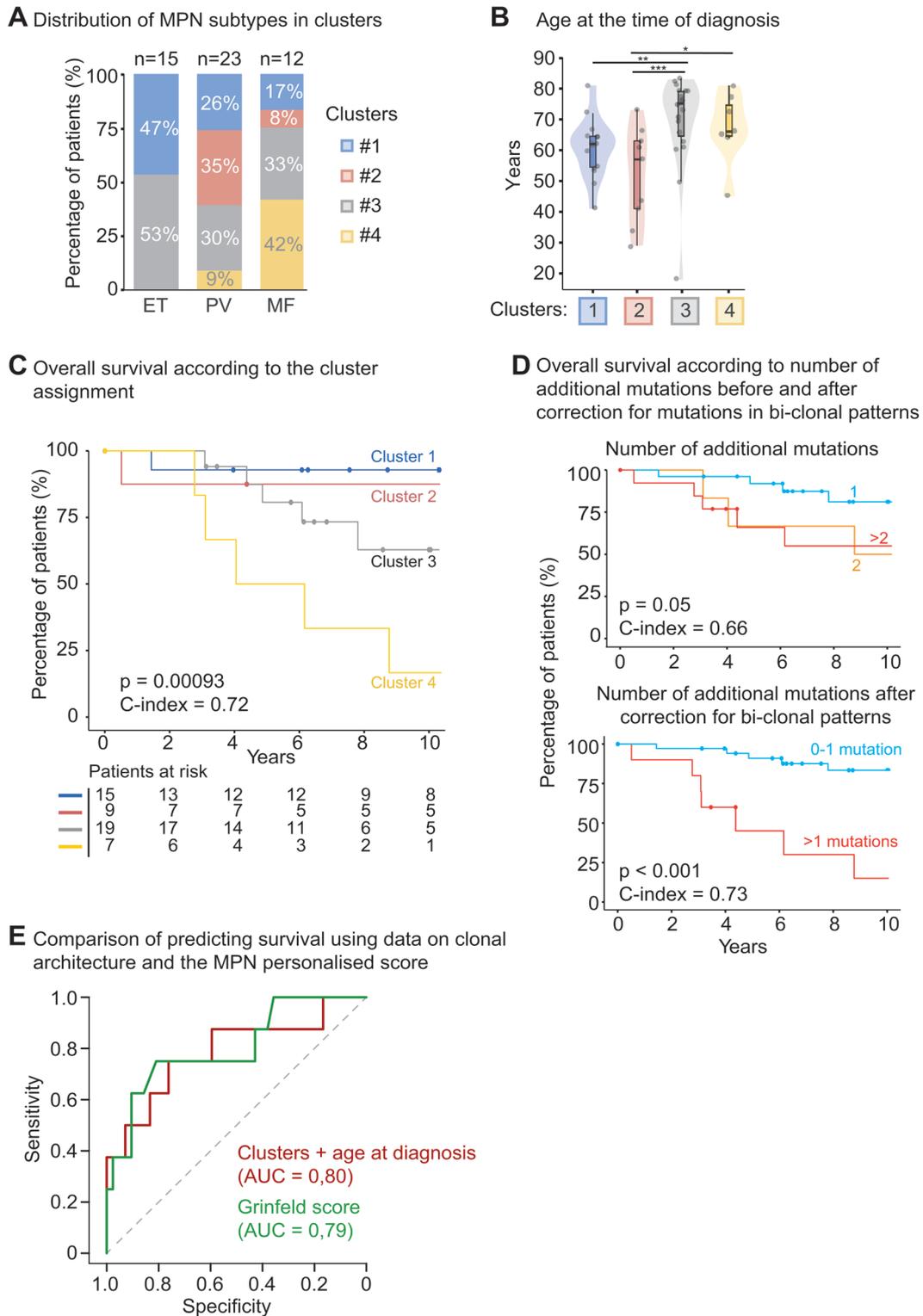


Figure 6. Clinical features of MPN patients assigned to clusters derived by clonal architecture data. (A) Frequencies of the 4 clusters within the MPN subtypes ET, PV, or MF. (B) Age at the time of diagnosis of patients within the 4 clusters. Differences were tested by a Kruskal-Wallis test followed by Dunn multiples tests. (C). Overall survival of MPN patients assigned to the 4 clusters. (D) Overall survival of MPN patients according to the number of additional gene mutations without correction for bi-clonal pattern (upper panel) and with correction for bi-clonal pattern (lower panel). Survival estimates were obtained with the Kaplan-Meier method and statistical tests performed using a log-rank test. The performance of prediction was accessed by Harrell's concordance index (C-index). (E) Comparison of survival predictions using data on clonal architecture vs the MPN personalized risk calculator. ROC curve were plotted, and the performance of each prognostic systems was evaluated using the AUC. AUC = area under the ROC curve; ET = essential thrombocythemia; MF = myelofibrosis; MPN = myeloproliferative neoplasms; PV = polycythemia vera; ROC = receiver operating characteristic.

The patients assigned to the 4 clusters also showed differences in clinical features. Patients with ET were present only in clusters 1 and 3 (Figure 5E), whereas patients with PV and MF were found in all 4 clusters, but MF was most frequent in cluster 4 (Figure 6A). Patients assigned to cluster 3 were on average older at the time of diagnosis than patients in clusters 1 or 2 (Figure 6B). Patients in cluster 4 had a reduced OS with a median OS of 5 years (Figure 6C). The hazard ratio for death in cluster 4 versus the other clusters was 5.32. Multivariate analysis confirmed that negative prognostic impact of cluster 4 on OS was independent of the subtype of MPN, the age at the time of diagnosis (Table 1), and the presence of high molecular risk (HMR) mutations included in the Mutation-Enhanced International Prognostic Scoring System (MIPSS70) risk score for PMF (ie, *ASXL1*, *EZH2*, *IDH1*, *IDH2*, *SRSF2*, and *SRSF2*) (Suppl. Table S4).

Evolution to AML or myelodysplastic syndrome was more frequently observed in patients assigned to cluster 4 (29% of patients) than in patients from other clusters (7%, 11%, and 11% for clusters 1, 2, and 3, respectively). The subtype of MPN, the number of HMR mutations, and the total number of additional mutations also had a prognostic impact in this cohort (Figure 6D, upper panel and Suppl. Figure S5). However, we found that clonal architecture classification performed better for predicting OS, as demonstrated by a higher C-index (0.72 versus 0.68, 0.68, and 0.66, respectively). Interestingly, despite the presence of additional mutations associated with high-risk, fewer deaths occurred in patients in cluster 1 than in other clusters, suggesting that mutations present in a clone separated from the clone carrying the disease driver mutation do not have an impact on prognosis. Indeed, when only mutations residing in the same clone as *JAK2*-V617F were counted, the predictive value of the number of mutations for survival improved, more clearly separating a high-risk and low-risk group (Figure 6D, lower panel).

Finally, we compared the predictions based on our clonal architecture classification with predictions by the web-based MPN personalized risk calculator from the Sanger Institute that also uses gene mutation data and is applicable to all MPN subtypes.²¹ Both methods were in good agreement for predicting the OS of patients in our cohort at 10 years of follow-up with an AUC = 0.79 for the Sanger risk calculator and 0.80 for our clonal architecture-based classification (Figure 6E). In some patients in cluster 1, with a high proportion of bi-clonal patterns, our clonal architecture-based classification performed better in predicting OS. A comparison with the MIPSS70 score was not feasible, because MIPSS70 is applicable only to PMF, and our cohort included only 11 PMF and 1 secondary MF patient.

DISCUSSION

The clonal architectures deduced from genotyping of colonies and from Tapestry scDNAseq of PBMCs showed very good overall concordance (Figure 2C). Genotyping by scDNAseq

showed higher sensitivity for mutations with low VAF due to the high number of single cells that were scored per patient sample. Furthermore, scDNAseq allowed identifying subclones defined by additional mutations not previously found by bulk NGS, because the methodology allows to include other genes or gene regions (eg, mutational hot spots) covered by sequencing and not only the location where the previously identified mutation lies (Figure 3). In principle, sequencing of DNA from single colonies could also include an expanded set of amplicons covering more genes. However, this would substantially increase the work and the costs and would make it even less applicable for larger cohorts of patients.

As expected, scDNAseq had more difficulties to distinguish between heterozygous and homozygous mutations. Allele drop-out artifacts are inherent to the scDNAseq methodology and their rate varies for different gene loci and between samples, in most cases ranging around 3%–10% of individual cells.¹⁸ By a manual curation, we identified and removed clones corresponding to allele drop-out, but we cannot exclude that some cells in these clusters could be true homozygotes. Another reason why some of the results between scDNAseq and genotyping of colonies differed could be that different starting cell populations were used for the 2 methodologies, that is, unfractionated PBMC versus peripheral blood CD34+ progenitors.

Other approaches that do not require expansion of cells in culture and can be applied to unfractionated cell populations have been developed for genotyping. The 10× chromium method uses RNA from single cells as template for sequencing the mutated part of the gene. The sensitivity is dependent on the expression levels of the corresponding mRNA and decreases with the distance from the 3'-end of the mRNA. The sensitivity can be improved by adding nested PCR for amplifying cDNA of the specific locus. This GoT approach applied to MPN allowed genotyping for *CALR* mutations, but did not work reliably for *JAK2*-V617F mRNA,^{22,23} which is expressed at lower levels than *CALR* and located far from the 3'-end.

Several methods for genotyping of fluorescence activated cell sorted (FACS) single cells have been described, which are less suited for high-throughput applications, because each cell has to be processed individually. A DNA-based approach that involved genome-wide DNA amplification of single-cell DNA followed by Taqman-based detection of previously defined mutations was applied to PMF patients.²⁴ The Smart-seq2 method used full-transcript RNA sequencing that avoids bias for the 3'-ends, but produced high rates of allele drop-out due to the nonrandom monoallelic nature of expression of the selected genes in single cells.²⁵ TargetSeq, a protocol combining full-transcript RNA sequencing with DNA-targeted genotyping in individual FACS-sorted cells resulted in a good efficiency of genotyping and low rate of allele drop-out,²⁶ but the number of individual cells and transcriptomes that can be analyzed is lower than with the GoT. The Tapestry scDNAseq methodology using encapsulation allowed genotyping several thousand single cells per patient sample with an acceptable rate of allele drop-out and is suited for high-throughput applications. However, this DNA-based methodology does not allow simultaneous sequencing of the transcriptome. Currently, the high costs of Tapestry scDNAseq is a limiting factor for large-scale use.

These technologies now allow to more efficiently determine the clonal structure of hematopoiesis in patients with multiple somatic mutations and allow to address to what degree the clonal architecture influences clinical features and outcome. We defined 4 distinct clusters by unsupervised analysis of data on clonal architecture from 50 MPN patients that were genotyped by Tapestry scDNAseq and/or single-cell colony assays. These 4 clusters differed from each other by the order of acquisition of the mutations, the total number of mutations, the complexity in branching and evolution of the subclones, and the presence or

Table 1
Multivariate Cox Model for Overall Survival

Variable	Hazard Ratio	95% CI	P Value
Age at diagnosis	1.06	1-1.12	0.049
Diagnosis (ET as reference)			
PV	0.37	0.08-1.83	0.22
Myelofibrosis	1.14	0.23-5.6	0.87
Clusters (clusters 1, 2, and 3 as reference)			
Cluster 4	5.8	1.4-24	0.015

CI = confidence interval; ET = essential thrombocythemia; PV = polycythemia vera.

absence of a subclone homozygous for *JAK2*-V617F (Figure 5). All ET patients were found in either cluster 1 or 3. In cluster 1, the majority of additional mutations represented a clone separate from *JAK2*-V617F, whereas in cluster 3 at least 1 additional mutation always preceded the acquisition of *JAK2*-V617F (Figure 5C).

Previous reports focusing on *TET2* or *DNMT3A* also found that in the majority of ET patients the *TET2* or *DNMT3A* mutation preceded the acquisition of *JAK2*-V617F.^{16,27} However, in the first study, bi-clonal patterns with separated clones were not reported for *TET2* and *JAK2*-V617F mutations.¹⁶ The follow-up study reported a few cases of bi-clonal pattern for *DNMT3A* and *JAK2*-V617F mutations in PV and PMF, but none in ET patients.²⁷ In cluster 2, *JAK2*-V617F was always the first event, and these patients were on average younger than those assigned to the other clusters (Figure 6B). Patients assigned to cluster 4 had always a detectable subclone homozygous for *JAK2*-V617F and overall carried a larger number of additional mutations without a preferred order of acquisition resulting in more complex subclonal structures than the other clusters.

A global pan-MPN genomic classification for prognosis stratification in all subtypes of MPN was proposed.^{21,28} Cluster 4 in our clonal architecture-based stratification was a risk factor predicting reduced OS and in multivariate analysis was independent of the MPN subtype and the age at diagnosis. The number of additional mutations was previously associated with worse prognosis in patients with MPN in general,¹³ and PMF in particular,²⁹ but in our study, the cluster classification performed better in predicting OS than the number of additional mutations (Figure 5D). The prediction power was improved when information on the bi-clonal structure was used and only additional mutations found in a common clone with the *JAK2*-V617F driver mutation were considered (Figure 6D). Our results suggest that also HMR mutations in the genes *ASXL1*, *EZH2*, *IDH1/2*, *SRSF2*, or *U2AF1*^{30,31} when present in a separated clone may not be associated with an adverse prognosis. The cooperative effects of additional mutations with *JAK2*-V617F has also been documented in several mouse models, and depended on their presence in the same hematopoietic cell.^{32–36}

Our results show that scDNAseq can be reliably applied to decipher the clonal architecture in patients with MPN that carry multiple gene mutations. This information can improve the molecular-based classification and prognostic stratification, in particular when gene mutations residing in clones separated from the clone carrying the disease driver mutation are subtracted. To clarify the potential impact of the order of acquisition of somatic mutations and subclonal structures on prognosis and response to therapies, larger cohorts of patients with MPN will need to be analyzed.

ACKNOWLEDGMENTS

We thank Stella Stefanova, Lorenzo Raeli, and Emmanuel Traunecker from the flow cytometry core facility of Department of Biomedicine for help in cell sorting, Christian Beisel from the Genomics Facility Basel (ETH Zurich, D-BSE located in Basel operated jointly with the University of Basel) for sequencing support, Jérémie Riou from the University of Angers for support with statistical analysis and members of the laboratory for helpful discussions and critical reading of our article.

AUTHOR CONTRIBUTIONS

DLP designed and performed research, analyzed data, and wrote the article. MSB and JRP collected and analyzed patient data. RN, SR, TAF, JS, HHS, and GMS, performed research. RCS designed research, analyzed data, and wrote the article.

DATA AVAILABILITY

For original data and reagents, please contact radek.skoda@unibas.ch.

DISCLOSURES

RCS is a scientific advisor and has equity in Ajax Therapeutics; he consulted for and received honoraria from Novartis and BMS/Celgene. All the other authors have no conflicts of interest to disclose.

SOURCES OF FUNDING

This work was supported by grants from the Société Française d'Hématologie (Aide à la Mobilité 2020) to DLP, and the Swiss National Science Foundation (31003A_166613, and 310030_185297) and Swiss Cancer Research Foundation (KFS-3655-02-2015 and KFS-4462-02-2018) to RCS.

REFERENCES

- James C, Ugo V, Le Couedic JP, et al. A unique clonal *JAK2* mutation leading to constitutive signalling causes polycythaemia vera. *Nature*. 2005;434:1144–1148.
- Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of *JAK2* in myeloproliferative disorders. *N Engl J Med*. 2005;352:1779–1790.
- Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase *JAK2* in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell*. 2005;7:387–397.
- Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase *JAK2* in human myeloproliferative disorders. *Lancet*. 2005;365:1054–1061.
- Scott LM, Tong W, Levine RL, et al. *JAK2* exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med*. 2007;356:459–468.
- Pikman Y, Lee BH, Mercher T, et al. *MPLW515L* is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med*. 2006;3:e270.
- Pardanani AD, Levine RL, Lasho T, et al. *MPL515* mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood*. 2006;108:3472–3476.
- Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of *calreticulin* in myeloproliferative neoplasms. *N Engl J Med*. 2013;369:2379–2390.
- Nangalia J, Massie CE, Baxter EJ, et al. Somatic *CALR* mutations in myeloproliferative neoplasms with nonmutated *JAK2*. *N Engl J Med*. 2013;369:2391–2405.
- Khoury JD, Solary E, Abila O, et al. The 5th edition of the World Health Organization Classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*. 2022;36:1703–1719.
- Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood*. 2022;140:1200–1228.
- Tefferi A, Guglielmelli P, Larson DR, et al. Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood*. 2014;124:2507–13; quiz 2615.
- Lundberg P, Karow A, Nienhold R, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood*. 2014;123:2220–2228.
- Guglielmelli P, Lasho TL, Rotunno G, et al. MIPSS70: Mutation-Enhanced International Prognostic Score System for Transplantation-Age Patients With Primary Myelofibrosis. *J Clin Oncol*. 2018;36:310–318.
- Schaub FX, Jager R, Looser R, et al. Clonal analysis of deletions on chromosome 20q and *JAK2*-V617F in MPD suggests that *del20q* acts independently and is not one of the predisposing mutations for *JAK2*-V617F. Multicenter Study Research Support, Non-U.S. Gov't. *Blood*. 2009;113:2022–2027.
- Ortmann CA, Kent DG, Nangalia J, et al. Effect of mutation order on myeloproliferative neoplasms. *N Engl J Med*. 2015;372:601–612.
- Miles LA, Bowman RL, Merlinsky TR, et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature*. 2020;587:477–482.
- Morita K, Wang F, Jahn K, et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun*. 2020;11:5327.
- Maslah N, Verger E, Giraudier S, et al. Single-cell analysis reveals selection of TP53-mutated clones after MDM2 inhibition. *Blood Adv*. 2022;6:2813–2823.
- Kralovics R, Teo SS, Li S, et al. Acquisition of the V617F mutation of *JAK2* is a late genetic event in a subset of patients with myeloproliferative disorders. *Blood*. 2006;108:1377–1380.

21. Grinfeld J, Nangalia J, Baxter EJ, et al. Classification and personalized prognosis in myeloproliferative neoplasms. *N Engl J Med*. 2018;379:1416–1430.
22. Nam AS, Kim KT, Chaligne R, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature*. 2019;571:355–360.
23. Van Egeren D, Escabi J, Nguyen M, et al. Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms. *Cell Stem Cell*. 2021;28:514–523.e9.
24. Mylonas E, Yoshida K, Frick M, et al. Single-cell analysis based dissection of clonality in myelofibrosis. *Nat Commun*. 2020;11:73.
25. Picelli S, Bjorklund AK, Faridani OR, et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*. 2013;10:1096–1098.
26. Rodriguez-Meira A, Buck G, Clark SA, et al. Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol Cell*. 2019;73:1292–1305.e8.
27. Nangalia J, Nice FL, Wedge DC, et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica*. 2015;100:e438–e442.
28. Grinfeld J, Nangalia J, Green AR. Molecular determinants of pathogenesis and clinical phenotype in myeloproliferative neoplasms. *Haematologica*. 2017;102:7–17.
29. Guglielmelli P, Lasho TL, Rotunno G, et al. The number of prognostically detrimental mutations and prognosis in primary myelofibrosis: an international study of 797 patients. *Leukemia*. 2014;28:1804–1810.
30. Vannucchi AM, Lasho TL, Guglielmelli P, et al. Mutations and prognosis in primary myelofibrosis. *Leukemia*. 2013;27:1861–1869.
31. Tefferi A, Guglielmelli P, Lasho TL, et al. MIPSS70+ version 2.0: mutation and karyotype-enhanced international prognostic scoring system for primary myelofibrosis. *J Clin Oncol*. 2018;36:1769–1770.
32. Abdel-Wahab O, Adli M, LaFave LM, et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell*. 2012;22:180–193.
33. Shimizu T, Kubovcakov L, Nienhold R, et al. Loss of Ezh2 synergizes with JAK2-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis. *J Exp Med*. 2016;213:1479–1496.
34. Yang Y, Akada H, Nath D, et al. Loss of Ezh2 cooperates with Jak2V617F in the development of myelofibrosis in a mouse model of myeloproliferative neoplasm. *Blood*. 2016;127:3410–3423.
35. Nagase R, Inoue D, Pastore A, et al. Expression of mutant Asxl1 perturbs hematopoiesis and promotes susceptibility to leukemic transformation. *J Exp Med*. 2018;215:1729–1747.
36. McKenney AS, Lau AN, Somasundara AVH, et al. JAK2/IDH-mutant-driven myeloproliferative neoplasm is sensitive to combined targeted inhibition (vol 128, pg 789 2018). *J Clin Invest*. 2018;128:4743–4743.