Type 1 interferon remodels normal and neoplastic hematopoiesis in human

Chhiring Lama[1], Shira Rosenberg[1], Andrea Kubas-Meyer[1], X. Amy Xie[2,3], Sara Moein[1], Neelang Parghi[1], Mansi Totwani[1], Mirca S. Saurty-Seerunghen[1], Mohamed Omar[1], Alicia Dillard[1], Nathaniel D. Omans[4,5,6], Neville Dusaj[4,5,7], Paulina Chamely[4,5], Eleni Mimitou[8], Peter Smibert[9], Heidi E. Kosiorek[10], Amylou C. Dueck[11], Rona Weinberg[12], Ronan Chaligne[13], Luigi Marchionni[1], Sanjay Patel[1], Paul Simonson[1], Dan A. Landau[4,5,6], Ronald Hoffman[13,14], Anna S. Nam[1,2,6]*

[1]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine; New York, NY, USA.
[2]Biochemistry, Structural Biology, Cell Biology, Developmental Biology and Molecular Biology Graduate Programs, Weill Cornell Medicine; New York, NY, USA.
[3]Memorial Sloan Kettering Cancer Center; New York, NY, USA.
[4]New York Genome Center; New York, NY, USA.
[5]Division of Hematology and Medical Oncology, Department of Medicine and Meyer Cancer Center, Weill Cornell Medicine; New York, NY, USA.
[6]Institute for Computational Biomedicine, Weill Cornell Medicine; New York, NY, USA.
[7]Tri-Institutional MD-PhD Program, Weill Cornell Medicine, Rockefeller University, Memorial Sloan Kettering Cancer Center; New York, NY, USA.
[8]Immunai; New York, NY, USA.
[9]10x Genomics; Pleasanton, CA, USA.
[10]Department of Biostatistics, Mayo Clinic Arizona; Scottsdale, AZ, USA.
[11]Department of Quantitative Health Sciences and Alliance Statistics and Data Management Center, Mayo Clinic, Rochester, MN, USA.
[12]New York Blood Center; New York, NY, USA.
[13]Myeloproliferative Neoplasms Research Consortium; New York, NY, USA.
[14]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai; New York, NY, USA.
*Corresponding author.

Corresponding author contact details:

Anna S. Nam, MD
Weill Cornell Medicine
Weiss Research Building
Rockefeller University Campus
1230 York Ave
New York, NY 10021
Email: shn9035@med.cornell.edu

Keywords
Single-cell, RNA-seq, multi-omics, type 1 interferon, hematopoiesis, myeloproliferative neoplasm

Abstract

Inflammatory cytokines perturb hematopoietic stem cell (HSC) homeostasis and modulate the fitness of neoplastic HSC clones in mouse models. However, the study of cytokines in human hematopoiesis is challenging due to the concerted activities of multiple cytokines across physiologic and pathologic processes. To overcome this limitation, we leveraged serial bone marrow samples from patients with *CALR*-mutated myeloproliferative neoplasms who were treated with recombinant interferon-alpha (IFNa). We interrogated baseline and IFNa-treated $CD34^+$ stem and progenitor cells using single-cell multi-omics platforms that directly link, within the same cell, the mutation status, whole transcriptomes and immunophenotyping or chromatin accessibility. We identified a novel IFNa-induced inflammatory granulocytic progenitor defined by expression and activities of RFX2/3 and AP-1 transcription factors, with evidence supporting a direct differentiation from HSCs. On the other hand, IFNa also induced a significant B-lymphoid progenitor expansion and proliferation, associated with enhanced activities of PU.1 and its co-regulator TCF3, as well as decreased accessibility of megakaryocytic-erythroid transcription factor GATA1 binding sites in HSCs. In the neoplastic hematopoiesis, the lymphoid expansion was constrained by a preferential myeloid skewing of the mutated cells, linked with increased myeloid proliferation and enhanced CEBPA and GATA1 activities compared to wildtype cells. Further, IFNa caused a downregulation of the TNFa signaling pathway, with downregulation of NFKB and AP-1 transcription factors. Thus, IFNa simultaneously initiated both – pro-inflammatory and anti-inflammatory – cell states within the same hematopoiesis, and its phenotypic impact varied as a function of the underlying HSC state and mutation status.

Introduction

Inflammation perturbs hematopoietic stem cell (HSC) homeostasis. The isolated effects of individual inflammatory cytokines, including type 1 interferons (IFNa/b)[1-3], type 2 IFN (IFNg)[4-6], TNFa[7,8], and IL-1[9-11], have been well-documented in mouse hematopoietic development. These cytokines display overlapping features, such as inducing cell cycle entry of HSCs and their differentiation, frequently toward the granulo-monocytic lineage[1,4-6,12-15]. However, these cytokines also feature distinct roles in the inflammatory milieu. Recent reports, for example, link germline genetic defects in type 1 IFN with exaggerated or severe COVID-19 response associated with neutrophilia[16,17]. However, deciphering the contributions of individual cytokines in human hematopoiesis is challenging due to the concerted activities of multiple cytokines in the setting of infection, inflammatory disorders or cancer. Moreover, HSC-enriched bone marrow biopsies are rare, as they are restricted to patients with suspected or diagnosed hematopoietic neoplasms or defects.

The use of recombinant IFNa for the treatment of patients with myeloproliferative neoplasms (MPN) therefore presents a unique opportunity to study the effects of isolated type 1 IFN in human HSCs. MPNs are often driven by a single somatic mutation in *CALR*, *JAK2* or *MPL*, and thus provide an informative model of an early clonal process within the hematopoietic system[18]. Clinically, these patients present with an overproduction of one or more myeloid cell lineages. IFNa remains a highly effective therapy for MPN, frequently resulting in the normalization of the patients' blood counts and even molecular response (i.e. reduction in variant allele frequencies), with evidence suggesting that IFNa may deplete clonal stem cells via induction of HSCs into cell cycle entry and exhaustion[12,18]. However, the effects of IFNa on HSCs appear to be genotype-specific as IFNa frequently induces molecular response in patients with *JAK2*-mutated MPN but not in those with *CALR*-mutated MPN[19]. Thus, the molecular effects undergirding the clinical improvements of *CALR*-mutated MPN through IFNa therapy remain an open arena of investigations.

While MPNs provide a useful model to study the effects of somatic mutations on human hematopoiesis, the precise examination of IFNa effects on mutated versus wildtype blood formation is challenging because the mutated cells are not distinguishable from the admixed wildtype cells by cell surface markers. We therefore leveraged single-cell multi-omics platforms that detect the mutation status and whole transcriptomes with immunophenotyping or chromatin

accessibility data, within the same thousands of cells. These innovative platforms build upon a method called Genotyping of Transcriptomes (GoT), in which we modified a high-throughput single-cell RNA-seq (scRNA-seq) platform to capture somatic genotype information within the same cells[20]. These methods allow us to overlay two hematopoietic differentiation landscapes—one mutated and the other wildtype—from the same individual, thus enabling a direct comparison between mutated and wildtype cells, before and after treatment. Focusing on *CALR*-mutated MPN, we applied these multi-modality single-cell methods to $CD34^+$ hematopoietic stem and progenitor cells (HSPC) from serial bone marrow sampling from patients who were treated with IFNa for at least one year, enabling us to define the transcriptional and epigenetic alterations induced by IFNa in normal and neoplastic human hematopoiesis.

Results

GoT-IM captures cell identity, somatic genotyping and treatment status information for thousands of $CD34^+$ cells from MPN patients treated with IFNa

To determine the *in vivo* effects of isolated IFNa on human normal and neoplastic hematopoiesis, we leveraged the GoT technology that simultaneously captures the mutation status and whole transcriptomes in thousands of single cells[20]. To overcome inter-patient variabilities in IFNa response, we applied GoT to FACS-isolated $CD34^+$ cells from serial (i.e. baseline and treated) bone marrow from individuals who were diagnosed with *CALR*-mutated essential thrombocythemia (ET), a subtype of MPN that exhibits increased megakaryopoiesis and platelet counts[21] (**Fig. 1a**). As these serial bone marrow specimens are exceedingly rare in clinical practice, we utilized cryopreserved specimens from our clinical trials MPN-RC-111 and -112 wherein patients were treated weekly with a pegylated form of IFNa[22,23] (**Fig. 1a**, n = 5 individuals, 4 baseline, 7 treated; additional five baseline samples were included from our previous work[20]; see **Supplementary Table 1** for patient and sample information). We further incorporated cell hashing[24] that enabled multiplexing baseline and IFNa-treated samples into the same scRNA-seq reactions via time-point specifying barcoded antibodies, to obviate technical batch effects between serial samples (**Fig. 1b**). We also advanced upon GoT by incorporating immunophenotyping (GoT-IM), through integration of Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq)[25], to link cell identities to cell surface protein expression (**Fig. 1b**). GoT provided genotyping data for the canonical *CALR* frameshift mutations for 81.5% of $CD34^+$

HSPCs (n = 40,668), consistent with our previously reported genotyping rates[20]. In this way, we obtained somatic genotyping, whole transcriptomes, immunophenotyping and treatment status for thousands of cells from the same GoT-IM experiment.

Integration of CD34$^+$ cells across baseline and IFNa treated samples reveals a novel inflammatory cell state

To determine the cell identities of the CD34$^+$ HSPCs consistently across samples and treatment status, we analytically separated the cells from each timepoint from the same individual into individual datasets (based on the Cell Hashing data), and integrated across all of the 16 samples based on the canonical correlation analytic framework[26] (n = 49,891 cells from 16 samples, **Fig. 1c, Supplementary Fig. 1a,b**). As single-cell gene expression provides high-resolution mapping of the HSPC identities, we clustered the cells based on gene expression data alone and annotated the clusters based on known canonical cell markers (**Supplementary Fig 2a,b**). To precisely identify the HSCs, we leveraged immunophenotyping via CITE-seq to identify the CD38$^-$, CD45RA$^-$, CD90$^+$ HSCs (**Fig. 1d, Supplementary Fig. 2c**). Consistently, the RNA expression of the HSC gene marker, *AVP*, was also elevated in this population (**Fig. 1d, Supplementary Fig. 2c,** see **Supplementary Table 2** for cell numbers). We observed the expected cell types, such as megakaryocytic progenitors (MkP), immature myeloid progenitors (IMP, consisting predominantly of phenotypically-defined common myeloid progenitors (CMP) and granulo-monocytic progenitors (GMP)[27]), and lymphoid progenitors (Pro-B, Large Pre-B, Small Pre-B cells). We also identified cell types not previously described in scRNA-seq datasets of human CD34$^+$ cells. Namely, we identified monocytic dendritic progenitors (MDP) which were distinct from the uni-lineage monocytic (MP), classic dendritic (cDCP) or plasmacytoid dendritic progenitors (pDCP). They were assigned as such based on high *SOX4* and *RUNX3* expression (**Supplementary Fig. 2a,b**), which were characteristic of the MDP population in mice[28]. Transcriptionally, the MDPs were closely related to multipotent lymphoid progenitors (MLPs, **Fig. 1c**), consistent with data demonstrating that MLPs can give rise to monocytes and dendritic cells, along with lymphoid cells, in human[29].

We also identified an unknown cluster (Cluster X), adjacent to the HSCs in the UMAP space (**Fig. 1c**), that resembled the IMPs immunophenotypically based on CD38^{mid}, CD90^{lo}, and CD45RA^{mid} expression (**Fig. 1d**). To elucidate the identity of Cluster X, we performed differential expression analysis between Cluster X and IMPs and identified a striking upregulation of the immediate early

response transcription factors (TF) of the AP-1 (*JUN, FOS, JUNB, FOSB, ATF3, FOSL1, MAFF*), KLF (*KLF2, KLF4, KLF6*), and NR4A (*NR4A1, NR4A2*) families (**Fig. 1e left, Supplementary Fig. 2d**, **Supplementary Table 3,** linear mixed model that explicitly models the effects of patient batch and treatment status, see online methods). We also observed a robust upregulation of *RFX2* and *RFX3* TFs, not well characterized in hematopoietic stem and progenitor cells (**Fig. 1e left, Supplementary Fig. 2d, Supplementary Table 3**). Upregulation of *CEBPB* and *CEBPD* (**Fig. 1e left, Supplementary Fig. 2d, Supplementary Table 3**), implicated in emergency granulopoiesis[30] and granulopoiesis under cellular stress[31] respectively, suggested a differentiation state into the granulocytic lineage. Interferon regulatory factor 1 (*IRF1*) was also upregulated (**Fig. 1e left, Supplementary Table 3**), indicating that the cluster was associated with IFNa treatment. Gene set enrichment analysis identified the upregulation of TNFa via NFKB pathway (Adj. P-val = 3.5×10^{-8}, Hallmark), and downregulation of the MYC targets (Adj. P-val = 6.8×10^{-6}, Hallmark, **Fig. 1e right, Supplementary Table 4**). In light of these findings, this novel cluster was named inflammatory granulocytic progenitor (IGP).

Strikingly, the distributions of these differentially upregulated TFs were highly enriched in the most quiescent HSCs with the highest overall expression of *AVP* and CD90 (HSC1, **Fig. 1f,g, Supplementary Fig. 2c**), driving the transcriptional similarities of the IGPs and HSC1 as revealed by their proximity on the UMAP space. To determine changes in gene expression that may drive the differentiation from HSC1 to IGPs, we compared IGPs to HSC1 and identified a reinforcement of the AP-1 and KLF family TF expression and downregulation of NR4A TFs (**Fig. 1g,h, Supplementary Table 3,4;** notably, only IFNa-treated cells were included in the differential expression analysis to remove the treatment variable altogether). NR4A1/2 have been demonstrated to maintain HSC quiescence[32-34]. Their downregulation relative to HSC1 is therefore consistent with the upregulation of cell cycle-related genes in the IGPs (**Fig. 1h, Supplementary Table 3,4**). *MPO*, a neutrophil-specific gene, and *CSF3R*, encoding the G-CSF receptor, were upregulated while the MHC class II genes (*CD74, HLA-DRA, HLA-DPA1, HLA-DRB1, HLA-DPB1*) were downregulated, further providing evidence for its differentiation into a neutrophilic lineage versus monocytic or dendritic cell lineages (**Fig. 1h, Supplementary Table 3**). Although RFX2/3 are not well characterized in hematopoiesis, other RFX members such as RFX1 and RFX8 have been demonstrated to regulate MHC class II expression[35-37]. Thus, the strong upregulation of *RFX2/3* expression may be induced to rapidly downregulate MCH class II genes for granulocytic

differentiation upon IFNa signaling. Of the interferon regulatory factors, *IRF1* was upregulated in IGP compared to HSC1 (**Fig. 1h, Supplementary Table 3**; both comparison groups under IFNa treatment) and IMP (**Fig. 1e, Supplementary Table 3**), highlighting IRF1 as the key regulatory factor among other IFNa-induced TFs in IGP development. As only mononuclear cells were cryopreserved in these patient cohorts, precluding the assessment of neutrophils, we leveraged an available dataset of immune cells under inflammatory conditions[38] (in this case, COVID-19 infection). We identified that the IGP-specific gene signature was highly specific to neutrophils, providing evidence against a general inflammatory signature and further supporting the predicted lineage fate (**Fig. 1i**). Furthermore, as the IGPs constitute a small minority of the progenitors (<1%), IGPs are not likely to contribute significantly to overall neutrophil pool during IFNa therapy, consistent with the absence of neutrophilia in patients treated with IFNa. Transcriptional similarity of IGPs to the most quiescent HSC1 population, especially with respect to the TF program, provides evidence for a novel route of granulocytic differentiation that bypasses the intermediate IMP states upon IFNa stimulation.

IFNa induces IGPs and lymphoid differentiation shift and proliferation

Next, we sought to determine the phenotypic and transcriptional impact of IFNa on hematopoiesis overall. As individual GoT-IM experiments represent sampling from the same individual at different time points, we clustered cells from each individual separately based on the gene expression data and identified that IFNa exerted strong transcriptional perturbations in the treated cells compared to baseline HSCs (**Fig. 2a, Supplementary Fig. 3a**). In an example case of patient IFN03 for whom three time points were available – (1) baseline, (2) treated for one year, and (3) treated for four years but off therapy for 3 weeks at time of collection – HSCs at year 1 displayed a striking transcriptional difference compared to baseline cells, whereas cells that have been off therapy for 3 weeks at year 4 were less distinct, consistent with a half-life of the pegylated-IFNa of ~2 weeks (**Fig. 2a**). Projection of the cell identity assignments (as in **Fig. 1c**) revealed that the cells clustered based on cell identity as well as based on treatment status (**Fig. 2a, Supplementary Fig. 3b**). These findings contrasted with the more subtle transcriptional impact of *CALR*-mutations, since mutated and wildtype cells are co-mingled throughout differentiation[20,39]. We previously demonstrated that the whole transcriptomic data alone could not distinguish mutated versus wildtype cells in early clonal processes such as *CALR*-mutated MPN[20] (also confirmed in *JAK2*-mutated MPN[40,41] and *DNMT3A*-mutated clonal hematopoiesis[42]), necessitating the

development of GoT. Given the strong transcriptional impact of IFNa, we focused on deciphering the effects of IFNa on the overall HSC population first, before dissecting the differential impact on mutated versus wildtype cells.

To determine how IFNa may impact the hematopoietic differentiation trajectories, we computed the proportion of stem and progenitor subsets within the $CD34^+$ compartment and identified that the IGPs were indeed expanded upon IFNa treatment, consistent with the novel identification of this cell state (**Fig. 2b,c**). IFNa also induced a significant expansion of the lymphoid progenitors, especially large Pre-B and small Pre-B cells, and diminution of the megakaryocytic-erythroid lineage progenitors, namely MkPs, erythroid progenitors (EP), and megakaryocytic-erythroid progenitors (MEP; **Fig. 2b,c;** fold change of cell frequencies from serial samples in **Fig. 2d**). Corroborating a priming toward lymphoid differentiation, HSCs exhibited an increased protein expression of CD38 and CD45RA after IFNa exposure (**Fig. 2e**). Interestingly, while IFNa alters mouse HSC marker Sca-1[2] (rendering difficult the quantification of progenitor cell frequency changes by flow cytometric assessment in mice), IFNa did not alter CD90 expression after treatment in the HSPCs (**Fig. 2e, Supplementary Fig. 3c**).

To validate the generalizability of the lymphoid expansion in the HSPCs, we re-analyzed clinical multi-parametric flow cytometry data of bone marrow aspirates from patients with early-phase MPN with *JAK2* or *CALR* mutations (n = 47 samples with no IFNa exposure; n = 9 samples with IFNa therapy). Indeed, we identified that the proportion of TdT^+, $CD19^+$ cells within $CD34^+$ cells increased in IFNa-treated bone marrow compared to non-IFNa exposed samples (**Fig. 2f)**. We also observed an increase in the proportion of $CD34^+$, TdT^+, $CD19^+$ cells of total viable cells analyzed, as well as an increase in the $CD34^-$, TdT^-, $CD19^+$ B-lymphocytes (**Fig. 2f)**, providing evidence for an active expansion of the lymphoid progenitors, rather than simply a proportional increase due to diminution of the megakaryocytic-erythroid lineage cells. The major shifts in progenitor output suggested that the clinical improvement in the patient's platelet count post-therapy (despite no significant reduction in the variant allele frequencies) may be due to the differentiation skewing away from the megakaryocytic lineage. Consistent with this hypothesis, the proportions of MEPs and MkPs in $CD34^+$ cells could model the patient's platelet counts ($P = 0.0071$, F-test, **Fig. 2g**), thus supporting a novel model of re-balancing the differentiation landscape as a mode of therapeutic efficacy of IFNa.

As IFNa has been shown to induce HSC and progenitors into cell cycle entry[43], we posited that differential induction of progenitor subtypes into proliferation may underlie the shifts in differentiation. Therefore, we examined the gene signatures for proliferation in the context of cell identity shown to be an accurate assessment of proliferation state[44]. Consistent with previous reports in mice[43], the proportions of HSCs expressing G2/M/S-phase genes were enhanced upon IFNa treatment (**Fig. 2h,i)**. Surprisingly, we identified a global increase in proliferation across the progenitor lineages, including granulo-monocytic, erythroid, and lymphoid subsets (**Fig. 2h,i)**. Nonetheless, the proportion of Pro-B cells in G2/M/S-phases demonstrated on average almost a 2-fold increase, compared to more modest increases in other lineages (**Fig. 2i)**, consistent with the observed expansion of the lymphoid progenitors.

IFNa downregulates TNFa and TGFb signaling

To determine the rewiring of other transcriptional programs by IFNa, we performed a differential expression analysis between baseline and treated $CD34^+$ cells, as a function of cell identity. We identified genes commonly regulated across multiple progenitor subsets upon IFNa administration, including the canonical IFNa genes, such as *ISG15, IFITM3, IFI6* and *EPSTI1* (**Fig. 2j, Supplementary Table 5)**. In HSCs, we identified a downregulation of *CXCR4* gene expression (**Fig. 2j, Supplementary Table 5)**, a key regulator of HSC quiescence[45], which may thus be involved in cell cycle entry of HSCs. Indeed, HSCs expressing high levels of *CXCR4* gene were less likely to be in cell cycle and enriched for baseline cells ($P < 2.2 \times 10^{-22}$, Fisher exact test) and conversely, HSCs in the cell cycle did not express high *CXCR4* gene expression (**Supplementary Fig. 4a**). In the MkPs, we identified a downregulation of *CD9* and *VWF*, closely associated with MkP differentiation[46-49], upon IFNa treatment (**Supplementary Fig. 4b**) which is consistent with the decrease in the proportion of MkP observed after IFNa treatment. We also observed a downregulation of *TGFB1* (**Supplementary Fig. 4b**), that encodes the pro-fibrotic cytokine TGFb established as one of the main inducers of myelofibrosis in patients with fibrotic-phase of MPN[50,51]. We and others have shown that MPN megakaryocytes indeed exhibit higher expression of *TGFB1*[20,50,51], and thus downregulation of *TGFB1* by IFNa may be a key mechanism of improvement in fibrosis frequently observed in patients treated with IFNa.

Next, to determine differentially regulated pathways, we performed gene set enrichment analysis using the canonical Hallmark gene sets in the IFNa-treated versus baseline cells and identified upregulation of the IFNa signaling pathway across the cell subsets as expected (**Fig. 2k,**

Supplementary Table 6). Interestingly, most cell types also upregulated genes within the IFNg pathway, but not the IGPs (**Fig. 2k, Supplementary Table 6**), consistent with a specific upregulation of type 1 interferon-associated *IRF1* in the IGPs. The analysis also confirmed the upregulation cell cycle-related pathways (G2M check point and E2F targets) (**Fig. 2k, Supplementary Table 6**). Interestingly, MYC targets were also upregulated (**Fig. 2k, Supplementary Table 6**), thus corroborating a previous report demonstrating that IFNa upregulates MYC protein, which was thus postulated to regulate the cell cycle entry of HSCs in mice[52]. Consistent with downregulation of *TGFB1* gene itself, gene set enrichment analysis between baseline and IFNa-treated cells revealed a global decrease in TGFb signaling pathway across multiple HSPC subsets, including downregulation of *THBS1* and *SERPINE1* (**Fig. 2k, Supplementary Table 6)**. In a subset of the progenitor states, P53 and KRAS signaling were downregulated (**Supplementary Table 6**).

Moreover, in contrast to the pro-inflammatory state of the IFNa-associated IGPs, we observed a downregulation of pro-inflammatory pathways, including the TNFa signaling via NFKB and inflammatory response pathways, across most of the progenitor subtypes but particularly in the cDCPs (**Fig. 2k-l, Supplementary Table 6)**. Downregulated genes in the TNFa pathway included *NFKB1*, *VEGFA*, *MAP3K8,* as well as *IL1B*, that encodes the pro-inflammatory cytokine IL-1b (**Fig. 2l, Supplementary Table 6)**. Expression of *IL1R1*, *TNFRSF1B* and *CXCL8* genes from the inflammatory response pathway were also downregulated (**Supplementary Table 6)**. Previously, induction of IFNa in mice have yielded mixed results demonstrating either an upregulation[2] or downregulation[53] of TNFa. Our findings show that in human hematopoiesis, IFNa exerts an anti-inflammatory response associated with a downregulation of TNFa signaling and therefore may serve as another key mode of disease amelioration via downregulating MPN-associated inflammation. However, given the induction of the IGP differentiation, these findings suggest that IFNa is able to induce both a *pro-inflammatory* state via IGP differentiation and an *anti-inflammatory* state via differentiation shifts toward lymphoid development and downregulation of TNFa and NFKB activities – both occurring in parallel within the same individual's hematopoiesis. In fact, we observed a significant overlap between the IFNa-downregulated genes in the TNFa signaling and the genes that are upregulated in the IGPs (**Fig. 2l**, **Supplementary Table 6,** $P = 1.6 \times 10^{-5}$, hypergeometric test). The inverse, i.e. the overlap of IFNa-upregulated genes and genes downregulated in IGPs, was also observed ($P = 1.73 \times 10^{-69}$, hypergeometric test, **Supplementary**

Tables 3,6). Consistently, the HSC-specific IFNa-upregulated gene signature (data presented in **Fig. 2j**) was significantly downregulated in the IGPs compared to the HSCs while the IFNa-downregulated genes were upregulated in the IGPs, with both cell groups under the same IFNa treatment (**Fig. 2m**). Altogether these data indicated that IFNa can trigger opposing cell states within the same hematopoiesis.

GoT-IM reveals differential impact of IFNa on *CALR*-mutated versus wildtype hematopoiesis

Having determined the overall effects of IFNa on the *CALR*-mutated and wildtype cells in aggregate, we leveraged the genotyping capacity of GoT-IM to determine how IFNa may differentially impact neoplastic versus wildtype HSPCs. Consistent with our previous results[20], the mutated and wildtype HSPCs were co-mingled throughout differentiation both at baseline and upon IFNa therapy (**Fig. 3a, Supplementary Fig. 5a)**. To assess how IFNa may skew differentiation distinctly in the *CALR*-mutated versus wildtype HSPCs, we determined the progenitor subset frequencies between mutant and wildtype cells. As we had previously demonstrated, the mutated cells were enriched in the MkPs at baseline (**Fig. 3b,c)**. After treatment, we observed the expansion of the lymphoid compartment in both wildtype and mutated cells (**Fig. 3b,d)**. However, the lymphoid expansion was constrained in the mutated compartment, due to the relative expansion of the granulo-monocytic and megakaryocytic-erythroid progenitors compared to wildtype HSPCs (**Fig. 3d)**. These findings provide direct evidence that IFNa exerts distinct phenotypic effects on HSPCs as a function of the mutation status in human.

Given that a primary effect of IFNa involves inducing robust cell cycle programs in HSPCs[1], we hypothesized that the differential increase in proliferation in mutated versus wildtype hematopoietic progenitor cells may contribute to the differential expansion of the progenitor subsets in *CALR*-mutated versus wildtype cells. At baseline, the *CALR*-mutation induced proliferation in the HSCs and the myeloid cell lineages (i.e. the granulo-monocytic and megakaryocytic-erythroid lineages) but not in the lymphoid progenitors (**Fig. 3e**), consistent with our previous reporting[20]. Intriguingly, IFNa boosted the proliferation of mutated HSCs to a greater degree than the wildtype counterparts (**Fig. 3e**), with a corresponding decrease in *CXCR4* gene expression that was more pronounced in the mutated cells (**Supplementary Fig. 5b**). Moreover, IFNa induced proliferation to a greater degree in the mutated cells compared to wildtype cells within the myeloid progenitor compartment, but not the lymphoid, that is, it was restricted to the

progenitor subsets in which the *CALR*-mutation induced proliferation at baseline (**Fig. 3e)**. Thus, the proliferative effects of IFNa appeared to be compounded by those of the underlying *CALR* mutation; in other words, the mutated cells were primed toward a greater proliferative response upon IFNa treatment. However, in contrast to the other myeloid progenitors, within the MkPs (wherein the *CALR* mutation exerts the greatest phenotypic effects due to the activation of the thrombopoietin receptor[21,54-57]), the proliferative effects of IFNa were modest in the mutated cells compared to those at baseline (**Fig. 3e**), suggesting that the rate of cell cycle entry of the mutated cells was already near or at the point of saturation due to the *CALR* mutation. We orthogonally validated that IFNa induced greater proliferative rates in the myeloid progenitors by performing multiplexed *in situ* fluorescent imaging of bone marrow paraffin-embedded sections from patients with *CALR* mutations with or without IFNa treatment (n = 5 with no IFNa, n = 4 with IFNa, **Fig. 3f, left**). We determined the protein expression of CD34, CD117, CD38, Ki67 and mutant CALR to assess the frequencies of $CD34^+$, $CD38^+$, $CD117^+$ myeloid progenitors that express Ki67, a gold standard of cell cycle entry[58]. Overall, agnostic to the mutation status, IFNa-exposed myeloid progenitors demonstrated a trending increase of proliferation even in this non-paired cohort (**Supplementary Fig. 5c**, P = 0.063, Wilcoxon rank sum). The availability of a *CALR*-mutant specific antibody[59] enabled us to distinguish mutated from wildtype cells. We found that the frequency of myeloid cells in cell cycle was increased in the treated mutated versus wildtype progenitors (**Fig. 3f, right**). Overall, these findings provide evidence that the differential proliferation states as a function of the mutational status may contribute to the relative expansion of the myeloid versus lymphoid progenitors.

Impact of IFNa on clonal fitness of *CALR*-mutated HSC

As IFNa induces cell cycle entry of mutated HSCs to a greater extent than wildtype HSCs, it may be postulated that IFNa preferentially exhausts mutated HSCs for clonal remission, as was hypothesized to be the case for stem cells with *JAK2* mutations[60] and *BCR-ABL* translocations[1]. However, while *JAK2*-mutated MPN frequently displays clonal stem cell depletion upon IFNa treatment, *CALR*-mutated MPN rarely demonstrates molecular response[19], indicating a genotype-specific response to inflammation. Consistent with these data, when we examined the clone size in the HSC compartment in the serial samples, we observed no significant shifts in the relative proportion, except in one patient IFN04 whose clone expanded significantly upon IFNa therapy (**Supplementary Fig. 6a**). Interestingly, we found that cells from IFN04 harbored another

mutation in *CALR*: a single nucleotide variant in the *CALR* allele (M131I, predicted to impact protein structure[61]) which is trans to the MPN-causing frameshift mutation (**Supplementary Fig. 6b**). At baseline, the double mutant clone remained subclonal to the dominant clone with the single canonical *CALR* mutation, consistent with bulk sequencing data (**Fig. 3g**). Upon IFNa therapy, the double-mutated clone overtook the neoplasm and the overall stem cell population (**Fig. 3g**). We have previously demonstrated that *CALR* mutations enhance the unfolded protein response (UPR), specifically the pro-survival IRE1-XBP1 axis[20], likely due to a heterozygous compromise of the critical chaperone protein encoded by *CALR*. Consistent with these data, we found that the additional insult to CALR activities in the double mutant clone induced an even greater UPR activation compared to single mutant HSCs (**Fig. 3h,i, Supplementary Tables 7,8**). Surprisingly, however, the predominant signatures of the double mutant HSCs were decreased TNFa and TGFb signaling and upregulation of IFN response genes which is the observed IFNa signature in single mutant HSCs compared to wildtype HSCs at baseline (**Fig. 3h,i, Supplementary Tables 7,8**). Intriguingly, the double mutant HSCs intrinsically exhibited high IFN signaling at baseline (**Fig. 3h,i, Supplementary Tables 7,8**), supporting a potential causal link between UPR and IFN activation[62]. Indeed, the double mutant HSCs exhibited high expression of the HSC-specific IFNa-upregulated gene signature (derived from data displayed in **Fig. 2j**) at baseline and even higher upon IFNa therapy compared to baseline (**Fig. 3j, left**). Similarly, the HSC-specific IFNa-downregulated gene signature was downregulated in the double mutant clone both at baseline and to a greater degree at treatment (**Fig. 3j, right**). These data further supported the concept that the degree of cellular response to IFNa varies upon the degree of priming toward that signal at baseline. Consistent with these data, unbiased clustering and dimensional reduction revealed that the double mutant clone at baseline clusters with the treated HSCs rather than with the other HSCs at baseline (**Supplementary Fig. 6c**). The striking similarity of the intrinsically-activated IFNa signaling in the double mutant cells at baseline to the extrinsic IFNa effects indicated that the predominant IFNa signaling signatures observed in IFNa-treated HSCs are largely direct rather than simply mediated by secondarily modulated cytokines (e.g. TNFa, TGFb). These findings also highlighted a genotype-specific modulation of HSC fitness by IFNa, specifically suggesting that IFNa selects for HSC clones with high IFNa signaling at baseline.

GoT-ATAC identifies transcription factor regulatory networks that govern IGP differentiation

Chromatin accessibility enables approximation of TF activity based on accessibility of the TF binding sites. Thus, in order to determine the regulatory networks that govern the IFNa-induced differentiation states, we expanded upon GoT to capture somatic mutation status, chromatin accessibility and whole transcriptomes by adapting the 10x Multiome platform that captures single-nuclei RNA-seq (snRNA-seq) and chromatin accessibility (snATAC-seq) to include somatic genotyping, i.e. GoT-ATAC (**Fig. 4a**). We applied GoT-ATAC to serial bone marrow $CD34^+$ cells (n = 24,862 total; 12% of total cells with genotyping data (n=2,873 cells)) from the same cohort of IFNa-treated patient samples (n = 4 baseline, 3 IFNa-treated samples). Notably, genotyping efficiency was significantly lower in GoT-ATAC compared to GoT-IM, due to the limited transcript pool in the nucleus versus whole cell. As in GoT-IM, we incorporated time-point specifying barcoded antibodies[24] to combine serial samples from the same individuals into a single experiment to remove technical batch effects (**Fig. 4a, Supplementary Fig. 7a-d**). We then identified the treatment status analytically and segregated the baseline and IFNa-treated samples to cluster the cells based on cell identity alone as we did for the GoT-IM data (**Supplementary Fig. 7e**). We clustered the cells based on the transcriptomic data and identified the cell states identified by GoT-IM, including the novel IGP cell state (**Fig. 4b, Supplementary Fig. 8a,b, Supplementary Table 9**). The activities of cell lineage specifying TF were consistent with the cell type assignments (**Supplementary Fig. 8c-d**). Furthermore, we integrated the cells based the snATAC-seq data, and identified the same stem and progenitor cell relationships, including the similarity of the IGPs to the quiescent HSCs, again providing evidence for a direct differentiation of IGPs from HSCs (**Fig. 4c, Supplementary Fig. 9a)**.

As the IGPs were characterized by a robust upregulation of TF gene expression, especially those of AP-1 and RFX2/3, we sought to determine whether the chromatin accessibility of their binding motifs was increased, as a reliable surrogate of TF activities. Strikingly, differential TF motif enrichment analysis between the IGPs and HSCs identified enhanced accessibility of the same TFs whose expression levels were elevated, including the motifs of the AP-1 family (FOS, JUN, JUND, JUNB, FOSL1, FOSL2), CEBPB/D, IRF1 and RFX2/3 (**Fig. 4d, Supplementary Table 10**). These data thus demonstrated an unusually high concordance of gene expression of the TFs and their target motif accessibility, suggesting a rapid induction of the transcriptional regulatory program. Enhanced chromatin accessibilities for small MAF factors (MAFK and MAFG) and their interacting partners NFE2L1/3 and BACH1/2 were also observed in the IGPs (**Fig. 4d,**

Supplementary Table 10). The differential motif analysis further revealed an upregulation of STAT2 and the proinflammatory REL of the NFKB complex (**Fig. 4d, Supplementary Table 10**), consistent with the observed upregulation of the TNFa via NFKB signaling pathway (**Fig. 1h**). As RFX2/3 is not well characterized in hematopoietic stem and progenitor cells, we determined which TFs upregulated the expression of these TFs. We determined TF motifs present in the regulatory peaks that correlated with RFX3 gene expression (i.e. linked peaks analysis)[63] and identified those of STAT2 and IRF1 (further supporting an IRF1-specific response) as well as FOSB and SPI1 (or PU.1), which were differentially upregulated at the gene expression level in IGPs (**Fig. 4e, Supplementary Table 11**). Similarly, the most significant regulatory peaks for RFX2 included motifs for PU.1, KLF, NR4A1/2 and AP-1 factors that were differentially expressed in the IGPs (**Supplementary Fig. 9b, Supplementary Table 11**). These data further supported the model in which HSCs with high expression of the AP-1 and other immediate early response factors were primed toward a robust upregulation of RFX2/3 gene expression upon induction of IRF1 and PU.1 activities with IFNa treatment. Further, to determine the regulatory networks that govern the robust upregulation of the AP-1 family TFs in the IGPs, we aggregated the significantly correlated peaks for the AP-1 TFs and assessed for motif enrichment. Among the top TF motifs were those of RFX2-4 (which have the same binding motifs), implicating RFX2/3 in the upregulation of the AP-1 members upon IGP differentiation (**Fig. 4f, Supplementary Table 12**). Thus, RFX2/3 and AP-1 TFs positively regulated the expression of one another, synergizing the IGP development. Finally, as other RFX members have been demonstrated to regulate MHC class II expression[35,64], we hypothesized that RFX2/3 may play a significant role in downregulating MCH class II genes in the IGPs. To test this, we determined the significantly linked peaks that negatively regulated *HLA-DRA* expression (**Fig. 4g, bottom**). We identified a distal regulatory region with four negatively regulating loci (**Fig. 4g, inset, Supplementary Table 13**). These peaks included motifs for IRF1 and STAT2 as well as immediate response factors including AP-1 and KLF families, but not RFX2/3 (**Fig. 4g, inset, Supplementary Table 13**). However, within the same negative regulatory region, we identified an IGP-specific peak that included the binding motif for RFX1-4, KLF factors and IRF1 (**Fig. 4g, top, Supplementary Table 13**). These findings identified the immediate response factors and IRF1/STAT2 as negative regulators of MHC class II genes across cell types, while RFX2/3 was highly specific for MHC class II downregulation during IGP differentiation. Notably, the motif RFX8, known to positively regulate MHC class II expression

was present in the positively regulating peaks, but not in the negatively regulating loci (**Supplementary Table 13)**. Genes positively regulated by RFX2/3 included *MPO*, consistent with neutrophilic differentiation, and genes involved in cell cycle entry, implicating RFX2/3 as key TFs in the overall regulatory network of IGPs (**Supplementary Table 14**).

PU.1 initiates IFNa-induced hematopoietic differentiation remodeling

Furthermore, the GoT-ATAC data confirmed the expansion of the lymphoid progenitors upon IFNa treatment (**Supplementary Fig. 10**). Thus, in order to determine the regulatory networks that governed the transcriptional rewiring and differentiation skewing induced by IFNa, we performed differential motif enrichment analysis between IFNa-treated and baseline HSCs. IFNa enhanced the activities of STAT2 and numerous interferon regulatory factors in HSCs, in contrast to the specific activity of IRF1 in IGPs (**Fig. 5a, Supplementary Table 15**). IFNa downregulated the motif accessibility of AP-1 TFs consistent with the downregulation of AP-1 TF gene expression associated with TNFa signaling via NFKB (**Fig. 2j**). Moreover, TWIST1 has been implicated in mediating the downregulation of TNFa upon type 1 IFN treatment[53]. Consistently, accessibilities of the TWIST1 motif were enhanced after treatment. Furthermore, the activity of TGFb induced factor homeobox 2 (TGIF2), which inhibits TGFb response genes, was also enhanced (**Fig. 5a, Supplementary Table 15**), highlighting TGIF2 as a key TF involved in the downregulation of the observed TGFb signaling after IFNa treatment. IFNa also downregulated the activities of cAMP responsive element binding (CREB) proteins (**Fig. 5a, Supplementary Table 15**), consistent with the observed downregulation of KRAS signaling (**Fig. 2k**).

Importantly, critical TFs involved in hematopoietic differentiation[65] were differentially regulated. The motif accessibility of PU.1 and RUNX1, essential for early lymphoid and granulomonocytic differentiation[65], was enhanced (**Fig. 5a, Supplementary Table 15**). Consistently, the activities of GATA1/2, demonstrated to be negatively regulated by PU.1[65], were downregulated as well as those of the other critical megakaryocytic-erythroid lineage factor, TAL1[66,67], consistent with the differentiation away from the megakaryocytic-erythroid lineages upon IFNa treatment. Furthermore, the accessibilities of the motifs of the critical early B-lymphoid differentiation factors TCF3/4 were enhanced, highlighting TCF3/4 as TFs that govern the differentiation toward lymphoid progenitors by IFNa. CEBPA, essential for granulo-monocytic development[68,69], was also upregulated in its motif accessibility. Importantly, the IRFs, RUNX1, CEBPA and TCF3 have been demonstrated to co-regulate target gene expression with PU.1[70], suggesting that the induction

of IRFs may have stabilized PU.1 activities, which in turn enhanced the activities of its co-regulating TFs, such as CEBPA and TCF3. To confirm these co-regulations in IFNa-induced hematopoiesis, we examined the synergistic activities of the different combinatorial TFs by measuring the excess variability of accessibility for peaks with both TF motifs (compared to peaks with one motif)[71]. Indeed, PU.1 exhibited high synergistic activities with the IRFs, RUNX1, CEBPA and TCF3, while displaying antagonism with GATA1 and TAL1 (**Fig. 5b**). These findings again highlighted PU.1 as the master regulator of IFNa-induced differentiation shifts, as was the case in IGP differentiation. To determine the activities of these TFs upon lymphoid differentiation, we assessed the motif accessibilities of the TFs across the early and late progenitors. The activities of PU.1, RUNX1 and CEBPA were enhanced upon IFNa signaling during early stages of hematopoiesis and diminished with lymphoid development, whereas the activities of TCF3 were pronounced in the latter stages of development (**Fig. 5c**).

To identify the TFs that may govern the observed myeloid skewing of the mutated versus wildtype hematopoietic development, we tested the differential motif enrichment of the IFNa-regulated lineage TFs (**Fig. 5a**) in mutated vs. wildtype HSCs and IMPs. We identified an upregulation of motif enrichment for CEBPA at baseline (**Fig. 5d**), suggesting that PU.1 activation upon IFNa, together with increased CEBPA activity, enhanced the granulo-monocytic differentiation in the mutated cells. Moreover, GATA1 activity was upregulated in the mutated MkPs and MEPs (**Fig. 5d**), highlighting GATA1 as the key TF that governed the preferential megakaryocytic-erythroid development.

Altogether, GoT-ATAC identified the key regulators that governed the IFNa-induced hematopoietic differentiation shifts, which varied as a function of the underlying mutational status. PU.1 was identified the master regulator that mediated the hematopoietic differentiation remodeling by IFNa in both the induction of the IGPs and lymphoid expansion (**Fig. 5e**). Thus, IFNa-induced PU.1 activities mediated both a pro-inflammatory state via IGP development and an anti-inflammatory state via lymphoid differentiation skewing, linked with opposing AP-1 activities, suggesting a strong dependency on the underlying HSC transcriptional and epigenetic state (**Fig. 5e**).

Discussion

Inflammation perturbs HSC homeostasis and differentiation across physiologic and pathologic conditions[3,14,72-74]. In mice, the contributions of the various inflammatory cytokines have been clarified by the administration of individual cytokines to mouse models[4,5,13,75-78]. While inflammation also impacts human hematopoiesis across conditions such as ageing, cancer, autoimmune disorders and infections[3,72,73,79], a careful dissection of the contribution of individual cytokines in humans is challenging. Germline genetic disorders have provided clues into the distinct roles of individual cytokines. For example, germline defects in genes involved in the type 1 IFN signaling cascade have been linked to exaggerated or severe COVID-19 response[16,79], providing evidence for anti-inflammatory effects of type 1 IFN. Indeed, early studies of administering type 1 IFN to patients have suggested clinical benefit[17,80].

To directly examine the effects of type 1 IFN in human hematopoiesis, we identified a unique clinical setting in which patients with MPN receive recombinant IFNa as therapy. As patients with MPN often present with an early clonal and stable disease, serial bone marrow sampling enriched with stem and progenitor cells is rare. To overcome this limitation, we leveraged cryopreserved bone marrow samples at baseline and after IFNa treatment from our clinical trials. We applied our innovative single-cell multi-omics platforms to FACS-isolated $CD34^+$ stem and progenitor cells to determine how IFNa impacts human hematopoiesis in normal and neoplastic blood development. We identified an HSC-specific IFNa signature that was defined by an upregulation of IFNa/g response genes (as expected) and downregulation of the TNFa and TGFb signaling pathways. In mouse studies, IFNa has been demonstrated to modulate TNFa expression[13]. However, different studies have yielded distinct results, either an upregulation[2] or downregulation[53] of TNFa signaling. Thus, our studies clarify that the predominant effect of IFNa in human HSPCs is the downregulation of TNFa and other proinflammatory pathways, consistent with the abovementioned genetic studies in patients with severe COVID-19 infections[16,79] and amelioration of the inflammatory state in MPN and multiple sclerosis upon type 1 interferon administration[17,19]. Coherently, TNFa protein expression was demonstrated to be decreased in patients with MPN who received IFNa treatment[81]. Our single-cell multi-omics data revealed that the downregulation of AP-1 and NFKB transcription factors played a key role in downregulating the pro-inflammatory signaling. The pro-fibrotic TGFb signaling was also broadly downregulated across progenitor subsets, associated with the downregulation of the *TGFB1* gene itself and

upregulation of the TGFb signal-inhibiting TGIF2 activity. These findings were consistent with the observed downregulation of TGFb receptor (*TGFBR1*) in mice with IFNa administration[2]. In this way, the coordinated activities of IFNa downregulated two key cellular programs involved in the MPN-associated pathology and thus improved patient disease state.

Unexpectedly, we identified a novel progenitor state, we termed inflammatory granulocytic progenitors (IGP), that was induced by IFNa, specifically through IRF1. Immunophenotypically, IGPs were similar to the IMPs. Transcriptionally and epigenetically, they were most similar to a subset of the HSCs with the highest expression of CD90 and *AVP* (indicating a quiescent state) and with high gene expression and enhanced motif accessibility for the AP-1 members and RFX2/3 TFs. Upon IGP development, the AP-1 and RFX2/3 TFs and their activities were robustly upregulated, as key defining features of this novel progenitor state. The upregulation of PU.1, CEBPB and CEBPD expression and motif accessibility (as well as of their targets *MPO* and *CSF3R*), combined with a downregulation of the MHC class II genes indicated their lineage fate into the neutrophil lineage. The specificity of the IGP gene signature to neutrophils versus other myeloid lineages also supported this lineage fate. However, they were transcriptionally and epigenetically distinct from the IMPs and neutrophil progenitors, suggesting that they may arise directly from the quiescent HSCs, as a novel route of neutrophil formation. The functional and genetic lineage studies of this rare population are challenging, particularly since the *in vitro* effects of IFNa on HSCs completely differ from its *in vivo* effects. *In vitro*, IFNa induces cell death in the HSCs[2] while it induces proliferation *in vivo*[1]. These findings suggest that IFNa effects on HSCs are dependent on the bone marrow microenvironment. Consistent with this inference, we identified a downregulation of *CXCR4* linked with their changes in proliferation status upon IFNa treatment. Moreover, as scRNA-seq of HSPCs from mice treated with the IFNa-inducing polyI:C[77] or infected with Mycobacterium tuberculosis[82] (a model of IFNa signaling) did not reveal a comparable population, there is a possibility that the corresponding IGP population may not be present in mice.

Nevertheless, the identification of this novel IGP population highlighted an intriguing phenomenon in human HSCs. Induction of the IGP differentiation through upregulation of the pro-inflammatory AP-1 and REL, together with PU.1, indicated that IFNa is able to initiate this pro-inflammatory and overall anti-inflammatory cell states. Intriguingly, the IFNa-induced HSC transcriptional signature and the IGP differentiation program utilized an overlapping

transcriptional program, but in the opposite direction, i.e. the IFNa-induced upregulated genes in HSCs were downregulated in the IGPs and vice versa. The AP-1 program was highlighted in our data as modulating this switch in response to IFNa. These findings overall highlighted non-genetic cell state heterogeneity of the HSC pool as a key mediator of distinct, even opposing, cellular responses to IFNa therapy. IFNa thus exposed the functional consequences of HSC heterogeneity. Another major discovery in our work was the differentiation remodeling of hematopoietic differentiation toward the lymphoid lineage by IFNa. While various differentiation skewing by IFNa has been reported in mice[3,60,82], the interpretation of the results is complicated by the IFNa-induced alteration of the HSC-marker Sca-1 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). As other inflammatory cytokines, such as TNFa, IL-1, and IFNg, have been demonstrated to induce granulo-monocytic differentiation[6,8,9,11,76], IFNa presents as a unique cytokine among the inflammatory milieu to balance the granulo-monocytic differentiation with its activities on lymphoid differentiation as another mode of dampening the pro-inflammatory response.

The reshaping of the differentiation landscape by IFNa provided a novel model of therapeutic efficacy in patients with MPN. As MPN presents primarily as a defect in homeostatic hematopoietic development, due to an abnormal differentiation skewing and increased proliferation of one or more of the myeloid lineages, IFNa-induced lymphoid differentiation and proliferation may re-balance the differentiation landscape. Just as in the differentiation of the IGPs, PU.1 – the master regulator of granulo-monocytic and B-lymphoid lineages – emerged in our data as the initiating regulator of IFNa-induced differentiation remodeling. Intriguingly, IFNa reshapes the major bifurcation divide in MPN, that is, from the JAK2/STAT5-mediated bifurcation at the myeloid (i.e. granulo-monocytic and megakaryocytic-erythroid) versus lymphoid commitment, to the PU.1-mediated bifurcation at granulo-monocytic and lymphoid versus megakaryocytic-erythroid commitments (see model in **Fig. 5e**). Furthermore, the upregulation of PU.1 and its cooperating TFs, such as CEBPA and TCF3, indicated that the IFNa-induced shifts in the progenitor populations were due to an active differentiation re-programming, rather than due to a preferential cell death of a particular lineage.

Our single-cell multi-omics methods uniquely enabled us to decipher the distinct effects of IFNa on *CALR*-mutated versus wildtype HSPCs. Our data revealed that while lymphoid expansion was

observed in both mutated and wildtype cells, the lymphoid differentiation was constrained by a pre-existing *CALR*-induced bias toward the myeloid lineages. The *CALR* mutation-induced proliferation of the myeloid lineages at baseline primed the mutated cells for an amplified response in the IFNa-induced proliferation. Further, our GoT-ATAC data revealed that the *CALR*-mutated HSPCs were primed toward the myeloid lineages via enhanced CEBPA and GATA1 activities at baseline. As PU.1 cooperates with CEBPA to induce the granulo-monocytic lineage and antagonizes GATA1 activities (thereby reducing megakaryocytic-erythroid lineage specifying programs), the upregulation of CEBPA and GATA1 activities in the mutated cells indicates a modulation of the PU.1 activities upon IFNa therapy and downstream phenotypic effects toward a greater myeloid expansion in the mutated versus wildtype cells. Altogether, these findings provided direct evidence in human that the extrinsic signaling by cytokines are constrained by the transcriptional rewiring from the underlying somatic mutation.

Evidence for genotype-specific cellular priming and response was further observed in serial sampling from an individual with a 'double-hit' in the *CALR* genes. While homozygous canonical mutations in CALR are quite rare and only observed in aggressive myeloid neoplasms with multiple genetic defects[21], the presence of the subclone with an additional single nucleotide variant in the trans *CALR* allele provided further insights into IFNa signaling in HSCs. Surprisingly, the double mutant clone exhibited a strong HSC-specific IFNa signature compared to the wildtype and single mutant clone at baseline and were transcriptionally more similar to the IFNa-treated cells compared to baseline cells (in stark contrast to the other samples in which cells clustered based on treatment status). Three important inferences could be derived from this index case. First, the double mutant cells exhibited a greater IFNa-signature upon IFNa treatment compared to the treated wildtype and single-mutant HSCs, further supporting the idea that the degree of HSC-response to IFNa treatment is commensurate to the priming toward IFNa-induced signaling prior to exposure (e.g. proliferation, differentiation and IFNa-specific gene expression). Second, the double mutant clone exhibited a strong fitness advantage against non-neoplastic and single-mutant HSCs upon IFNa, by taking over both the neoplastic and the overall HSC population, suggesting that baseline type 1 IFN activity may enhance HSC fitness upon the IFNa challenge. Finally, in the double mutant cells, we observed not only the upregulation of the IFNa-positively induced genes, but also the reduction of the downregulated genes in the TNFa and TGFb pathways. These data demonstrated that the predominant effects of IFNa on HSCs is direct (and not secondary to

other mediators). While the overall reduction of TNFa and TGFb production in the other bone marrow cells has likely reinforced the downregulation of these pathways in the HSCs, the strong downregulation of TNFa and TGFb pathways in the mutation (i.e. intrinsically)-induced IFN signaling demonstrates that the predominant effects are direct. These data are supported by an elegant experiment in mice[1]. In a chimeric mouse model consisting of mostly IFNa/b receptor (IFNAR) deficient cells, the minority of HSCs with intact IFNAR demonstrated a robust cell cycle entry upon IFNa administration, establishing that IFNa exerts profound effects directly on HSCs[1]. Overall, we identified a human HSC signature of IFNa response characterized by a re-modeling of the differentiation landscape and an intriguing phenomenon of opposing transcriptional programs triggered by the same extrinsic perturbation, as a function of the underlying HSC states. These data thus highlighted mediators of functionally significant stem cell heterogeneity that may be manipulated for therapeutic advantage.

Acknowledgments

The work was enabled by the Weill Cornell Epigenomics Core and Flow Cytometry Core. D.A.L. is supported by the Burroughs Wellcome Fund Career Award for Medical Scientists, Valle Scholar Award, and the National Institutes of Health Director's New Innovator Award (DP2-CA239065). N.D. is supported by a F30 Predoctoral Fellowship from the NHLBI of the National Institutes of Health (F30HL156496) and by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the National Institutes of Health under award number T32GM007739 to the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program. A.S.N. is supported by the Burroughs Wellcome Fund Career Award for Medical Scientists, National Institutes of Health Director's Early Independence Award (DP5 OD029619-01) and Starr Cancer Consortium.

Competing interests

D.A.L. has served as a consultant for Abbvie and Illumina, and is on the Scientific Advisory Board of Mission Bio and C2i Genomics; D.A.L. has received prior research funding from BMS and Illumina unrelated to the current manuscript.

Corresponding author: Anna S. Nam.

Correspondence and requests for materials should be addressed to A.S.N.

Figure 1

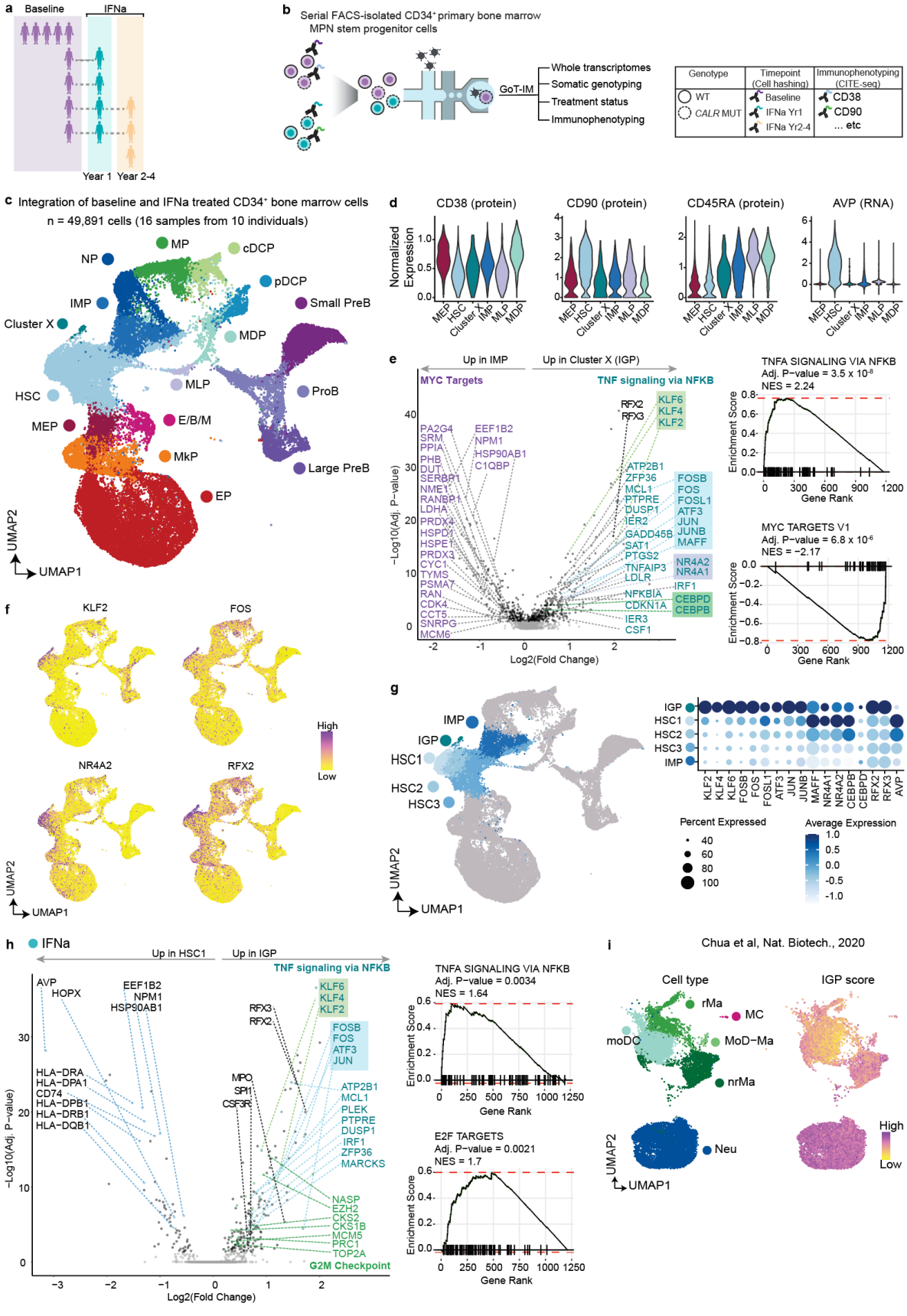

a
Baseline
IFNa
Year 1
Year 2-4
b
Serial FACS-isolated CD34+ primary bone marrow MPN stem progenitor cells
GoT-IM
Whole transcriptomes
Somatic genotyping
Treatment status
Immunophenotyping
Genotype
WT
CALR MUT
Timepoint (Cell hashing)
Baseline
IFNa Yr1
IFNa Yr2-4
Immunophenotyping (CITE-seq)
CD38
CD90
... etc
c
Integration of baseline and IFNa treated CD34+ bone marrow cells
n = 49,891 cells (16 samples from 10 individuals)
MP
cDCP
NP
pDCP
IMP
Small PreB
Cluster X
MDP
MLP
HSC
ProB
MEP
E/B/M
MkP
Large PreB
EP
UMAP2
UMAP1
d
CD38 (protein)
CD90 (protein)
CD45RA (protein)
AVP (RNA)
Normalized Expression
MEP
HSC
Cluster X
IMP
MLP
MDP
e
Up in IMP
Up in Cluster X (IGP)
MYC Targets
TNF signaling via NFKB
PA2G4
SRM
PPIA
PHB
DUT
SERBP1
NME1
RANBP1
LDHA
PRDX4
HSPD1
HSPE1
PRDX3
CYC1
TYMS
PSMA7
RAN
CDK4
CCT5
SNRPG
MCM6
EEF1B2
NPM1
HSP90AB1
C1QBP
RFX2
RFX3
KLF6
KLF4
KLF2
ATP2B1
ZFP36
MCL1
PTPRE
DUSP1
IER2
GADD45B
SAT1
PTGS2
TNFAIP3
LDLR
NFKBIA
CDKN1A
IER3
CSF1
FOSB
FOS
FOSL1
ATF3
JUN
JUNB
MAFF
NR4A2
NR4A1
IRF1
CEBPD
CEBPB
−Log10(Adj. P−value)
Log2(Fold Change)
TNFA SIGNALING VIA NFKB
Adj. P−value = 3.5 x 10−8
NES = 2.24
MYC TARGETS V1
Adj. P−value = 6.8 x 10−6
NES = −2.17
Enrichment Score
Gene Rank
f
KLF2
FOS
NR4A2
RFX2
High
Low
UMAP2
UMAP1
g
IMP
IGP
HSC1
HSC2
HSC3
UMAP2
UMAP1
KLF2
KLF4
KLF6
FOSB
FOS
FOSL1
ATF3
JUN
JUNB
MAFF
NR4A1
NR4A2
CEBPB
CEBPD
RFX2
RFX3
AVP
Percent Expressed
Average Expression
h
IFNa
Up in HSC1
Up in IGP
TNF signaling via NFKB
AVP
HOPX
EEF1B2
NPM1
HSP90AB1
HLA−DRA
HLA−DPA1
CD74
HLA−DPB1
HLA−DRB1
HLA−DQB1
RFX3
RFX2
MPO
SPI1
CSF3R
KLF6
KLF4
KLF2
FOSB
FOS
ATF3
JUN
ATP2B1
MCL1
PLEK
PTPRE
DUSP1
IRF1
ZFP36
MARCKS
NASP
EZH2
CKS2
CKS1B
MCM5
PRC1
TOP2A
G2M Checkpoint
−Log10(Adj. P−value)
Log2(Fold Change)
TNFA SIGNALING VIA NFKB
Adj. P−value = 0.0034
NES = 1.64
E2F TARGETS
Adj. P−value = 0.0021
NES = 1.7
Enrichment Score
Gene Rank
i
Chua et al, Nat. Biotech., 2020
Cell type
IGP score
rMa
MC
moDC
MoD−Ma
nrMa
Neu
High
Low
UMAP2
UMAP1

Figure 1. Integration of Genotyping of Transcriptomes with immunophenotyping and treatment status identifies a novel cell state in baseline and IFNa-treated MPN stem and progenitor cells. a. Representation of primary bone marrow samples at baseline and after IFNa treatment. **b.** Schematic of Genotyping of Transcriptomes with Immunophenotyping (GoT-IM), via CITE-seq and treatment status via Cell Hashing. MPN, myeloproliferative neoplasms; WT, wildtype; MUT, mutated. **c.** Uniform manifold approximation and projection (UMAP) of $CD34^+$ cells (n = 49,891 cells) from MPN samples (n = 16 samples from 10 individuals), overlaid with cell type assignment. HSC, hematopoietic stem cells; IMP, immature myeloid progenitors; NP, neutrophilic progenitors; MP, monocytic progenitors; cDCP, classic dendritic progenitors; pDCP, plasmacytoid dendritic progenitors; MDP, monocytic dendritic progenitors; MLP, multipotent lymphoid progenitors; E/B/M, eosinophil/basophil/mast cells; MkP, megakaryocytic progenitors; EP, erythroid progenitors; MEP, megakaryocytic-erythroid progenitors. **d.** Violin plots showing normalized expression of HSC-defining protein and RNA markers. **e.** Volcano plot showing genes differentially expressed (DE) between IGPs (Cluster X) and IMPs identified via linear mixed modeling with/without cluster identity. Left panel: Volcano plot of DE genes; Genes highlighted in purple represent genes enriched in the MYC pathway and those in blue enriched in the TNFa signaling via NFKB; Boxes represent transcription factors (TF) of the AP-1 (blue), KLF (light green), and NR4A (purple) families and genes involved in emergency or stress-induced granulopoiesis (dark green). Right panel: Pre-ranked gene set enrichment analysis using the MSigDB Hallmark collection. IGP, inflammatory granulocytic progenitors. **f.** UMAP showing expression levels of representative TF that are upregulated in IGPs. **g.** Left panel: UMAP highlighting IGP, IMP and the HSC subclusters. Right panel: Dot plot showing expression levels of upregulated TFs in IGPs (vs. IMPs) **h.** DE genes between IGPs and HSC1 identified via linear mixed modeling with/without cluster identity. Only IFNa-treated cells were included in the DE analysis to remove the treatment variable altogether. Left panel: Volcano plot of DE genes; Genes highlighted in blue represent genes enriched in the TNFa signaling via NFKB and those in green enriched in G2M Checkpoint pathway; Boxes represent transcription factors (TF) of the AP-1 (blue) and KLF (light green) families. Right panel: Pre-ranked gene set enrichment analysis using the MSigDB Hallmark collection. **i.** UMAP of myeloid immune cells under COVID-19 infection displaying previously-assigned cell types (left panel) and projection of the IGP gene signature score (right panel). scRNA-seq data from an available dataset of immune cells from COVID-19

patients[38]. Neu, neutrophil; moDC, MD dendritic cell, rMa, Resident macrophage; nrMa, non-resident macrophage; MC, mast cell; MoD-Ma, MD macrophage.

Figure 2

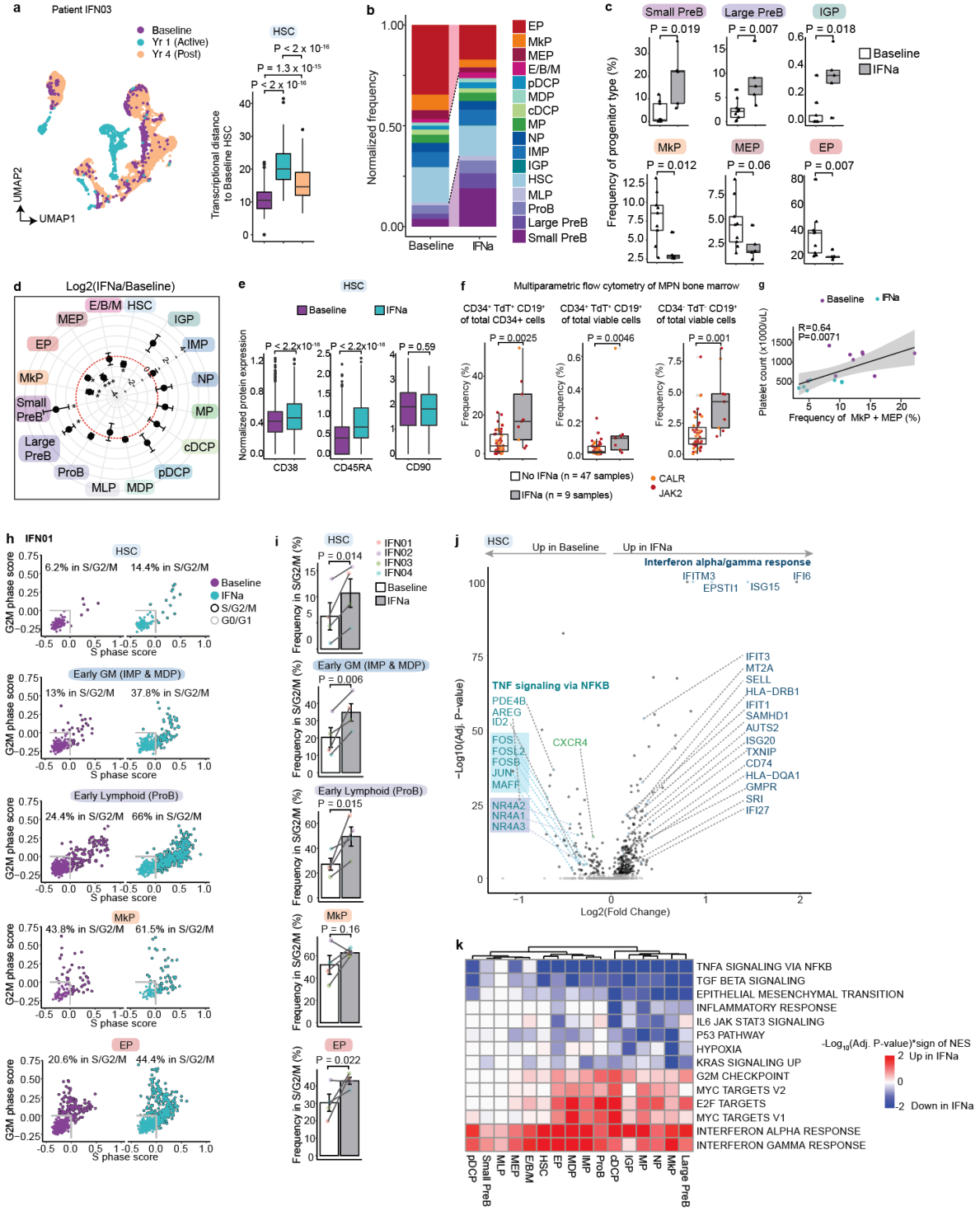

a
Patient IFN03
Baseline
Yr 1 (Active)
Yr 4 (Post)
UMAP2
UMAP1
HSC
P < 2 x 10-16
P = 1.3 x 10-15
P < 2 x 10-16
Transcriptional distance to Baseline HSC
b
Normalized frequency
Baseline
IFNa
EP
MkP
MEP
E/B/M
pDCP
MDP
cDCP
MP
NP
IMP
IGP
HSC
MLP
ProB
Large PreB
Small PreB
c
Small PreB
P = 0.019
Large PreB
P = 0.007
IGP
P = 0.018
Baseline
IFNa
MkP
P = 0.012
MEP
P = 0.06
EP
P = 0.007
Frequency of progenitor type (%)
d
Log2(IFNa/Baseline)
E/B/M
HSC
MEP
IGP
EP
IMP
MkP
NP
Small PreB
MP
Large PreB
cDCP
ProB
pDCP
MLP
MDP
e
HSC
Baseline
IFNa
P < 2.2x10-16
P < 2.2x10-16
P = 0.59
Normalized protein expression
CD38
CD45RA
CD90
f
Multiparametric flow cytometry of MPN bone marrow
CD34+ TdT+ CD19+ of total CD34+ cells
CD34+ TdT+ CD19+ of total viable cells
CD34- TdT- CD19+ of total viable cells
P = 0.0025
P = 0.0046
P = 0.001
Frequency (%)
No IFNa (n = 47 samples)
IFNa (n = 9 samples)
CALR
JAK2
g
Baseline
IFNa
R=0.64
P=0.0071
Platelet count (x1000/uL)
Frequency of MkP + MEP (%)
h
IFN01
HSC
6.2% in S/G2/M
14.4% in S/G2/M
Baseline
IFNa
S/G2/M
G0/G1
Early GM (IMP & MDP)
13% in S/G2/M
37.8% in S/G2/M
Early Lymphoid (ProB)
24.4% in S/G2/M
66% in S/G2/M
MkP
43.8% in S/G2/M
61.5% in S/G2/M
EP
20.6% in S/G2/M
44.4% in S/G2/M
G2M phase score
S phase score
i
HSC
P = 0.014
IFN01
IFN02
IFN03
IFN04
Baseline
IFNa
Early GM (IMP & MDP)
P = 0.006
Early Lymphoid (ProB)
P = 0.015
MkP
P = 0.16
EP
P = 0.022
Frequency in S/G2/M (%)
j
HSC
Up in Baseline
Up in IFNa
Interferon alpha/gamma response
IFITM3
EPSTI1
ISG15
IFI6
TNF signaling via NFKB
PDE4B
AREG
ID2
FOS
FOSL2
FOSB
JUN
MAFF
NR4A2
NR4A1
NR4A3
CXCR4
IFIT3
MT2A
SELL
HLA-DRB1
IFIT1
SAMHD1
AUTS2
ISG20
TXNIP
CD74
HLA-DQA1
GMPR
SRI
IFI27
-Log10(Adj. P-value)
Log2(Fold Change)
k
TNFA SIGNALING VIA NFKB
TGF BETA SIGNALING
EPITHELIAL MESENCHYMAL TRANSITION
INFLAMMATORY RESPONSE
IL6 JAK STAT3 SIGNALING
P53 PATHWAY
HYPOXIA
KRAS SIGNALING UP
G2M CHECKPOINT
MYC TARGETS V2
E2F TARGETS
MYC TARGETS V1
INTERFERON ALPHA RESPONSE
INTERFERON GAMMA RESPONSE
-Log10(Adj. P-value)*sign of NES
Up in IFNa
Down in IFNa
pDCP
Small PreB
MLP
MEP
E/B/M
HSC
EP
MDP
IMP
ProB
cDCP
IGP
MP
NP
MkP
Large PreB
l
cDCP
TNFA SIGNALING VIA NFKB
Enrichment Score
Adj. P-value = 5.6x10-7
NES = -2.22
Gene Rank
Leading edge genes
Upregulated genes in IGP (Fig. 1e)
AREG
SAT1
CEBPD
PDE4B
DUSP4
G0S2
NR4A3
VEGFA
GPR183
CD83
KYNU
IER3
BCL2A1
MAP3K8
NR4A2
FOSB
FOSL2
IL7R
GEM
CEBPB
NAMPT
NFKB1
SGK1
NR4A1
CD44
JUN
PHLDA1
SOCS3
CXCL2
NFKBIA
IL1B
GADD45B
PTPRE
ID2
SOD2
CDKN1A
SLC2A3
ABCA1
PLPP3
DUSP2
m
IFNa
P < 2.2 x 10-16
P < 2.2 x 10-16
IFNa HSC up module score
IFNa HSC down module score
HSC
IGP

Figure 2. IFNa induces lymphoid differentiation and downregulation of TNFa and TGFb signaling. a Left panel: UMAP showing IFNa treatment time-points for a representative experiment that includes three time points from a patient (IFN03, n = 7329 cells). Right panel: Box plot showing transcriptional distance measurements between HSCs from each time point to HSCs from baseline (IFN03). **b.** Normalized cell type frequencies at baseline and IFNa treatment. Cells from each sample were down-sampled to the same number (n = 500 cells from each sample, n = 9 baseline samples, n = 5 treated samples). **c.** Box plot showing normalized cell type frequencies at baseline and IFNa treatment (n = 9 baseline samples, n = 5 treated samples). P-values from linear mixed model with/without treatment status. **d.** Fold change of IFNa versus baseline cell frequencies in individuals with serial sampling (n = 8 samples from 4 patients). P-values from linear mixed model with/without treatment status. **e.** Box plot showing normalized protein expression on HSCs before and after treatment with IFNa. P-values from linear mixed model with/without treatment status. **f.** Box plot showing cell frequencies of B-lymphoid progenitors and B-cells from bone marrow of patients with early stage MPN treated or not with IFNa (n = 9 and 47 samples, respectively), as determined by multiparametric flow cytometry. P-values from Wilcoxon rank sum test, two-sided. **g.** Platelet counts versus frequencies of MEPs and MkPs (n = 14 samples, P-value from F-test). **h.** Cell cycle gene expression in HSCs and progenitor cells (representative patient IFN01). **i.** Frequencies of cells in G2/M/S phase as assessed in panel **h** (n = 8 samples from 4 patients). P-values from linear mixed model with/without treatment status. **j.** Volcano plot showing DE genes between baseline and IFNa-treated HSCs via linear mixed modeling with/without treatment status. Genes highlighted in blue-green are enriched in the TNFa signaling via NFKB and those in blue enriched in the IFNa/g response, identified by pre-ranked gene set enrichment analysis using the MSigDB Hallmark collection; Boxes represent transcription factors (TF) of the AP-1 (blue) and NR4A (purple) families. **k.** Heatmap showing results of the pre-rank gene set enrichment analysis of genes DE before and after treatment with IFNa across HSC and progenitor subsets. Values show the sign of the normalized enrichment score (NES) multiply by -log10(Adjusted P-value). **l.** Pre-rank gene set enrichment analysis of DE genes before and after treatment with IFNa in cDCPs, showing downregulation of TNFa signaling via NFKB and the leading-edge genes. Genes highlighted in blue represent those upregulated in the IGP vs. IMP. **m.** Box plot showing HSC-specific IFNa-induced signature score in IFNa-treated HSCs and

IGPs. Score calculated using upregulated or downregulated genes (left and right panels, respectively).

Figure 3

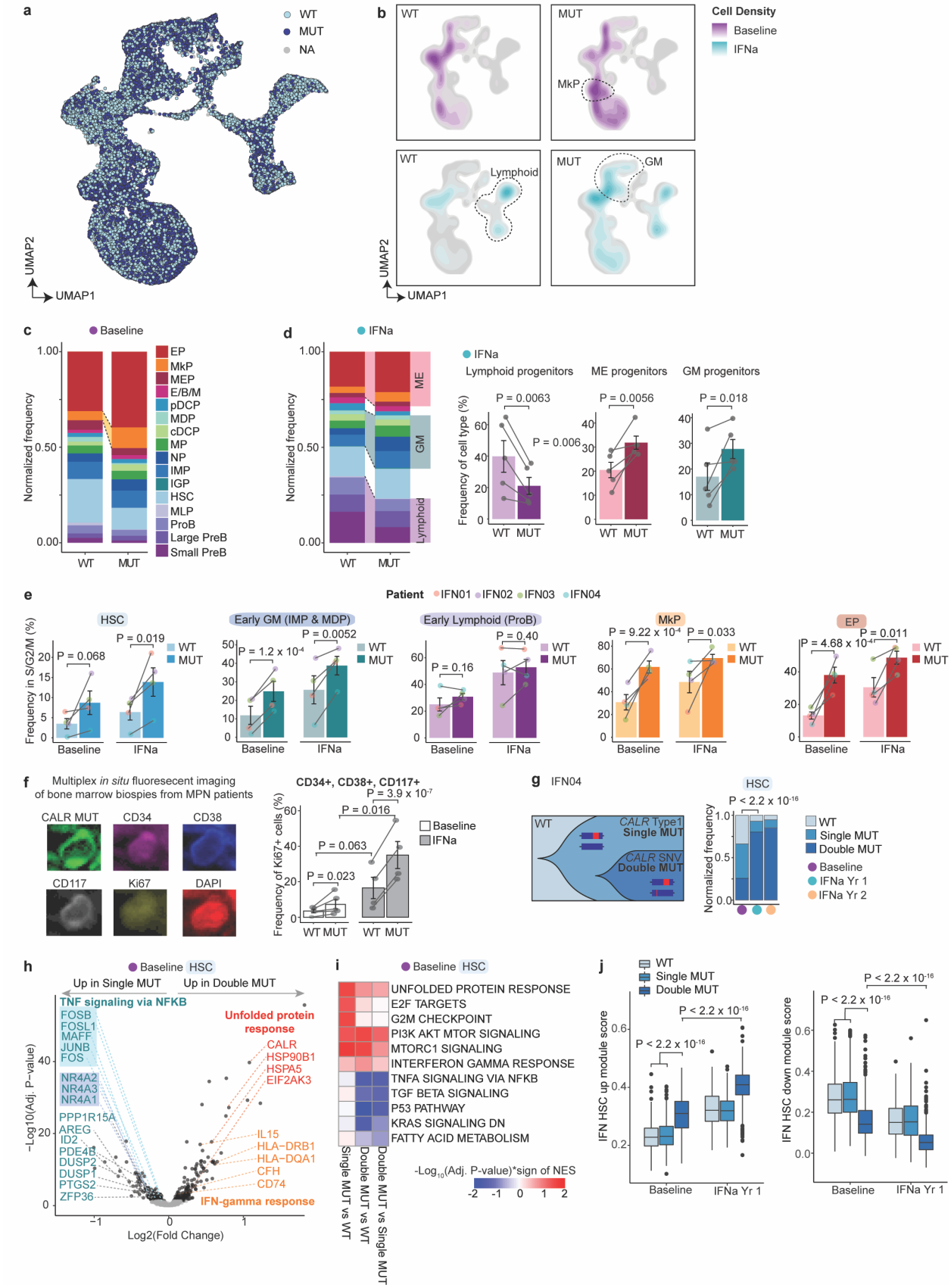

a
WT
MUT
NA
UMAP2
UMAP1
b
WT
MUT
MkP
Lymphoid
GM
Cell Density
Baseline
IFNa
c
Baseline
Normalized frequency
EP
MkP
MEP
E/B/M
pDCP
MDP
cDCP
MP
NP
IMP
IGP
HSC
MLP
ProB
Large PreB
Small PreB
d
IFNa
ME
GM
Lymphoid
Lymphoid progenitors
ME progenitors
GM progenitors
Frequency of cell type (%)
P = 0.0063
P = 0.006
P = 0.0056
P = 0.018
e
Patient IFN01 IFN02 IFN03 IFN04
HSC
Early GM (IMP & MDP)
Early Lymphoid (ProB)
MkP
EP
Frequency in S/G2/M (%)
P = 0.068
P = 0.019
P = 1.2 x 10-4
P = 0.0052
P = 0.16
P = 0.40
P = 9.22 x 10-4
P = 0.033
P = 4.68 x 10-4
P = 0.011
Baseline
IFNa
f
Multiplex in situ fluoresecent imaging of bone marrow biospies from MPN patients
CALR MUT
CD34
CD38
CD117
Ki67
DAPI
CD34+, CD38+, CD117+
Frequency of Ki67+ cells (%)
P = 3.9 x 10-7
P = 0.016
P = 0.063
P = 0.023
Baseline
IFNa
g
IFN04
WT
CALR Type1 Single MUT
CALR SNV Double MUT
HSC
P < 2.2 x 10-16
Normalized frequency
WT
Single MUT
Double MUT
Baseline
IFNa Yr 1
IFNa Yr 2
h
Baseline HSC
Up in Single MUT
Up in Double MUT
TNF signaling via NFKB
FOSB
FOSL1
MAFF
JUNB
FOS
NR4A2
NR4A3
NR4A1
PPP1R15A
AREG
ID2
PDE4B
DUSP2
DUSP1
PTGS2
ZFP36
Unfolded protein response
CALR
HSP90B1
HSPA5
EIF2AK3
IL15
HLA-DRB1
HLA-DQA1
CFH
CD74
IFN-gamma response
-Log10(Adj. P-value)
Log2(Fold Change)
i
Baseline HSC
UNFOLDED PROTEIN RESPONSE
E2F TARGETS
G2M CHECKPOINT
PI3K AKT MTOR SIGNALING
MTORC1 SIGNALING
INTERFERON GAMMA RESPONSE
TNFA SIGNALING VIA NFKB
TGF BETA SIGNALING
P53 PATHWAY
KRAS SIGNALING DN
FATTY ACID METABOLISM
Single MUT vs WT
Double MUT vs WT
Double MUT vs Single MUT
-Log10(Adj. P-value)*sign of NES
j
WT
Single MUT
Double MUT
IFN HSC up module score
IFN HSC down module score
P < 2.2 x 10-16
Baseline
IFNa Yr 1

Figure 3. IFNa differentially regulates mutated vs. normal hematopoietic development. a. UMAP of $CD34^+$ cells (n = 49891cells) with mutation status highlighted for wildtype (WT; n = 17461), *CALR*-mutant (MUT; n = 23207) or unassigned (NA; n = 9223) cells. **b.** UMAP showing densities of mutated vs. wildtype cells at baseline and after IFNa treatment. **c.** Normalized cell frequencies of mutated (n = 15258) vs. wildtype (n = 10838) cells at baseline (n = 9 samples; cells down-sampled to 500 for each sample). **d.** Normalized cell frequencies of mutated (n = 7949) vs. wildtype (n = 6623) cells at IFNa-treatment (n = 5 samples). Left panel: Cells were down-sampled to 500 for each sample in the aggregate analysis. Right panel: Cell frequencies of wildtype vs. mutated cells by progenitor group. **e.** Box plot showing frequencies of cells in G2/M/S phase as assessed in Fig. 2**i** (n = 8 serial samples from 4 patients). P-values from linear mixed model with/without mutation status. **f.** Multiplex *in situ* fluorescence imaging (Vectra Polaris) of bone marrow biopsy sections from MPN patients treated (n = 4) or not (n = 5) with IFNa. Left panel: Representative images. Right panel: Bar plots showing frequencies of proliferating myeloid cells as assessed by Ki67 expression. P-values from linear mixed model with/without treatment status or mutation status **g.** Left panel**:** Schematic of clonal structure of MPN from patient IFN04. Right panel: Bar plot of normalized mutant cell frequencies across treatment time points. **h.** Volcano plot showing DE genes between single mutant and double mutant HSCs at baseline. DE genes identified using linear mixed modeling with/without treatment status. Genes highlighted in blue-green are enriched in the TNFa signaling via NFKB, those in red are enriched in the unfolded protein response and those in orange in the IFN-gamma response. Boxes represent TF of the AP-1 (blue) and NR4A (purple) families. Enrichment identified by pre-ranked gene set enrichment analysis using the MSigDB Hallmark collection. **i.** Heatmap showing results of the pre-rank gene set enrichment analysis of genes DE between HSC clones at baseline. Values correspond to the sign of the normalized enrichment score (NES) times the -log10(Adjusted P-value). **j.** Box plot showing HSC-specific IFNa-induced signature score in HSC clones at baseline and after treatment. Score calculated using upregulated or downregulated genes (left and right panels, respectively).

Figure 4

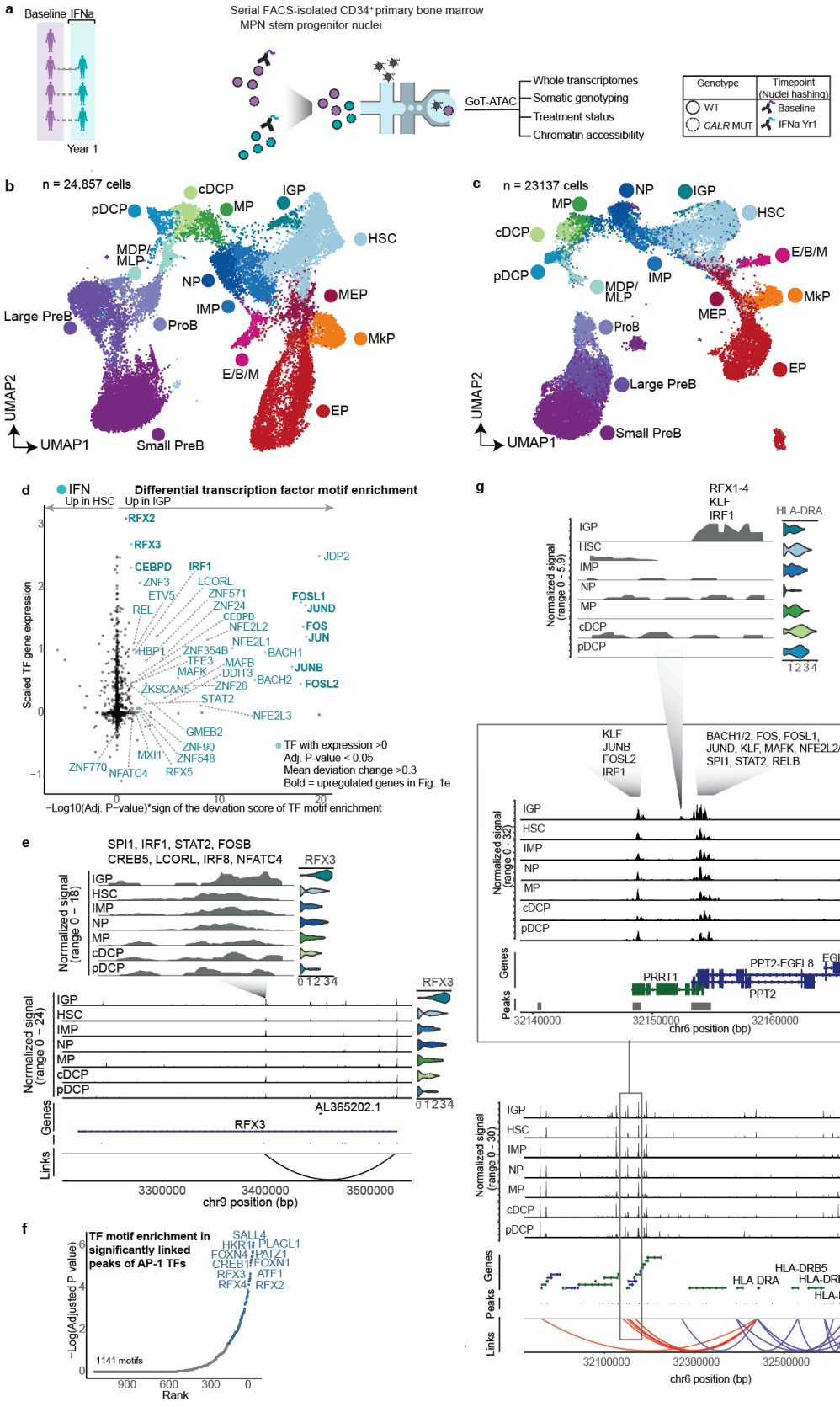

a
Baseline
IFNa
Year 1
Serial FACS-isolated CD34+ primary bone marrow MPN stem progenitor nuclei
GoT-ATAC
Whole transcriptomes
Somatic genotyping
Treatment status
Chromatin accessibility
Genotype
Timepoint (Nuclei hashing)
WT
CALR MUT
Baseline
IFNa Yr1
b
n = 24,857 cells
cDCP
IGP
pDCP
MP
HSC
MDP/ MLP
NP
MEP
IMP
Large PreB
ProB
MkP
E/B/M
EP
Small PreB
UMAP2
UMAP1
c
n = 23137 cells
NP
IGP
MP
HSC
cDCP
E/B/M
pDCP
IMP
MDP/ MLP
MkP
MEP
ProB
EP
Large PreB
Small PreB
UMAP2
UMAP1
d
IFN
Differential transcription factor motif enrichment
Up in HSC
Up in IGP
RFX2
RFX3
JDP2
CEBPD
IRF1
ZNF3
LCORL
ETV5
ZNF571
FOSL1
REL
ZNF24
JUND
CEBPB
NFE2L2
FOS
JUN
NFE2L1
HBP1
ZNF354B
BACH1
TFE3
MAFB
JUNB
MAFK
DDIT3
BACH2
FOSL2
ZKSCAN5
ZNF26
STAT2
NFE2L3
GMEB2
ZNF90
MXI1
ZNF548
ZNF770
NFATC4
RFX5
TF with expression >0
Adj. P-value < 0.05
Mean deviation change >0.3
Bold = upregulated genes in Fig. 1e
Scaled TF gene expression
−Log10(Adj. P−value)*sign of the deviation score of TF motif enrichment
e
SPI1, IRF1, STAT2, FOSB
CREB5, LCORL, IRF8, NFATC4
RFX3
Normalized signal (range 0 – 18)
IGP
HSC
IMP
NP
MP
cDCP
pDCP
Normalized signal (range 0 – 24)
Genes
Links
AL365202.1
RFX3
chr9 position (bp)
f
TF motif enrichment in significantly linked peaks of AP-1 TFs
SALL4
HKR1
PLAGL1
FOXN4
PATZ1
CREB1
FOXN1
RFX3
ATF1
RFX4
RFX2
1141 motifs
−Log(Adjusted P value)
Rank
g
RFX1-4
KLF
IRF1
HLA-DRA
Normalized signal (range 0 - 5.8)
KLF
JUNB
FOSL2
IRF1
BACH1/2, FOS, FOSL1,
JUND, KLF, MAFK, NFE2L2/3
SPI1, STAT2, RELB
Normalized signal (range 0 - 32)
Genes
Peaks
AGPAT1
EGFL8
PPT2-EGFL8
PRRT1
PPT2
chr6 position (bp)
Normalized signal (range 0 - 30)
HLA-DRB5
HLA-DRA
HLA-DRB1
HLA-DQB1
HLA-DQA1
Links
chr6 position (bp)

Figure 4. Regulatory networks of inflammatory granulocytic progenitors. a. Representation of primary bone marrow samples at baseline and after IFNa treatment. Schematic of GoT-ATAC. **b.** UMAP of $CD34^+$ cells based on snRNA-seq data from MPN samples (n = 24,857 cells, 7 samples from 4 individuals), overlaid with cell type assignment. **c.** UMAP of $CD34^+$ cells based on snATAC-seq data overlaid with cell type assignment n = 23,137 cells, 7 samples from 4 individuals). **d.** Motif enrichment and expression of Transcription Factors (TF) in the IGPs relative to HSCs from IFNa-treated samples. Gene expression derived from GoT-IM data. **e.** Chromatin accessibility tracks of regulatory regions of RFX3. **f.** Ranked TF motif enrichment of positively-regulating loci of the AP-1 members, relative to background peaks. **g.** Chromatin accessibility tracks of regulatory regions of *HLA-DRA1* (bottom), distal region enriched with negatively-regulating loci (inset), and IGP-specific regulatory locus (top).

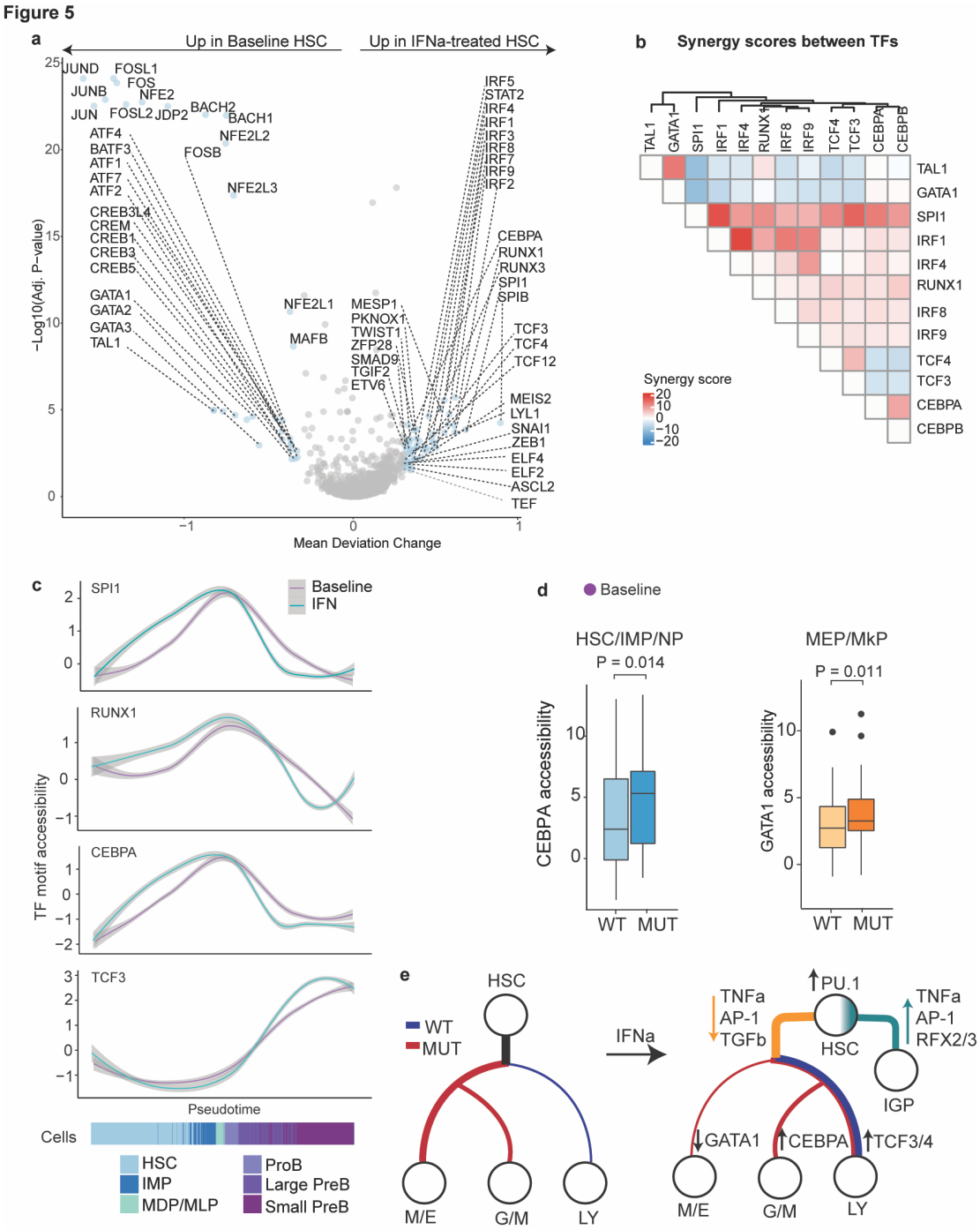


Figure 5. PU.1 is the master regulator of IFNa-mediated remodeling of hematopoiesis. a. Volcano plot showing TF motif differentially enriched in IFNa-treated versus baseline HSCs (ChromVar). P-values from Wilcoxon rank sum test with Benjamini-Hochberg FDR-correction. **b.** Heatmap showing synergy scores between TFs as assessed by measuring the excess variability of accessibility for peaks with both TF motifs (compared to peaks with one motif)[83] **c.** TF motif accessibility across stem and progenitor subsets in lymphoid development between baseline and

IFNa-treated cells. **d.** TF motif enrichment of CEBPA (left) and GATA1 (right) in mutated versus wildtype cells at baseline. **e.** Model of IFNa-induced remodeling of hematopoiesis. At baseline, MPN stem cells show a skewing toward the myeloid lineages. IFNa induces PU.1 activity that regulates downstream differentiation activities. In a minority population of HSCs with high RFX2/3 and AP-1 activities at baseline, PU.1 and IRF1 induce their differentiation into the IGPs, associated with high TNFa signaling. PU.1 also mediates skewing of hematopoiesis toward the lymphoid lineage via co-activity with TCF3 and toward the granulo-monocytic lineage, especially in the mutated cells, via CEBPA. Downregulation of TNF-a and TGFb signaling are also observed in the HSCs upon IFNa treatment.

Supplementary Information

Supplementary Table 1. Summary of patients' clinical history and mutation status for samples IFN01-IFN04 and IFN06-IFN08.

Supplementary Table 2 (related to Fig. 1c). Table of number of cells identified for each cell type in GoT-IM.

Supplementary Table 3 (related to Fig. 1e,h). Differential gene expression analysis between IGP versus IMP and IGP versus HSC1 cell type was performed via a linear mixed model.

Supplementary Table 4 (related to Fig. 1e,h). Gene set enrichment analysis of genes differentially expressed between IGPs versus IMPs, IGPs versus HSC1was performed against MSigDB Hallmark gene sets.

Supplementary Table 5 (related to Fig. 2j). Differential gene expression analysis between baseline and IFNa-treated HSPC subtypes was performed via a linear mixed model.

Supplementary Table 6 (related to Fig. 2k,l). Gene set enrichment analysis of genes differentially expressed between baseline and IFNa-treated HSPC subtypes was performed against MSigDB Hallmark gene sets.

Supplementary Table 7 (related to Fig. 3h). Differential gene expression analysis between single mutant and double mutant HSCs from IFN04 at baseline was performed via logistic regression framework.

Supplementary Table 8 (related to Fig. 3i). Gene set enrichment analysis of genes differentially expressed between HSC clones from IFN04 was performed against MSigDB Hallmark gene sets.

Supplementary Table 9 (related to Fig. 4b). Table of number of cells identified for each cell type in each sample using GoT-ATAC.

Supplementary Table 10 (related to Fig. 4d). Differential transcription factor motif enrichment analysis between IGPs and HSCs after IFNa treatment was performed using Wilcoxon rank sum test.

Supplementary Table 11 (related to Fig 4e). Motifs identified in peaks linked with RFX2/3 via motif scanning with FIMO (v5.4.1).

Supplementary Table 12 (related to Fig 4f). Overenrichment motif analysis for peaks linked with AP-1 genes using FindMotif function (Signac, v1.7.0).

Supplementary Table 13 (related to Fig 4g). Motifs identified in peaks linked with MHC class II genes via motif scanning with FIMO (v5.4.1).

Supplementary Table 14: Table with genes that are identified as being positively regulated by RFX2 and RFX3 using Gene-Peak cis-association .

Supplementary Table 15 (related to Fig. 5a). Differential transcription factor motif enrichment analysis between baseline and IFNa-treated HSCs was performed using Wilcoxon rank sum test.

Supplementary Table 16: List of antibodies used for CITE-seq.

Supplementary Table 17: Primer sequences used in GoT-IM and GoT-ATAC.

References

1. Essers MA, Offner S, Blanco-Bose WE, et al. IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature*. Apr 16 2009;458(7240):904-8. doi:10.1038/nature07815
2. Pietras EM, Lakshminarasimhan R, Techner J-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*. 2014;211(2):245-262. doi:10.1084/jem.20131043 PMID - 24493802
3. Haas S, Hansson J, Klimmeck D, et al. Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. *Cell Stem Cell*. 2015;17(4):422-434. doi:10.1016/j.stem.2015.07.007 PMID - 26299573
4. Baldridge MT, King KY, Boles NC, Weksberg DC, Goodell MA. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. *Nature*. Jun 10 2010;465(7299):793-7. doi:10.1038/nature09135
5. de Bruin AM, Demirel Ö, Hooibrink B, Brandts CH, Nolte MA. Interferon-γ impairs proliferation of hematopoietic stem cells in mice. *Blood*. May 02 2013;121(18):3578-85. doi:10.1182/blood-2012-05-432906
6. Matatall KA, Shen CC, Challen GA, King KY. Type II interferon promotes differentiation of myeloid-biased hematopoietic stem cells. *Stem Cells*. Nov 2014;32(11):3023-30. doi:10.1002/stem.1799
7. Pronk CJ, Veiby OP, Bryder D, Jacobsen SE. Tumor necrosis factor restricts hematopoietic stem cell activity in mice: involvement of two distinct receptors. *J Exp Med*. Aug 01 2011;208(8):1563-70. doi:10.1084/jem.20110752
8. Yamashita M, Passegué E. TNF-α Coordinates Hematopoietic Stem Cell Survival and Myeloid Regeneration. *Cell Stem Cell*. 2019;doi:10.1016/j.stem.2019.05.019
9. Ueda Y, Cain DW, 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*. May 15 2009;182(10):6477-84. doi:10.4049/jimmunol.0803961
10. Weisser M, Demel UM, Stein S, et al. Hyperinflammation in patients with chronic granulomatous disease leads to impairment of hematopoietic stem cell functions. *J Allergy Clin Immunol*. 07 2016;138(1):219-228.e9. doi:10.1016/j.jaci.2015.11.028
11. Pietras EM, Mirantes-Barbeito C, Fong S, et al. Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. *Nat Cell Biol*. 06 2016;18(6):607-18. doi:10.1038/ncb3346
12. Sato T, Onai N, Yoshihara H, Arai F, Suda T, Ohteki T. Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon–dependent exhaustion. *Nature Medicine*. 2009;15(6):696-700. doi:10.1038/nm.1973 PMID - 19483695
13. Huys L, Van Hauwermeiren F, Dejager L, et al. Type I interferon drives tumor necrosis factor-induced lethal shock. *J Exp Med*. Aug 31 2009;206(9):1873-82. doi:10.1084/jem.20090213
14. King KY, Goodell MA. Inflammatory modulation of HSCs: viewing the HSC as a foundation for the immune response. *Nature Reviews Immunology*. 2011;11(10):685-692. doi:10.1038/nri3062 PMID - 21904387
15. Craver BM, El Alaoui K, Scherber RM, Fleischman AG. The Critical Role of Inflammation in the Pathogenesis and Progression of Myeloid Malignancies. *Cancers (Basel)*. Apr 03 2018;10(4)doi:10.3390/cancers10040104

16. Zhang Q, Bastard P, Liu Z, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*. 2020;370(6515):eabd4570. doi:10.1126/science.abd4570 PMID - 32972995
17. Schreiber G. The Role of Type I Interferons in the Pathogenesis and Treatment of COVID-19. *Frontiers in Immunology*. 2020;11:595739. doi:10.3389/fimmu.2020.595739 PMID - 33117408
18. Mosca M, Hermange G, Tisserand A, et al. Inferring the dynamics of mutated hematopoietic stem and progenitor cells induced by IFNα in myeloproliferative neoplasms. *Blood*. 2021;138(22):2231-2243. doi:10.1182/blood.2021010986 PMID - 34407546
19. Czech J, Cordua S, Weinbergerova B, et al. JAK2V617F but not CALR mutations confer increased molecular responses to interferon-α via JAK1/STAT1 activation. *Leukemia*. 2019;33(4):995-1010. doi:10.1038/s41375-018-0295-6 PMID - 30470838
20. 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
21. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med*. Dec 19 2013;369(25):2379-90. doi:10.1056/NEJMoa1311347
22. Yacoub A, Mascarenhas J, Kosiorek H, et al. Pegylated interferon alfa-2a for polycythemia vera or essential thrombocythemia resistant or intolerant to hydroxyurea. *Blood*. 10 31 2019;134(18):1498-1509. doi:10.1182/blood.2019000428
23. Mascarenhas J, Kosiorek HE, Prchal JT, et al. A randomized phase 3 trial of interferon-α vs hydroxyurea in polycythemia vera and essential thrombocythemia. *Blood*. 05 12 2022;139(19):2931-2941. doi:10.1182/blood.2021012743
24. Stoeckius M, Zheng S, Houck-Loomis B, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol*. 12 19 2018;19(1):224. doi:10.1186/s13059-018-1603-1
25. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. Sep 2017;14(9):865-868. doi:10.1038/nmeth.4380
26. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. May 2015;33(5):495-502. doi:10.1038/nbt.3192
27. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. Apr 2017;19(4):271-281. doi:10.1038/ncb3493
28. Miller JC, Brown BD, Shay T, et al. Deciphering the transcriptional network of the dendritic cell lineage. *Nat Immunol*. Sep 2012;13(9):888-99. doi:10.1038/ni.2370
29. Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, Dick JE. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol*. Jul 2010;11(7):585-93. doi:10.1038/ni.1889
30. Hirai H, Zhang P, Dayaram T, et al. C/EBPbeta is required for 'emergency' granulopoiesis. *Nat Immunol*. Jul 2006;7(7):732-9. doi:10.1038/ni1354
31. Pawar SA, Shao L, Chang J, et al. C/EBPδ deficiency sensitizes mice to ionizing radiation-induced hematopoietic and intestinal injury. *PLoS One*. 2014;9(4):e94967. doi:10.1371/journal.pone.0094967
32. Land RH, Rayne AK, Vanderbeck AN, et al. The orphan nuclear receptor NR4A1 specifies a distinct subpopulation of quiescent myeloid-biased long-term HSCs. *Stem Cells*. Jan 2015;33(1):278-88. doi:10.1002/stem.1852

33. Hou Y, Li W, Sheng Y, et al. The transcription factor Foxm1 is essential for the quiescence and maintenance of hematopoietic stem cells. *Nat Immunol*. Aug 2015;16(8):810-8. doi:10.1038/ni.3204
34. Freire PR, Conneely OM. NR4A1 and NR4A3 restrict HSC proliferation via reciprocal regulation of C/EBPα and inflammatory signaling. *Blood*. 03 08 2018;131(10):1081-1093. doi:10.1182/blood-2017-07-795757
35. Reith W, Herrero-Sanchez C, Kobr M, et al. MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. *Genes Dev*. Sep 1990;4(9):1528-40. doi:10.1101/gad.4.9.1528
36. Pugliatti L, Derré J, Berger R, Ucla C, Reith W, Mach B. The genes for MHC class II regulatory factors RFX1 and RFX2 are located on the short arm of chromosome 19. *Genomics*. Aug 1992;13(4):1307-10. doi:10.1016/0888-7543(92)90052-t
37. Siegrist CA, Mach B. Antisense oligonucleotides specific for regulatory factor RFX-1 inhibit inducible but not constitutive expression of all major histocompatibility complex class II genes. *Eur J Immunol*. Nov 1993;23(11):2903-8. doi:10.1002/eji.1830231126
38. Chua RL, Lukassen S, Trump S, et al. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nature Biotechnology*. 2020;38(8):970-979. doi:10.1038/s41587-020-0602-4 PMID - 32591762
39. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med*. Dec 19 2013;369(25):2391-2405. doi:10.1056/NEJMoa1312542
40. Mead AJ, Chowdhury O, Pecquet C, et al. Impact of isolated germline JAK2V617I mutation on human hematopoiesis. *Blood*. May 16 2013;121(20):4156-65. doi:10.1182/blood-2012-05-430926
41. Van Egeren D, Kamaz B, Liu S, et al. Transcriptional differences between JAK2-V617F and wild-type bone marrow cells in patients with myeloproliferative neoplasms. *Exp Hematol*. 03 2022;107:14-19. doi:10.1016/j.exphem.2021.12.364
42. Nam AS, Dusaj N, Izzo F, et al. Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. *Nat Genet*. Sep 22 2022;doi:10.1038/s41588-022-01179-9
43. Essers MAG, Offner S, Blanco-Bose WE, et al. IFNα activates dormant haematopoietic stem cells in vivo. *Nature*. 2009;458(7240):904-908. doi:10.1038/nature07815 PMID - 19212321
44. Hodis E, Torlai Triglia E, Kwon JYH, et al. Stepwise-edited, human melanoma models reveal mutations' effect on tumor and microenvironment. *Science*. 04 29 2022;376(6592):eabi8175. doi:10.1126/science.abi8175
45. Nie Y, Han YC, Zou YR. CXCR4 is required for the quiescence of primitive hematopoietic cells. *J Exp Med*. Apr 14 2008;205(4):777-83. doi:10.1084/jem.20072513
46. Le Naour F, Francastel C, Prenant M, Lantz O, Boucheix C, Rubinstein E. Upregulation of CD9 expression during TPA treatment of K562 cells. *Leukemia*. Aug 1997;11(8):1290-7. doi:10.1038/sj.leu.2400721
47. Clay D, Rubinstein E, Mishal Z, et al. CD9 and megakaryocyte differentiation. *Blood*. Apr 01 2001;97(7):1982-9. doi:10.1182/blood.v97.7.1982
48. Månsson R, Hultquist A, Luc S, et al. Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity*. Apr 2007;26(4):407-19. doi:10.1016/j.immuni.2007.02.013

49. Sanjuan-Pla A, Macaulay IC, Jensen CT, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*. Oct 10 2013;502(7470):232-6. doi:10.1038/nature12495
50. Ciurea SO, Merchant D, Mahmud N, et al. Pivotal contributions of megakaryocytes to the biology of idiopathic myelofibrosis. *Blood*. Aug 01 2007;110(3):986-93. doi:10.1182/blood-2006-12-064626
51. Terui T, Niitsu Y, Mahara K, et al. The production of transforming growth factor-beta in acute megakaryoblastic leukemia and its possible implications in myelofibrosis. *Blood*. Apr 01 1990;75(7):1540-8.
52. Ehninger A, Boch T, Uckelmann H, et al. Posttranscriptional regulation of c-Myc expression in adult murine HSCs during homeostasis and interferon-α-induced stress response. *Blood*. Jun 19 2014;123(25):3909-13. doi:10.1182/blood-2013-10-531038
53. Sharif MN, Šošić De, Rothlin CV, et al. Twist mediates suppression of inflammation by type I IFNs and Axl. *The Journal of Experimental Medicine*. 2006;203(8):1891-1901. doi:10.1084/jem.20051725 PMID - 16831897
54. Marty C, Pecquet C, Nivarthi H, et al. Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. *Blood*. Mar 10 2016;127(10):1317-24. doi:10.1182/blood-2015-11-679571
55. Nivarthi H, Chen D, Cleary C, et al. Thrombopoietin receptor is required for the oncogenic function of CALR mutants. *Leukemia*. Aug 2016;30(8):1759-63. doi:10.1038/leu.2016.32
56. Elf S, Abdelfattah NS, Chen E, et al. Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation. *Cancer Discov*. Apr 2016;6(4):368-81. doi:10.1158/2159-8290.CD-15-1434
57. Kollmann K, Warsch W, Gonzalez-Arias C, et al. A novel signalling screen demonstrates that CALR mutations activate essential MAPK signalling and facilitate megakaryocyte differentiation. *Leukemia*. Apr 2017;31(4):934-944. doi:10.1038/leu.2016.280
58. Gerdes J, Stein H, Pileri S, et al. Prognostic relevance of tumour-cell growth fraction in malignant non-Hodgkin's lymphomas. *Lancet*. Aug 22 1987;2(8556):448-9. doi:10.1016/s0140-6736(87)90977-9
59. Stein H, Bob R, Dürkop H, et al. A new monoclonal antibody (CAL2) detects CALRETICULIN mutations in formalin-fixed and paraffin-embedded bone marrow biopsies. *Leukemia*. Jan 2016;30(1):131-5. doi:10.1038/leu.2015.192
60. Mullally A, Bruedigam C, Poveromo L, et al. Depletion of Jak2V617F myeloproliferative neoplasm-propagating stem cells by interferon-α in a murine model of polycythemia vera. *Blood*. May 02 2013;121(18):3692-702. doi:10.1182/blood-2012-05-432989
61. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics*. 06 01 2019;35(11):1978-1980. doi:10.1093/bioinformatics/bty897
62. Hu F, Yu X, Wang H, et al. ER stress and its regulator X-box-binding protein-1 enhance polyIC-induced innate immune response in dendritic cells. *Eur J Immunol*. Apr 2011;41(4):1086-97. doi:10.1002/eji.201040831
63. Ma S, Zhang B, LaFave LM, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell*. 11 12 2020;183(4):1103-1116.e20. doi:10.1016/j.cell.2020.09.056
64. Seguín-Estévez Q, De Palma R, Krawczyk M, et al. The transcription factor RFX protects MHC class II genes against epigenetic silencing by DNA methylation. *J Immunol*. Aug 15 2009;183(4):2545-53. doi:10.4049/jimmunol.0900376

65. Burda P, Laslo P, Stopka T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia*. 2010;24(7):1249-1257. doi:10.1038/leu.2010.104 PMID - 20520638
66. Hall MA, Curtis DJ, Metcalf D, et al. The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc Natl Acad Sci U S A*. Feb 04 2003;100(3):992-7. doi:10.1073/pnas.0237324100
67. Mikkola HK, Klintman J, Yang H, et al. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature*. Jan 30 2003;421(6922):547-51. doi:10.1038/nature01345
68. Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc Natl Acad Sci U S A*. Jan 21 1997;94(2):569-74. doi:10.1073/pnas.94.2.569
69. Radomska HS, Huettner CS, Zhang P, Cheng T, Scadden DT, Tenen DG. CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors. *Mol Cell Biol*. Jul 1998;18(7):4301-14. doi:10.1128/MCB.18.7.4301
70. Chavez JS, Rabe JL, Loeffler D, et al. PU.1 enforces quiescence and limits hematopoietic stem cell expansion during inflammatory stress. *bioRxiv*. 2020:2020.05.18.102830. doi:10.1101/2020.05.18.102830
71. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods*. Oct 2017;14(10):975-978. doi:10.1038/nmeth.4401
72. Weeks LD, Marinac CR, Redd RA, et al. Age-related Diseases of Inflammation in Myelodysplastic Syndrome and Chronic Myelomonocytic Leukemia. *Blood*. 2021;139(8):1246-1250. doi:10.1182/blood.2021014418 PMID - 34875037
73. Caiado F, Pietras EM, Manz MG. Inflammation as a regulator of hematopoietic stem cell function in disease, aging, and clonal selection. *Journal of Experimental Medicine*. 2021;218(7):e20201541. doi:10.1084/jem.20201541 PMID - 34129016
74. Trowbridge JJ, Starczynowski DT. Innate immune pathways and inflammation in hematopoietic aging, clonal hematopoiesis, and MDS. *Journal of Experimental Medicine*. 2021;218(7):e20201544. doi:10.1084/jem.20201544 PMID - 34129017
75. Murray PJ, Young RA, Daley GQ. Hematopoietic remodeling in interferon-gamma-deficient mice infected with mycobacteria. *Blood*. Apr 15 1998;91(8):2914-24.
76. Chen C, Liu Y, Zheng P. Mammalian target of rapamycin activation underlies HSC defects in autoimmune disease and inflammation in mice. *J Clin Invest*. Nov 2010;120(11):4091-101. doi:10.1172/JCI43873
77. Fast EM, Sporrij A, Manning M, et al. External signals regulate continuous transcriptional states in hematopoietic stem cells. *eLife*. 2021;10:e66512. doi:10.7554/elife.66512 PMID - 34939923
78. Niu H, Fang G, Tang Y, et al. The function of hematopoietic stem cells is altered by both genetic and inflammatory factors in lupus mice. *Blood*. Mar 14 2013;121(11):1986-94. doi:10.1182/blood-2012-05-433755
79. Hadjadj J, Yatim N, Barnabei L, et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*. 2020;369(6504):718-724. doi:10.1126/science.abc6027 PMID - 32661059

80. Darazam IA, Shokouhi S, Pourhoseingholi MA, et al. Role of interferon therapy in severe COVID-19: the COVIFERON randomized controlled trial. *Scientific Reports*. 2021;11(1):8059. doi:10.1038/s41598-021-86859-y PMID - 33850184
81. Tashi T, Swierczek S, Kim SJ, et al. Pegylated interferon Alfa-2a and hydroxyurea in polycythemia vera and essential thrombocythemia: differential cellular and molecular responses. *Leukemia*. 08 2018;32(8):1830-1833. doi:10.1038/s41375-018-0080-6
82. Khan N, Downey J, Sanz J, et al. M. tuberculosis Reprograms Hematopoietic Stem Cells to Limit Myelopoiesis and Impair Trained Immunity. *Cell*. 10 29 2020;183(3):752-770.e22. doi:10.1016/j.cell.2020.09.062
83. Granja JM, Klemm S, McGinnis LM, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol*. 12 2019;37(12):1458-1465. doi:10.1038/s41587-019-0332-7
84. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. Jun 13 2019;177(7):1888-1902 e21. doi:10.1016/j.cell.2019.05.031
85. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Cornell University; 2018.
86. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. 03 2017;27(3):491-499. doi:10.1101/gr.209601.116
87. Tirosh I, Izar B, Prakadan SM, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. Apr 8 2016;352(6282):189-96. doi:10.1126/science.aad0501
88. Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*. 2016:060012. doi:10.1101/060012
89. Gellert P, Segal CV, Gao Q, et al. Impact of mutational profiles on response of primary oestrogen receptor-positive breast cancers to oestrogen deprivation. *Nat Commun*. 11 09 2016;7:13294. doi:10.1038/ncomms13294
90. Weirauch MT, Yang A, Albu M, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. Sep 11 2014;158(6):1431-1443. doi:10.1016/j.cell.2014.08.009
91. Carey CD, Gusenleitner D, Lipschitz M, et al. Topological analysis reveals a PD-L1-associated microenvironmental niche for Reed-Sternberg cells in Hodgkin lymphoma. *Blood*. 11 30 2017;130(22):2420-2430. doi:10.1182/blood-2017-03-770719
92. Patel SS, Lipschitz M, Pinkus GS, et al. Multiparametric in situ imaging of NPM1-mutated acute myeloid leukemia reveals prognostically-relevant features of the marrow microenvironment. *Mod Pathol*. 07 2020;33(7):1380-1388. doi:10.1038/s41379-020-0498-z
93. Rosenthal J, Carelli R, Omar M, et al. Building Tools for Machine Learning and Artificial Intelligence in Cancer Research: Best Practices and a Case Study with the PathML Toolkit for Computational Pathology. *Mol Cancer Res*. 02 2022;20(2):202-206. doi:10.1158/1541-7786.MCR-21-0665
94. Martinez K, Cupitt J. VIPS – a highly tuned image processing software architecture. In Proceedings of IEEE International Conference on Image Processing 2: Genova; 2005. p. 574-577.

Methods

Patient samples

The study was approved by the local ethics committee and by the Institutional Review Board (IRB) of Weill Cornell Medicine. The study was conducted in accordance with the Declaration of Helsinki protocol, and all patients provided informed consent. Cryopreserved bone marrow mononuclear cells or peripheral blood mononuclear cells were obtained from patients with *CALR*-mutated essential thrombocythemia (ET) treated with weekly pegylated IFN-alfa2a during clinical trials MPN-RC-111 (NCT01259817)[22] and MPN-RC-112 (NCT01258856)[23] (n = 4 baseline and 7 treated samples collected from 5 individuals, see **Supplementary Table 1** for clinical and sample information). Baseline ET samples from our previous work[20] were included as additional baseline controls.

Cell preparation

Cryopreserved bone marrow mononuclear cells were thawed and stained using standard procedures with the surface antibody CD34-PE (clone AC136, dilution 1:50, Miltenyi Biotec) and DAPI (Sigma-Aldrich), according to manufacturer's protocol. To eliminate experimental batch effects, cells were labelled simultaneously with hashing antibodies with timepoint-identifying barcodes as described[15] using Hashtag Antibodies 1, 2 or 3 (TotalSeq-A, clone LNH-94, Biolegend) for GoT-IM. To link cell identities to expression of cell surface proteins, cells were also incubated with CITE-seq antibodies[17] according to manufacturer protocol (TotalSeq-A, Biolegend, see **Supplementary Table 16** for information on antibodies). Cells were subsequently sorted for $DAPI^{-}$, $CD34^{+}$ cells using BD Influx at the Weill Cornell Medicine flow cytometry core.

GoT-IM

In order to simultaneously capture genotyping data and whole transcriptomic data, Genotyping of Transcriptomes (GoT) was performed by adapting the 10x Genomics platform as previously described[18]. FACS-sorted $CD34^{+}$ cells for each time point from the same individual were pooled, targeting a total of 10,000-15,000 cells. The standard 10x Genomics Chromium 3′ (v.3 or v.3.1 chemistry) was implemented according to the manufacturer's recommendations up to the cDNA amplification step (10X Genomics, Pleasanton, CA). After cDNA amplification and SPRI cleanup, 10x scRNA-seq and ADT/HTO libraries were generated as recommended. A portion of the cDNA

was used for somatic genotyping as previously described[20]. Briefly, to capture the somatic genotypes of cells, cDNA was amplified with a locus-specific amplification (10 PCR cycles), using the generic forward SI-PCR primer and the following locus-specific reverse primer for the CALR mutations (see **Supplementary Table 17** for primer sequences). The amplified locus-specific cDNAs are then cleaned using SPRI purification to remove unincorporated primers. Finally, the targeted amplicon libraries are generated through a PCR performed with a generic forward PCR primer together with an RPI-x primer (**Supplementary Table 17**). The targeted amplicon libraries were spiked into the remainder of the gene expression and immunophenotyping libraries to be sequenced together on a NovaSeq (Illumina, San Diego, CA). The cycle settings were as follows: 28 cycles for Read 1, 90 cycles for Read 2, 8 cycles for i7 sample index and 8 cycles for i5 sample index.

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

For single-cell GoT-IM data from IFN01-IFN05 patient samples, the pooled scRNA-seq, CITE-seq and hashing libraries were processed with Cell Ranger (v6.1.1 and v6.1.2) using cellranger-multi pipeline (v1). The reads were aligned to the human genome GRCh38 with default parameters. The Seurat package[26] (v4.1.0) was used to perform unbiased clustering of the $CD34^+$ sorted cells from patient (v4.1.0). In brief, for individual datasets, cells with UMI > or < 3 standard deviations from the mean UMI or mitochondrial gene percentage > 10%, were filtered (Supplementary Figure 1a) . The HTO data was normalized with centered log-ratio (CLR) transformation and used to assign the time-points for each experiment[24]. The cells from each time point were analytically separated into individual datasets. These individual datasets, together with the baseline ET samples from our previous work[20], were integrated and underwent batch-correction within Seurat, which implements canonical correlation analysis and the principles of mutual nearest neighbor[84]. Recommended settings were used for the integration (30 canonical correlation vectors for canonical correlation analysis in the FindIntegrationAnchors function and 30 principal components for the anchor weighting procedure in IntegrateData function). Principal component analysis was performed using variable genes using recommended settings (i.e. top 2000 variable genes using variance stabilizing transformation)[84]. The first statistically significant 30 principal components were used as inputs to the UMAP algorithm for cluster visualization[85]. 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 of the two comparison sets as input, and log-transformed fold change of 0.25 as the threshold, using Wilcoxon rank sum test). We identified 35 clusters in the integrated data, which were annotated according to canonical lineage markers identified previously in single-cell RNA-seq data of normal hematopoietic progenitor cells[27]. These clusters were collapsed into 15 main progenitor subsets based on expression of levels of these canonical markers (**Supplementary Fig. 2a,b**). The CITE-seq data was normalized using the CLR transformation, and used to identify the HSCs (**Supplementary Fig. 2c**)

Ironthrone-GoT for processing targeted amplicon sequences and mutation calling

Analysis of the GoT library was carried out as described previously[20] using Ironthrone pipeline V2.1[42]. Amplicon reads (Read 2) were screened for the presence of the primer sequence and the shared sequence (i.e. the expected sequence between the primer sequence and the mutation locus). Reads (Read 1) from GoT-IM experiments were also assessed for matching to the cell barcode list of the 10x dataset. A mismatch of 20% was allowed for all sequence matching steps. Only UMIs with at least 2 or more supporting reads were retained for final genotyping assignments, after the UMI collapse algorithm[42,86]. Filtered cells were then genotyped as follows: cells with at least one mutant UMI were categorized as mutant cells whereas cells with no mutant UMI and at least one wildtype UMI were identified as wildtype.

Gene module scoring, differential expression and gene set enrichment analysis

For examining gene and gene module expression (e.g. HSC-specific IFNa-induced gene signature), the function AddModuleScore was used to calculate the relative expression of the genes for each cell within the Seurat package was calculated using the AddModuleScore function)[84]. To calculate the module expression of cell-cycle related genes, G2M phase and S phase marker genes were used as available in Seurat with CellCycleScoreing function. Briefly, control gene module expressions were calculated and subtracted from the average gene module expression of interest, as previously described[87]. All analyzed genes were classified based on average expression into 24 bins, and for each gene in the module, 100 control genes are randomly selected from the same expression bin as the gene of interest[87]. For statistical analysis, if cell types are being compared, cell type status was entered as the fixed effect and subjects as random effects in a linear mixed

model. If the wildtype and mutated cells are being compared, genotype status was entered as the fixed effect and subjects as random effects in a linear mixed model. P-values were obtained by likelihood ratio tests of the full model with the fixed effect against the model without the fixed effect.

For differential gene expression testing within an individual (e.g. **Fig. 3h**) between two groups were calculated using FindMarkers function using the logistic regression framework. The tested genes included the top 2,000 variable genes from the CCA integration, which were filtered for those expressed in at least 10% of either group. In aggregated differential gene expression analysis, the two groups (e.g. treated versus baseline) via the linear mixed model framework. For each gene, treatment status was entered as the fixed effect and subjects as random effects. P-values were obtained by likelihood ratio tests of the full model with the fixed effect against the model without the fixed effect. Pathway enrichment was performed via a pre-ranked gene set enrichment approach (ranking based on the sign of the fold change * -log10(P-value)) using the msigdbr (v7.2.1) and fgsea (v1.12.0) R packages, using the canonical Hallmark pathway genes from MsigDB[88].

GoT-ATAC

Cryopreserved bone marrow mononuclear cells were thawed and stained using standard procedures (30 min, 4°C) with the surface antibody CD34-PE (clone AC136, dilution 1:50, MACS) and DAPI (Sigma-Aldrich), according to manufacturer's protocol. Cells were subsequently sorted for $DAPI^{-}$, $CD34^{+}$ cells using BD Influx at the Weill Cornell Medicine flow cytometry core. Nuclei were isolated from $DAPI^{-}$, $CD34^{+}$ cells according to 10x Genomics Demonstrated Low Cell Input Nuclei Isolation protocol. Lysis buffer was prepared following manufacturer's recommendations and then split into aliquots for each serial sample. Either TotalSeq-A Anti-Nuclei Pore Complex Proteins Hashtag 9 or 10 Antibody (1uL at 1:5 dilution; BioLegend) was added to each aliquot of lysis buffer. Low-input nuclei isolation was otherwise performed following manufacturer's recommendations. Subsequently, nuclei from each time point were counted and pooled together at approximately equal proportions. For IFN07, additional nuclei from the IFNa-treated sample were available to be run on a separate lane. Single-nucleus

gene expression (GEX) and chromatin accessibility libraries were constructed from the pooled nuclei according to the Chromium Next GEM Single Cell Multiome User Guide (10X Genomics). Genotyping libraries targeting the *CALR* mutant transcripts were constructed from the remaining amplified cDNA, similar to the original GoT method. For each PCR, 12.5uL Kapa HiFI HotStart Ready Mix was mixed with 0.75uL of 10uM forward primer, 0.75uL of 10uM reverse primer and nuclease-free water for a total reaction volume of 25uL. In the first PCR, 3uL cDNA was re-amplified with Partial TSO and Partial Read 1 primers, using the following PCR condition: 98°C for 3min; 3 cycles of 98°C for 15sec, 67°C for 20sec and 72°C for 1min; 72°C for 1min. The re-amplified sampled was purified and concentrated via 0.7X SPRI cleanup, eluting it into 10uL Buffer EB. To pre-enrich the *CALR* mutation locus, a gene-specific PCR was performed with 3uL of cleaned re-amplified cDNA and Partial Read 1 and gene-specific primers (gene-specific, see **Supplementary Table 17** for primer sequences). The following PCR condition was used: 98°C for 3min; 11 cycles of 98°C for 20sec, 60°C for 20sec, and 72°C for 2min; 72°C for 2min. After 0.7X SPRI cleanup, cDNA was eluted into 10uL Buffer EB. *CALR* locus-specific amplification was then performed with 3uL of cleaned gene-specific amplified cDNA and SI-PCR and loci-specific Primers, using the PCR condition: 98°C for 3min; 11 cycles of 98°C for 20sec, 60°C for 20sec, and 72°C for 2min; 72°C for 2min. A 0.7X SPRI cleanup was performed, and cDNA was eluted into 11uL Buffer EB. Finally, to construct the targeted amplicon library, loci-amplified cDNA was mixed with P5 Generic and RPxx indexing primers and amplified with the PCR condition: 98°C for 3min; 5 cycles of 98°C for 15sec, 60°C for 20sec, and 72°C for 1min; 72°C for 1min. The constructed library was cleaned via 0.8X SPRI cleanup and eluted into 12uL Buffer EB.

At the cDNA amplification stage of the Chromium Next GEM Single Cell Multiome protocol, supernatant from the 0.6X size selection was retained and was used to generate the hashing libraries (see **Supplementary Table 17** for TruSeq DNA D7xx_s primers used) as per HTO protocol[25] with the following modification. For the hashing library construction step, the PCR reaction was prepared with 0.65uL of 10uM SI-PCR primer, 0.65uL of 10uM TruSeq DNA D7xx_s primer, 11.25uL cleaned supernatant and 12.5 uL Kapa HiFI HotStart Ready Mix (Roche, Basel, Switzerland). Hashing, gene expression and genotyping libraries were pooled and sequenced together on a NovaSeq (Illumina) with cycle settings: 28 cycles for Read 1, 90 cycles for Read 2, 10 cycles for i7 sample index and 10 cycles for i5 sample index. The ATAC library

was sequenced separately on a NovaSeq, with cycle settings: 50 cycles for Read 1 and Read 2, 8 cycles for i7 sample index and 24 cycles for i5 sample index.

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

For single-nuclei GoT-ATAC data from IFN01 and IFN06-08 patient samples, 10x data were processed using Cell Ranger (v6.1.1 and v6.1.2). Multi-omic nuclear data for snATAC-seq and snRNA-seq were processed together with Cell Ranger Arc (v2.0.0). snRNA-seq data was also combined with cell hashing data (HTO) and run using the Cell Ranger Multi pipeline (v1). The reads were aligned to the human reference genome GRCh38. The downstream analysis of the processed data was performed using Seurat (v4.1.0)[26] and Signac (v1.5.0)[89] packages. Ironthrone-GoT protocol mentioned above was used to determine mutation calling within the GoT-ATAC assay. For the ATAC analysis, we called peaks on individual samples using MACS2 peak caller[63]. Gene annotations from EnsDb.Hsapiens.v75 and motifs annotations from Cis-BP[90] for TF binding motifs were utilized. Cells with blacklist ratio >0.02, TSS enrichment <2 and nucleosome signal >4 were filtered out. For the RNA data, cells with UMI > or < 3 standard deviations from the mean UMI or mitochondrial gene percentage > 25%, were filtered. The nuclei hashing data processing was performed as for the GoT-IM data. As nuclear hashing is known to be noisy[89] the nuclear hashing data was used in combination with cell clustering data, as the cells cluster based on treatment status. snRNA-seq data integration and and cell type assignment was performed as described. 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 dimensional reduction using latent semantic indexing (LSI). The first 2:30 dimensions were retained and batch-correction was perfomed with runHarmony (Harmony, v0.1.0) which iteratively learns cell-specific linear correction function to account for batch effect.

Identification of distal regulatory elements with gene-peak cis-association

For each GoT-ATAC sample, we examined all ATAC peaks within ± 500kb of all annotated TSS to identify regulatory networks of genes using LinkPeaks function[63]. Pearson correlation between gene expression and accessibility of the peaks located in the window was calculated after correcting for bias arising from GC content, overall accessibility and peak size. Recommended

settings were implemented (200 background peaks per peak with similar GC content and accessibility, P-value < 0.05 and min.cell = 10).

Motif enrichment analysis

Per-cell TF motif activity score (chromatin accessibility) was calculated by running chromVAR 1.18.0[71]. We used the curated Cis-BP motif database[90] which contains 1141 human TF motif PFMs. The function matchMotifs was first called to identify which peaks contain which motifs (P-value = 5×10^{-5}). A set of background peaks that are similar to a peak in GC content and average accessibility was internally picked and used for normalizing the deviation scores. Deviation Z-scores, namely bias-corrected deviation z-scores in accessibility from the expected accessibility based on the average of all the cells, were then calculated for each TF motif and each cell.

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 given genomic feature set, 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. ChromVAR was used to compute the synergy between pairs of TF motifs, where synergy is defined as the excess variability of chromatin accessibility for peaks sharing both motifs compared to a random sub-sample of the same size of peaks with one motif. High synergy usually indicated that cooperative binding relationship between pairs of TFs. The function getAnnotationSynergy was called to calculate synergy scores.

Retrospective flow cytometry data analysis

Retrospective flow cytometry data analysis was performed in accordance with relevant guidelines, regulations and approval by the Institutional Review Board at Weill Cornell Medicine (IRB #1007011151). Patient flow cytometry data selected corresponded to patients with myeloproliferative neoplasms and which had the same antibody panel analyzed by flow cytometry. The antibody panel chosen for evaluation was a modified version of the EuroFlow AML/MDS tube #4[51, 52] that consisted of the following antibody-fluorophore pairs, in addition to

forward scattering and side scattering pulse area and width measurements (FSC-A, FSC-H, SSC-A, and SSC-H): cytoplasmic TdT/FITC (clone HT-6, Agilent/Dako, cat. F7139), CD56/PE (clone C5.9, Cytognos, cat. CYT-56PE), CD34/PerCP-Cy5.5 (clone 8G12, BD Biosciences, cat. 347213), CD117/PE-Cy7 (clone 104D2D1, Beckman Coulter/Immunotech, cat. IM3698), CD7/APC (clone GP40 [Leu-9], Invitrogen, cat. 17-0079-42), Fixable Viability Stain 700 (BD Biosciences, cat. 564997), CD19/APC-H7, HLA-DR/Pacific Blue (clone SJ25C1, BD Biosciences, cat. 643078), and CD45/V500 (clone 2D1, BD Biosciences, cat. 347213). Data was collected using BD LSR II flow cytometers, with approximately 500,000 events collected per antibody panel per sample to generate the raw data FCS files.

Custom software was developed using python, FlowKit[53] and umap-learn[54–56] to detect the antibody panels that were used to generate each raw data FCS file and to determine the unused flow cytometer channels that should be disregarded using the self-contained metadata for each file. Subsamples from each FCS file were combined to make an "ensemble FCS file" that could be used to create the UMAP embedding that could be applied to all of the individual files. Each subsample consisted of the same number of randomly-selected flow cytometer events such that the combined total number of events was approximately 250,000 for each unique antibody panel. The various channels were normalized and processed using UMAP to calculate the normalization constants and UMAP embedding that were then applied to all FCS files of the given antibody panel. Of note, the UMAP calculations included the forward scatter height (FSC-H), side scatter height (SSC-H), and each of the defined fluorescence channels. The normalization factors and UMAP embedding were then applied to all of the individual files. Modified FCS files were created that included the UMAPs as additional channels for subsequent evaluation and gating using FlowJo software (v10.8.1, FlowJo LLC, Ashland, Oregon, USA).

Using FlowJo, appropriate gates based on the UMAP plot were determined using the ensemble FCS file. Additional standard gating was also performed using the original data channels (gating using the other channels is essential to determine the identities of the various cell clusters within the UMAP plots). UMAP gates were based on data after gating out doublets and non-viable cells via standard gating approaches. Once the UMAP gates were determined to adequate satisfaction (sufficient segregation of cell subpopulations and verified to encompass cells of the same or similar type), they were then applied to all of the FCS files of the given antibody panel. The FCS files were divided into an interferon-treated cohort (n = 9) and a non-interferon-treated cohort (n =

47). The ratios and absolute numbers of cells, as well as other summary statistics were then calculated, and the values exported as CSV files. Statistics included numbers of $CD34^+$ blasts; $CD19^+$, $cTDT^+$, $CD34^+$ lymphoid progenitors; $CD19^+$ lymphocytes; CD19-negative lymphocytes/NK cells; and monocytes. Relevant distributions of cells for each cohort are plotted, and the Wilcoxon p-statistic was calculated for various compared distributions.

Multiplexed Immunofluorescence

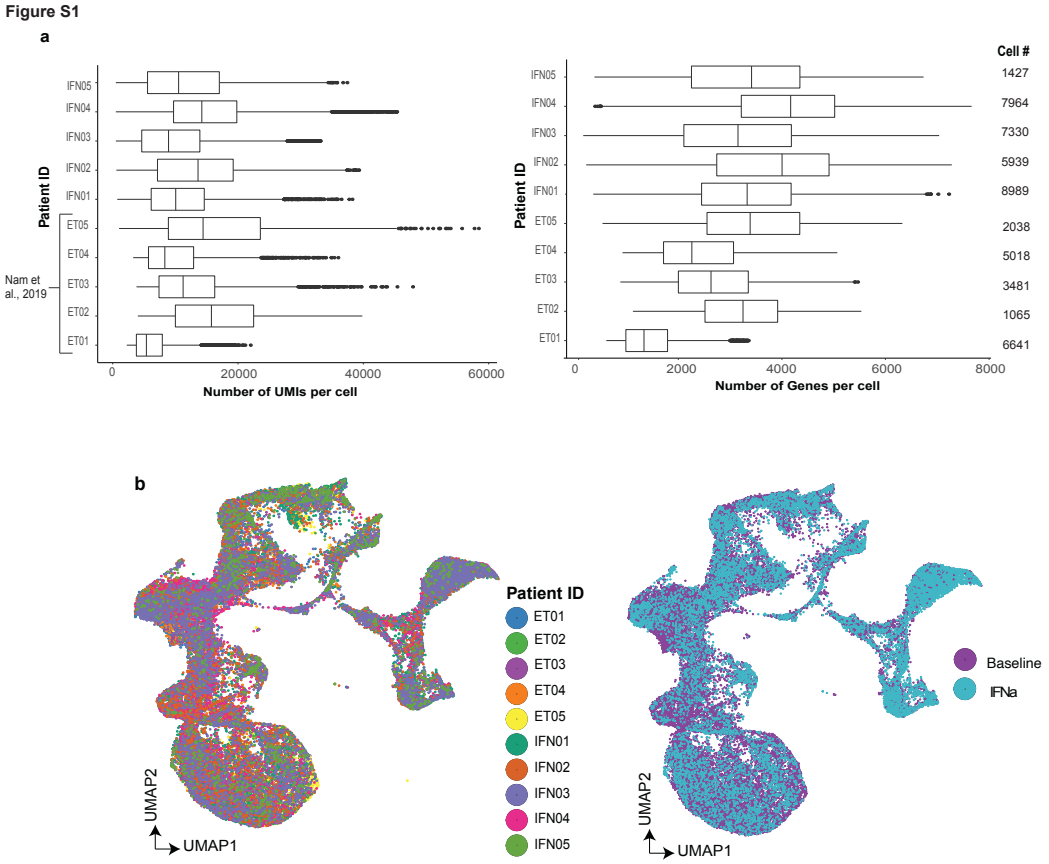
Multiplexed immunofluorescence (mIF) was performed using the Opal system (Akoya Biosciences) by staining 4 micron-thick Bouin-fixed, paraffin-embedded whole-tissue sections from decalcified human bone marrow core biopsy specimens in a Bond RX automated tissue stainer (Leica Biosystems, Buffalo Grove, IL), as described previously[91,92]. Briefly, tissue sections were first deparaffinized prior to EDTA-based antigen retrieval (Leica ER2 solution, 20min). A cyclical staining protocol was then performed, 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) to execute stripping of primary/secondary antibody complexes between staining cycles. 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 applied to all cases, in a sequential order as shown: 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). Slides were cover-slipped using ProLong™ Diamond Antifade Mountant (Invitrogen). Whole slide scans were subsequently obtained at 20X magnification using the Vectra Polaris Automated Quantitative Pathology Imaging System (Akoya Biosciences) to generate a collection of tiled images, which were subsequently spectrally unmixed in InForm (v2.4.8, Akoya Biosciences). Unmixed tiles were finally fused together in HALO (v3.3.2541.231, Indica Labs) to generate a multi-layered TIFF image file for each sample, which was used in downstream analyses.

Image Analysis with PathML

Vectra whole-slide images (WSIs) were digitized using digital whole-slide scanners and stored in tiff format. WSIs of bone marrow sections were captured (n = 9 samples). Each sample was stained based on 8 cell markers including DAPI, mutant-specifc CALR, CD38, CD117, TdT, CD34, Ki67, and auto-fluorescence. Using Fiji, the brightness/contrast is adjusted from the Image menu. To analyze the Fiji-preprocessed WSIs, a machine-learning based package namely pathML was applied to images[93]. The PathML framework is based on PyThrch and TensorFlow deep learning libraries to segment the images. For tiling, the Slide.setTileProperties() method was run for extracting and storing each tile into a single image file. WSI is read using the pyvips library[94] within PathML. Next, the DenseNet CNN is trained based on the default parameters of pathML to detect the tissue, artefact and background regions from a robust dataset containing various tissues and species types[93]. PathML package provides a function for quantifying the images based on image intensity. Using that function, a counts matrix is generated per each cell marker based on the intensity of marker. To remove the noise from the counts matrix data, any cell with 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.

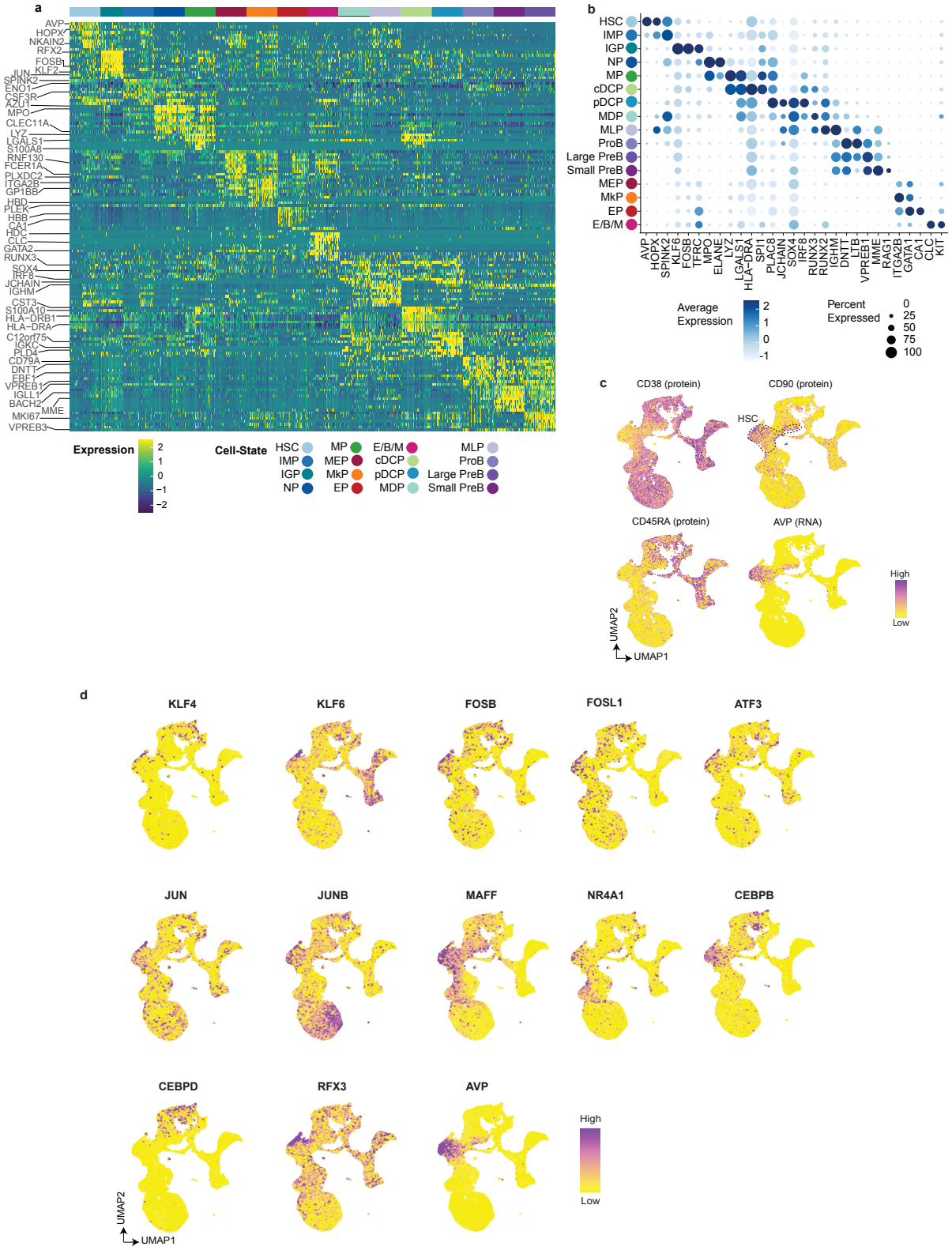
Data Availability. Data generated in this study will be available upon publication.

Supplementary Figures



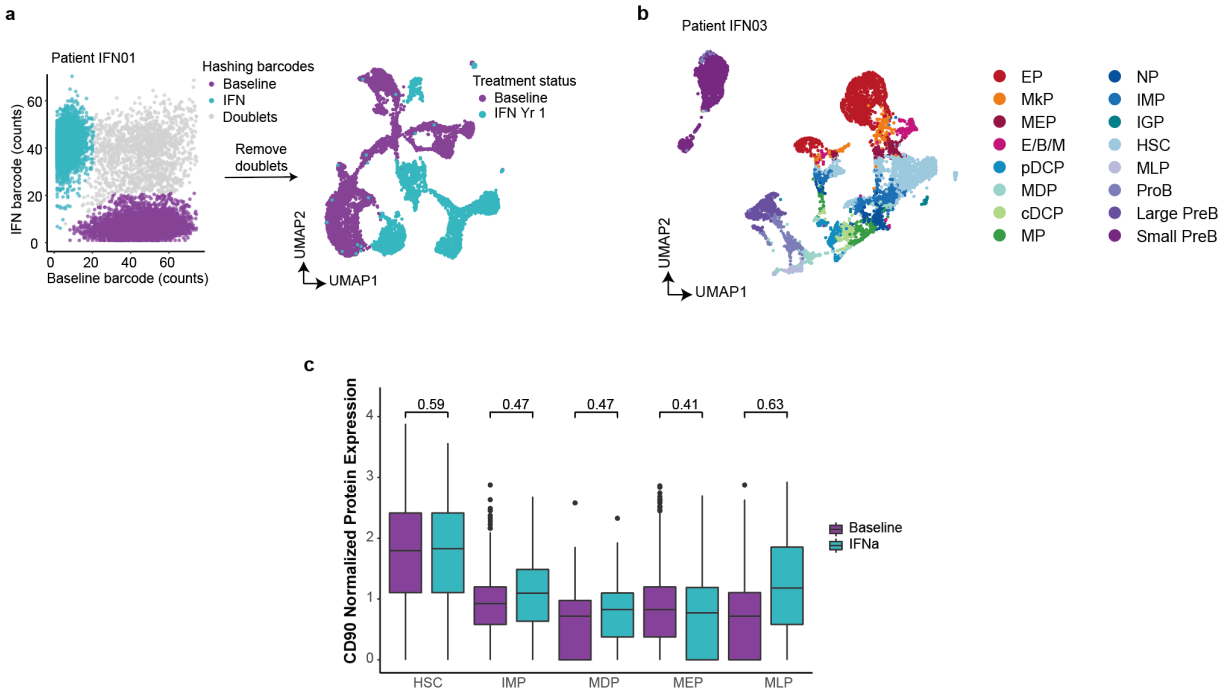
Supplementary Figure 1 (related to Figure 1): GoT-IM captures genotyping, immunophenotyping, and whole transcriptomes in thousands of CD34$^+$ cells at baseline and treatment with IFNa. a. Box plots showing number of UMIs (left panel) and genes (right panel) detected per cell in sorted CD34$^+$ hematopoietic progenitors from each patient after filtering based on Quality Control (QC) metrics. Cells profiled using GoT-IM (Genotyping of Transcriptomes integrated with immunophenotyping). **b.** UMAP of sorted CD34$^+$ progenitors (n = 49891 cells) highlighting MPN samples (n = 10 individuals) (left panel) and treatment status (right panel) after integration with Seurat package.

Figure S2

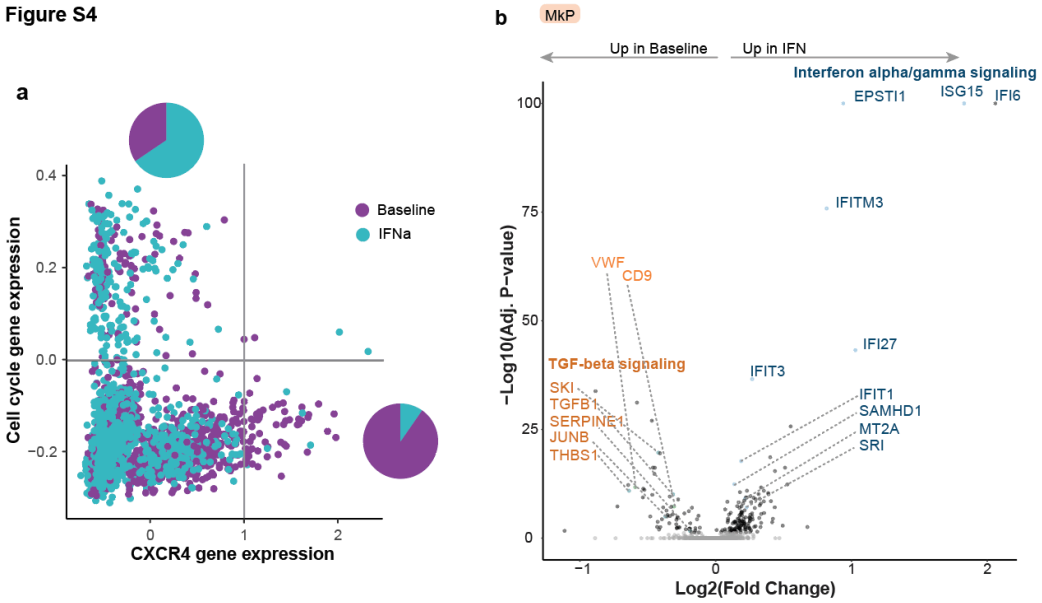

a
AVP
HOPX
NKAIN2
RFX2
FOSB
JUN
KLF2
SPINK2
ENO1
CSF3R
AZU1
MPO
CLEC11A
LYZ
LGALS1
S100A8
RNF130
FCER1A
PLXDC2
ITGA2B
GP1BB
HBD
PLEK
HBB
CA1
HDC
CLC
GATA2
RUNX3
SOX4
IRF8
JCHAIN
IGHM
CST3
S100A10
HLA−DRB1
HLA−DRA
C12orf75
IGKC
PLD4
CD79A
DNTT
EBF1
VPREB1
IGLL1
BACH2
MME
MKI67
VPREB3
Expression
2
1
0
−1
−2
Cell-State
HSC
IMP
IGP
NP
MP
MEP
MkP
EP
E/B/M
cDCP
pDCP
MDP
MLP
ProB
Large PreB
Small PreB
b
HSC
IMP
IGP
NP
MP
cDCP
pDCP
MDP
MLP
ProB
Large PreB
Small PreB
MEP
MkP
EP
E/B/M
AVP
HOPX
SPINK2
KLF6
FOSB
TFPI
MPO
ELANE
LYZ
LGALS1
HLA−DRA
SPI1
PLAC8
JCHAIN
SOX4
IRF8
RUNX3
RUNX2
IGHM
DNTT
LTB
VPREB1
MME
RAG1
ITGA2B
GATA1
CA1
CLC
KIT
Average Expression
2
1
0
-1
Percent Expressed
0
25
50
75
100
c
CD38 (protein)
CD90 (protein)
HSC
CD45RA (protein)
AVP (RNA)
High
Low
UMAP2
UMAP1
d
KLF4
KLF6
FOSB
FOSL1
ATF3
JUN
JUNB
MAFF
NR4A1
CEBPB
CEBPD
RFX3
AVP
High
Low
UMAP2
UMAP1

Supplementary Figure 2 (related to Figure 1): GoT-IM of baseline and IFNa treated samples reveal a novel cell state. a. Heatmap of top 10 differentially expressed genes for each progenitor cell type. Cells of each cell type were down-sampled to the same number (n = 100 cells per cluster) **b.** Dot-plot showing expression levels of cell type specific gene markers in each progenitor subset. **c.** UMAP of sorted CD34+ progenitors (n = 49891 cells) highlighting CD38, CD90 and CD45RA protein expression and *AVP* RNA expression. **d.** UMAP showing gene expression levels of differentially upregulated TFs in IGP compared to IMPs**.**

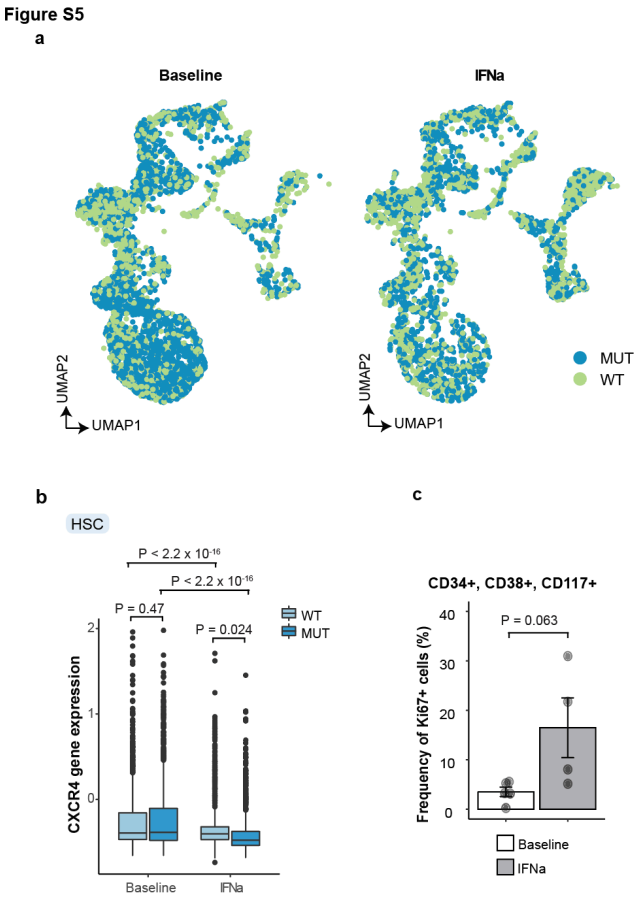
Figure S3



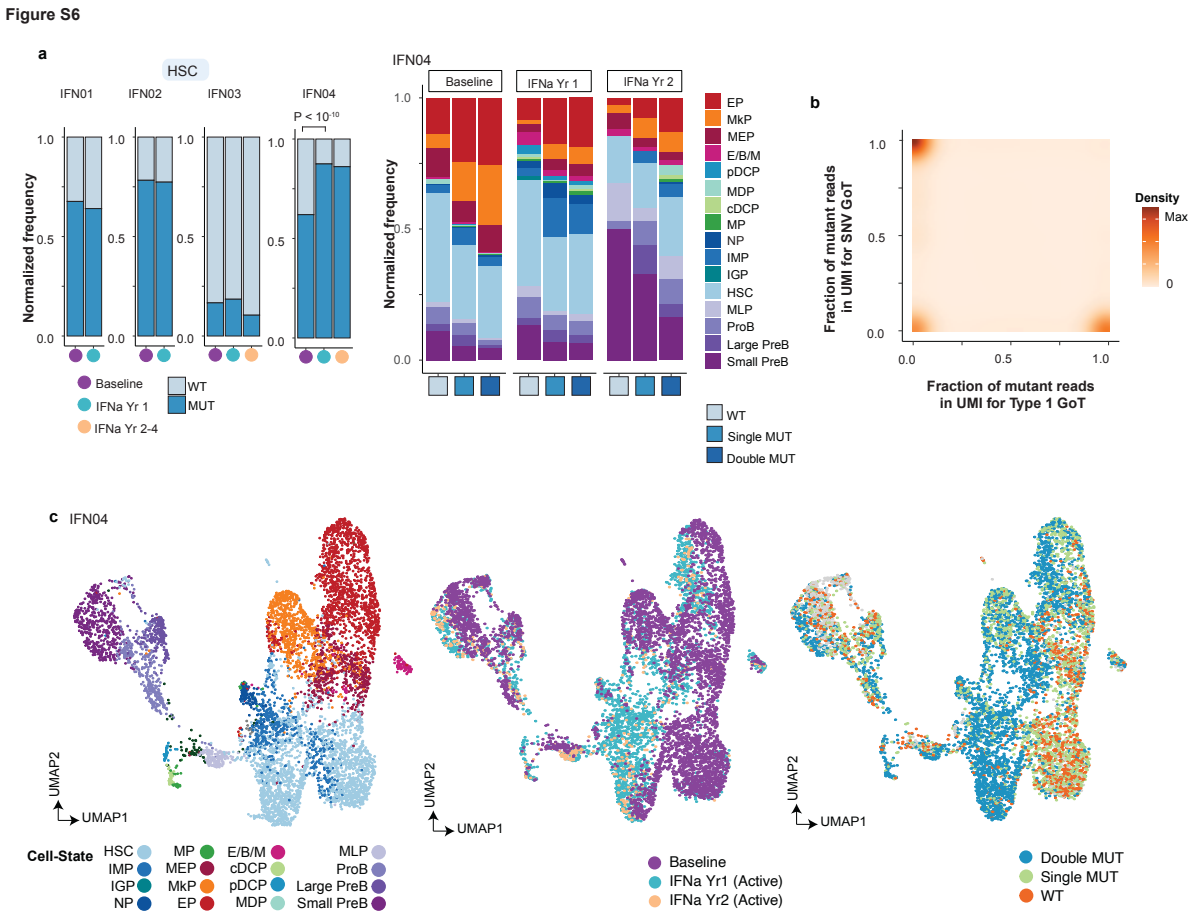
Supplementary Figure 3 (related to Figure 2). IFNa exerts strong transcriptional perturbations in HSPCs. a. Left panel: Analytical time point assignment (data demultiplexing). Using the time point specifying barcodes, baseline and IFNs-treated cells were separated. Cells in which both barcodes are detected are considered as doublets and excluded. Right panel: UMAP of $CD34^+$ cells from a representative experiment that includes two time points (IFN01, n = 5225 baseline cells, 3764 IFNa treated year 1 cells), highlighting treatment status. **b.** UMAP of $CD34^+$ cells from patient IFN03 highlighting cell type assignments (n = 727 baseline cells, 1135 IFNa treated cells at year 1, 5467 IFNa treated cells at year 4). **c.** Box plots showing normalized CD90 protein expression levels in $CD34^+$ progenitors before and after treatment with IFNa. P-values from linear mixed model with/without treatment status.



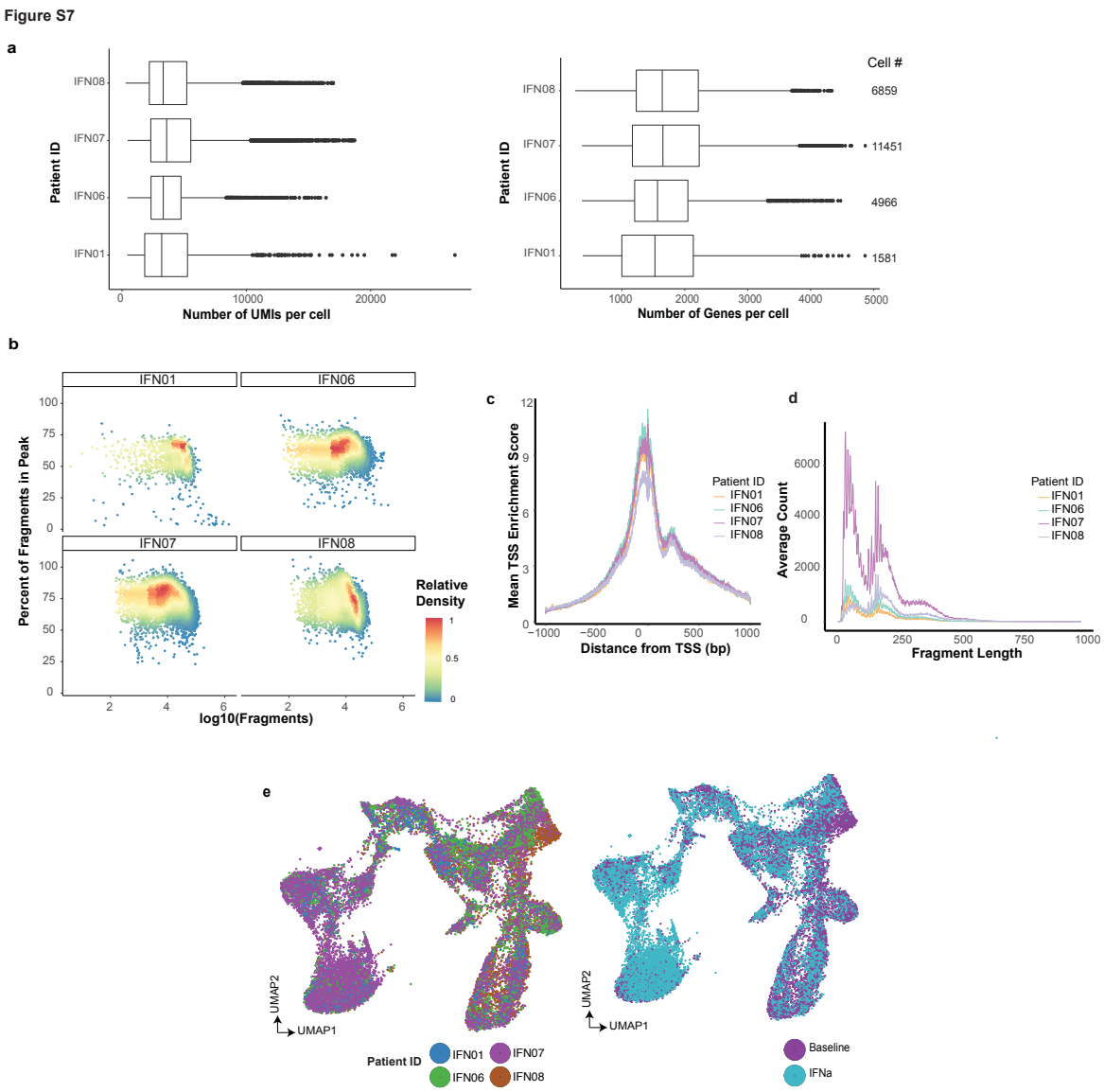
Supplementary Figure 4 (related to Figure 2). IFNa induces HSC proliferation and downregulation of TNFa and TGFb signaling. a. Expression of *CXCR4* vs. cell cycle gene expression in HSCs before and after IFNa treatment. Pie charts show frequencies of baseline versus treated cells in cell cycle-low, *CXCR4*-high (n = 523 cells) and those in cell cycle-high, *CXCR4*-low (n = 573 cells) populations. **b.** Volcano plot showing DE genes between baseline and IFNa-treated MkPs. DE genes identified using linear mixed modeling with/without treatment status. Genes highlighted in orange are enriched in the TGFb signaling and those in blue enriched in the IFNa/g response. Enrichment based on pre-ranked gene set enrichment analysis (GSEA) using the MSigDB Hallmark collection.



Supplementary Figure 5 (related to Figure 3). IFNa induces proliferation differentially in the mutated versus wildtype cells. a. UMAP of sorted CD34^{+} stem and progenitors at baseline and after IFNa treatment. *CALR* mutation status highlighted. Cells from each sample were down-sampled to the same number for each treatment status (n = 500 cells from each treatment status per sample) **b.** Box plot showing normalized *CXCR4* expression levels in HSCs at baseline (n = 2655 wildtype; n = 2030 *CALR*-mutated) and after IFNa treatment (n = 1169 wildtype; n = 2655 C*ALR*-mutated). P-values from linear mixed model with/without treatment status or genotype. **c.** Frequencies of proliferating myeloid cells before and after IFNa treatment, as assessed by Ki67 expression. P-values from Wilcoxon rank sum test.

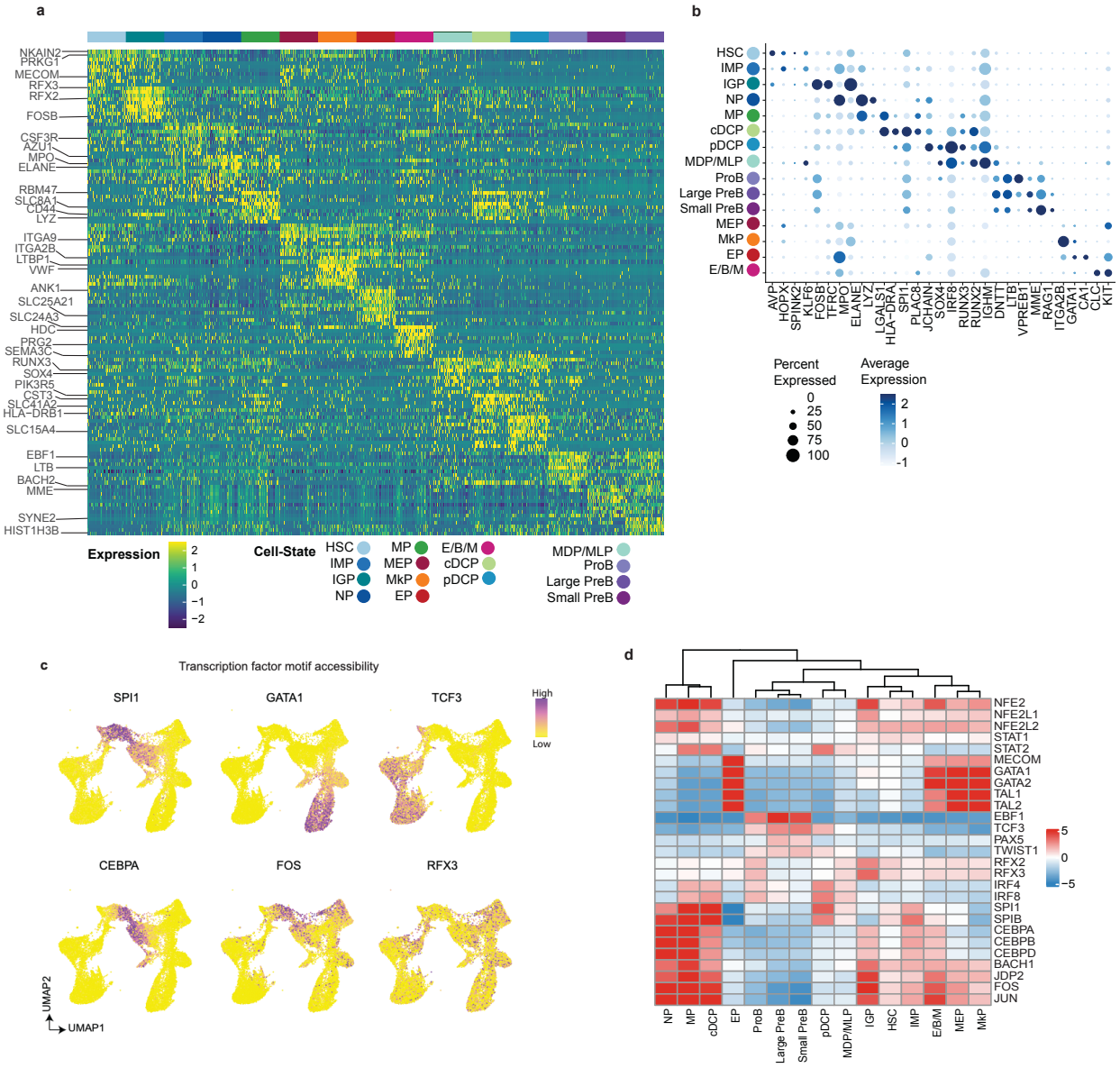


Supplementary Figure 6 (related to Figure 3). Clonal HSCs with high IFNa signaling at baseline expand upon IFNa treatment. a. Left panel: Normalized frequencies of HSCs at each time-point for IFN01 (n = 8989 cells), IFN02 (n = 5939 cells), IFN03 (n = 7329 cells) and IFN04 (n = 7964 cells). P-values from linear mixed model comparing different treatment time-points. Right panel: Normalized frequencies of each progenitor subset among WT, single MUT (Type 1 *CALR*) and double MUT (Type 1 and SNV mutations in *CALR*) cell populations at each timepoint for IFN04. **b.** Relative density of proportion of Type 1-mutant reads versus fraction of SNV-mutant reads in UMIs captured via GoT, showing mutual exclusivity. **c.** UMAP of IFN04 patient based on scRNA-seq data highlighting cell types, treatment time-points and mutation status (from left to right, n = 7964 cells).

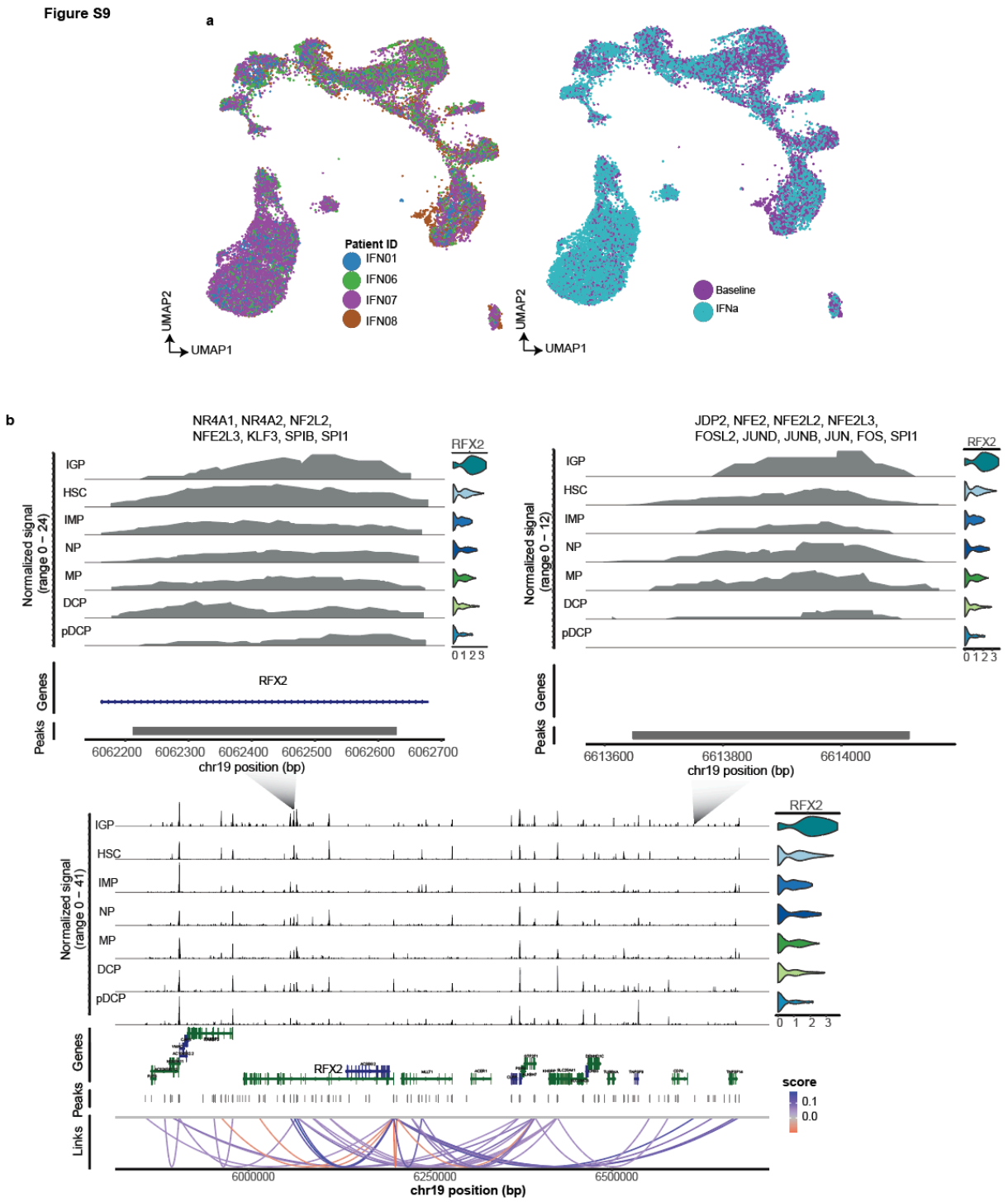


Supplementary Figure 7 (related to Figure 4). GoT-ATAC captures genotyping, snRNA-seq and snATAC-seq data for CD34$^+$ progenitors. a. Box plots showing number of UMIs (left) and genes (right) detected per cell in sorted CD34$^+$ hematopoietic stem and progenitors from each patient after filtering based on Quality Control (QC) metrics. Cells profiled using GoT-ATAC. **b.** Density plot comparing percentage of snATAC fragments within peaks to the total number of fragments detected per sample. **c.** Distribution of mean TSS enrichment score at each position relative to the TSS per sample (n = 4 patients, 7 samples). **d.** Average distribution of fragment length per sample (n = 4 patients, 7 samples). **e.** UMAP of sorted CD34$^+$ stem and progenitors (n = 24857 cells), with patient and treatment status highlighted (left and right panels, respectively).

Figure S8



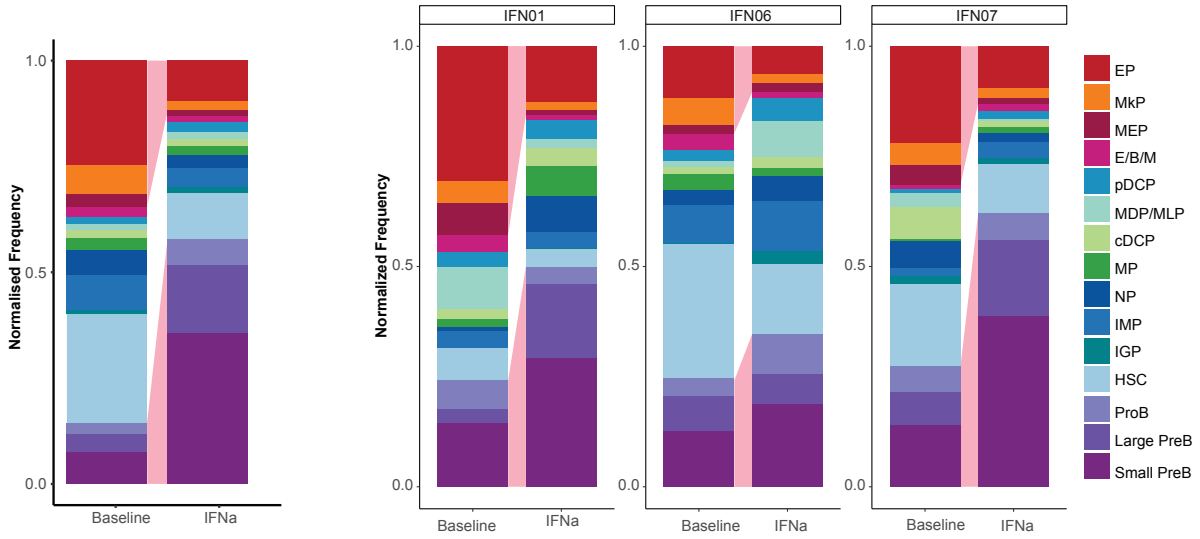
Supplementary Figure 8 (related to Figure 4). GoT-ATAC identifies the same progenitor subsets identified in the GoT-IM data. a. Heatmap of top 10 differentially expressed genes for each progenitor cell type. Cells of each cell type were down-sampled to the same number (n = 100 cells per cluster)**. b.** Dot-plot showing expression levels of cell type-specific gene markers in each progenitor subset. **c.** UMAP highlighting TF motif accessibility (n = 24857 cells). TF accessibility scores added with AddChromatinScore function in Signac. **d.** Heatmap showing cell type specific TF accessibility scores.



Supplementary Figure 9 (related to Figure 4). TF activities in inflammatory granulocytic progenitors. a. UMAP of CD34^{+} cells (n =23137 cells, 7 samples, 4 patients) based on snATAC-seq data overlaid with patient ID (left) and treatment time-point (right). **b.** Chromatin accessibility

tracks of regulatory regions of *RFX2* (bottom), distal region enriched with the two most significant positively regulating loci (top-left, top-right).

Figure S10



Supplementary Figure 10 (related to Figure 4). IFNa induces lymphoid differentiation. Left panel: Normalized cell frequencies of progenitor subsets at baseline and after IFNa treatment. Cells from each treatment status and individual were down-sampled to the same number (n = 100 cells per treatment status per sample). Right panel: Cell frequency distribution as in left panel for patients IFN01, IFN06 and IFN07 (n = 100 cells per treatment status).