Single cell long-read genotyping of transcripts reveals discrete mechanisms of clonal evolution in post-myeloproliferative neoplasm acute myeloid leukemia.

Julian Grabek[1,2*], Jasmin Straube[1,2*], Leanne Cooper[1], Rohit Haldar[1], Ranran Zhang[1,2], Inken Dulige[1,3], Matthew Barker[1,2], Will Gatehouse[1], Helen Christensen[1], Gerlinda Amor[1], Victoria Y. Ling[1,2,4], Caroline McNamara[2,5], David M. Ross[6,7], Andrew Perkins[8,9], Megan J. Bywater[1,2ϕ] & Steven W. Lane[1,2,5ϕ]

Affiliations:

[1]Cancer Research Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.
[2] The University of Queensland, St Lucia, Brisbane, QLD, Australia.
[3]University of Lübeck, Lübeck, Schleswig-Holstein, Germany.
[4]Department of Haematology, Princess Alexandra Hospital, Brisbane, Australia.
[5]Cancer Care Services, Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia.
[6]Department of Haematology, Royal Adelaide Hospital, Adelaide, SA, Australia.
[7]Acute Leukaemia Laboratory, Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, SA, Australia.
[8]Alfred Hospital, Melbourne, Victoria, Australia.
[9]Australian Centre for Blood Diseases, Monash University, Melbourne, Victoria, Australia.

*these authors contributed equally to the study

ϕco-corresponding authors:

megan.bywater@qimrb.edu.au

steven.lane@qimrb.edu.au

Words 3977/4000

Abstract (207/250)

Myeloproliferative neoplasms (MPNs) are caused by acquired mutations in hematopoietic stem and progenitor cells (HSPCs). The acquisition of additional mutations like *TP53* and the overall mutational burden influence a patient's risk of disease progression toward lethal post-MPN acute myeloid leukemia (AML).

Recent technological advancements in linking single-cell gene expression with genotype have improved our understanding of tumor heterogeneity. However, current

methodologies have limitations in simultaneously genotyping low-expression genes (such as JAK2) alongside other pathogenic loci.

To address this, we developed a novel long read genotyping pipeline of cDNA transcriptomes called LOTR-Seq, which can genotype the full length of 30 expressed genes at once. Using LOTR-Seq, we genotyped HSPCs at the *JAK2V617* locus in 9,075 single cells from eight patients with chronic phase MPN (CP-MPN) and in 5,016 cells from four patients with post-MPN AML. We then linked the mutations to the single cell transcriptome of 29,712 *JAK2V617F*-driven CP-MPN cells and 16,895 post-MPN AML cells.

In our analysis of post-MPN AMLs, we identified nine mutated loci across six genes (*JAK2, IDH1/2, TP53, SRSF2, U2AF1*) and linked these mutations to specific transcriptional phenotypes. Overall, LOTR-Seq has provided us with unparalleled insights into the pathogenesis of the *JAK2V617F* mutation in HSPCs in CP-MPN and the effects of secondary mutations in post-MPN AML.

Introduction

Hematopoietic stem cells (HSC) are the most primitive population of blood cells, with the unique ability to give rise to both mature myeloid and lymphoid lineages in addition to retaining the ability to self-renew. There are distinct sub-populations within the immunophenotypically defined hematopoietic stem and progenitor cell (HSPC) compartment that show transcriptional evidence of myeloid and lymphoid lineage priming[1] and variable self-renewal capacity[2]. Hematopoietic lineage priming can therefore be identified even in early stem cells prior to loss of self-renewal and full lineage commitment[2].

Myeloproliferative Neoplasms (MPN) are clonal blood cancers characterized by the overproduction of mature, functional myeloid elements. MPNs are caused by driver mutations within the HSC compartment[3], with the most common being *JAK2*V617F[4]. *JAK2*V617F gives rise to a clonal stem cell population[5] and results in pathology mediated by mature myeloid cells. Exactly how this clonal outgrowth effects the transcriptional diversity of the HSPC compartment is unclear.

In contrast to MPN, acute myeloid leukemias (AML) are characterized by the overproduction of immature myeloid cells. *JAK2*V617F mutations also characterize a molecular sub-type of acute myeloid leukemia (AML) that occurs subsequent to a diagnosis of an MPN. Genetic complexity is a common feature of *JAK2*V617F post-MPN AML, with many patients having multiple additional mutations identified at the time of leukemic transformation[6]. The process of disease progression from a chronic phase MPN to post-MPN AML is highly variable and presumed to be dictated by the acquisition of mutations in HSPCs additional to the MPN driver and subsequent clonal expansion[7,8].

We sought to understand the process that underlies disease progression from MPN to AML using a combined approach of single cell RNA-Seq to measure gene expression, together with a novel single cell pipeline to genotype a broad panel of transcripts expressed from loci commonly mutated in post-MPN AML (LOTR-Seq). Using this approach, we were able to identify clonal evolution and interrogate the impact of additional mutations within *JAK2* mutant HSPCs through quantification of the effect on gene expression, lineage priming and differentiation during the transformation to post-MPN AML.

Materials and methods

Ethics approval and Primary Human Sample processing

Eight chronic phase MPN and four post-MPN AML patient bone marrow (BM) samples were obtained from the Royal Brisbane Women's Hospital and from South Australian Cancer Research Biobank (SACRB) analyzed with approvals from QIMR Berghofer HREC (p1382). Patients provided written informed consent for sample collection, storage and use in research. Detailed patient information can be found in Supplemental Table 1. Patient BM samples were stained with CD3-PE/Cy7, CD34-PE, CD14-FITC and CD15-APC (Supplemental Table 2). CD34+ HSPCs were FACS sorted and single cell 3' RNA sequencing was performed with the 10x Chromium v3.1 followed by short read sequencing on the Illumina NextSeq 500/550 Platform. CD34+ BM bulk genomic DNA capture panel sequencing and mutation calling, was performed as described in Magor et al 2016[9].

Long read sequencing of transcripts: LOTR-seq

We first designed a capture panel on the Roche HyperDesign platform targeting all exonic regions of 30 recurrently mutated MPN/AML genes: *ASXL1*, *BCOR*, *CALR*, *CBL*, *CSF3R*, *DNMT3A*, *EZH2*, *FLT3*, *GATA2*, *IDH1*, *IDH2*, *JAK2*, *KIT*, *KRAS*, *MPL*, *MYC*, *NPM1*, *NRAS*, *PPM1D*, *PTPN11*, *RUNX1*, *SF3B1*, *SH2B3*, *SRSF2*, *STAG2*, *TET2*, *TP53*, *TERT*, *U2AF1* and *ZRSR2*. 10x barcoded cDNA underwent pre-capture PCR and read extension with custom primers (supplemental Table 3)[10]. Amplified and extended cDNA was used to capture genes of interest using the default the KAPA HyperCap Workflow v3 followed by post-capture PCR amplification. Further detailed information can be found in the supplemental methods. Oxford Nanopore Technology (ONT) ligation kits LSK110 or LSK114 were used in accordance with manufacturer's protocols with modifications to the incubation time from 10 minutes to 20 minutes[11] to improve adapter ligation and ONT Flow Cell loading amount (50-100 fmol, supplemental Table 4).

Single cell short read sequencing processing

MPN and post-MPN AML sequenced reads were adapter trimmed and processed through cellranger (v6.0.1, or MPN10 v3.0.2) with GRCh38 genome build to obtain a counts per

single cell and gene matrix. 10x single cell RNA-seq data from healthy human CD34+ BM HSPCs from Ainciburu et al. 2023[12] were downloaded from GEO (supplemental Table 5). Healthy, CP-MPN and post-MPN AML samples were preprocessed in R (v3.6.3) following the Seurat v4.3 pipline[13]. In brief, cells were retained if their mitochondrial RNA content was below 15% and they had more than 200 expressed genes. Each sample was normalized and the top 2000 variable features identified. For sample integration the top 2000 integration features were detected (`SelectIntegrationFeatures()`) followed by identification of top 2000 integration anchors (`FindIntegrationAnchors()`). The anchors were then used as input for the `IntegratedData function` with default parameters. Integrated data were cell cycle scored ('CellCycleScoring()') for G1, S and G2M phase and cell cycle was regressed using G1 and S phase. Integrated and cell cycle regressed data were used as input for dimension reduction using principal component analysis (PCA). Uniform Manifold Approximation and Projection (UMAP) dimensional reduction was run on the first 30 principal components with default parameters. Clustering was performed by shared nearest neighbor graph construction (`FindNeighbors`) on the first two UMAP components, followed by optimizing the modularity function (`FindClusters`) with resolution 0.1. Clusters were annotated according to key lineage marker gene expression. For analyzing lineage skewing between healthy and chronic phase and its correlation with mutation fraction within the compartments, clusters were broadly grouped by adding up clusters into HSC/LMPP (HSC+LMPP), Mega/Ery-Primed (MEP+MkP+e-Ery+l-Ery+Ery), Gran/Mono-Primed (GMP+Mono+Baso+DC) and Lymphoid-Primed (CLP+B-primed). Single cell gene set enrichment analysis (ssgsea) was performed on normalized counts using R v4.4.1 and GSVA (v1.52.3) R package. The KEGG JAK-STAT signaling pathway was downloaded as part of the c2.all.v2025.1.Hs.symbols.gmt from the human molecular signature database (MSigDB) collection[14-16]. Pseudo time prediction and cell fate scoring was performed using Palantir (v1.3.0)[17] within python (v3.11.4).

ONT sequencing, read processing and mutational profiling analysis

Base calling and chimeric read separation

Raw fast5 signals were converted using ONTs pod5 python package (v0.1.21). Basecalling was performed with ONTs dorado basecaller (v0.7.1) with `--no-trim` option

using transformer models specific for the sample Ligation Kit and Platform detailed in supplemental Table 4. Chimeric read trimming was perform using porechop (v0.2.4; https://github.com/rrwick/Porechop) with custom adapter settings added to adapter.py code ` Adapter('PCR adapters 10x', start_sequence=('PCR_10x_start', 'CTACACGACGCTCTTCCGATCT'), end_sequence= ('PCR_10x_end', 'AGATCGGAAGAGCGTCGTGTAG')) and settings `--extra_middle_trim_bad_side 0 --min_split_read_size 200 --extra_middle_trim_good_side -25 --extra_end_trim -25`.

Barcode deconvolution and mapping

Reads were assigned a 10x barcode using FLAMES[18] allowing for 2 mismatches. Fastqs were split into individual barcode fastqs using zgrep to identify FLAMES assigned barcodes in fastq header line. Deconvoluted reads were mapped against GRCh38 genome build using minimap2 (v2.27)[19] with option `-L -ax splice`. Secondary, supplementary alignments in addition to unmapped reads were filtered using samtools (v1.20)[20] with `-F 2820` option. Since chimeric reads remained we only retained primary alignments without supplementary alignments using sambamba (v1.0.1) “-F [SA]==null" to reduce false positive/negative rates.

Mutation calling

For single nucleotide variants (SNVs) calling we used bcftools (v1.19) mpileup function on processed bams to count bases at gene loci of interest (supplemental table 6 and 7). Insertion/Deletion (INDEL) calling was performed only over known mutation regions using samtools (v1.20) mpileup with `-d 0` option and VarScan2 (v2.4.2)[21] mpileup2indel with options` --min-coverage 0 --min-reads2 0 --min-avg-qual 0 --output-vcf 1 --variants 1`. Expressed variant allele frequency (eVAF) was calculated in R as the ratio of the number of reads with bases matching the mutant allele vs the number of total reads at the loci. Finally, eVAFs were assigned to a single cell transcriptome by matching barcodes recovered by long read and short read sequencing.

Post-MPN AML differentiation stage assessment

For raw counts of bulk RNA-Seq data from 45 CP MPN, 34 post-MPN AML and 11 healthy controls were downloaded from GEO with GSE283710[22]. Single cell RNA-Seq data were

downloaded from GEO with GSE226340[8]. *JAK2V617F TP53* post-MPN AML (n=11) and healthy controls (n=5) were extracted and raw counts pseudo-bulked. Both data were normalized to counts per million and tmm using methods implemented in the edgeR R package (v 4.2.2) followed by log2 transformation. Differentiation stage scoring was performed as described in Zheng et al 2025[23]. AML differentiation scores from bulked and pseudo-bulked RNASeq data were centered and patient-specific scores Euclidean distance underwent hierarchical clustering using hclust function stats R base package with default parameters. Data were visualized with the heatmap3.R package adopted from https://github.com/obigriffith/biostar-tutorials/blob/master/Heatmaps/heatmap.3.R. Single cell RNA-Seq raw counts from 11 post-MPN AML *JAK2V617F TP53* bi-allelic loss[8] and our post-MPN AML samples were projected on the Bone marrow map as described in Zeng et al 2025[23] https://github.com/andygxzeng/BoneMarrowMap (accessed 27/04/2025).

Results

HSPC heterogeneity is preserved in chronic phase MPN but with lineage skewing

We first sought to determine whether the expression of *JAK2*V617F alters the transcriptional heterogeneity of the CD34+ HSPC compartment. To do this we compared the single cell transcriptional landscape of CD34+ HSPC enriched populations (Figure 1A) from normal human bone marrow samples[12] (supplemental Figure 1A; n=6) to that of CD34+ HSPCs from eight patients with *JAK2*V617F-driven CP-MPN (Figure 1B). Transcriptional heterogeneity was largely preserved in CP-MPN CD34+ cell populations, comprising HSC/LMPPs, megakaryocyte-erythroid-, lymphoid- or granulocyte-monocyte-primed populations and consistent with that seen in the healthy CD34+ populations (Figure 1B-C; supplemental Figure 1B). However, CP-MPN CD34+ populations demonstrated a higher percentage of cells exhibiting megakaryocyte/erythroid-primed gene expression profiles (Figure 1D-E: Mega/Ery-Primed mean±SD: 22.5±6.9% healthy vs 34.1±10.3% CP-MPN, p=0.028; supplemental Figure 1C).

These findings demonstrate that, consistent with healthy CD34+ HSPCs, transcriptional lineage-priming is present in primitive hematopoietic cell populations in CP-MPN; however, CP-MPN exhibits an HSPC compartment with a higher abundance of cells with a megakaryocyte/erythroid transcriptional phenotype, consistent with the pathology of CP-MPN, characterized by the overproduction of mature megakaryocytes and erythrocytes.

MEP lineage priming in HSPCs is linked to the expansion of *JAK2*V617F mutant cells

We were interested to determine how clonal heterogeneity within CP-MPN altered this lineage bias within the CD34+ HSPC population. To identify *JAK2*V617F-mutated cells within CP-MPN CD34+ HSPCs, we developed a technique for long read genotyping of transcriptomes (LOTR-Seq). LOTR-Seq uses the single cell barcoded cDNA derived by

the 10x Chromium platform to capture the cDNA of target genes, like *JAK2*, followed by ONT long read sequencing (Figure 2A). Barcode deconvolution can then be performed to correlate expressed mutational information to the transcriptome.

When LOTR-Seq was performed on the CP-MPN CD34+ HSPCs, we were able to demonstrate that *JAK2*V617F pseudo-bulked eVAF strongly correlated with both genomic VAF (r=0.9; p=0.005; supplemental Figure 2A) and short read derived eVAF (r=0.96; p<000.1; supplemental Figure 2B) of the CD34+ compartment, validating the reproducibility of mutation calling from long read sequencing. Indeed, LOTR-Seq enabled us to genotype up to 51% of cells at the *JAK2*V617 locus, improving upon previous techniques[24,25]. Notably, the percentage of genotyped cells at the *JAK2*V617 locus correlated with sequencing depth (r=0.8, p=0.0006; supplemental Figure 2C).

We then assessed the contribution of *JAK2*V617F-mutated cells to the CD34+ HSPC compartment using the density distribution of *JAK2*V617F-mutated and *JAK2*-wild type cells within the clusters identified by uniform manifold approximation and projection (UMAP) (Figure 2B). In CP-MPNs *JAK2*V617F-mutated cells were most abundant in the Mega/Ery-primed CD34+ compartment. Furthermore, the percentage of *JAK2*V617F mutated HSPCs within the Mega/Ery-primed CD34+ compartment correlated with the relative size of the Mega/Ery-primed compartment (R=0.75; p=0.031; Figure 2C) and inversely correlated with that of the Gran/Mono-primed compartment (R=-0.79; p=0.02). These findings suggest that *JAK2*V617F either skews lineage bias towards the Mega/Ery fate or confers a proliferative advantage exclusively to Mega/Ery-primed HSPCs. To resolve this, we used Palantir analysis of differentiation in pseudotime, to examine the effect of *JAK2*V617F on HSPC on cell fate potential[17]. Here, HSPCs containing the *JAK2*V617F mutation show a higher probability of converging towards an erythroid cell fate, evidenced by higher average erythroid cell fate score compared to both healthy CD34+ and JAK2 wild type cells (Figure 2D-E). This supports the finding that *JAK2*V617F favors erythroid differentiation. Furthermore, both *JAK2V617F*-mutated and *JAK2*-wild type mega-erythroid primed HSPCs have a significantly lower proportion of cells in G1, and higher proportions in S phase of cell cycle, however only *JAK2V617F* mutant cells

contained a greater proportion of cells in G2M phase (Figure 2F, supplemental Figure 2D).

Finally, we sought to determine whether the *JAK2*V617F mutation results in altered JAK-STAT signaling within MPN CD34+ cells. Compared to healthy CD34+, CP-MPN CD34+ demonstrated increased JAK-STAT signaling across all compartments (Figure 2G, supplemental 2E) in both *JAK2*V617F, *JAK2*-wild type cells, suggesting that the activation of JAK-STAT signaling within HSPC is also regulated by cell-extrinsic factors, such as increased activation of cytokine pathways[26].

These data validate the single-cell genotyping approach to identify *JAK2*V617F mutant vs *JAK2* WT HSPCs within individual CP-MPN samples. *JAK2*V617F mutated cells demonstrate increased proliferation and preferential expansion towards an erythroid lineage.

Transformation to post-MPN AML is characterized by the loss of stem cell transcriptional heterogeneity driven by genetic complexity

We next sought to determine whether the LOTR-Seq pipeline could be used to investigate the relationship between genetic and transcriptional heterogeneity within the HSPC compartment during progression from CP-MPN to post-MPN AML. In dramatic contrast to the transcriptional heterogeneity observed in CP-MPN, the CD34+ HSPC compartment in post-MPN AML was largely dominated by a single transcriptional profile (Figure 3A-B). For example, MPN10 AML exhibits a dominant HSC/LMPP-like transcriptional signature, representing 79.4% of all cells (Figure 3A), while MPN20 AML shows a dominant Mega/Ery primed-like transcriptional signature (84.8%; Figure 3B).

A limitation of existing single cell RNA-Seq and genotyping protocols is the inability to simultaneously genotype across the numerous loci that are recurrently mutated in post-MPN AML with sufficient coverage and representation to draw inferences on cells with multiple mutations within the CD34+HSPC compartment[25]. The LOTR-Seq pipeline overcomes this limitation by using potent enrichment for a panel of 30 genes, enabling the genotyping of expressed genes across a range from 0% to 100% (supplemental

Figure 3A). Notably, 19 genes show a higher percentage of genotyped cells as the sequencing depth is increased. We therefore sought to apply this technology to determine the relationship between the genomic architecture of post-MPN AML and the transcriptional consequences linked with disease progression.

In MPN10, where we identified a dominant, HSC/LMPP-like expanded population in post-MPN AML, LOTR-Seq enabled us to identify the MPN-driver *JAK2*V617F mutation together with an *IDH2*R140W mutation and *SRSF2*P95 in frame deletion (Figure 3A,C-E, supplemental Figure 3B). These were identified in 56.7%, 92.4 %, 83.4% of genotyped cells, respectively, and expressed at high VAF in the HSC/LMPP-like CD34+ cells.

In the second post-MPN AML, MPN20, where we identified a dominant Mega/Ery primed-like HSPC population, this expanded population contains both *JAK2*V617F (89.6% of genotyped cells) and a biallelic *TP53*C238F mutation (93.9% of genotyped cells) (Figure 3B, F-G, supplemental Figure 3C). This identification of a Mega/Ery-signature in *JAK2*V617F multihit-*TP53* post-MPN AML is consistent with data from other published human samples[8] and from murine models of post-MPN AML with this same combination of mutations[27,28].

Altogether, these data validate the use of LOTR-Seq in clonal analysis of human post-MPN AML and demonstrate that post-MPN AMLs exhibit loss of HSPC heterogeneity with the expansion of a dominant transcriptional cluster that appears to be driven by the identity of co-existing secondary mutations.

Post-MPN AML demonstrates lineage restricted, recurrent differentiation priming

Recent findings suggest that molecular subtypes of de novo AML are defined by unique transcriptional profiles that belie their lineage bias[23,29,30]. Our initial analysis of the HSPC compartment in the two MPN patients following transformation to AML suggests that post-MPN AML also exhibits lineage priming analogous to that seen in *de novo* AML[23,30]. In order to determine if these findings can be applied more generally to post-MPN AML, we analyzed bulk RNA-Seq of 34 CD34+ post-MPN AML samples[22] together with single cell RNA-Seq CD34+ HSPCs of a further 11 post-MPN AML cases[8]. Here we adopted

methods developed by Zeng et al 2025[23] to assess differentiation stages in post-MPN AML. Both methods of projecting AML on a detailed, healthy bone marrow hierarchical hematopoiesis map[23] (Figure 4A-B) and AML differentiation stage scoring (Figure 4C) suggest that even genetically homogenous subgroups, such as *JAK2V617F* with multihit *TP53*, can exhibit expansion in different lineage-primed compartments. In the 45 post-MPN AML samples, 51% showed an immature differentiation stage (HSC/MPP/MEP), with HSC/LMPP/early lymphoid stage or MEP/MKP/erythroid stage observed in 31% and 18%, respectively. Compared to adult de novo AML, as reported by Zeng et al 2025[23] (n=28 of 139), post-MPN AML exhibited a significantly higher overrepresentation of HSC/MPP states (n=23 of 45, Chi-Square p=0.007). Striking, there was an absence of a promonocytic/monocytic biased subgroup in *JAK2*V617F mutated post-MPN AML, a dominant subgroup in de novo AML[23] (Figure 4C). These data demonstrate that post-MPN AML is not only genetically distinct from de novo AML, but also phenotypically and transcriptionally distinct with a more restricted differentiation repertoire.

High-risk mutations in post-MPN AML are already present in chronic phase disease.

Post-MPN AML is intrinsically chemotherapy resistant and novel strategies are urgently required to prevent the progression of high-risk CP-MPN to AML. We therefore sought to determine whether LOTR-Seq could identify pre-leukemic high-risk populations within CD34+ cells from patients with CP-MPN. To do this, we analysed two patients with matched CP-MPN and post-MPN AML samples. In the first patient's samples (Figure 5A-F), we identified the *JAK2*V617F mutation in 73.5% of *JAK2* genotyped cells in CP-MPN and 72.2% in post-MPN AML (Figure 5B, E). Additionally, deep sequencing analysis of the CP-MPN CD34+ cells identified a rare subpopulation of co-mutated *JAK2*V617F and *IDH1*R132S cells, present in 8.2% of *IDH1*-genotyped cells in CP-MPN, which expanded to 64.5% in post-MPN AML (Figure 5C, F). In contrast to what was observed in the analysis of the previous post-MPN AML samples (Figure 4), the expansion of the double mutated population did not result in the complete abrogation of transcriptional heterogeneity, however it resulted in relative expansion of the HSC/LMPP clusters with partially maintained representation of the more committed lineage-primed cells.

The co-occurrence of *JAK2*V617F and *IDH1*R132S clones in chronic phase with expansion of the double mutated clone during AML progression suggests that this leukemia was driven by the linear clonal evolution of the double mutant cells (Figure 5G). Thus, the LOTR-Seq protocol can identify genetic complexity at the single cell level, enabling the identification of pre-leukemic clones that later give rise to post-MPN AML.

It has been observed that in 16-50% of post-MPN AML cases, the MPN driver is absent at the time of leukemia transformation, even when the original disease contained a mutation in a canonical MPN driver gene [31,32,6]. The origin of this MPN-driver negative post-MPN AML is incompletely understood. We reasoned that this category of post-MPN AML must have a different mechanism of transformation compared to the classical linear clonal evolution model.

To examine this further, we used LOTR-Seq to analyse paired, sequential samples from a patient who presented with *JAK2*V617F CP-MPN and *JAK2*V617F-negative post-MPN AML (Figure 6 A-H, supplemental Figure 4A-D). Here, the CP-MPN contained a dominant clone with *TP53*Y236C mutations in addition to a *JAK2L393V* mutation, previously described to result in gain of function[33] (Figure 6C, supplemental Figure 4B). Within the *TP53*Y236C, *JAK2L393V* population, there was further heterogeneity with two identifiable sub-clonal populations containing either *TP53*R282W (Figure 6D) or *U2AF1S34Y* and *JAK2*V617F (supplemental Figure 4C; Figure 6B). With evolution to AML, the *TP53*Y236C, *JAK2L393V TP53*R282W containing cells demonstrated selective clonal advantage and was dominant in the post-MPN AML stage (Figure 6H). Both *TP53* mutations identified are located within the DNA-binding domain (supplemental Figure 5A), and thus are predicted to confer a loss of function phenotype. Importantly, we were also able to demonstrate that the *TP53* mutations occurred on independent alleles in the same cell and exhibited random monoallelic expression (i.e. somatically acquired in trans; supplemental Figure 5B-C). These findings identify leukemic transformation that arose from a genetically complex, high-risk ancestral clone and demonstrates parallel clonal evolution resulting in *JAK2*V617F negative AML (Figure 5I; supplemental Figure 4E).

In summary, using this novel approach to combine single cell transcriptional analysis and panel-enriched genotyping of transcripts from loci commonly associated with myeloid

malignancies, we have been able to demonstrate that the transformation of CP-MPN to post-MPN AML is associated with the acquisition of recurrent secondary mutations that drive loss of transcriptional heterogeneity and the immortalization of a genotypically defined, transformed progenitor cell population.

Discussion

Secondary AML arising from antecedent MPN is associated with 2 or more genetic mutations in the vast majority of patients[6]. It is well established within the field, that post-MPN AML can arise from CP-MPN and is driven by these additional mutations. This, combined with the knowledge that CP-MPN arises from, and is maintained by, HSCs compels us to examine how these high-risk genetic lesions can arise within CP-MPN and lead to leukemia. Single cell RNA-sequencing approaches have been developed that provide a high-resolution map of transcriptional changes within cancers and have been able to unveil numerous insights into disease biology. Until now, these single cell approaches have had limited capacity to genotype individual cells across the spectrum of genetic loci that are recurrently mutated in post-MPN AML[24,25,34].

Here, we describe a novel pipeline combining 10x single cell RNA-sequencing of CP-MPN and post-MPN AML, together with panel-based enrichment and long read genotyping at the vast majority of genetic loci frequently mutated in MPN. We have demonstrated that individual, high risk cells can be found within CP-MPN, and that these cells have the capacity to expand and become clonally dominant *in vivo*, giving rise to post-MPN AML. The utility of this approach is broad and clinically useful. Firstly, LOTR-Seq can be used as a monitoring tool for high-risk patients with CP-MPN, to evaluate for clonal evolution during treatment, or to study the interaction between treatment and subclones in order to select the most appropriate treatments. Eventually, it is hoped that the identification of genetic susceptibilities using single cell analysis may be leveraged into new treatments that can prevent the development of post-MPN AML. Secondly, this is the only technology that is able to determine conclusively whether *TP53* mutations are monoallelic or biallelic in MPN, as this is highly predictive of prognosis[35]. The impact of *TP53* mutations in human disease confers a very poor prognosis in secondary AML and

is associated with chemotherapy resistance[6,8,28]. Biologically, we noted that the presence of a *TP53* mutation within the *JAK2V617F* mutant cell showed a dominant transcriptional signature of a megakaryocyte-erythroid primed cell. There was distinct loss of heterogeneity in the HSPC compartment in addition to the emergence of a MEP-like transcriptional signature, consistent with recent descriptions of MPN transforming to secondary AML in humans[8] and corroborating the findings demonstrated in mouse models[27,28].

These data can be used to inform or validate other laboratory models of MPN and post-MPN AML. For example, the presence of *IDH1*/2 mutations in CP-MPN has previously been identified as conferring high risk of progression to secondary AML[36,37]. Through demonstrating the presence of the *IDH* mutation in the *JAK2V617F* cells, we showed that this genotype was associated with an LMPP-like transcriptional signature in primary human disease. A mouse model of combined *Jak2V617F* mutant with *Idh2R140Q* mutation demonstrated expansion of the lineage-negative LSK cell population (equivalent to the stem and multipotent cell compartment in humans). Whilst the MPN disease was maintained through the LT-HSC compartment in these mouse models through serial transplantation, the double mutant mice showed the greatest expansion of the MPP fraction of the LSK compartment[38].

The presence of *JAK2V617F* negative secondary AML arising from a *JAK2V617F* mutant MPN shows a varied incidence among secondary AML with the reported literature showing rates from 16% up to 50%[31,32]. Although it has been proposed that *JAK2V617F* can be lost during this process[31], our data shows the presence of *JAK2V617F* negative *TP53* mutant clone in chronic phase disease suggesting transformation occurs from an antecedent or independent ancestral clone. In this example, the presence of an independent kinase activating mutation in JAK2 has clear therapeutic implications.

In aggregate, the potential for this approach is its broader applicability to a diagnostic and therapeutic setting for clonal identification and tracking in response to treatment.

Acknowledgement

We are grateful for the assistance of the QIMR Berghofer, flow cytometry and sequencing facilities and helpful comments from members of the Lane Lab including Stacey Anderson. We thank Rachel Thjissen for fruitful discussion on ONT barcode recovery and primer sequences and ONT read extension protocols. We gratefully acknowledge the support of Neil Herron and the Herron Family Trust, and the Gordon and Jessie Gilmour Family Trust as dedicated supporters of leukaemia research in Queensland. We gratefully acknowledge the contribution of the MPN Alliance Australia for the valuable patient perspective they provided for this study. SWL was funded by an NHMRC Investigator Grant (1195987) and by a philanthropic donation from the MPN Alliance of Australia. ID was supported by a PROMOS, DAAD stipend. JS was funded by Cancer Council Queensland fellowship (2025829). JG was supported by a HSANZ/Leukaemia Foundation PhD scholarship. This research was supported by the Australian Cancer Research Foundation Centre for Optimised Cancer Therapy.

Authorship

SWL, MJB, and JS conceptualized, designed and supervised, wet lab experiments and bioinformatics analysis. JG, LC, RH, RZ, SA performed wet lab experiments. JS, JG, ID and MB performed bioinformatical analysis. CM, DR and SWL coordinated primary patient sample collection. GA, HC, WG coordinated patient sample information collection, consenting and storage. VYL helped with gene panel design. AP performed genomic DNA NGS analysis of primary patient samples. JG, JS, MJB and SWL wrote the manuscript. All authors contributed to and edited the manuscript.

Conflict of Interest Disclosure

SWL, MJB and JS have received research funding from Bristol Myers Squibb for unrelated projects. We acknowledge receipt of reagents from Pharmaessentia for unrelated projects. MJB has received research funding from Cylene Pharmaceuticals. DMR has consulted for or received honoraria from GSK, Jubilant, Keros, Merck, Menarini, Novartis, Prelude, and Takeda, for unrelated projects. SWL has consulted for Abbvie, Novartis, Astellas and GSK, also for unrelated projects.

Data Availability Statement

The 10x sequencing raw data in this study will be deposited in the NCBI Gene Expression Omnibus prior to peer-reviewed publication. Processed data will be made available via figshare/zendo. Code is available via github https://github.com/JStrau/LOTRSeq. ONT raw data due to its size will be made available up on request.

References

1. Challen GA, Boles NC, Chambers SM, Goodell MA. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell stem cell*. Mar 5 2010;6(3):265-78. doi:10.1016/j.stem.2010.02.002
2. Yamamoto R, Morita Y, Ooehara J, et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*. Aug 29 2013;154(5):1112-1126. doi:10.1016/j.cell.2013.08.007
3. Jamieson CH, Gotlib J, Durocher JA, et al. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation. *Proceedings of the National Academy of Sciences of the United States of America*. Apr 18 2006;103(16):6224-9. doi:10.1073/pnas.0601462103
4. Prins D, González Arias C, Klampfl T, Grinfeld J, Green AR. Mutant Calreticulin in the Myeloproliferative Neoplasms. *HemaSphere*. Feb 2020;4(1):e333. doi:10.1097/hs9.0000000000000333
5. Beer PA, Delhommeau F, LeCouédic JP, et al. Two routes to leukemic transformation after a JAK2 mutation-positive myeloproliferative neoplasm. *Blood*. Apr 8 2010;115(14):2891-900. doi:10.1182/blood-2009-08-236596
6. McNamara CJ, Panzarella T, Kennedy JA, et al. The mutational landscape of accelerated- and blast-phase myeloproliferative neoplasms impacts patient outcomes. *Blood Adv*. Oct 23 2018;2(20):2658-2671. doi:10.1182/bloodadvances.2018021469
7. Calabresi L, Carretta C, Romagnoli S, et al. Clonal dynamics and copy number variants by single-cell analysis in leukemic evolution of myeloproliferative neoplasms. *American journal of hematology*. Oct 2023;98(10):1520-1531. doi:10.1002/ajh.27013
8. Rodriguez-Meira A, Norfo R, Wen S, et al. Single-cell multi-omics identifies chronic inflammation as a driver of TP53-mutant leukemic evolution. *Nature genetics*. Sep 2023;55(9):1531-1541. doi:10.1038/s41588-023-01480-1
9. Magor GW, Tallack MR, Klose NM, et al. Rapid Molecular Profiling of Myeloproliferative Neoplasms Using Targeted Exon Resequencing of 86 Genes Involved in JAK-STAT Signaling and Epigenetic Regulation. *J Mol Diagn*. Sep 2016;18(5):707-718. doi:10.1016/j.jmoldx.2016.05.006
10. Thijssen R, Tian L, Anderson MA, et al. Single-cell multiomics reveal the scale of multilayered adaptations enabling CLL relapse during venetoclax therapy. *Blood*. Nov 17 2022;140(20):2127-2141. doi:10.1182/blood.2022016040
11. Wei S, Williams Z. Rapid Short-Read Sequencing and Aneuploidy Detection Using MinION Nanopore Technology. *Genetics*. Jan 2016;202(1):37-44. doi:10.1534/genetics.115.182311
12. Ainciburu M, Ezponda T, Berastegui N, et al. Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. *Elife*. Jan 11 2023;12doi:10.7554/eLife.79363

13. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021/06/24/ 2021;184(13):3573-3587.e29. doi:https://doi.org/10.1016/j.cell.2021.04.048

14. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. Oct 25 2005;102(43):15545-50. doi:10.1073/pnas.0506580102

15. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. Dec 23 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004

16. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260

17. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir. *Nature biotechnology*. Apr 2019;37(4):451-460. doi:10.1038/s41587-019-0068-4

18. Tian L, Jabbari JS, Thijssen R, et al. Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. *Genome biology*. Nov 11 2021;22(1):310. doi:10.1186/s13059-021-02525-6

19. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191

20. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. Feb 2021;10(2)doi:ARTN giab008

10.1093/gigascience/giab008

21. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. Mar 2012;22(3):568-76. doi:10.1101/gr.129684.111

22. Kong T, Laranjeira ABA, Letson CT, et al. RSK1 is an exploitable dependency in myeloproliferative neoplasms and secondary acute myeloid leukemia. *Nat Commun*. Jan 16 2025;16(1):492. doi:10.1038/s41467-024-55643-7

23. Zeng AGX, Iacobucci I, Shah S, et al. Single-cell Transcriptional Atlas of Human Hematopoiesis Reveals Genetic and Hierarchy-Based Determinants of Aberrant AML Differentiation. *Blood Cancer Discov*. Jul 1 2025;6(4):307-324. doi:10.1158/2643-3230.BCD-24-0342

24. 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*. Mar 4 2021;28(3):514-523.e9. doi:10.1016/j.stem.2021.02.001

25. Nam AS, Kim KT, Chaligne R, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature*. Jul 2019;571(7765):355-360. doi:10.1038/s41586-019-1367-0

26. Austin RJ, Straube J, Bruedigam C, et al. Distinct effects of ruxolitinib and interferon-alpha on murine JAK2V617F myeloproliferative neoplasm hematopoietic stem cell populations. *Leukemia*. Apr 2020;34(4):1075-1089. doi:10.1038/s41375-019-0638-y

27. Li B, An W, Wang H, et al. BMP2/SMAD pathway activation in JAK2/p53-mutant megakaryocyte/erythroid progenitors promotes leukemic transformation. *Blood*. Jun 23 2022;139(25):3630-3646. doi:10.1182/blood.2021014465

28. Rampal R, Ahn J, Abdel-Wahab O, et al. Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. *Proceedings of the National Academy of Sciences of the United States of America*. Dec 16 2014;111(50):E5401-10. doi:10.1073/pnas.1407792111

29. Zeng AGX, Bansal S, Jin L, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat Med*. Jun 2022;28(6):1212-1223. doi:10.1038/s41591-022-01819-x

30. van Galen P, Hovestadt V, Wadsworth Ii MH, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*. Mar 7 2019;176(6):1265-1281 e24. doi:10.1016/j.cell.2019.01.031

31. Venton G, Courtier F, Charbonnier A, et al. Impact of gene mutations on treatment response and prognosis of acute myeloid leukemia secondary to myeloproliferative neoplasms. *American journal of hematology*. Mar 2018;93(3):330-338. doi:10.1002/ajh.24973

32. Theocharides A, Boissinot M, Girodon F, et al. Leukemic blasts in transformed JAK2-V617F-positive myeloproliferative disorders are frequently negative for the JAK2-V617F mutation. *Blood*. Jul 1 2007;110(1):375-9. doi:10.1182/blood-2006-12-062125

33. Lanikova L, Babosova O, Swierczek S, et al. Coexistence of gain-of-function JAK2 germ line mutations with JAK2V617F in polycythemia vera. *Blood*. Nov 3 2016;128(18):2266-2270. doi:10.1182/blood-2016-04-711283

34. 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*. Mar 21 2019;73(6):1292-1305 e8. doi:10.1016/j.molcel.2019.01.009

35. Tefferi A, Abdelmagid M, Loscocco GG, et al. TP53 Mutations in Myeloproliferative Neoplasms: Context-Dependent Evaluation of Prognostic Relevance. *American journal of hematology*. Apr 2025;100(4):552-560. doi:10.1002/ajh.27609

36. Pardanani A, Lasho TL, Finke CM, Mai M, McClure RF, Tefferi A. IDH1 and IDH2 mutation analysis in chronic- and blast-phase myeloproliferative neoplasms. *Leukemia*. Jun 2010;24(6):1146-51. doi:10.1038/leu.2010.77

37. Tefferi A, Lasho TL, Abdel-Wahab O, et al. IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis. *Leukemia*. Jul 2010;24(7):1302-9. doi:10.1038/leu.2010.113

38. McKenney AS, Lau AN, Somasundara AVH, et al. JAK2/IDH-mutant-driven myeloproliferative neoplasm is sensitive to combined targeted inhibition. *J Clin Invest*. Feb 1 2018;128(2):789-804. doi:10.1172/JCI94516

Figure legends

Figure 1. scRNAseq identifies expanded megakaryocyte-erythroid primed cells in chronic phase MPN. (A) Schematic of MPN bone marrow hematopoietic stem and progenitor cell (HSPC) enrichment followed by 10x Chromium single cell separation and 3’ cDNA barcoding. (B) Chronic phase MPN UMAPs (n=4, polycythemia vera, PV; n=1, essential thrombocythemia, ET; n=1 primary myelofibrosis, PMF; n=2 secondary myelofibrosis, sMF) of short read sequencing derived whole transcriptomics of HSPCs. Clusters annotated by gene expression associated with HSC, hematopoietic stem cells;

LMPP, lympho-myeloid multipotent progenitor; CLP, common lymphoid progenitors; GMP, granulocyte-monocyte progenitors; DC; Dendritic Cell progenitors; Baso, eosinophil/basophil/mast cells progenitors; MEP, megakaryocyte-erythroid progenitor; MkP, megakaryocyte progenitor, e-Ery; early erythroid and l-Ery, late erythroid priming. (C) Pooled healthy sample UMAPs (n=6) showing the expression of key lineage-defining transcriptional markers including HSC-CRHBP, LMPP-HOPX, B-primed-VPREB1, myeloid-MPO, and monocyte-LYZ, DC-SPIB, basophil-LMO4, erythroid-KLF1, megakaryocyte-ITGA2B. (D) Schematic of summarized clusters analysed in (E). (E) Bargraph with mean (black dot) and standard deviation (black line) comparing the percentage (%) of clustered cells in healthy (n=6) vs chronic phase (CP) MPN (n=8). Each grey or blue dot represents data from an individual sample. P-value derived from pairwise, two-sided Welch t-test.

Figure 2. Target enrichment and long read sequencing demonstrates the impact of mutational status on the gene transcription in chronic phase MPN. (A) Schematic showing the long read genotyping of transcriptomes (LOTR-Seq) pipeline that enriches single cell cDNA of 30 MPN/AML associated genes followed by long read sequencing and genotyping of transcripts, with an example of one MPN patient's UMAP colored by *JAK2V617F* expressed variant allele frequency (eVAF). (B) UMAPs of the CD34+ HSPC single cells (light grey) of n=8 MPNs with the density distribution of *JAK2V617F* mutation (red), *JAK2* wild type (blue) and non-genotyped cells (NA; grey). (C) Scatterplots of HSC/LMPP and key lineage-primed compartment percentage (%; x-axis) and the % of *JAK2V617F* genotyped cells (y-axis) within the compartment (n=8 chronic phase MPN). Pearson's correlation coefficient R and test of association p-value. The blue line represents the linear regression line with a 95%-confidence interval (grey band). (D) Smoothed regression lines of the Palantir predicted erythroid cell fate score (y-axis) in dependency of the pseudotime (x-axis) for n=5 healthy, n=8 CP-MPN *JAK2V617F* and *JAK2* wildtype cells. (E) Bargraph of the mean erythroid cell fate for each sample comparing n=5 healthy to n=8 CP-MPN *JAK2V617F* and *JAK2* wild type (*JAK2WT)* summarised in (D). Black dot and line represent the group mean and standard deviation, respectively. P-values derived from pairwise, two-sided Welch t-tests. (F) Bargraph of the percentage (%) of cell cycle phase (G1, S, G2M) for the

Mega/Ery-Primed compartment comparing n=6 healthy to n=8 CP-MPN *JAK2V617F* and *JAK2* wild type (*JAK2WT)*. Black dot and line represent the group mean and standard deviation, respectively. P-values derived from pairwise, two-sided Welch t-tests. (F) Bargraph of the KEGG JAK-STAT signaling pathway average single cell gene set enrichment score comparing n=6 healthy to n=8 MPN *JAK2V617F* and *JAK2* wild type (*JAK2WT*). Black dot and line represent the group mean and standard deviation, respectively. P-values derived from pairwise, two-sided Welch t-tests.

Figure 3. Loss of heterogeneity occurs in leukemic transformation and is driven by additional mutations in the *JAK2V617F* mutant clone. (A-B) UMAPs of post-MPN AML samples (A) MPN10 and (B) MPN20 showing transcriptionally defined cells clusters and a bargraph depicting the relative percentage of cells per cluster, (C-E) expressed variant allele frequencies (eVAF) of mutations (C) *JAK2V617F* (D) *IDH2R140W* and (E) *SRSF2P95del*. (F-G) UMAPs of post-MPN AML sample MPN20 showing expressed variant allele frequencies (eVAF) of mutations (F) *JAK2V617F* and (G) *TP53C238F*. Grey colored dots indicate non-genotyped cells.

Figure 4. Post-MPN AML exhibit lineage-biased hematopoiesis (A) Bone marrow hematopoietic stem cell atlas UMAP with cell type annotations as generated by Zeng et al. 2025. (B) UMAPs highlighting key lineage gene expression markers HSC-*CRHBP*, LMPP-*HOPX*, erythroid-*KLF1*, megakaryocyte-*ITGA2B*, Baso-*LMO4*, and B-primed-*VPREB1*. (C) Projection of our generated n=3 *JAK2V617F* post-MPN AML single cell samples and n=11 *JAK2V617F* post-MPN AMLs from Rodriguez-Meira et al. 2023. (D) Hierarchical clustering of lineage-defining gene expression scores (y-axis) of n= 11 *JAK2V617F* post-MPN AML and n=6 healthy samples from Rodriguez-Meira et al. 2023, n=45 chronic phase MPN, n=34 post-MPN AML and n=11 healthy Kong et al 2025 (x-axis).

Figure 5. Leukemic transformation from chronic phase MPN showing a linear pattern of clonal expansion of a high risk *JAK2V617F* and *IDH1R132S* mutant clone. UMAPs of a patient at (A-C) chronic phase MPN (SA2CP) and (D-E) post-MPN AML (SA2LT), showing (A,D) transcriptionally defined clusters with bargraphs showing the percentage of cells in the transcriptionally defined clusters and (B-C,E-F) expressed

variant allele frequencies (eVAF) of mutations (B,E) *JAK2V617F* and (C,F) *IDH1R132S*. Grey colored dots indicate non-genotyped cells. (G) Fish plot illustrating the linear clonal evolution of the *JAKV617F* and *IDH1R132S* mutant clone.

Figure 6. Leukemic transformation from chronic phase MPN showing a parallel clonal evolution of a *JAK2V617F-negative TP53-multihit* clone. UMAPs of a patient at (A-D) chronic phase MPN (SA1CP) and (E-H) post-MPN AML (SA1LT), showing (A,E) transcriptionally defined clusters and bargraphs with the percentage of cells in the clusters and (B-D,F-H) expressed variant allele frequencies (eVAF) of mutations (B,F) *JAK2V617F* and (C,G) *TP53Y236C* and (D,H) *TP53R282W*. Grey colored dots indicate non-genotyped cells. (I) Fish plot illustrating the parallel clonal evolution of the *JAK2V617F* and *TP53R282W* mutant clone.

Supplementary Figure 1. Lineage-defining gene expression markers in CD34+ HSPCs. (A) UMAPs of n=6 healthy CD34+ bone marrow single cells derived from short read sequencing whole transcriptomics (B) Dotpot of average percentage of lineage defining markers expressing cells in healthy samples (n=6, x-axis) within the annotated clusters (y-axis). (C) Bargraph of the relative cell cluster percentage (y-axis) of each analyzed sample (x-axis; n=6 healthy; n=8 chronic phase MPN).

Supplementary Figure 2. Scatter plots with linear regression lines of (A-C) Nanopore long read expressed variant allele frequency (eVAF; x-axis) compared to (A) the next-generation sequencing derived (NGS) genomic VAF (gVAF; y-axis) and (B) pseudobulked eVAF (y-axis) and, (C) Nanopore sequenced gigabases PASS (x-axis) and the percentage (%) of genotyped cells (y-axis). Statistics displayed are the Pearson's product-moment correlation coefficient r and test for association derived p-value. (D) Bargraph of the percentage of cell cycle phase (G1, S, G2M) for each compartment comparing n=6 healthy to n=8 MPN *JAK2V617F* and *JAK2* wild type (*JAK2WT)*. Black dot and line represent the group mean and standard deviation, respectively. P-values derived from pairwise, two-sided Welch t-tests. (E) Bargraph of the KEGG JAK-STAT signaling pathway average single cell gene set enrichment (ssgsea) score comparing n=6 healthy to n=8 MPN *JAK2V617F* and *JAK2* wild type

(*JAK2WT)*. Black dot and line represent the group mean and standard deviation, respectively. P-values derived from pairwise, two-sided Welch t-tests.

Supplementary Figure 3. (A) Table of the median percentage (%) of genotyped cells at MPN/AML mutated genes hotspot locations for n=10 MinION sequenced samples and n=6 PromethION sequenced samples. P-values derived from a pairwise, two-sided Welch t-test. (B-C) Euler diagrams depicting the number of mutated cells with genotyping information for all 2-3 loci by LOTR-Seq and their co-occurrence for samples (A) MPN10 and (B) MPN20.

Supplemental Figure 4. UMAPs of a patient at (A-B) chronic phase MPN (SA1CP) and (C-D) post-MPN AML (SA1LT), displaying expressed variant allele frequencies (eVAF) of mutations (A,C) *U2AF1S34Y* and (B,D) *JAK2L393V*. Grey colored dots indicate non-genotyped cells. (E) Hierarchical clustering using ward.D2 method of the Euclidean distance of eVAFs of mutated genes *JAK2L393V, TP53R282W, TP53Y236C, U2AF1S34Y and JAK2V617F* of SA1CP and SA1LT.

Supplementary Figure 5. (A) Lollipop plot extracted from cBioportal depicting the location of *TP53R282W* and *TP53Y236C* in the DNA binding domain of TP53. (B) IGV screen shot of the mapped ONT long reads from SA1LT of a single cell with *TP53R282W* and *TP53Y236C* mutations on independent reads. (C) Scatter and density plot of the expressed variant allele frequency (eVAF) of *TP53Y236C* (x-axis) and *TP53R282W* (y-axis). (D) SA1LT CD34+ HSPC single cell gene expression projected on Zeng et al. 2025 bone marrow map.

Supplementary Methods

Fluorescence Activated Cell Sorting (FACS) Flow Cytometry

Primary human samples were suspended in staining buffer (PBS with 2% FCS and Pen-Strep). Fluorochromes CD3-PE/Cy7, CD34-PE, CD14-FITC and CD15-APC (supplemental Table 2) were added to the sample at 1µL per $1x10^6$ cells, in order to separate T-cells, hematopoietic stem and progenitor cells, mature monocytes and mature myeloid cells, respectively.

Excess antibody was removed with washing and the sample was re-suspended in sorting buffer (PBS with 2% FCS, PS and 1:20,000 Sytox Blue). Cell Sorting was performed on the Aria Instrument (BD Sciences). Following cell sorting a cell stock with a concentration of 1000-1200 cells/µL was used for 10x Chromium Single Cell Encapsulation.

Pre and Post-capture PCR Amplification of cDNA

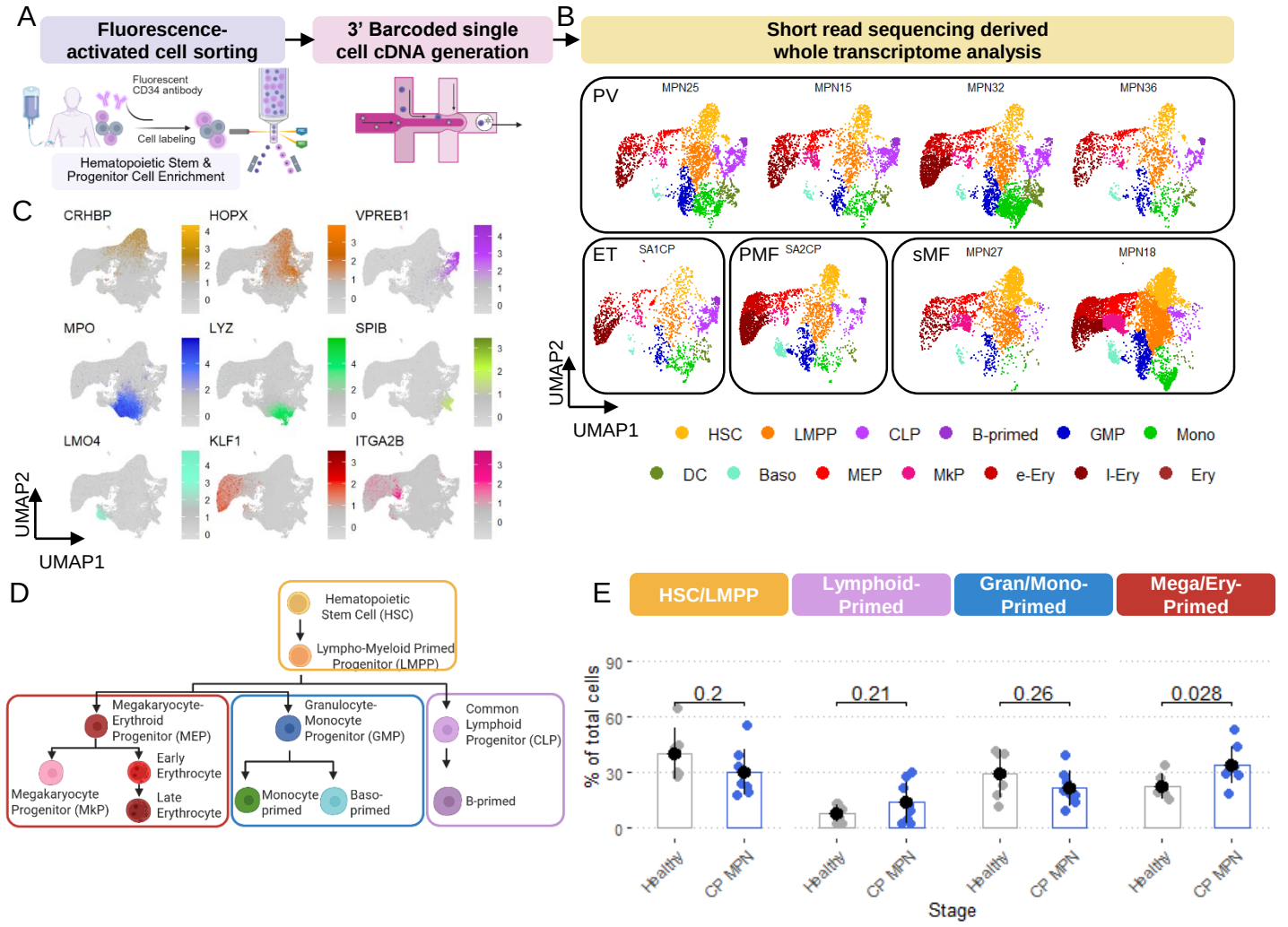
PCR amplification was performed to generate sufficient starting template for target enrichment (based on manufacturer recommendation of minimum of 1000ng). The PCR reaction mix contained 30µL of 10x barcoded cDNA sample, 10µL of partial TSO (5µM), 10µL of R1 primer (5µM) (supplemental Table 3) and 50µL of KAPA HotStart HiFi ReadyMix (Roche, Ref 07958927001).

Pre-caputre PCR was performed with the following incubation times of 98°C for 3 mins then the PCR cycle commences at 98°C for 20 secs followed by 63°C for 30 secs then 72°C for 4.5 mins. The cycle number varied based on starting amount with between 12-24 cycles required. The final extension step was 72°C for 3:00 mins.

The sample underwent AMPure Bead clean up at 1X with 80% Ethanol. The cDNA was eluted into PCR grade water and quantified by Qubit (Invitrogen, ThermoFisher Scientific).

The post capture PCR Master Mix contains 25µL of KAPA HiFi HotStart ReadyMix (2X) with 2.5 µL of 5µM R1 primer and 2.5µL of 5µM partial TSO primer (supplemental Table 3). The 30µL of master mix was added to the 20µL of bead bound cDNA from the post capture bead wash.

Figure 1


A
Fluorescence-activated cell sorting
3' Barcoded single cell cDNA generation
B
Short read sequencing derived whole transcriptome analysis
Fluorescent CD34 antibody
Cell labeling
Hematopoietic Stem & Progenitor Cell Enrichment
PV
MPN25
MPN15
MPN32
MPN36
ET
SA1CP
PMF
SA2CP
sMF
MPN27
MPN18
UMAP2
UMAP1
HSC
LMPP
CLP
B-primed
GMP
Mono
DC
Baso
MEP
MkP
e-Ery
l-Ery
Ery
C
CRHBP
HOPX
VPREB1
MPO
LYZ
SPIB
LMO4
KLF1
ITGA2B
UMAP2
UMAP1
D
Hematopoietic Stem Cell (HSC)
Lympho-Myeloid Primed Progenitor (LMPP)
Megakaryocyte-Erythroid Progenitor (MEP)
Early Erythrocyte
Megakaryocyte Progenitor (MkP)
Late Erythrocyte
Granulocyte-Monocyte Progenitor (GMP)
Monocyte primed
Baso-primed
Common Lymphoid Progenitor (CLP)
B-primed
E
HSC/LMPP
Lymphoid-Primed
Gran/Mono-Primed
Mega/Ery-Primed
% of total cells
0.2
0.21
0.26
0.028
Healthy
CP MPN
Stage

Supplemental Figure 1

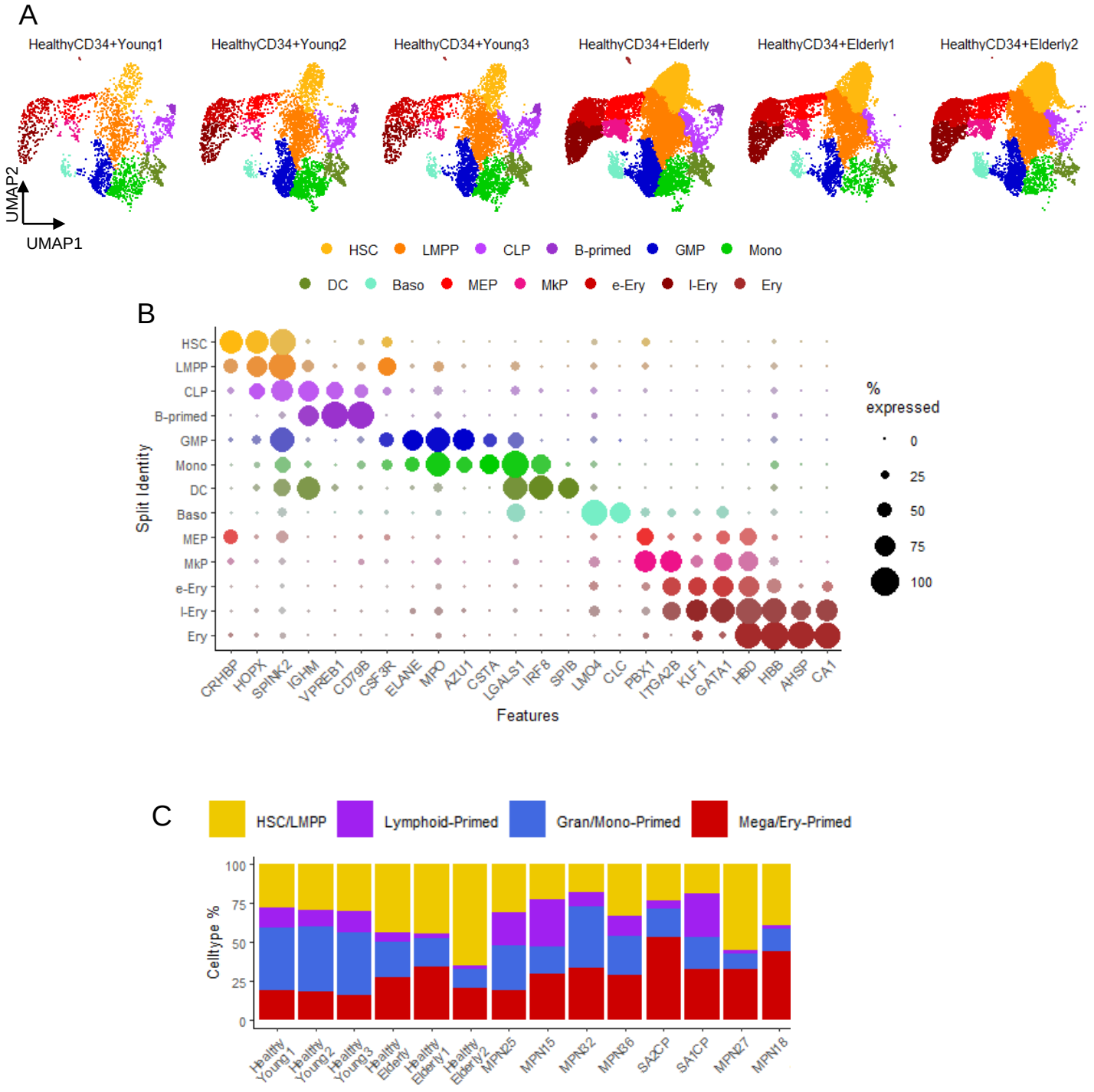
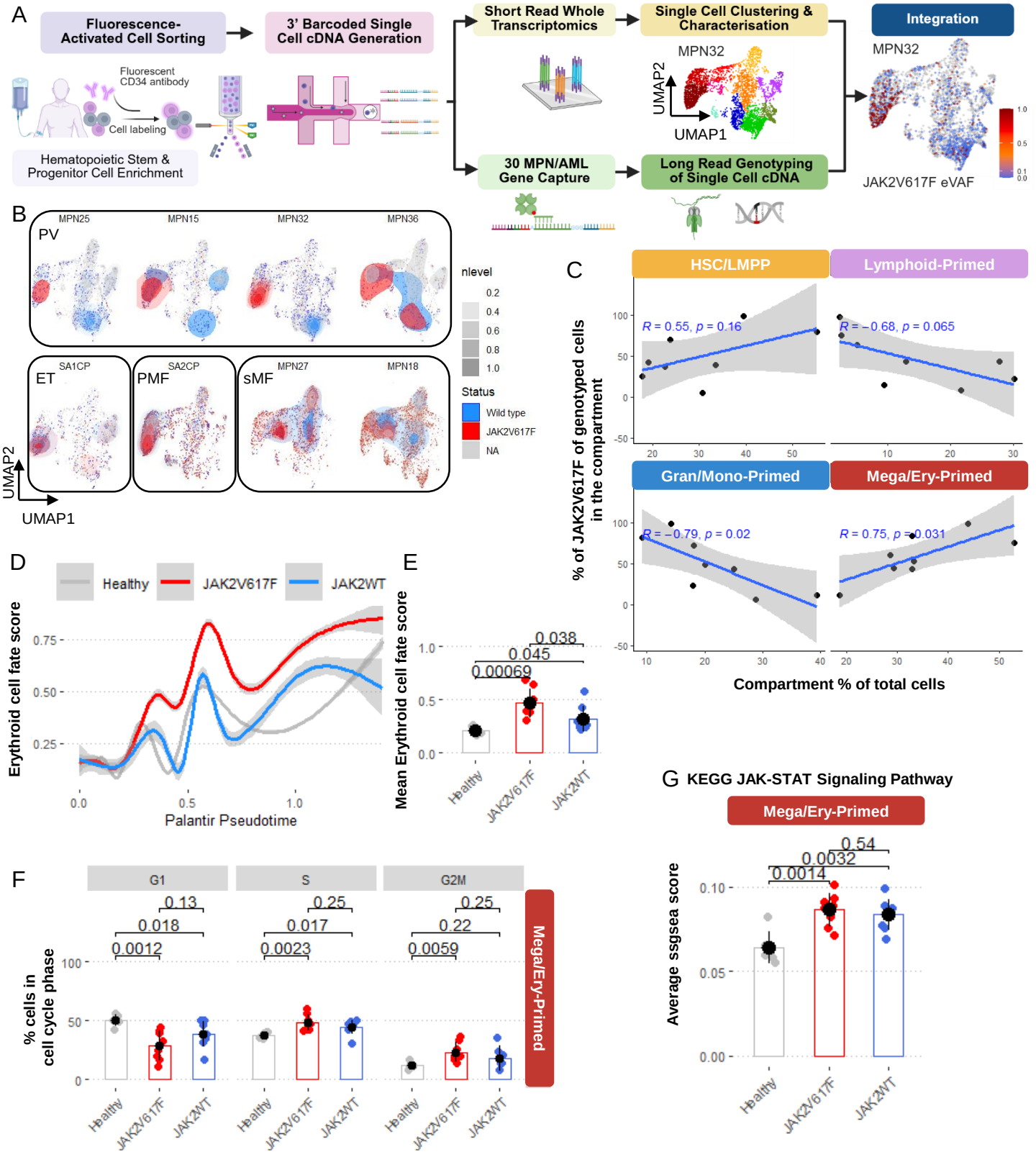


Figure 2


A
Fluorescence-Activated Cell Sorting
3' Barcoded Single Cell cDNA Generation
Fluorescent CD34 antibody
Cell labeling
Hematopoietic Stem & Progenitor Cell Enrichment
Short Read Whole Transcriptomics
Single Cell Clustering & Characterisation
MPN32
UMAP2
UMAP1
Integration
MPN32
JAK2V617F eVAF
30 MPN/AML Gene Capture
Long Read Genotyping of Single Cell cDNA
B
MPN25
MPN15
MPN32
MPN36
PV
SA1CP
ET
SA2CP
PMF
MPN27
MPN18
sMF
UMAP2
UMAP1
nlevel
0.2
0.4
0.6
0.8
1.0
Status
Wild type
JAK2V617F
NA
C
HSC/LMPP
Lymphoid-Primed
Gran/Mono-Primed
Mega/Ery-Primed
R = 0.55, p = 0.16
R = −0.68, p = 0.065
R = −0.79, p = 0.02
R = 0.75, p = 0.031
% of JAK2V617F of genotyped cells in the compartment
Compartment % of total cells
D
Healthy
JAK2V617F
JAK2WT
Erythroid cell fate score
Palantir Pseudotime
E
Mean Erythroid cell fate score
0.038
0.045
0.00069
Healthy
JAK2V617F
JAK2WT
F
G1
S
G2M
Mega/Ery-Primed
% cells in cell cycle phase
0.13
0.018
0.0012
0.25
0.017
0.0023
0.25
0.22
0.0059
Healthy
JAK2V617F
JAK2WT
G KEGG JAK-STAT Signaling Pathway
Mega/Ery-Primed
Average ssgsea score
0.54
0.0032
0.0014
Healthy
JAK2V617F
JAK2WT

Supplemental Figure 2

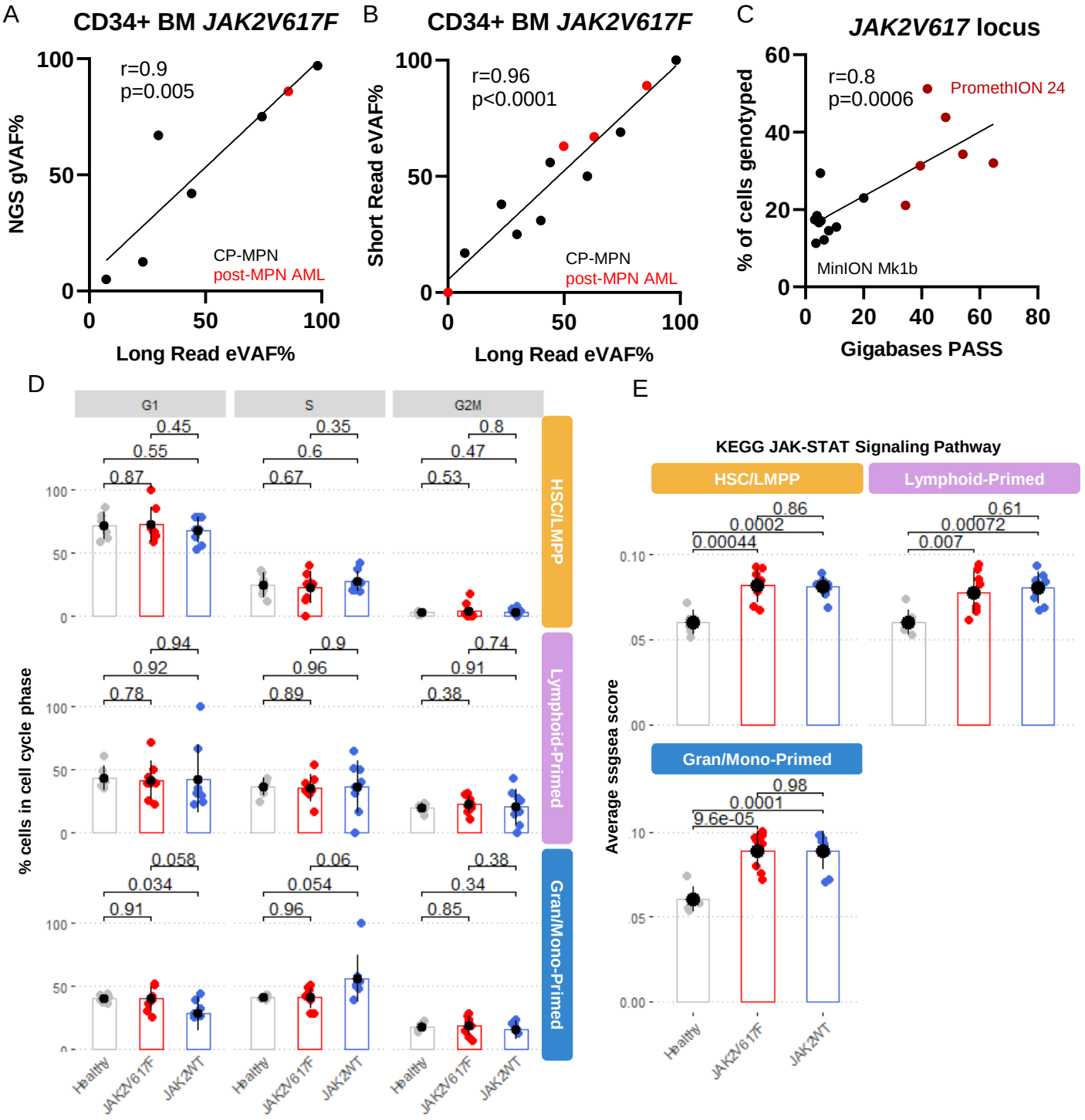
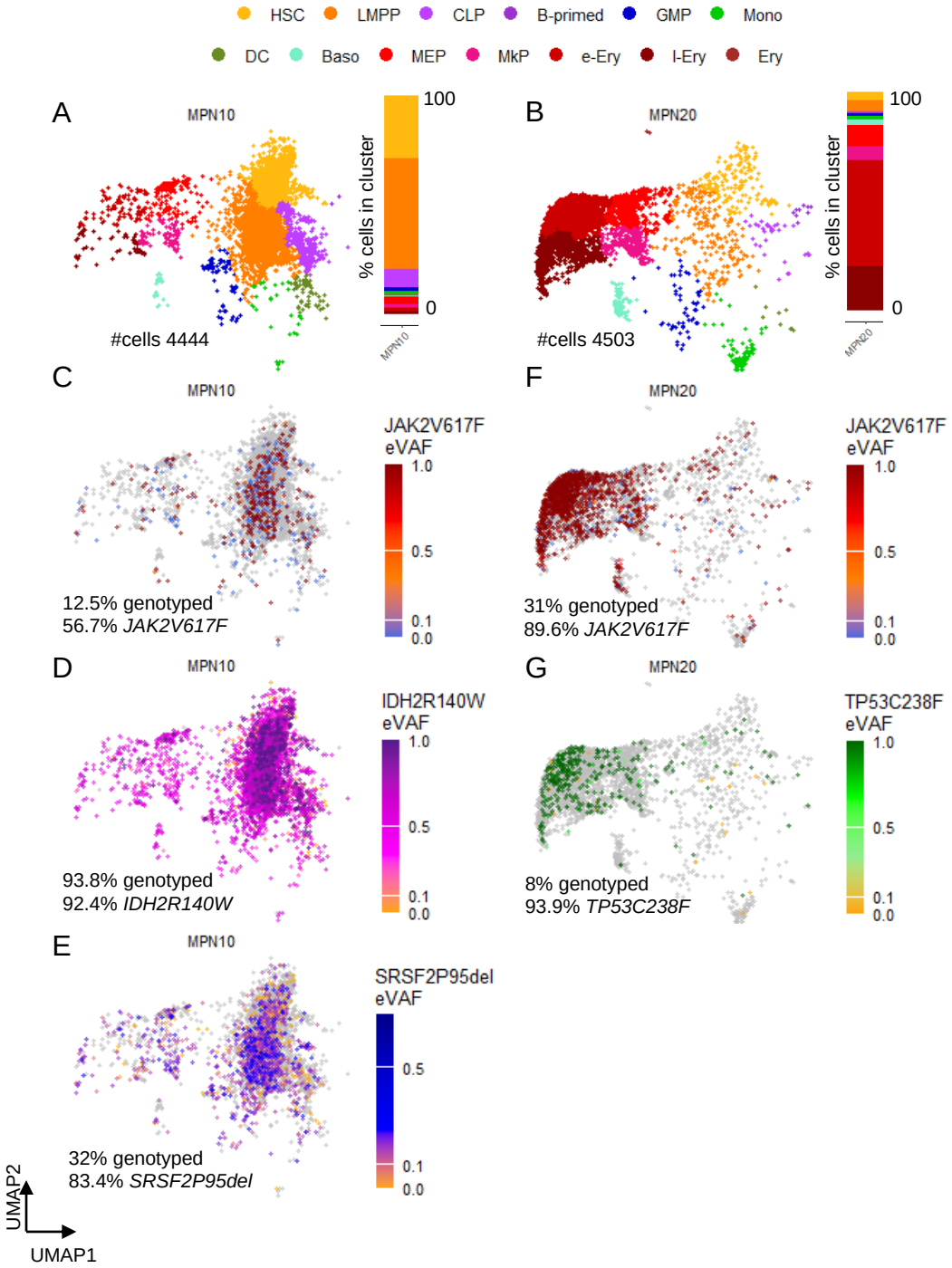


Figure 3


HSC
LMPP
CLP
B-primed
GMP
Mono
DC
Baso
MEP
MkP
e-Ery
l-Ery
Ery
A
MPN10
#cells 4444
% cells in cluster
100
0
MPN10
B
MPN20
#cells 4503
% cells in cluster
100
0
MPN20
C
MPN10
JAK2V617F
eVAF
1.0
0.5
0.1
0.0
12.5% genotyped
56.7% JAK2V617F
F
MPN20
JAK2V617F
eVAF
1.0
0.5
0.1
0.0
31% genotyped
89.6% JAK2V617F
D
MPN10
IDH2R140W
eVAF
1.0
0.5
0.1
0.0
93.8% genotyped
92.4% IDH2R140W
G
MPN20
TP53C238F
eVAF
1.0
0.5
0.1
0.0
8% genotyped
93.9% TP53C238F
E
MPN10
SRSF2P95del
eVAF
0.5
0.1
0.0
32% genotyped
83.4% SRSF2P95del
UMAP2
UMAP1

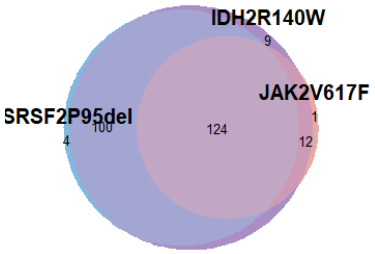
Supplemental Figure 3

A

	Median MinION	Median PromethION	
Gigabases PASS	5.25	45.15	
Assessed Gene Loci	Median MinION % of genotyped cells	Median PromethION % of genotyped cells	P-value Welch t-test, two-sided: MinION vs PromethION
NPM1W288	93.2	97.9	**0.0108**
U2AF1S34	90.9	94.1	0.0672
CALRK385	83.0	89.5	**0.0287**
PTPN11A72	58.5	85.3	**<0.0001**
IDH2R140	77.3	84.5	0.0626
KRASG12	60.3	83.7	**0.0002**
SRSF2P95	71.0	82.0	**0.0080**
SF3B1K666	37.4	67.8	**<0.0001**
ZRSR2R19	32.8	61.2	**0.0219**
MYCP74	46.8	53.4	0.5358
DNMT3AR882	32.1	52.7	**0.0209**
GATA2A318	22.9	49.6	0.1743
RUNX1R201	28.8	48.5	**0.0194**
NRASG12	25.9	48.4	**0.0042**
EZH2Y646	22.6	38.2	**0.0040**
IDH1R132	19.2	37.7	**0.0411**
TET2R1440	17.0	35.3	**0.0263**
JAK2V617	16.8	33.2	**0.0066**
FLT3D835	25.4	31.8	0.2886
SH2B3E208	22.0	28.3	0.0631
PPM1DQ150	11.8	21.9	**0.0028**
KITD816	9.0	19.4	**0.0070**
CSF3RT618	8.1	18.9	0.0564
STAG2R614	7.4	16.4	**0.0150**
TP53C238	10.4	15.9	0.0512
MPLW515	1.5	5.5	0.1840
BCORS336	1.6	5.0	**0.0082**
CBLC404	1.8	3.0	0.3707
ASXL1G642	1.2	2.8	**0.0134**

B

MPN10



C

MPN20

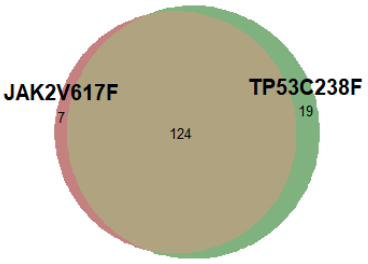
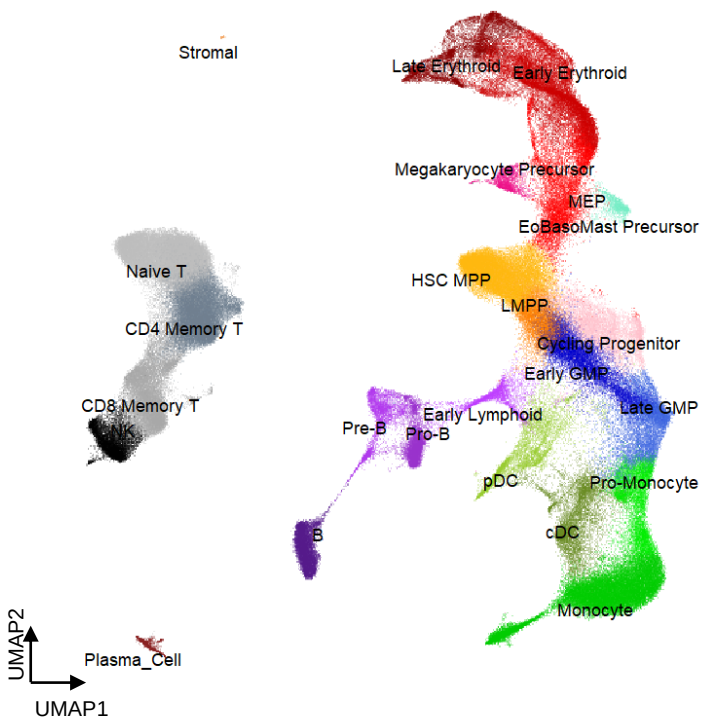


Figure 4

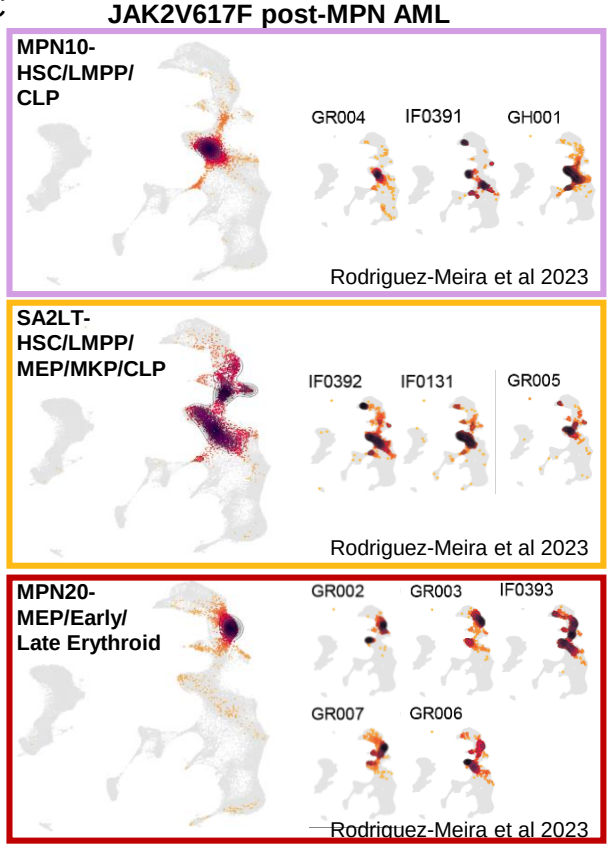
A

Bone Marrow Map - Zeng et al. 2025

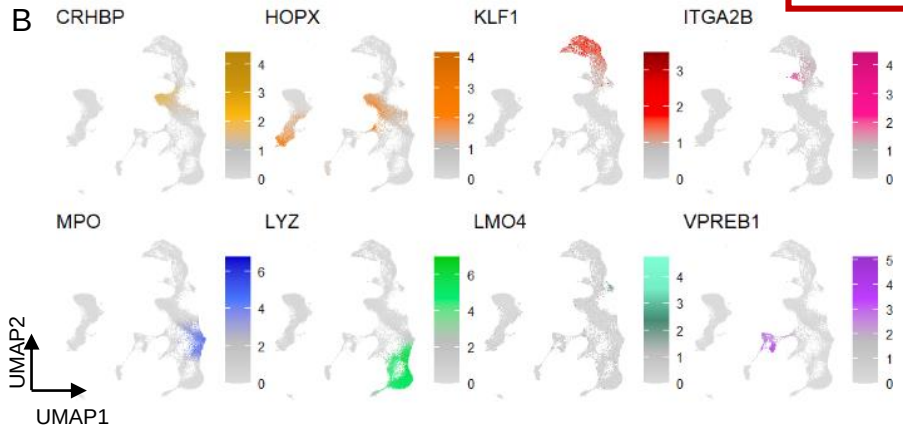


C

JAK2V617F post-MPN AML



B



D

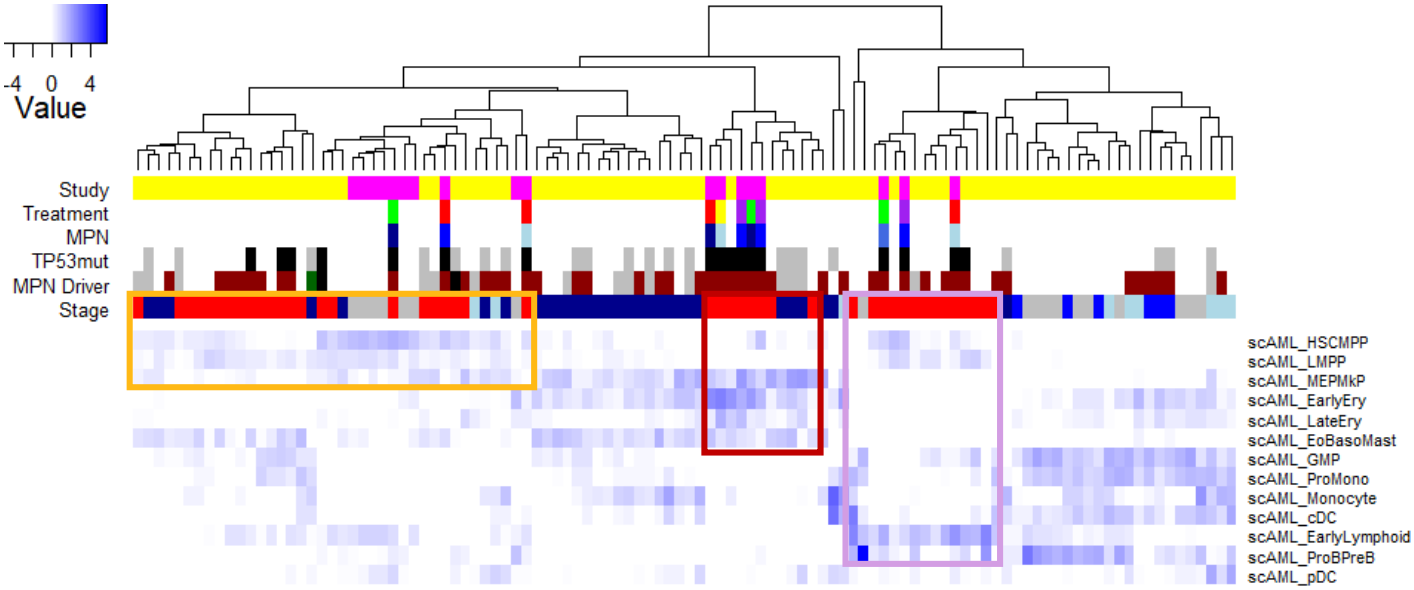


Figure 5

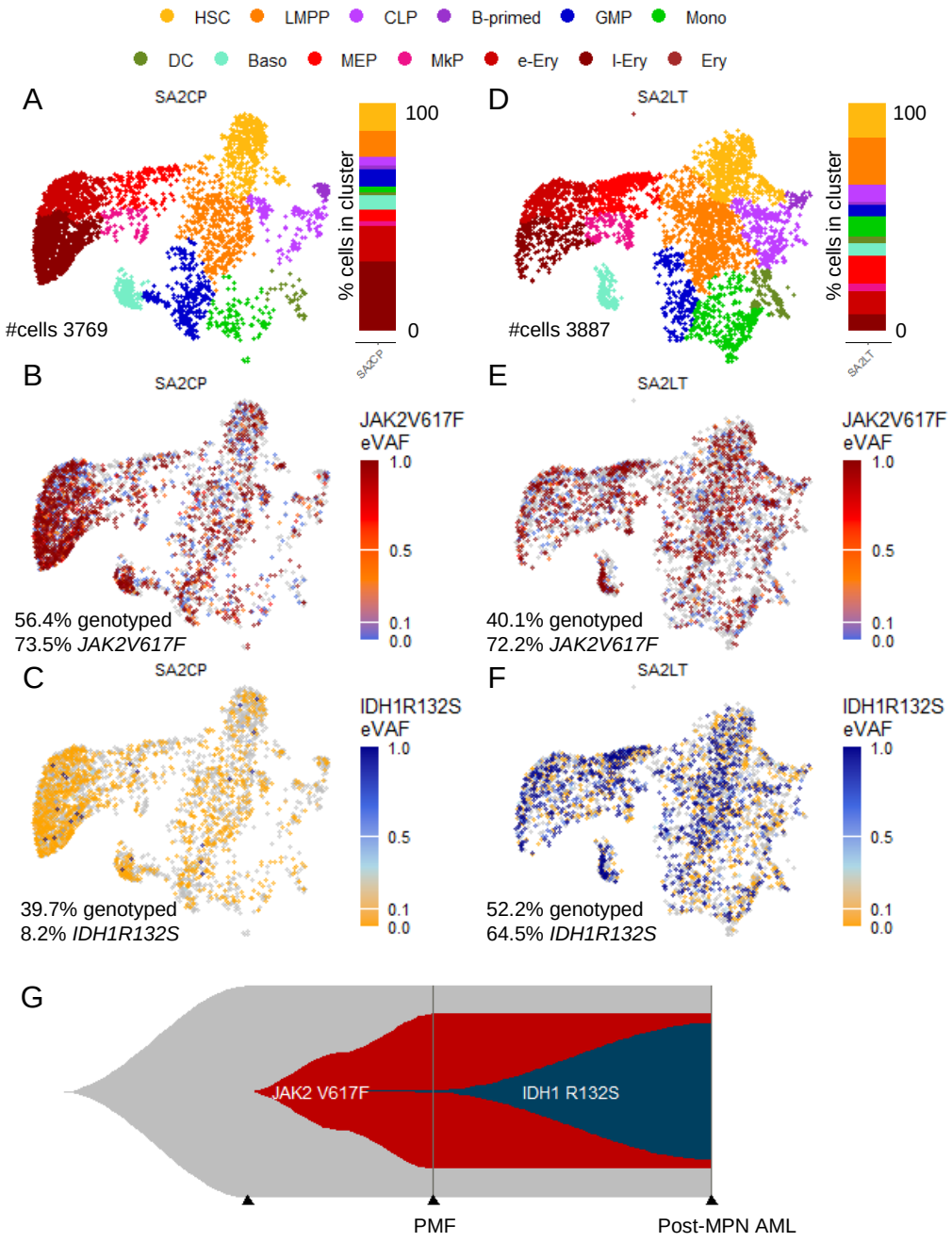
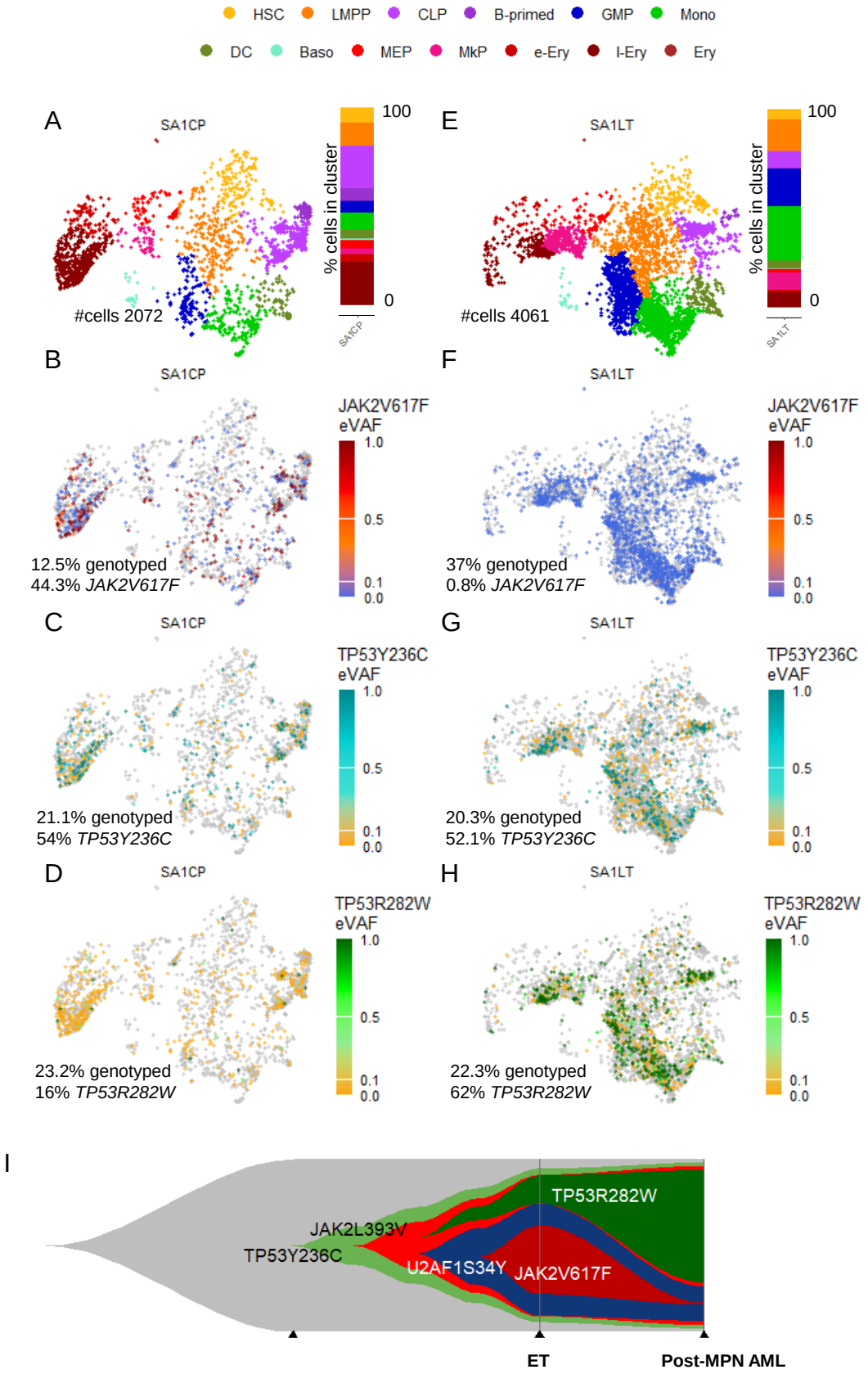
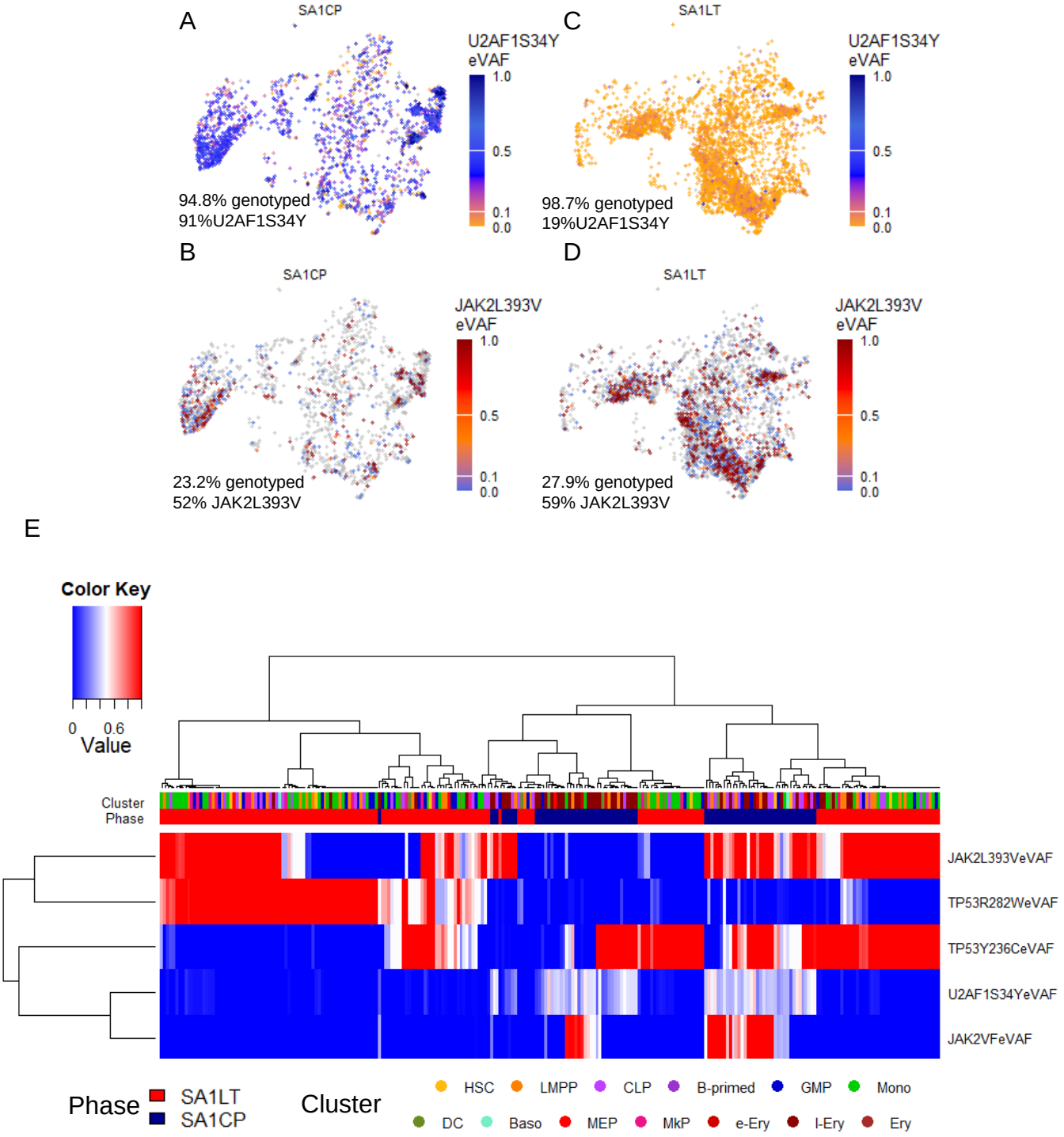

HSC
LMPP
CLP
B-primed
GMP
Mono
DC
Baso
MEP
MkP
e-Ery
l-Ery
Ery
A
SA2CP
#cells 3769
% cells in cluster
100
0
D
SA2LT
#cells 3887
B
JAK2V617F eVAF
1.0
0.5
0.1
0.0
56.4% genotyped
73.5% JAK2V617F
E
40.1% genotyped
72.2% JAK2V617F
C
IDH1R132S eVAF
39.7% genotyped
8.2% IDH1R132S
F
52.2% genotyped
64.5% IDH1R132S
G
JAK2 V617F
IDH1 R132S
PMF
Post-MPN AML

Figure 6

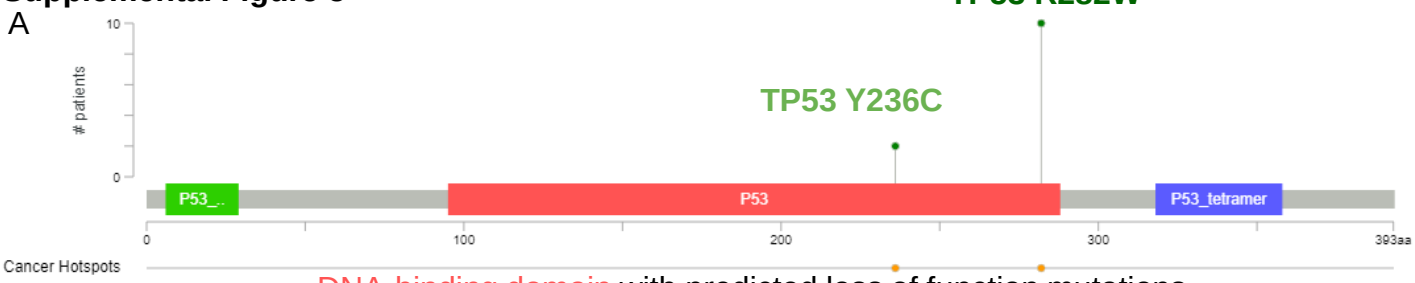

HSC
LMPP
CLP
B-primed
GMP
Mono
DC
Baso
MEP
MkP
e-Ery
l-Ery
Ery
A
SA1CP
% cells in cluster
100
0
#cells 2072
E
SA1LT
#cells 4061
B
JAK2V617F eVAF
1.0
0.5
0.1
0.0
12.5% genotyped
44.3% JAK2V617F
F
37% genotyped
0.8% JAK2V617F
C
TP53Y236C eVAF
21.1% genotyped
54% TP53Y236C
G
20.3% genotyped
52.1% TP53Y236C
D
TP53R282W eVAF
23.2% genotyped
16% TP53R282W
H
22.3% genotyped
62% TP53R282W
I
TP53R282W
JAK2L393V
TP53Y236C
U2AF1S34Y
JAK2V617F
ET
Post-MPN AML

Supplemental Figure 4



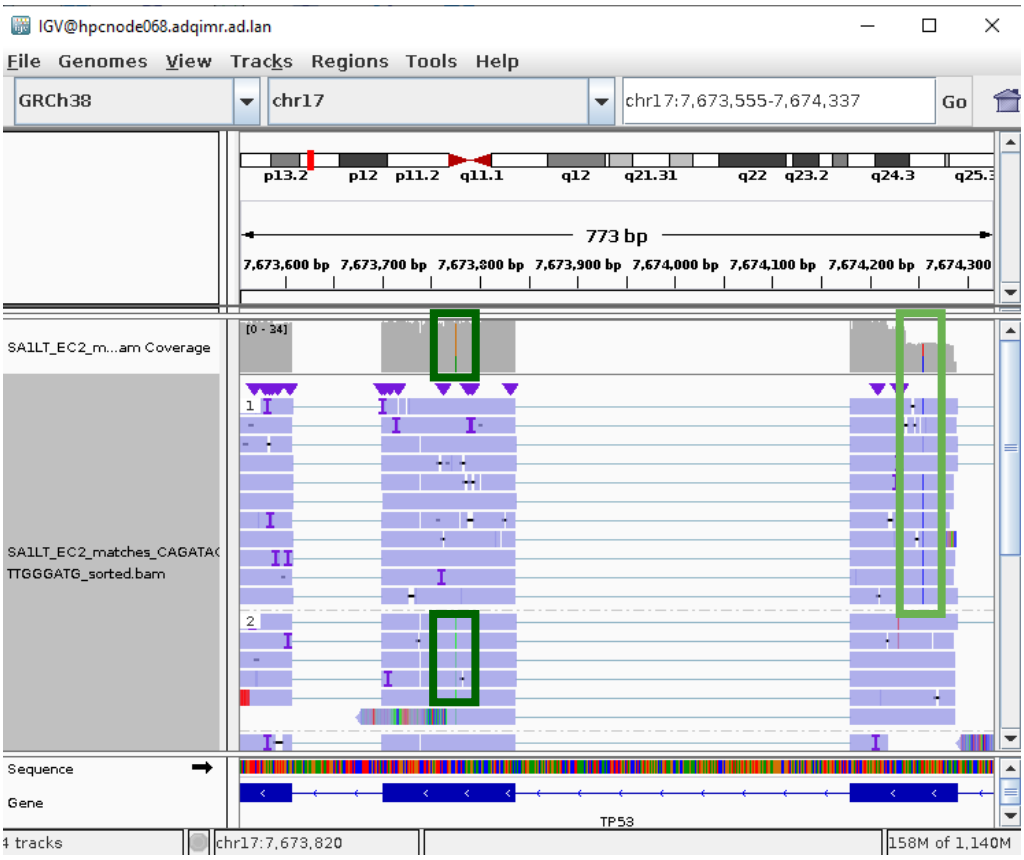
Supplemental Figure 5

A

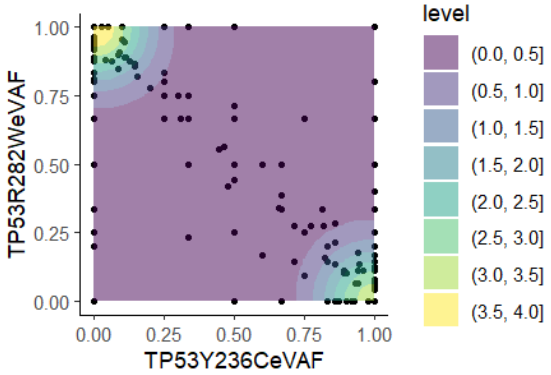


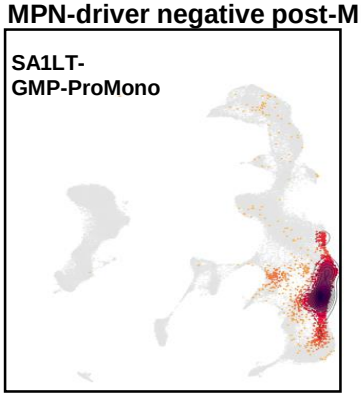
DNA-binding domain with predicted loss of function mutations

B



C



D

MPN-driver negative post-MPN AML