nature genetics

Supplementary information

https://doi.org/10.1038/s41588-026-02751-3

Type 1 interferon perturbates clonal competition by reshaping human blood development

In the format provided by the authors and unedited

Supplementary Note

Supplementary Methods

Multiplexed Immunofluorescence

Multiplexed immunofluorescence (mIF) was performed using the Opal system (Akoya Biosciences) by staining 4 micron-thick Bouin-fixed, paraffin-embedded whole-tissue sections from decalcified human bone marrow core biopsy specimens in a Bond RX automated tissue stainer (Leica Biosystems, Buffalo Grove, IL), described previously.[2,3] Briefly, tissue sections were first deparaffinized prior to EDTA-based antigen retrieval (Leica ER2 solution, 20min), then stained with horseradish peroxidase-mediated deposition of tyramide-Opal fluorophore constructs (Akoya Biosciences) in each cycle, with intervening application of heat, citrate-based epitope retrieval solution (Leica ER1) and Bond Wash Solution (Leica). Finally, 4', 6-diamidino-2-phenylindole (Spectral DAPI, Akoya Biosciences) was applied per provided protocols to label nuclei. The following panel of primary antibody/fluorophore pairs was used: 1) Opal 480/anti-mutant CALR (1:120, CAL2, Dianova), 2) Opal 520/anti-CD38 (1:50, 38C03 (SPC32), Invitrogen), 3) Opal 570/anti-CD117 (1:100, D3W6Y, Cell Signaling), 4) Opal 620/anti-TdT (1:8, SEN28, Invitrogen), 5) Opal 690/anti-CD34 (1:100, QBEND/10, Invitrogen), 6) Opal 780/anti-Ki67 (Ready-to-use, MM1, Leica). Whole slide scans were obtained at 20X magnification using the Vectra Polaris Automated Quantitative Pathology Imaging System (Akoya Biosciences).

Single-Cell Differentiation Assay

One day before sorting HSCs, flat-bottom 96-well plates were coated with 60µL 0.2% gelatine (Sigma) per well for 1 hour and then excess removed. Murine MS5 stroma cells were plated at a density of 1500 cells/well in 100µL Myelocult H5100 medium (Stem Cell Technologies), 1% Penicillin-Streptomycin (Pen/Strep, 10,000 U/mL, Gibco) and 1% glutamine (Gibco). On the day of the assay, the medium was changed to 100µL/well Myelocult H5100 medium, 1% Pen/Strep and 1% glutamine supplemented with the following cytokines: IL-2 10ng/mL, IL-6 20ng/mL, IL-7 20ng/mL, SCF 100ng/mL, TPO 50ng/mL, G-CSF 20ng/mL, FLT3L 10ng/mL and GM-CSF 20ng/mL (all from Bio-Techne). Cryopreserved BMMCs (n = 3 baseline and 4 IFNα-treated patient samples) were thawed and $CD34^+$ cells isolated using the EasySep™ Human CD34 Positive Selection Kit II (StemCell Technologies #17856) following manufacturer's protocol. $CD34^+$ isolated cells were then stained with FITC CD45RA (1:50), APC-Cy7 CD34 (1:200), APC CD90 (1:50), Pe-Cy7 FLT3 (1:100), PerCP-Cy5.5 HLA-DR (1:100), and BV786 CD41 (1:100) (Supplementary Table 34 for antibody details). $CD34^+$ $CD90^+$ cells were index-sorted by FACS directly into 96-well plates with pre-plated MS5 stromal cells. Eight days after sorting, 70µL of media were removed from the top of the plate without disturbing the colonies, and 170µL of IMDM medium (Gibco) with 10% BIT 9500 (StemCell Technologies), 1% Pen/Strep, 1% glutamine, supplemented with IL-2 10ng/mL, IL-3 20ng/mL, IL-6 10ng/mL, IL-7 10ng/mL, SCF 100ng/mL, TPO 50ng/mL, G-CSF 10ng/mL, FLT3L 10ng/mL, SDF1 5ng/mL and 2-ME (1.8µL per 50mL) were added. Colonies were analyzed under the microscope 16-18 days after sorting and all visible colonies were detached by pipetting from the stromal cell layer and transferred into a 96-well U-bottom plate using a plate filter (Pall AcroPrep), to prevent the carryover of MS5 cells. Cells were stained for 1 hour at 4ºC with FITC CD66b (1:100), PE CD16 (1:100), PerCP-Cy5.5 CD184 (1:100), Pe-Cy7 CD83 (1:100), APC CD14 (1:100), V500 CD45 (1:100), BV650 CD71 (1:100), and BV711 CD33 (1:100) with 50µL/well antibody mix and washed afterwards with 200µL PBS and 2.5% FBS. Immunophenotype of the colonies was assessed in a BD LSRFortessa Analyzer.

Flow cytometric analysis was performed using FlowJo and OMIQ. To generate differentiation plots, OMIQ was used to gate live cells and calculate UMAP dimensionality reduction based on expression of CD14, CD34, CD45, CD71, CD33, CD16, and CD66b after integrating cells across every colony using recommended settings. A total of 177 colonies were included in the analysis from 5 samples (after filtering for colonies with <350 live cells). GEMM colonies showed $CD71^+$, $CD14^+$, and $CD66b^+$ populations. GM colonies showed both $CD14^+$ and $CD66^+$populations (at least 1% each). Neutrophil-only colonies showed <1% $CD14^+$ cells and at least 75% $CD66b^+$ cells. To increase the stringency of the neutrophil-only assignment, a minimum of 4000 events was required for this assignment. The remainder of the colonies were labeled as early myeloid (EM) colonies showing a large $CD33^+$ population (>40%) that lacked either CD14 or CD66b expression.

Lentiviral Constructs for Overexpression Experiments

Lentiviral overexpression (OE) vectors and their corresponding control vectors were designed and obtained from VectorBuilder. All constructs included a lentivirus backbone (pLV) with gene-specific inserts, a fluorophore and an antibiotic resistance insert. The RFX2-OE insert consisted of human RFX2 ORF (NM_000635.4) driven by the human phosphoglycerate kinase (hPGK) high-expression promoter, with CMV-EGFP-T2A-Puro for selection. The RFX3-OE insert consisted of human RFX3 ORF (NM_001377999.1) driven by the hPGK promoter, with CMV-mCherry-T2A-Puro for selection. The control vectors contained the hPGK promoter driving an empty ORF stuffer, followed by CMV-mCherry-T2A-Puro or CMV-EGFP-T2A-Puro.

Lentivirus Production

HEK-293T cells were seeded at a density of $2x10^6$ cells in DMEM, 10% FBS and 1% Pen/Strep in a 10cm plate. After 24 hours, media was changed to DMEM and 10% FBS. 9µg OE or control plasmid was added to 1mL Opti-MEM (ThermoFisher) containing 3µg pMD2.G plasmid and 8µg psPAX2 plasmid and incubated with Lipofectamine 2000 (ThermoFisher) in 1mL of Opti-MEM for 15min at room temperature. Mixture was then added dropwise to the cells and incubated at 37°C for 24 hours. Media containing lentivirus was collected over the next 48 hours. Lentivirus was then concentrated using Lenti-X Concentrator (Clontech #631231) according to manufacturer's protocol. Viral titer was determined using the qPCR Lentivirus Titer Kit (Applied Biological Materials #LV900).

Lentiviral RFX2 and RFX3 Overexpression of $CD34^+$ Cells

After isolation, UCB $CD34^+$ cells were plated in 96-well round-bottom plates (Falcon) in StemSpan™ SFEM II media (StemCell Technologies) with StemSpan™ $CD34^+$ Expansion Supplement (StemCell Technologies) and 1% Pen/Strep. After 24 hours in culture, cells were spun down and resuspended in minimal media consisting of StemSpan™ SFEM II media with StemSpan™ $CD34^+$ Expansion Supplement, 1% Pen/Strep, 10µM prostaglandin E2 (StemCell Technologies #72192) and 100ng/µL poloxamer 407 (Millipore Sigma #P2164030). Cells were split into 2 technical replicates for each transduction condition. Lentivirus or PBS control was added at an MOI of 100 directly into the well. Cells were spinoculated at 300g, 32°C for 1 hour. After the spin, cells were incubated in same lentivirus-containing media for 24 hours at 37°C before resuspension in fresh media consisting of StemSpan™ SFEM II media with StemSpan™ $CD34^+$ Expansion Supplement and 1% Pen/Strep. After 48 hours of incubation, cells were stained with CD34-APC antibody (1:50) (Supplementary Table 34). FACS was then used to isolate the $CD34^+$ population in non-transduced control, the $CD34^+$ $EGFP^+$ populations in the RFX2-OE and EGFP-control conditions, and the $CD34^+$ $mCherry^+$ populations in the RFX3-OE and mCherry-

control conditions. Independent transduction experiments (n = 3 for RFX2, and n = 4 for RFX3) were completed with new UCB units.

Non-transduced, mCherry-control, and RFX3-OE $CD34^+$ cells were stained with Cell Hashing antibodies (HTO 1-3 TotalSeq-A, BioLegend) for 30min at 4°C. Cells were then washed with FACs Buffer three times, pooled, and counted for loading. scRNA-sequencing was performed using the 10x Genomics 3' v3.1 platform according to the manufacturer's protocol.

For CFU assays, MethoCult™ H4034 Optimum (StemCell Technologies #04044) was supplemented with 10ng/mL human FLT3L and IL-6 (Miltenyi-Biotec) and 1% Pen/Strep. 250 cells per group (non-transduced, EGFP-Control, RFX2-OE, mCherry-Control, and RFX3-OE) were added to 3mL MethoCult total and vortexed thoroughly. Tubes were then left at room temperature until bubbles rose to the surface. 1.25mL of MethoCult with cell suspension was transferred per well of a 6-well plate (2 technical replicates) via blunt-edge 16g needle. Plates were incubated at 37°C for 14 days after which colonies were counted and identified by morphology. CFU assay was completed on all the independent transduction experiments.

Transient overexpression of RFX3 in $CD34^+$ Cells

To achieve transient overexpression, we used the P3 Primary Cell 4D-Nucleofector® X Kit L (Lonza Bioscience Cat # V4XP-3024) and the 4D-Nucleofector® Unit X (Lonza Biosciences) to nucleofect $CD34^+$ cells with RFX3-OE, mCherry-Control, or manufacturer-provided EGFP positive control vectors. The same RFX3-OE and mCherry control plasmids were used for the lentiviral transduction experiments. We followed the Amaxa™ 4D-Nucleofector™ Protocol for Unstimulated Human $CD34^+$ Cells. Briefly, cryopreserved human UCB CD34p cells were thawed rapidly at 37°C and cultured overnight in 96-well round-bottom plates (Falcon) in StemSpan™ SFEM II media (StemCell Technologies) with StemSpan™ $CD34^+$ Expansion Supplement (StemCell Technologies) and 1% Pen/Strep. After 24 hours, cells were centrifuged at 300g for 10min at room temperature. Supernatant was carefully removed and cells were resuspended gently in room temperature 4D-Nucleofector™ Solution and separated into aliquots of 2 technical replicates per condition. 5µg of plasmid was added to each reaction in 100µL cuvettes. EO-100 program was used for nucleofection, after which cells were recovered from the cuvettes and plated in 96-well round-bottom plates in StemSpan media with $CD34^+$ Expansion Supplement. 48 hours post-transfection, cells were collected and stained with CD34-APC antibody (1:50) and subsequently sorted by FACS for $CD34^+$ $mCherry^+$ cells. CFUs were plated as described above. Three independent biological replicates were completed with different units of UCB.

GoT-IM scRNA-seq data processing, alignment, cell-type classification and clustering

For individual datasets, cells with UMI > or < 3 standard deviations from mean UMI and mitochondrial gene percentage >10% were filtered (see Extended Data Fig. 1b and Extended Data Fig. 3g). The HTO data were normalized with centered log-ratio (CLR) transformation and used for time-point assignment for each experiment.[1] The cells from each time-point were analytically separated into individual datasets based on HTO counts. These individual datasets (in case of $CD34^+$ samples, together with baseline ET samples from our previous work[4]), were integrated within Seurat, which implements reciprocal principal component analysis (RPCA) and the principles of mutual nearest neighbor as per recommended settings (30 principal components (PCs) for the anchor determining procedure in IntegrateLayers function).[5] Principal component analysis (PCA) was performed using the top 2000 variable genes identified with variance-stabilizing transformation.[5] The first 30 statistically significant PCs were used in the UMAP

algorithm for visualization.[6] Clusters were manually assigned based on differentially expressed genes using the FindAllMarkers function using default settings (using top 2000 variable genes, in a minimum of 10% of cells in either group, and log-transformed fold change of 0.25, using Wilcoxon rank sum test). The clusters were annotated according to canonical lineage markers identified previously in single-cell RNA-seq data of normal hematopoietic progenitor cells[7] into 16 main progenitor subsets. For dimensionality reduction and visualization of individual experiments, top 3000 variable genes were included for principal component analysis (PCA), and the first 30 statistically significant PCs were used as inputs to the UMAP algorithm. The CITE-seq data after CLR normalization was used to distinguish cell types and identify HSCs.

Gene module scoring, differential expression and gene set enrichment analysis

For statistical analysis involving cell identities (e.g., IMP vs GMP), cell type was entered as the fixed effect and samples as random effects in a linear mixed model (LMM). For IMP and GMP comparison, treatment status was also included as a fixed effect. When comparing genotype statuses (e.g., *CALR* mutant versus wildtype), a LMM was conducted with genotype status as the fixed effect and samples as random effects. P-values were obtained from likelihood ratio tests of the full model with the effect in question against the model without the effect.

The tested genes for differential gene expression analysis included the top 2,000 variable genes from the integration, among which those expressed in at least 10% of either group were kept. Among $CD34^+$ cells with GoT-IM data, IFN01-IFN10 samples at baseline and active treatment were included for the aggregated differential expression analysis. For each gene, the variable in question (i.e., treatment or mutation status) was entered as the fixed effect and samples as random effects. P-values were obtained from likelihood ratio tests with or without the effect in question. Significant genes, with abs(avg_log2FC)> 0.1 and adjusted p-value <0.05, were used for further analysis such as module scoring for HSC-specific IFNα-induced upregulated or downregulated genes.

For examining gene module expression (e.g., HSC-specific IFNα-upregulated or downregulated gene signature), the function AddModuleScore within the Seurat package[8] was used to calculate the relative expression of the genes for each cell. To calculate the module expression of cell-cycle-related genes, G2M phase and S phase marker genes were used as available in Seurat with CellCycleScoring function. Briefly, expression of a control gene module was calculated and subtracted from the average gene module expression of interest, as previously described[9]. All analyzed genes were classified based on average expression into 24 bins, and for each gene in the module, 100 control genes were randomly selected from the same expression bin as the gene of interest[9]. For the overall IMP module score, the cells were first scored for the upregulated and downregulated genes; the downregulated gene score was subtracted from the upregulated gene score to obtain the overall IMP score.

Data preprocessing, alignment and cell type identification for GoT-ATAC

Gene annotations from EnsDb.Hsapiens.v86 and motif annotations from Cis-BP[10] for transcription factor binding motifs were utilized. Cells with blacklist ratio >0.02, TSS enrichment <2 and nucleosome signal >4 were filtered out. For RNA data, cells with UMI > or < 3 standard deviations from mean UMI or mitochondrial gene percentage > 25%, were filtered. The nuclei hashing data processing was the same as for GoT-IM data. As nuclear hashing is known to be noisier than cell hashing,[11] hashing data was used in combination with cell clustering data, as the cells cluster based on treatment status. After the cells were segregated analytically based on time-point, snRNA-seq

data were integrated, where they underwent batch correction with canonical correlation analysis (CCA) within Seurat and the principles of nearest neighbor.[5] Recommended settings were used (30 canonical correlation vectors for CCA in the FindIntegrationAnchors function and 30 PCs for the anchor weighting procedure in the IntegrateData function). PCA and cell type assignments were performed as described in GoT-IM. Integration via the ATAC-seq data was performed by normalizing the merged counts using first term frequency inverse document frequency (TFIDF) normalization with RunTFDIF, followed by linear dimensionality reduction using latent semantic indexing (LSI). The first 2:30 dimensions were retained, and batch-correction was performed with runHarmony (Harmony, v0.1.0), which iteratively learns a cell-specific linear correction function to account for batch effect. The first LSI component was removed as it represents variation due to sequencing depth rather than biological differences.[11]

A requirement of at least four reads supporting a UMI was implemented. To capture genotyping reads within the snRNA-seq data, we called *CALR* variants using the gene expression BAM derived from the standard Multiome workflow. BAMQL[12] and the CIGAR string of the BAM file were used to parse reads containing the expected wildtype and mutant sequences. Cells expressing at least one mutant UMI were categorized as mutant, and cells with just wildtype UMI were assigned as wildtype. Genotyping assignments were then appended to calls made by IronThrone.

Motif enrichment analysis in GoT-ATAC

To perform differential motif enrichment analysis, within a sample and on the deviation z-score computed by chromVAR, we applied the function FindMarkers in Signac (Wilcoxon Rank Sum test), where the average difference in z-score between the groups was calculated. For integrated data, we combined the P-values (Fisher's method) and calculated weighted mean deviation score across individual samples. P-values were adjusted by the Benjamini-Hochberg method.

To find over-enriched motifs for a group of genomic features, the FindMotifs function was used, accounting for accessibility and GC content bias by selecting 5000 accessible background peaks with similar GC content for each feature set. To identify motifs enriched in a singular genomic range, FIMO (v5.4.1) from MEME suite[13] was used to scan for Cis-BP transcription factor motifs along the nucleotide sequence from the human reference genome GRCh38 with a p-value threshold of 0.0001. We de-prioritized zinc fingers (ZNFs) in the list of motifs specific to genomic regions, as each individual ZNF binds to a three-bp motif leading to more frequent matches and higher match scores.[14] Synergy between pairs of transcription factor motifs was defined as the excess variability of chromatin accessibility for peaks sharing both motifs compared to a random subsample of the same size of peaks with one motif. High synergy usually indicates a cooperative binding relationship between pairs of transcription factors. The function getAnnotationSynergy from chromVAR was called to calculate synergy scores.[15]

GoT-ChA data processing

After filtering for high-quality cells based on recommended parameters, we performed cell type transfer labels from the CALR-mutated GoT-ATAC data to the JAK2-mutated $CD34^+$ cells using the Seurat function FindTransferAnchors[5] to identify pairs of cells—one each from the reference and the query—with the closest chromatin accessibility profiles. The similarity is determined with a nearest-neighbor approach between cells in the reference and query datasets. We performed label transfer according to recommended settings by identifying significant peaks in both datasets, normalizing counts, running singular value decomposition (SVD), performing dimensionality reduction, identifying anchors, and transferring cell labels from the reference data to the query

data. For datasets without sufficient overlapping peaks with the reference datasets, cell types were annotated based on cell marker gene and transcription factor activity scores

Retrospective flow cytometry data analysis

Data processing involved selecting the panel with CD34, TdT and CD19 antibodies, as more than one antibody panel may have been run on a given sample. A UMAP embedding was created that corresponded to the antibody panel and could be applied to each FCS file so that cells of a given immunophenotype were co-localized in the UMAP plots for all the FCS files. The ensemble UMAP was calculated using a subsampling of each of the FCS files to present the full spectrum of immunophenotypes to the UMAP algorithm for calculating the embedding. The various channels used in the UMAP embedding included the forward scatter height (FSC-H), side scatter height (SSC-H), and channels corresponding to each of the fluorescent markers. The normalization factors and UMAP embeddings were then applied to all the individual files.

Using FlowJo, appropriate gates were determined based on the clusters in the UMAP plots and their expression of protein markers. Additional standard gating was also performed using the original data channels. UMAP gates were determined following exclusion of doublets and non-viable cells via standard flow cytometry gating approaches.

Multiplexed Immunofluorescence Image Analysis with PathML

Images were loaded and divided into equal-sized tiles on which we ran our preprocessing pipeline. This pipeline starts with coercing the tile shape into the standard x,y,c format, followed by segmentation. Nuclear and cellular segmentation were performed using the Mesmer[16] deep learning segmentation model implemented in PathML, with DAPI and autofluorescence used as nuclear and cell membrane markers, respectively. Subsequently, we used the 'QuantifyMIF' function from PathML to convert the segmented images into a count matrix, which includes the intensity of each marker in each segmented cell along with cell coordinates, size, and eccentricity. To remove the noise from the count matrix data, any cell with raw intensity less than 50 for DAPI and auto-fluorescence markers was excluded from the analysis. Next, the thresholds were obtained to find the positive and negative expression level for each cell marker. The thresholds were manually set for each marker based on examination by a board-certified hematopathologist.

CUT&RUN Assay data analysis

CUT&RUN data were downsampled to equalize the number of reads across all conditions within the sample replicate to account for differences in sequencing depth. Among the down-sampled reads, low-quality reads were filtered out using Trimmomatic[17], resulting reads were aligned with the hg38 genome with Bowtie2[18], and PCR duplicates were removed with SAMtools[19] with a previously described workflow.[20] Aligned reads were then used for peak calling with MACS3[21] with a q-value threshold of 0.01 and reads from IgG antibody as the control.

Additional Notes on Statistics and Reproducibility

LMM was implemented using the lme4 R package (v.1.2-1). LMMs were generated with/without cell mutational status, treatment status, or cell types, as specified in figure legends. We included patient samples as random effects in our statistical comparisons. P-values from the likelihood ratio test between two nested models were calculated with the Stats R package (v.3.5.1). Benjamini-Hochberg FDR-correction was used unless specified otherwise.

For all box plots presented, the box represents the interquartile range (IQR); upper and lower whiskers represent the largest and smallest values within 1.5xIQR; the central line represents the median. Dots represent outlier values or data value distributions. For all violin plots, the violin represents the kernel probability density of the data, and dots represent the observed values.

Supplementary Discussion

Discussion of current study findings in the context of published literature

In mouse studies, IFNα has been demonstrated to modulate TNFα expression[22] with data suggesting either an upregulation[23] or downregulation[24] of TNFα by IFNα. Our studies clarify that a predominant effect of sustained IFNα in human HSPCs is the downregulation of TNFα signaling via AP-1 and NF-κB, consistent with amelioration of the pro-inflammatory state in MPN and multiple sclerosis upon type 1 IFN administration.[25,26] The observed downregulation of the pro-inflammatory NF-κB activity in HSPCs and lymphoid differentiation shift may also be consistent with prior observations that patients with deficiencies in type 1 IFN response (due to germline inborn errors of type 1 IFN or antibodies against IFNα) exhibit severe and life-threatening COVID-19.[27-29] The deficiency in type 1 IFN response was associated with elevated TNF and IL6 mediated via NF-kB.[27] While the primary role of type 1 IFN may be to control the virus itself, the exaggerated pro-inflammatory NF-kB-mediated response may also potentially be due to the lack of anti-inflammatory effects of IFNα. Coherently, *TNF* expression was demonstrated to be decreased in patients with MPN who received IFNα treatment.[30] Downregulation of NF-kB pathway in HSCs and upregulation of a similar NF-kB-mediated IMP program revealed that HSC heterogeneity may underlie polarized anti- and pro-inflammatory states to regulate specialized immune function, whereas prior studies highlighted distinct lineage outputs downstream of HSC heterogeneity.[31,32]

The pro-fibrotic TGFβ signaling was also broadly downregulated across progenitor subsets, through a coordinated downregulation of the *TGFB1* gene itself and upregulation of the TGFβ signal inhibiting TGIF2 activity. In this way, IFNα downregulated two key cellular programs of NF-κB and TGFβ involved in MPN pathology, potentially underlying improved disease states. Moreover, as NF-κB and TGFβ signaling are both implicated in HSC quiescence,[33-36] the anti-inflammatory effects of IFNα were also linked with HSC exit from quiescence. The robust upregulation of TGFβ signaling in HSCs following resolution of an acute IFNα exposure supported data in mice of re-entry into quiescence to protect from HSC depletion following inflammation-induced cell cycle entry.[23]

Other cytokines, such as TNFα, IL-1, and IFNγ, have been demonstrated to induce granulo-monocytic differentiation.[37-41] IFNα presents as a unique cytokine among the inflammatory milieu to balance the granulo-monocytic differentiation with its positive regulation of B-lymphoid differentiation. While various differentiation skewing by IFNα has been reported in mice,[42-45] the interpretation of the results is complicated by the alteration of the HSC-marker Sca-1 upon IFNα exposure in mice (i.e. induction of Sca-1 in the CMPs, GMPs and MEPs, resulting in their inclusion within the Sca-1$^+$, Kit$^+$ HSC/multipotent progenitor compartment).[23] A study in newborn mice identified that IFNα treatment impaired lymphoid differentiation at pro-B and pro-T-cell stages, potentially due to impaired IL-7 signaling[45]. In our data, IL-7 signaling was not impaired by IFNα treatment. Such paradoxical differences may potentially be due to experimental setup-related

distinctions, such as differing developmental stages (newborn versus adult) at which hematopoiesis was probed, reflecting the highly context dependent nature of IFNα effects.

The ability of IFNα to upregulate PU.1 expression *in vitro* suggested that the differentiation modulating effects of IFNα may be largely direct. Moreover, as the three genetic clones from IFN02 share the same microenvironment, the similarity of the intrinsically activated IFN signaling in the double mutant cells to extrinsic IFNα effects suggested that the effects of IFNα on HSCs may be largely direct rather than secondarily mediated by downstream IFNα effects. These findings are consistent with the landmark study from Essers et al. that elegantly demonstrated the ability of IFNα to directly activate HSCs into cell cycle entry, via an *in vivo* cell mixing study in which only a small minority of the bone marrow cells harbored intact IFNα receptors.[46] Differential gene expression analysis between treated and baseline HSCs identified downregulation of CXCR4 and upregulation of MYC pathway as potentially underlying IFNα-induced HSC cell cycle entry. Enhanced MYC activity was also observed in mice HSCs exposed to type 1 IFN.[47]

Differential HSC clone activation into cell cycle entry could not fully explicate the IFNα-induced clonal dynamics. These findings suggested that the relative fitness of clonal HSCs via programs already in place at homeostasis produced no significant net difference upon therapy. Our data pointed to the differential rate of IMP development as an avenue to perturb clonal dynamics, which may be generalizable to other early neoplastic hematopoietic states, such as CH. *DNMT3A*-mutated HSCs upregulated IMP^{DN} signature, suggesting *DNMT3A*-mutated HSCs may be poised to resist IMP development. These results were consistent with a prior study in a mouse model of *DNMT3A*-mutated clonal hematopoiesis subjected to *M. avian* infection that elicited an IFN-driven inflammatory response.[48] Although IMPs were not specifically identified, *Dnmt3a*-knocked out HSCs resisted differentiation and expressed reduced levels of IMP transcription factors, such as AP-1 subunits and *Nr4a1*.[48]

Limitations and future directions

All IFNα-treated samples included IMPs except for IFN02 Year 2 and IFN07 Year 1 samples due to the small number of overall $CD34^+$ cells captured (541 and 346, respectively). However, the frequency of IMP was heterogeneous across patient samples, potentially due to patient-specific factors (e.g. co-mutations, treatment history, prior infections, clinical response, etc). The following results from our data suggest that IMPs may represent an alternate path of myeloid differentiation distinct from GMPs: (1) the transcription factor network of IMPs was most similar to that of a subset of HSCs (HSC1) and markedly distinct from the GMP state under the same IFNα-activated microenvironment; (2) relatedly, RNA velocity measurements predicted their derivation from HSC1; and (3) by chromatin state, IMPs could be traced contiguously from HSCs and are discontiguous from GMPs. These data provide useful corroborating evidence, as several known cell state transitions of this blood developmental system are represented in the UMAP embeddings of scRNA-seq and snATAC-seq data and are predicted by RNA velocity measurements. However, to definitively determine the origins and differentiation trajectories of IMPs, we would require a means to genetically lineage trace these cells (e.g. via lentiviral introduction of molecular barcodes[49,50]). Toward this end, future directions include developing an *in vivo* model of IMPs (mouse or xenograft) and genetically tracing their lineage trajectory. Developing an *in vivo* model of IMPs would also enable us to determine the mechanisms of IMP development and identify

whether IFNα directly induces IMPs in a subset of poised HSCs or IMP generation is mediated by another molecule with or subsequent to IFNα activation. Such an *in vivo* model would also facilitate a platform to causally determine the relationship between clonal dynamics and IMP differentiation rates.

Our results suggested that inflammatory neutrophils may be generated from poised HSCs, whereas neutrophils are typically thought to be polarized into an inflammatory state after lineage-commitment, not simultaneously. The upregulation neutrophil lineage markers (e.g. *MPO*, *CSF3R*) and downregulation of antigen presenting molecules suggested a neutrophilic differentiation bias of IMPs. Concerted expression and chromatin accessible binding sites of immediate early response transcription factors and emergency granulopoiesis transcription factors (CEBPB/D), nominated IMPs as a means to rapidly generate inflammatory neutrophils (which harbor the IMP transcription factor signature). Single cell colony output of neutrophil only colonies enriched in IFNα-treated bone marrow suggested an early uni-lineage commitment by HSCs, supporting this model. As additional support for a neutrophilic-biased differentiation program of IMPs, overexpression of *RFX3* induced an IMP-like population and increased CFU-granulocytic colonies. However, lineage tracing methods via an *in vivo* model of IMPs would allow us to trace the lineage outputs of IMPs and the relative contributions of neutrophils from IMPs versus GMPs. Nonetheless, the identification of the IMP population itself indicates that HSCs traverse an alternate differentiation course. Further studies are also required to determine the role of IFNα-induced inflammatory neutrophils (Neu1). Identification of IMPs highlighted HSCs as a potential source of neutrophil heterogeneity that has recently been appreciated in modulating cancer development and immunotherapy response.[51-53] While its physiologic relevance to clinical response data is complicated by the overall reduction of neutrophil counts by IFNα,[54] we may posit that Neu1 may contribute to the known side effects of IFNα therapy, such as fever, malaise and other flu-like symptoms, as well as exacerbation or precipitation of autoimmune disorders, such as vasculitis or lupus nephritis wherein neutrophils may play a key role.[55-58] IMPs and Neu1 may also contribute to the pathologic processes of inflamm-aging and auto-inflammatory disorders (especially type 1 interferonopathies) in non-neoplastic conditions.

It is possible that IFNα may select for clones that are already lymphoid biased, particularly in the *CALR*-wildtype HSC compartment. The selection of the *CALR*-double mutant clone from patient IFN02 that displayed marked megakaryocytic differentiation bias and its concurrent shift toward lymphoid differentiation upon IFNα demonstrated that IFNα is able to shift the differentiation trajectories of HSCs. Due to the technical limitations of tracking the clones within the wildtype HSC pool, we cannot exclude the possibility that IFNα is concurrently selecting for HSC clones that are already lymphoid biased. Another parameter we could not rigorously assess was the contribution of apoptosis to clonal dynamics in primary HSCs, due to the reservation of these rare samples for the single-cell multi-omics studies and the challenges of deconvoluting biological versus technical sources of cell death in these cryopreserved samples.

Limitations of our study include the restricted functional studies to determine the sufficiency and requirements of key factors such as PU.1 and RFX3 in mediating IFNα-induced phenotypes. While the role of PU.1 in regulating lymphoid differentiation is well-established, genetic loss-of-function studies of PU.1 in primary HSCs would establish its requirement in lymphoid differentiation upon IFNα treatment. Furthermore, while RFX3 appeared to play a key role for IMP differentiation downstream of the granulo-monocytic versus lymphoid divide, *RFX3* overexpression upregulated

IL-7R, suggesting a possible impact on lymphoid differentiation. Further studies are required to determine other potential roles of RFX3 in hematopoiesis.

Supplementary figure legends

Extended Data Fig. 1. GoT with immunophenotyping reveals inflammatory myeloid progenitors. a. Platelet counts at baseline and one year of treatment for 7 patients with paired samples included in the study (top) and 12 patients whose samples were not available (bottom) from MPN-RC-111/112 trials. **b.** UMIs (left), genes (middle), genotyping UMIs (right) per cell in $CD34^+$ cells from 13 patients after filtering based on quality control (QC) metrics. **c.** Representative time-point assignment by barcode demultiplexing (n = 8,696 cells; top); HTO expression for baseline and IFNα-treated (bottom). **d.** UMAP of $CD34^+$ cells (n = 65,452 cells, samples from this study and Nam et al., 2019) highlighted by patient ID (n = 13, top) and treatment status (bottom), insets show IMPs. **e.** UMAP highlighting enrichment of treated versus baseline (top) and MUT versus WT (bottom). **f.** UMAP of $CD34^+$ cells from Nam et al., 2019 (n = 14,872 cells) highlighted by patient ID (n = 3, top) and cell type clusters (bottom). **g.** UMAP of $CD34^+$ cells of 10 patients from MPN-RC-111/112 trials (n = 50,580 cells) labeled by patient ID (top) and cell type (bottom). **h.** Top differentially expressed genes per cell type after down-sampling (100 cells per cluster). **i.** Median scaled expression of HSPC protein markers from representative IFN03. **j.** Cell type-specific gene expression. **k.** UMAP of $CD34^+$ HSPCs (n = 65,452 cells) highlighting expression of HSC-defining markers. **l.** IMP percentage of $CD34^+$ cells from 7 paired samples at baseline (n = 17 cells) and IFNα (n = 98 cells). **m.** Cell numbers at baseline and IFNα treatment per sample (n = 11 baseline samples, n = 13 IFNα). P-value from two-sided Wilcoxon rank sum test. **n**. Differentially expressed genes between IFNα-treated GMPs and IMPs. P-values from two-sided likelihood ratio test of linear mixed models (LMM) with/without cluster identity (Methods). Blue and purple represent genes enriched in TNFα signaling via NF-κB and MYC targets, respectively. For **a** and **l**, P-values from two-sided likelihood ratio test of LMM with/without treatment status. For all figures, box plots show median (center line), interquartile range (box), 1.5 × IQR (whiskers), and outliers (points).

Extended Data Fig. 2. IFNα induces lymphoid differentiation and coordinates anti- and pro-inflammatory programs. a. Cell type percentages at baseline and IFNα (n = 14 paired samples from 7 individuals). **b.** IL-7 pathway module expression per cell. **c.** For each cell type, cell cycle gene expression (left subpanel with representative IFN01, cell numbers in Supplementary Table 2), and percentages of cells in G2/M/S phase (right subpanel; n = 9 baseline and 7 treated samples). **d.** Percentages of cells in G2/M/S phase at baseline and IFNα (n = 10 paired samples from 5 individuals). **e.** HSC module expression (based on differentially expressed genes in HSC cluster, see Extended Data Fig. 2a) and canonical stem/progenitor marker protein expression. P-values from two-sided F-test, Pearson correlation. **f.** UMAP of $CD34^+$ HSPCs overlaid with cell types from representative IFN04 (n = 7,282 cells). **g.** UMAP of IFN02 $CD34^+$ cells highlighting treatment status (n = 4,326 baseline cells, 2,912 IFNα-treated cells at year 1, and 541 IFNα-treated cells at year 2, left). Transcriptional distance measurements (Euclidean distance of the first thirty principal components) between HSCs from each time-point and HSCs at baseline (n = 1061 baseline, 725 IFNα year 1, 93 IFNα year 2, right). P-values from Wilcoxon rank sum test, two-sided. **h.** Differentially expressed genes between baseline and IFNα-treated MkPs. Genes enriched

in the TGFβ signaling (black) and in the IFNα/γ response (orange). **i.** CD184 expression in HSCs, GMPs, MLPs, MEPs and MDPs at baseline and IFNα (n = 5319 baseline, 3769 IFNα cells). **j.** Differentially expressed genes between baseline and HU-treated HSPCs highlighting genes in cellular response to DNA damage (light-blue), $IFN\alpha^{UP}$ genes (teal) and $IFN\alpha^{DN}$ genes (purple). **k.** Pre-ranked GSEA comparing baseline and HU-treated HSPCs. **l.** Heatmap showing pre-ranked GSEA comparing baseline and during IFNα treatment in HSCs using ImmuneSigDB gene sets (pathways description, right). **m.** Pre-ranked GSEA comparing IFNα treatment versus baseline in cDCP. Genes in teal are those upregulated in IMP versus GMP. FDR-adjusted P-values from an adaptive multilevel split Monte Carlo scheme (pre-ranked GSEA). Unless otherwise noted, P-values from two-sided likelihood ratio test of LMM with/without treatment status.

Extended Data Fig. 3. IMPs represent an alternative differentiation pathway. a. UMAPs showing transcription factor expression. **b.** UMAP highlighting IMP, GMP and HSC subclusters with RNA velocity-based trajectory for IFNα-treated cells with directionality derived from PAGA (left) and RNA velocity values (right). **c.** Differentially expressed genes between IFNα-treated IMPs and HSC1. In blue, genes enriched in TNFα signaling via NF-κB and in green, genes in G2M checkpoint (box representation as Fig. 1e). **d.** Pre-ranked GSEA of G2M checkpoint and E2F targets. **e.** Cell cycle gene expression in IFNα-treated HSC1 (n = 183 cells) versus IMPs (n = 136 cells). **f.** Proposed model of HSC versus HSC1 differentiation (top-left). UMAPs from flow-cytometric data showing representative colonies from single-cell differentiation assay (bottom-left). Normalized proportion of colonies (top-right) and frequency (bottom-right) including early myeloid, mixed granulocytic monocytic (GM), mixed granulocytic, monocytic, early myeloid (GEMM), and neutrophil only from sorted baseline (n = 3) and IFNα-treated (n = 4) samples. **g.** UMIs (top panel) and gene counts (bottom panel) per cell in immune cells after quality control filtering. **h.** UMAP of immune cells (n = 59,912 cells) overlaid with patient ID (top; n = 7 patients) and treatment status (bottom). **i.** Heatmap of top differentially expressed genes per cell type after down-sampling to 75 cells per cluster. **j.** UMAP of $CD34^+$ and $CD34^-$ cells from representative samples (n = 4 individuals, IFN09-IFN12) overlaid with cell type assignment (top) and RNA velocity vectors (bottom). **k.** Expression of IFNα response, TGFβ pathway and TNFα signaling (Hallmark) and HSC1 versus HSC2/3 signature in baseline, during and post-IFNα in HSC2/3 (n = 64 baseline, 325 IFNα active, 412 post-IFNα cells), HSC1 (n = 6 baseline, 19 IFNα active, 21 post-IFNα cells) and IMPs (n = 15 IFNα active, 22 post-IFNα cells) from representative IFN05. P-values from two-sided likelihood ratio test of LMM with/without treatment status. **l.** Aging-signature upregulated (left) and downregulated gene expression (right) in IFNα-treated HSC2/3 (n = 4648 cells), HSC1 (n = 183 cells) and IMPs (n = 136 cells). Unless otherwise stated, P-value from two-sided likelihood ratio test of LMM with/without cluster identity.

Extended Data Fig. 4. GoT-ATAC captures genotyping, snRNA-seq and snATAC-seq data in $CD34^+$ HSPCs from MPN patients treated with IFNα. a. Box plots showing number of UMIs (left) and genes (right) detected per cell in sorted $CD34^+$ hematopoietic stem and progenitors from each patient (n = 4) after filtering based on quality control (QC) metrics (Methods) from GoT-ATAC experiments. Box plots show the median (center line), the interquartile range (box), whiskers extending to the most extreme values within 1.5 × IQR, and outliers as individual points. **b.** Density plot comparing percentage of snATAC fragments within peaks to the total number of

fragments detected per sample (n = 7 samples from 4 individuals, additional IFN03 IFNα-treated cells also sequenced separately). **c.** Distribution of mean TSS enrichment score at each position relative to the TSS per sample. **d.** Average distribution of fragment length per sample. **e.** Heatmap showing HTO expression level for baseline and IFNα-treated cells from representative IFN01 (n = 1,471 cells). **f.** Genotyping UMI counts detected per cells. **g.** Correlation of the pseudobulk VAF calculated based on all genotyped $CD34^+$ HSPCs by single cell genotyping and the bulk DNA VAF on unfractionated peripheral blood obtained during the clinical study. P = 0.0105 from two-sided generalized linear model, Spearman R = 0.5. Shading denotes 95% confidence interval. **h.** UMAP of sorted $CD34^+$ stem and progenitors (n = 23,137 cells, 7 samples from 4 individuals), with cell type (left), patient ID (middle) and treatment status (right) using weighted-nearest neighbor (WNN) analysis of snRNA-seq and snATAC-seq data (Methods). **i.** Heatmap of top 15 differentially expressed genes for each HSPC type. Cells of each progenitor type were downsampled to the same number (n = 100 cells per cluster). **j.** Dot plot showing expression levels of cell type specific gene markers in each progenitor subset. **k.** UMAP based on weighted nearest neighbor (WNN) analysis (n = 23,137 cells) highlighting transcription factor motif accessibility. Transcription factor accessibility scores added with AddChromatinModule function in Signac. **l.** Heatmap showing cell type specific transcription factor accessibility scores.

Extended Data Fig. 5. Transcription factor activities in inflammatory myeloid progenitors. **a.** Ranked transcription factor motifs in the most significant individual positive regulatory peak of RFX3 (top) and RFX2 (bottom, first peak: upstream of TSS, second peak: downstream of TSS) identified using motif scanning with FIMO for a representative sample (IFN03; Methods). transcription factors ordered based on position of motifs in peak for all tracks in the figure. P-value obtained from FIMO and FDR-corrected. **b.** Chromatin accessibility track (left) of the regulatory region of *RFX3* (representative example from IFN03, bottom). Violin plots (right) display gene expression level of *RFX3*. **c.** Chromatin accessibility tracks of regulatory regions of *RFX2* (representative example from IFN03, bottom) and distal region enriched with the two most significant positively regulating loci (top-left, top-right). Violin plots display gene expression level of *RFX2*.

Extended Data Fig. 6. RFX3 is a key transcription factor of IMP development. a. Chromatin accessibility tracks of regulatory regions of *HLA-DRA1* (bottom), distal region enriched with negative regulatory loci (inset), and IMP-specific regulatory locus (top, representative example from IFN07). Violin plots display gene expression level of *HLA-DRA1*. transcription factors ordered based on position of motifs in peaks. **b.** Box plots showing normalized expression of HLA DR protein expression per cells. P-values from two-sided likelihood ratio tests of LMM with/without cell type identity. Cells numbers included in Fig. 1c legend. **c.** *RFX2* mRNA levels (right, n = 1 experiment) and *RFX3* mRNA levels (left, n = 2 independent experiments) in CFU colony cells grown from transduced *RFX2*-overexpressing or *RFX3*-overexpressing $CD34^+$ cells compared to matched controls. mRNA levels calculated by RT-PCR and correspond to *RFX2* or *RFX3* Ct values normalized to TBP Ct values. Data are presented as mean ± SD **d.** Normalized percentage of erythroid (BFU-E), granulocytic (CFU-G), granulo-monocytic (CFU-GM), monocytic (CFU-M) and myeloid (CFU-GEMM) colonies grown from *RFX2*-overexpressing or *RFX3*-overexpressing $CD34^+$ umbilical cord blood (UCB) cells in methylcellulose-based CFU assays compared to matched control $CD34^+$ UCB cells. P-values from two-sided likelihood ratio

test linear-mixed modeling (LMM) with/without *RFX3* overexpression. Data are presented as mean ± SEM. **e.** Box plots showing number of UMIs (left) and genes (right) detected per cell in sorted $CD34^+$ cord blood cells after filtering based on quality control (QC) metrics (n = 23137 cells). **f.** UMAP of sorted $CD34^+$ cord blood cells (n = 2,609 cells) highlighted by transduction status. **g.** UMAP of *RFX3*-overexpressing and mCherry (control) subsets (n = 1,520 cells) highlighted by their status. **h.** UMAP of *RFX3*-overexpressing and control $CD34^+$ cells showing gene expression levels of cell type-specific gene markers for HSPCs. **i.** UMAP of *RFX3*-overexpressing and control, highlighting cell type assignments.

Extended Data Fig. 7. PU.1 is a master regulator of IFNα-mediated lymphoid differentiation. **a.** Frequencies of B-lymphoid progenitors and B-cells from bone marrow with early-stage MPN treated with IFNα (n = 9) or HU (n = 10) or untreated (n = 33) by multiparametric flow cytometry. P-values from Wilcoxon rank sum test, two-sided. **b.** B-cell frequencies in G2/M/S phase at baseline and IFNα (n = 4 patients, n = 7 samples (4 baseline, 3 treated). For all bar plots, bars represent the mean, and error bars indicate SEM. **c.** Normalized cell percentages of progenitor subsets at baseline and IFNα (left). Cells from each treatment status and individual were downsampled to 100 cells. Cell frequencies for each patient (right). **d.** Normalized expression of *TCF3* (top) and *EBF1* (bottom) per cell at baseline and IFNα (Year 1). **e.** Chromatin accessibility tracks of regulatory regions of *SPI1* (representative IFN07, bottom-left), distal region enriched with two most significant positively regulating loci (top-left). Violin plots display *SPI1* expression. Ranked transcription factor motif enrichment of positive regulatory peaks of *SPI1* across three samples; P-values from hypergeometric test (right, Methods). **f.** *SPI1* expression in HSCs, GMPs, MLPs, MEPs and MDPs at baseline (n = 6,256) and IFNα (n = 4,575). **g.** Normalized mRNA levels in K562 treated with IFNα with/without IFNAR1-antibody *in vitro* for 24 hours (RT-qPCR, n = 6). **h.** Representative western blot showing PU.1 protein levels in K562 treated with IFNα *in vitro* (left). Log-fold change of PU.1 protein levels based on luminescence intensity (n = 4, right). **i.** STAT5 pathway module expression (Wierenga_STAT5A_targets_group2) in WT (n = 2,652) and MUT HSC (n = 1,996). P-values from two-sided likelihood ratio test of LMM with/without genotype status. **j.** Proportion of MkPs in complete or partial response (n = 15 timepoints from 8 patients). P-value from two-sided Wilcoxon rank sum test. **k.** Mean *SPI1* expression in early stem and progenitors (HSC, IMP, MEP, MLP) versus platelet counts. P-value from likelihood ratio test of LMM with/without mean *SPI1* expression, shading denotes 95% confidence interval. Unless otherwise stated, P-value from two-sided likelihood ratio test of LMM with/without treatment status.

Extended Data Fig. 8. *CALR* mutations modify the effects of IFNα signaling. **a.** UMAP of $CD34^+$ cells at baseline and IFNα with *CALR* mutation status. Cells were downsampled to 1,000 cells per mutation status and sample. **b.** Scatter plot showing pathways from pre-ranked GSEA comparing MUT versus WT cells at baseline and after IFNα treatment. Values show the sign of the NES*-log10(Adjusted P-value). Pathways in red are present in HSCs. P-value from two-sided generalized linear model, Pearson correlation. **c.** Pre-ranked GSEA of differentially expressed genes between MUT and WT cells at baseline and after IFNα treatment. FDR-adjusted P-values from an adaptive multilevel split Monte Carlo scheme (pre-ranked GSEA). **d.** Platelet counts relative to percentages of MUT MEPs and MkPs. P-value from two-sided F-test, Pearson

correlation. Shading denotes 95% confidence interval. **e.** Normalized mutant cell percentage at baseline versus IFNα (only clusters with ≥10 genotyped cells per sample were included). **f.** Top five hits from ranked transcription factor motif enrichment of peaks from CUT&RUN of PU.1. P-values from Fisher's exact test (FDR-adjusted). **g.** Differential transcription factor motif enrichment between baseline and IFNα-treated, focusing on exclusive PU.1 peaks in WT (left) and MUT (right; HOMER). **h.** Difference in distal and proximal PU.1 peak numbers between control and IFNα-treated UT7-MPL. Peaks <500 bp from TSS considered proximal. **i.** Differential transcription factor motif enrichment between MUT and WT-exclusive PU.1 peaks. **j.** Normalized proportions of MUT and WT HSCs at each time-point (n = 8 individuals with at least 2 time-points). **k.** Percentages of MUT and WT cells in G2/M/S phase as assessed in Fig. 1k for all $CD34^+$ GoT-IM samples (n = 9 baseline and 7 IFNα year 1 samples, top) and paired samples (n = 12 samples from 5 individuals, bottom). P-values derived from two-sided likelihood ratio test of LMM with/without treatment status. **l.** Normalized *CXCR4* gene expression per cell in HSCs. P-values from two-sided likelihood ratio test of LMM with/without mutational status. **m.** $Ki67^+$ myeloid cell percentage in bone marrow sections from MPN patients with (n = 4) or without (n = 5) IFNα treatment. P-value from Wilcoxon rank sum test, two-sided.

Extended Data Fig. 9. IFNα perturbs clonal dynamics via IMP differentiation. a. Density of proportion of Type 1-mutant reads versus SNV-mutant reads in UMIs captured via GoT. **b.** Normalized frequencies of cell types among WT, single MUT (Type 1 *CALR*) and double MUT cell populations (n = 7,779 cells). **c.** Normalized proportion of baseline HSCs in cell cycle phases. P-values from Fisher's exact test between mutation status and cell-cycle entry status (G2/M/S vs G1). **d.** UMAP of $CD34^+$ cells highlighting cell types (left), treatment (middle) and mutation status (right). **e.** Genes differentially expressed between single mutant and double mutant HSCs at baseline (n = 890 genotyped HSCs). P-values from logistic regression model (Methods). Genes in blue are enriched in TNFα signaling via NF-κB, red in unfolded protein response, and orange in IFNγ response (box representation as Fig. 1e). **f.** Pre-ranked GSEA of differentially expressed genes between HSC clones at baseline. **g.** UMAP of HSCs and IMPs from IFN02 (n = 1,895 cells) overlaid with cell type (left), treatment + mutation status (middle), and module score for $IFN\alpha^{UP}$ genes (right). **h.** IMP gene signature scores in IFN02 cell types (n = 1879 HSC, 17 IMP, 605 GMP, top). P-values from Wilcoxon rank sum test, two-sided. Gene expression levels of upregulated transcription factors in IMPs in IFN02 (bottom). **i.** IMP-specific signature score in HSC clones at baseline in IFN02 (n = 319 WT, 391 MUT, 272 double MUT cells). P-values from Wilcoxon rank sum test, two-sided. **j.** Scatter plot of clone size change following IFNα versus difference in normalized frequency of MUT HSC1 and IMPs of all MUT cells. P-value from two-sided LMM, Pearson correlation. Shading denotes 95% confidence interval. **k.** Differentially expressed genes between MUT and WT in IMP in patients showing molecular response (left) and those whose mutated clones expanded (right). P-values from two-sided likelihood ratio test of LMM with/without genotype (Methods). Genes in blue enriched in TNFα signaling via NF-κB (* denotes overexpression in IMPs) and genes in green are in IFNα response gene set.

Extended Data Fig. 10. Genotyping in snATAC-seq (GoT-ChA) of *JAK2*-mutated primary MPN HSPCs at baseline and following IFNα therapy. a. Density plot comparing percentage of snATAC fragments within peaks to the total number of fragments detected per sample (n = 6

samples from 3 individuals). **b.** Distribution of mean TSS enrichment score at each position relative to the TSS per sample. **c.** Average distribution of fragment length per sample. **d.** Heatmap showing cell type specific transcription factor accessibility scores. **e.** Normalized cell proportion of progenitor subsets at baseline and after IFNα treatment from each GoT-ChA sample. **f.** Genotyping DNA fragment counts per cell. (n = 682 cells IFN14 baseline, n = 1188 IFN14 IFNα, n = 493 IFN15 baseline, n = 79 IFN15 IFNα, n = 1852 IFN16 baseline, n = 1536 IFN16 IFNα). **g.** Scatter plot showing correlation of the pseudobulk VAF calculated based on all genotyped HSPCs using GoT-ChA and the bulk DNA VAF on unfractionated peripheral blood collected during clinical trials. Spearman's R = 0.7. Shading denotes 95% confidence interval. **h.** Normalized cell type proportion of WT versus *JAK2* MUT cells at baseline and following IFNα-treatment from each GoT-ChA sample.

Supplementary References

1. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* **19**, 224 (2018).
2. Carey, C.D. *et al.* Topological analysis reveals a PD-L1-associated microenvironmental niche for Reed-Sternberg cells in Hodgkin lymphoma. *Blood* **130**, 2420-2430 (2017).
3. Patel, S.S. *et al.* Multiparametric in situ imaging of NPM1-mutated acute myeloid leukemia reveals prognostically-relevant features of the marrow microenvironment. *Mod Pathol* **33**, 1380-1388 (2020).
4. Nam, A.S. *et al.* Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **571**, 355-360 (2019).
5. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
6. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (Cornell University, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*, 2018).
7. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271-281 (2017).
8. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
9. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-96 (2016).
10. Weirauch, M.T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431-1443 (2014).
11. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021).
12. Masella, A.P. *et al.* BAMQL: a query language for extracting reads from BAM files. *BMC Bioinformatics* **17**, 305 (2016).

13. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-8 (2011).

14. Persikov, A.V. *et al.* A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res* **43**, 1965-84 (2015).

15. Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).

16. Greenwald, N.F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol* **40**, 555-565 (2022).

17. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).

18. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).

19. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).

20. Bale, M. jlabflow: DSL2 nextflow hub for Bulk Sequencing Assays for JLab. (GitHub).

21. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

22. Huys, L. *et al.* Type I interferon drives tumor necrosis factor-induced lethal shock. *J Exp Med* **206**, 1873-82 (2009).

23. Pietras, E.M. *et al.* Re-entry into quiescence protects hematopoietic stem cells from the killing effect of chronic exposure to type I interferonsIFN-1 effects on HSC function. *The Journal of Experimental Medicine* **211**, 245-262 (2014).

24. Sharif, M.N. *et al.* Twist mediates suppression of inflammation by type I IFNs and Axl. *The Journal of Experimental Medicine* **203**, 1891-1901 (2006).

25. Czech, J. *et al.* JAK2V617F but not CALR mutations confer increased molecular responses to interferon-α via JAK1/STAT1 activation. *Leukemia* **33**, 995-1010 (2019).

26. Schreiber, G. The Role of Type I Interferons in the Pathogenesis and Treatment of COVID-19. *Frontiers in Immunology* **11**, 595739 (2020).

27. Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718-724 (2020).

28. Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).

29. Bastard, P. *et al.* Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**(2020).

30. Tashi, T. *et al.* Pegylated interferon Alfa-2a and hydroxyurea in polycythemia vera and essential thrombocythemia: differential cellular and molecular responses. *Leukemia* **32**, 1830-1833 (2018).

31. Fast, E.M. *et al.* External signals regulate continuous transcriptional states in hematopoietic stem cells. *Elife* **10**(2021).

32. Rodriguez-Fraticelli, A.E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212-216 (2018).

33. Ohta, M., Greenberger, J.S., Anklesaria, P., Bassols, A. & Massagué, J. Two forms of transforming growth factor-beta distinguished by multipotential haematopoietic progenitor cells. *Nature* **329**, 539-41 (1987).

34. Stein, S.J. & Baldwin, A.S. Deletion of the NF-κB subunit p65/RelA in the hematopoietic compartment leads to defects in hematopoietic stem cell function. *Blood* **121**, 5015-24 (2013).

35. Zhao, M. *et al.* Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nat Med* **20**, 1321-6 (2014).

36. Scandura, J.M., Boccuni, P., Massagué, J. & Nimer, S.D. Transforming growth factor beta-induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation. *Proc Natl Acad Sci U S A* **101**, 15231-6 (2004).

37. Yamashita, M. & Passegué, E. TNF-α Coordinates Hematopoietic Stem Cell Survival and Myeloid Regeneration. *Cell Stem Cell* (2019).

38. Pietras, E.M. *et al.* Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. *Nat Cell Biol* **18**, 607-18 (2016).

39. Ueda, Y., Cain, D.W., Kuraoka, M., Kondo, M. & Kelsoe, G. IL-1R type I-dependent hemopoietic stem cell proliferation is necessary for inflammatory granulopoiesis and reactive neutrophilia. *J Immunol* **182**, 6477-84 (2009).

40. Matatall, K.A., Shen, C.C., Challen, G.A. & King, K.Y. Type II interferon promotes differentiation of myeloid-biased hematopoietic stem cells. *Stem Cells* **32**, 3023-30 (2014).

41. Chen, C., Liu, Y. & Zheng, P. Mammalian target of rapamycin activation underlies HSC defects in autoimmune disease and inflammation in mice. *J Clin Invest* **120**, 4091-101 (2010).

42. Mullally, A. *et al.* Depletion of Jak2V617F myeloproliferative neoplasm-propagating stem cells by interferon-α in a murine model of polycythemia vera. *Blood* **121**, 3692-702 (2013).

43. Haas, S. *et al.* Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. *Cell Stem Cell* **17**, 422-434 (2015).

44. Khan, N. *et al.* M. tuberculosis Reprograms Hematopoietic Stem Cells to Limit Myelopoiesis and Impair Trained Immunity. *Cell* **183**, 752-770.e22 (2020).

45. Lin, Q., Dong, C. & Cooper, M.D. Impairment of T and B Cell Development by Treatment with a Type I Interferon. *Journal of Experimental Medicine* **187**, 79-87 (1998).

46. Essers, M.A. *et al.* IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature* **458**, 904-8 (2009).

47. Ehninger, A. *et al.* Posttranscriptional regulation of c-Myc expression in adult murine HSCs during homeostasis and interferon-α-induced stress response. *Blood* **123**, 3909-13 (2014).

48. Hormaechea-Agulla, D. *et al.* Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNgamma signaling. *Cell Stem Cell* **28**, 1428-1442 e6 (2021).

49. Rodriguez-Fraticelli, A.E. *et al.* Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* **583**, 585-589 (2020).

50. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D. & Klein, A.M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020).

51. Gungabeesoon, J. *et al.* A neutrophil response linked to tumor control in immunotherapy. *Cell* **186**, 1448-1464 e20 (2023).

52. Hirschhorn, D. *et al.* T cell immunotherapies engage neutrophils to eliminate tumor antigen escape variants. *Cell* **186**, 1432-1447 e17 (2023).

53. Ng, M.S.F. *et al.* Deterministic reprogramming of neutrophils within tumors. *Science* **383**, eadf6493 (2024).

54. Barbui, T. *et al.* Neutrophil-to-Lymphocyte ratio as surrogate for JAK2V617F suppression and event-free survival in polycythemia vera. *Blood Cancer Journal* **15**, 132 (2025).

55. Samuelsson, J. *et al.* A phase II trial of pegylated interferon α-2b therapy for polycythemia vera and essential thrombocythemia. *Cancer* **106**, 2397-2405 (2006).

56. Masarova, L. *et al.* Pegylated interferon alfa-2a in patients with essential thrombocythaemia or polycythaemia vera: a post-hoc, median 83 month follow-up of an open-label, phase 2 trial. *Lancet Haematol* **4**, e165-e175 (2017).

57. Bennett , L. *et al.* Interferon and Granulopoiesis Signatures in Systemic Lupus Erythematosus Blood. *Journal of Experimental Medicine* **197**, 711-723 (2003).

58. Michailidou, D., Mustelin, T. & Lood, C. Role of Neutrophils in Systemic Vasculitides. *Front Immunol* **11**, 619705 (2020).