Supplementary material

Inferring the dynamics of mutated hematopoietic stem and progenitor cells induced by IFNα in myeloproliferative neoplasms

Matthieu Mosca[1,2,3,4*], Gurvan Hermange[3,5*], Amandine Tisserand[1,2,4,6*], Robert Noble[7,8,9,10*], Christophe Marzac[1,2,3,11], Caroline Marty[1,2,3,4,], Cécile Le Sueur[7], Hugo Campario[12], Gaëlle Vertenoeil[13], Mira El-Khoury[1,2,4], Cyril Catelain[2,14], Philippe Rameau[2,14], Cyril Gella[11], Julien Lenglet[15], Nicole Casadevall[1,16], Rémi Favier[17], Eric Solary[1,2,3,18,19], Bruno Cassinat[20,21], Jean-Jacques Kiladjian[20,22], Stefan N. Constantinescu[13], Florence Pasquier[1,2,18], Michael E. Hochberg[8,23], Hana Raslova[1,2,3], Jean-Luc Villeval[1,2,3], François Girodon[12,24], William Vainchenker [1,2,3,4,25], Paul-Henry Cournède[3,5], Isabelle Plo [1,2,3,4]

Table of contents

Online methods:

Genotyping and fragment analysis

Genomic DNA from granulocytes was isolated and genotyping was performed as previously described, by Taqman allelic discrimination for *MPL*^{W515K} and *JAK2*^{V617F} [1,2] and by fragment sizing for *CALR*^{m} using the BigDye Terminator chemistry (Life Technologies) on a Genetic Analyzer.

Next-generation targeted sequencing

Libraries were obtained from 200 ng of genomic DNA using HaloPlex Target Enrichment System (Agilent technologies) and sequencing was performed using a MiSeq sequencer (Illumina), according to the manufacturer's protocols. Results were analyzed after alignment of the reads using two dedicated pipelines, SOPHiA DDM® (Sophia Genetics) and an in-house software GRIO-Dx®. For all samples, an average depth exceeding 200X for > 90% of the target regions was required. All pathogenic variants were manually checked using Integrative Genomics Viewer software.

A panel of 77 genes designed for the diagnosis of myeloid malignancies was designed as follow *ABL1* (Chr9; NM_007313 ;exons4-9), *ADGRV1*(chr5 ; NM_032119 ; exons1-90), *ANKRD26* (Chr10 ; NM_014915 ; exon1 et 5'UTR), *APC (chr5 ; NM_001127510;exons1-15), ASXL1(chr20;NM_015338;exons11-12), ASXL2(chr2;NM_018263;exons11-12), ATG2B (Chr14 ; 20 polymorphismes), ATM(chr11;NM_000051;exons2-63), ATRX (chrX ; NM_000489 ; exons1-35), BCOR (chrX; NM_001123385;exons3-16), BCORL1 (chrX ; NM_001184772 ; exons1-13), BRAF (Chr7 ; NM_004333 ; exon15), CALR(Chr19;NM_004343;exon9), CBL(chr11;NM_005188;exons4, 8-9,12,16),CEBPA(chr19;NM_ 001287424;exon1 codant), CHEK2 (chr22 ; NM_001005735 ;exons3-16), CREBBP (chr16 ; NM_004380 ; exons1-31), CSF3R(chr1; NM_156039 ; exons14 et 17), CUX1(chr7 ; NM_001202543 ; exons1-24), DDX41(chr5;NM_016222;exons1-17), DDX54 (chr12 ; NM_001111322 ; exons1-20), DHX29 (chr5 ; NM_019030 ; exons1-27), DIS3 (chr13 ; NM_014953 ; exons1-21), DNMT3A(chr2; NM_022552;exons2-23), EED (chr11 ; NM_001308007 ; exons1-13), EP300 (chr22 ; NM_001429 ; exons1-31), EPOR (chr19 ; NM_000121 ; exon 8), ERBB4 (chr2 ; NM_005235 ; exons1-28), ETNK1(chr12 NM_018638; exon3), ETV6 (chr12 ; NM_001987 ; exons1-8), EZH2(chr7;NM_004456;exons2-20), FLT3(chr13;NM_004119;exons8-21), GATA1 (chrX ; NM_002049 ; exon2-4), GATA2(chr3;NM_032638;exons4-6 ; c.1017_462-703), HRAS(chr11; NM_005343;exons2-3 et Codon 117dans exon4), IDH1(chr2; NM_005896, Codons 130-134 dans exon4), IDH2(chr15; NM_002168 ; exon4), JAK2(chr9;NM_004972;exons3-25 ; 46/1 Haplotype ; c.*122G>A), KDM6A (ChrX; NM_001291415 ; exons1-30), KMT2A(chr11;6 polymorphismes), KIT(chr4;NM_000222;exons2, 8-17), KRAS(chr12; NM_033360 ; exons2 et 4 ; codons57-64 dans exon3), MPL(chr1;NM_005373;exons1-12), MYC(chr8; NM_002467 ;*

exon2),NF1(chr17;NM_001042492;exons1-58), NFE2 (chr12; NM_001261461; exons3-4), NPM1 (chr5;NM_002520;exon11), NRAS (chr1; NM_002524; exons2-3), PHF6 (chrX ;NM_001015877 ;exons2-10), PPM1D (chr17; NM_003620; exon6), PRPF40B (chr12; NM_001031698 ; exons1-26), PRPF8 (chr17; NM_006445; exons2-43), PTPN11(chr12;NM_002834 ;exons3-4;8;11-13), RAD21 (chr8; NM_006265; exons2-14), RUNX1 (chr21; NM_001001890; exons3-9), SETBP1 (chr18; NM_015559; exon4), SETD2 (chr3 ; NM_014159; exons5-9), SF1 (chr11 ; NM_004630 ; exons1-13), SF3A1(chr22 ; NM_005877 ; exons1-16), SF3B1(chr2;NM_012433;exons13-16), SH2B3(chr12; NM_005475 ; exons 2-8), SIMC1 (chr5 ;NM_001308195 ; exons2-10), SMC1A(chrX ;NM_6306;exons1-25), SMC3(chr10;NM_5445 ;exons1-29), SRP72 (chr4 ; NM_006947 ; exons1-19), SRSF2(chr17; NM_001195427 ; exon1), STAG2(chrX;NM_001042749;exons3-35), SUZ12 (chr17 ; NM_015355 ; exons1-16), TERC (chr3 ; NR_001566), TERT (chr5 ; NM_198253 ; exons1-16), TET2(chr4;NM_001127208;exons3-11), THPO (chr3 ; NM_001289998 ; 5'UTR),TP53(chr17; NM_001126112 ; exons2-11 ; NM_001126113_Exon 10; NM_001126114_Exon 10), U2AF1(chr21;NM_001025203; codon 34 dans exon2 et codon 157 dans exon 6), U2AF2(chr19 ;NM_007279 ; exons1-12), WT1(chr11;NM_024426;exons1-4 ; 7-9), ZRSR2(chrX;NM_005089;exons1-11).

COMPARTMENTAL MODEL

To understand the effect of treatment at the stem cell level, we constructed a mathematical model. We had no experimental observations on these stem cells except that they contributed a small amount to the compartment of HSC-enriched progenitors, which was, however, insufficient to draw robust conclusions. We therefore chose to infer the latent process of IFNα-targeting mutated stem cell dynamics. The inference method required constructing a mathematical model that could explain the observed experimental dynamics without overparameterization. The scientific literature is rich in hematopoietic models, either deterministic or stochastic, to describe normal or abnormal hematopoiesis or study treatment effects (see for example, [3,4]). Among them, we find compartmental models, as in Marciniak et al. or in Colijn & Mackey [5,6]. These latter models are convenient when dealing with data because experimental observations can directly be associated with compartments, allowing model calibration. Furthermore, they enable comparison between groups of individuals with shared characteristics. Given its many advantages, this family of models is applied to a wide range of systems, from cell populations to human populations, such as in epidemiology.

We chose to design a tailor-made compartmental model describing the effect of IFNα on the mutated hematopoietic cells. Among the existing models, some of them attempt to explain the mechanisms of action of a treatment in the stem cell compartment, but mainly in the case of chronic myeloid leukemia

[7,8]. We therefore built a new model with the objective to leverage prior biological knowledge of the disease and test hypotheses concerning the action of IFNα.

One brick for one cellular type

First, we considered wild-type, heterozygous, and homozygous mutated cells separately, without considering any interaction between their clones. In terms of notation, we add appropriates indices – none, het or hom – for wild-type, heterozygous and homozygous mutated cells, respectively. For a given cell-type, we defined a progenitor compartment and a mature one, as done, for example, by Michor et al. [9]. Mature cells are fully differentiated; they no longer encounter cell divisions and die at a rate δ_m. This mature compartment matches the experimental measures of the Variant Allele Frequency (VAF) among granulocytes (after normalisation). In our model, compartments did not correspond to homogeneous sets of cells, as we know that hematopoiesis is a complex process where a large number of heterogeneous cells proliferate and differentiate. Thus, the progenitor compartment in our model gathers several distinct progenitors that experimentally correspond to the cells sorted according to the surface markers: $CD90^-CD34^+CD38^+$, $CD90^-CD34^+CD38^-$ and $CD90^+CD34^+CD38^-$. We modeled them as cells that originate from a HSC, encountered several divisions (modeled by the parameter κ_i) and that will finally exit their compartment at the differentiation rate δ_i. To account for cells undergoing several rounds of proliferation and differentiation before becoming fully differentiated, we introduced the parameter κ_m. Both compartments directly correspond to data.

We also introduced two stem cell compartments. HSC are heterogeneous cells for which we had little knowledge; we had to model their dynamics based on some assumptions. First, we divided them into two functional compartments: either the HSC is quiescent (or inactive) or active and thus can directly contribute to hematopoiesis. A stem cell can go from quiescent to active and vice versa. Parameters γ and β model these exchanges between compartments. If active, we assume that the HSC might be recruited at a rate α to contribute to hematopoiesis and divide. We assume that three division types occur at different rates (or probabilities). First, the HSC may divide asymmetrically (with probability p_1) and give rise to one HSC and one differentiated cell that will further proliferate and become a progenitor. Second, the HSC may divide symmetrically (with probability p_2) and give rise to two HSC. Finally, the HSC may undergo a differentiating division and give rise to two differentiated cells (with probability $p_0=1-p_2-p_1$). At the population level, the parameter $\Delta = p_2 - p_0$ models the balance between the divisions of different types and characterizes a mutated clone's behavior at the stem cell level. If $\Delta=0$, we have homeostatic conditions; if strictly positive, there is an invasion; and exhaustion if negative. We considered nearly homeostatic conditions for $JAK2^{V617F}$ and MPL^m patients before the treatment starts. Following the idea of Michor et al. [9], we considered that IFNα modifies some model parameters, resulting in new equilibrium conditions and consequently, the start of a long-term

dynamic that will reach this new state. In terms of notation, we will add a star symbol (*) for parameters impacted by IFNα.

For *CALR*^{m} patients, because mutated clones expand much more rapidly than in *JAK2*^{V617F} and *MPL*^{m} patients, we further assumed that mutated HSC might have a substantial advantage at the stem cell level and so the homeostatic approximation before treatment might be inaccurate. We additionally simplified the model by assuming that *CALR*^{m} mutated HSC stayed active [10,11].

The model for a given type (*wt*, *het*, or *hom* for the considered mutation) is schematized in figure 3A. For *JAK2*^{V617F} and *MPL*^{m} patients, the system of equations to describe the dynamics of a given hematopoietic cell population (*wt*, *het* or *hom*) is as follows:

$$\begin{cases} \frac{dN_1(t)}{dt} & = -\gamma N_1(t) + \beta N_2(t) \\ \frac{dN_2(t)}{dt} & = \gamma N_1(t) + (\alpha\Delta - \beta) N_2(t) \\ \frac{dN_i(t)}{dt} & = \alpha(1-\Delta)\kappa_i N_2(t) - \delta_i N_i(t) \\ \frac{dN_m(t)}{dt} & = \delta_i \kappa_m N_i(t) - \delta_m N_m(t) \end{cases} \quad (1)$$

With $N_1(t)$, $N_2(t)$, $N_i(t)$ and $N_m(t)$ the numbers of cells, of a given type, in compartments 1, 2, progenitors and mature respectively at time *t*. Appropriate indices will be used whether we refer to *wt*, *het* or *hom* cells. The proposed ODE (Ordinary Differential Equations) system was linear, and we could derive an analytical solution.

Considering all cell types

In our data, we did not have the numbers of *wt*, *het* or *hom* cells in each compartment, but only their relative proportions.

To take this into account, we considered that each population of a given type follows an ODE system as described in eq. (1). Then we considered as outputs of our model no longer the numbers of cells but the proportions of immature heterozygous and homozygous cells respectively $z_{het}(t) = \frac{N_{het,i}(t)}{N_i(t)+N_{het,i}(t)+N_{hom,i}(t)}$ and $z_{hom}(t) = \frac{N_{hom,i}(t)}{N_i(t)+N_{het,i}(t)+N_{hom,i}(t)}$, as well as the mature cell VAF $y(t) = \frac{0.5 \cdot N_{het,m}(t)+N_{hom,m}(t)}{N_m(t)+N_{het,m}(t)+N_{hom,m}(t)}$ where every contribution of heterozygous cells is only counted for one half.

For convenience, we defined $k_{i,het}$ such that $\kappa_{i,het} = k_{i,het}\kappa_i$ and likewise $k_{i,hom}$, $k_{m,het}$ and $k_{m,hom}$.

Effect of IFNα and initial conditions

Following the idea of Michor et al.[9], we considered that IFNα acts by modifying the values of some parameters in the model.

We define time *t=0* at the start of treatment. Before that time, Equation (1) remained valid, but we considered for *JAK2*^{V617F} and *MPL*^{m} patients that the homeostatic conditions were satisfied, i.e. the

system was in a quasi-stationary state. It is of course verified for *wt* cells as soon as $\Delta=0$, which gave us the following initial conditions: $N_1(0) = \frac{\beta}{\beta+\gamma} N_{HSC}$, $N_2(0) = \frac{\gamma}{\beta+\gamma} N_{HSC}$, $N_i(0) = \frac{\kappa_i \alpha}{\delta_i} N_2(0)$ and $N_m(0) = \frac{\kappa_m \delta_i}{\delta_m} N_i(0)$ with N_{HSC} the total wild-type HSC number considered constant at all times. Mutated cells disrupt homeostasis by invading the stem cell compartment. But since the invasion happens over a long time (30 years or more according to Hirsch et al. [12]), we assumed for *JAK2*^{V617F} and *MPL*^{m} patients that $\Delta_{het} \approx \Delta_{hom} \approx 0^+$. We introduced $\eta_{het} = \frac{N_{1,het}(0)+N_{2,het}(0)}{N_{HSC}}$ and $\chi_{het} = \frac{N_{2,het}(0)}{N_{1,het}(0)+N_{2,het}(0)}$. We could then express the initial conditions for *het* cells: $N_{1,het}(0) = \eta_{het}(1-\chi_{het})N_{HSC}$, $N_{2,het}(0) = \chi_{het}\eta_{het}N_{HSC}$, $N_{i,het}(0) = \frac{\kappa_{i,het}\alpha_{het}}{\delta_{i,het}} N_{2,het}(0)$ and $N_{m,het}(0) = \frac{\kappa_{m,het}\delta_{i,het}}{\delta_{m,het}} N_{i,het}(0)$. The same goes for *hom* cells.

In *CALR*^{m} patients, as the *CALR* mutation is associated with a selective advantage at the progenitor level, faster clonal expansion than *JAK2*^{V617F}, and higher initial clonal fractions among immature cells, the homeostatic approximation is less accurate [10,11,13]. As already discussed, for these patients, the model was simplified not to take into account inactive mutated HSC (so that $\chi_{het} = \chi_{hom} = 0$). We also made an additional assumption on the initial conditions, by considering that $\Delta_{het} = \Delta_{hom} > 0$ and that the parameter value would be the same for all *CALR*^{m} patients. This parameter was estimated through the hierarchical Bayesian framework made for *CALR*^{m} patients. With this assumption, the initial condition for $N_{i,het}(0)$ and $N_{i,hom}(0)$ was slightly modified accordingly.

From $t=0$, the patient is under treatment. IFNα will then modify the values of the parameters, potentially in different ways depending on the cell type. In terms of notation, we add the superscript * to the parameters impacted by the drug. We define k_i^* such that $\kappa_i^* = k_i^* \kappa_i$ and similarly for $k^*_{i,het}$, $k^*_{i,hom}$, k^*_m , $k^*_{m,het}$ and $k^*_{m,hom}$.

From $t>=0$, eq. (1) remains valid with new parameters, and there is an equilibrium shift. This induces a new dynamic. From our data, we then inferred the unknown model parameters.

DATA AND OBSERVATION MODEL

To rigorously account for measurement error and other sources of uncertainty in the data, we defined statistical models for Bayesian inference.

Notations

We have a set of patients that we denote: $\mathcal{P} = \{1, \cdots, N\}$. For each patient *j*, we have observations (both progenitors and mature cells) measured at several times from the start of the treatment. Let $\mathcal{I}_j$ be the set of observation times for a given patient *j*. The whole data set is given by:

$$\mathcal{D} = \{\mathcal{D}_j\}_{j\in\mathcal{P}} = \left\{ \left(t_{i,j}, \hat{n}^{het}_{i,j}, \hat{n}^{hom}_{i,j}, \hat{n}_{i,j}, \hat{y}_{i,j}\right)_{i\in\mathcal{I}_j} \right\}_{j\in\mathcal{P}}$$

Where, for a given patient *j*: $t_{i,j}$ is its i^{th} observation time; $\hat{y}_{i,j}$ is the measured VAF among mature cells (granulocytes) at this time; $\hat{n}_{i,j}, \hat{n}^{het}_{i,j}, \hat{n}^{hom}_{i,j}$ represent respectively the *wt*, *het* and *hom* numbers of progenitors that have been sampled and genotyped from progenitor-derived colonies at time t_i.
In terms of notation, a hat above a quantity indicates that this is an experimental observation.
When referring to progenitor cells, we considered the cells characterized by the following surface markers pooled together: $CD90^-CD34^+CD38^+$, $CD90^-CD34^+CD38^-$ and $CD90^+CD34^+CD38^-$.

Mature cells

At time t_i (in this paragraph, for the sake of clarity, we omit the patient index *j*), the measured VAF $\hat{y}_i$ among mature cells corresponds to an unknown real fraction y_i. We have $\hat{y}_i, y_i \in [0,1]$.
Conventional uncertainty models are additive or multiplicative Gaussian noises that are not entirely appropriate in our case. Indeed, in the case where $y_i=1$ for example, both noise models would authorize $\hat{y}_i \neq 1$. But if the true VAF among mature cells is equal to 1, experimentally we would not expect to detect non-mutated alleles in the blood sample. Thus, we choose to generalize them as follows:

$$\hat{y}_i | y_i \sim \mathcal{N}(y_i, \sigma^2(y_i)) \quad (2)$$

with $\sigma^2 : [0,1] \longrightarrow \mathbb{R}$. By choosing for the function σ^2 a constant, we find ourselves with an additive noise and by choosing $\sigma^2(y_i) = \sigma_m^2 y_i^2$ we find ourselves with a multiplicative noise.
In order to have a noise symmetrical around $y_i = 1/2$ and that vanishes for $y_i = 0 = 1$, we choose:

$$\sigma^2(y_i) = y_i(1-y_i)\sigma_m^2 \quad (3)$$

where σ_m^2 is a parameter to be estimated.

Immature cells

To model the sampling noise related to the mutated clonal fractions (CF) of progenitors, we assume that we have randomly drawn with replacement immature cells from the patient's body.
This approach is used, for example, by Catlin et al. [14] and allows us to model the uncertainty by a multinomial distribution. For a large number of immature cells in the body, the approach is almost identical to that of Xu et al.[3] who consider a multivariate hypergeometrical law used to model sampling without replacement.
Let's consider that at date t_i (for a given patient *j*, again, we drop the *j* index for clarity), the real proportions of heterozygous and homozygous immature cells are respectively $z_{i,het}$ and $z_{i,hom}$ (and for the wild type, $z_i = 1 - z_{i,het} - z_{i,hom}$).

From the set of immature cells, of unknown but very large number, we draw a number $\hat{N}_i := \hat{n}_{i,het} + \hat{n}_{i,hom} + \hat{n}_i$ of cells. Among these cells, we have exactly $\hat{n}_{i,het}$ heterozygous and $\hat{n}_{i,hom}$ homozygous mutated cells. Since they are random variables, they follow a multinomial law:

$$\mathbb{P}\left[\hat{n}_i = n_1, \hat{n}_{i,het} = n_2, \hat{n}_{i,hom} = n_3 | z_{i,het}, z_{i,hom}\right] = \frac{(n_1+n_2+n_3)!}{n_1!n_2!n_3!} z_i^{n_1} z_{i,het}^{\ n_2} z_{i,hom}^{\ n_3} \quad (4)$$

Excluded and included data & patients

A whole cohort of 48 patients has been followed. Several inclusion criteria were used for the parameter estimation and then the statistical analyses as presented in Figure S1. To calibrate the model, we included almost all patients (40 patients for which we have at least 3 data timepoints) so that even patients with few data points bring information through our population-based framework. Then, for further statistical analyses, we compared only patients for which we had sufficient data to obtain precise inferences and avoid overfitting risks. At this stage, additional excluded *JAK2*^{V617F} patients are those with no more than five progenitor measures. With this criterium, we mainly excluded intolerant patients for whom we had only three or four data points, as well as the only patient in our cohort who had five data points. This last patient was patient #21, who was followed during the course of treatment but for which we lacked information at the start of the therapy.

All *CALR*^{m} patients used for the model calibration are also used for our statistical analyses. Indeed, for those patients, because we use a simplified model with a fewer number of parameters, the risk of overfitting is lower than in the case of *JAK2*^{V617F} patients.

Finally, after the use of these inclusion criteria, still almost 90% of all our data are used for the model-based statistical analysis.

Besides, patients might have homozygous and heterozygous mutated HSC, but not in the same proportions. When we analyzed not only the HSC VAF but also the zygosity, for example when comparing the effect of IFNα on homozygous *versus* heterozygous cells, it was necessary to exclude patients whose subclones exhibited too low of CF overtime. Most patients with *CALR*^{m} MPN (2/12 cases harbor homozygous subclones) in our cohort had only heterozygous mutated cells. For J*AK2*^{V617F} patients, we defined two subgroups (that are overlapping), one made of patients with a sufficient proportion of homozygous subclones, the second made of patients with a sufficient proportion of heterozygous subclones. The intersection between both subgroups being not empty. A *JAK2*^{V617F} patient was considered as carrying homozygous (respectively heterozygous) subclones when >7% of homozygous (respectively heterozygous) progenitors were identified from at least one of the collected samples. Using this criterium, 10 out of 20 *JAK2*^{V617F} patients carry homozygous subclones and 17 out of 20 carry heterozygous subclones. To be more precise, 8 out of 20 patients belong to the *JAK2*^{V617F} subgroup of patients carrying both homozygous and heterozygous subclones,

9 out of 20 to the group of patients carrying (only) heterozygous subclones and 2 out of 20 to the group of patients carrying (only) homozygous subclones. One patient belongs to none of these two groups since his progenitor CF is lower than 7% both for homozygous and heterozygous cells. But this patient was still considered in previous analyses relying on the progenitor VAF.

The data before the start of the treatment, when available, were considered as initial data, due to our assumption of quasi-stationary state before treatment. To note that patient #1 was followed before *t=0*. In his case, the initial time in our model corresponds to the first date for which we had an information about the posology the patient received.

STATISTICAL METHOD

Hierarchical Bayesian framework

We considered a population $\mathcal{P} = \{1, \cdots, N\}$ of patients (*JAK2*^{V617F}, *CALR*^{m} or *MPL*^{m}).

We denote by $\theta = \left\{\theta^{(j)}\right\}_{j\in\mathcal{P}}$ the set of all model parameters with:

$$\theta^{(1)} = \left(\theta_1^{(1)}, \cdots, \theta_P^{(1)}\right)$$

$$\vdots$$

$$\theta^{(N)} = \left(\theta_1^{(N)}, \cdots, \theta_P^{(N)}\right)$$

where *P* is the number of parameters to estimate for each patient and *N* is the number of patients in the considered population.

In a non-hierarchical Bayesian framework, each parameter vector $\theta^{(j)}$ is estimated individually, considering only the data of patient *j*. This approach can result in overfitting. To improve robustness, we used a hierarchical framework, which tends to reduce variance between patients and to produce more precise inferences [15,16]. Before having any information about different patients, we assume that their parameter values fit a certain type of distribution. The hierarchical framework thus accounts for inter-individual variability (with the probability distribution) and the similarity in the population (the fact that the distribution is common to all individuals in the population). The distribution parameters are called hyperparameters.

The effect of this hierarchical component is shown in Figure S4. In this figure, we show at the top the posterior distribution of one of the model parameters, estimated through a non-hierarchical Bayesian approach for different patients. We can see that, for some patients, there is high variance in the parameter posterior distribution, because the patient has few data points and/or the dynamics are uninformative. Below, we show results from the corresponding hierarchical inference method. The posterior distributions of the different patients are closer to the population average, this latter having been learned via the hierarchical framework. Furthermore, the variability and uncertainty in the results is reduced.

Technically, this population approach was implemented by considering a hierarchical model with one level and by introducing some hyperparameters $\tau = (\tau_1, \cdots, \tau_P)$ and $\sigma^2 = (\sigma_1^2, \cdots, \sigma_P^2)$ so that:

$$\forall j \in \mathcal{P}, \forall i \in \{1, \cdots, P\}, \theta_i^{(j)} | \tau_i, \sigma_i^2 \sim \mathcal{N}_{c,i}(\tau_i, \sigma_i^2) \quad (5)$$

where $\mathcal{N}_{c,i}$ represents a truncated (over a range that depends on the considered parameter *i*) Gaussian distribution.

We could then express the posterior distribution of the parameters and hyperparameters:

$$\begin{aligned} \left[\theta^{(1)}, \cdots, \theta^{(N)}, \tau, \sigma^2 | \mathcal{D}\right] &\propto \left[\mathcal{D} | \theta^{(1)}, \cdots, \theta^{(N)}, \tau, \sigma^2\right] \left[\theta^{(1)}, \cdots, \theta^{(N)}, \tau, \sigma^2\right] \\ &\propto \prod_{j \in \mathcal{P}} \left(\left[\mathcal{D}_j | \theta^{(j)}\right] \left[\theta^{(j)} | \tau, \sigma^2\right] \right) [\tau] \left[\sigma^2\right] \end{aligned} \quad (6)$$

In eq. 6, the expression of the priors simplifies by assuming independence of the components for the parameters and hyperparameters vectors. For a given patient $j \in \mathcal{P}$, we have:

$$\begin{aligned} \left[\theta^{(j)} | \tau, \sigma^2\right] &= \prod_{i \in \{1, \cdots, P\}} \left[\theta_i^{(j)} | \tau_i, \sigma_i^2\right] \\ [\tau] &= \prod_{i \in \{1, \cdots, P\}} [\tau_i] \\ \left[\sigma^2\right] &= \prod_{i \in \{1, \cdots, P\}} \left[\sigma_i^2\right] \end{aligned}$$

Likelihood

In eq. 6, For patient *j*, the likelihood is given by: $\left[\mathcal{D}_j | \theta^{(j)}\right]$. In the following, we omit for clarity the patient's index *j*. The likelihood, i.e. the probability of the data given the parameters, can be expressed by:

$$[\mathcal{D}|\theta] \overset{def}{=} \mathbb{P}[\mathcal{D}|\theta] \overset{(1)}{=} \prod_{i \in \mathrm{I}} \mathbb{P}[\left(\hat{n}_i^{het}, \hat{n}_i^{hom}, \hat{n}_i, \hat{y}_i\right) | \theta] \overset{(2)}{=} \prod_{i \in \mathrm{I}} \mathbb{P}[\left(\hat{n}_i^{het}, \hat{n}_i^{hom}, \hat{n}_i\right) | \theta] \; \mathbb{P}[\hat{y}_i | \theta] \quad (7)$$

with (1) because conditionally on θ, the measures are independent, and (2) because we considered that the measures for immature and mature cells were obtained through two independent measurements. The two last terms were computed using the noise model of the immature (eq. 4) and mature cells (eq. 2) respectively, as presented previously.

Metropolis-Hasting within Gibbs

Since we did not have an analytical expression for the posterior distribution in eq. 6, we approximated the distribution by sampling from it with a Markov Chain Monte Carlo (MCMC) algorithm, more specifically the Metropolis-Hastings (MH) within Gibbs. Conditionally on the hyperparameters, the parameters of the patients were sampled independently from the other patients and based on the MH

algorithm. This method relies on a proposal distribution, for which we chose a multivariate normal distribution with zero mean and covariance matrix Σ.

When the parameter space dimension is large, the Metropolis-Hastings algorithm often proves inefficient, as it becomes complicated to define a covariance matrix that allows a good convergence of the algorithm. Adaptive algorithms have been proposed to circumvent this problem [17]. Here we propose an alternative and simpler method which proved very efficient in practice. For each patient, we started by learning the covariance matrix of the proposal distribution and by choosing a starting point of the Markov chain located at the maximum a posteriori parameter values using the CMA-ES method [18] as detailed below.

We implemented the methods using the programming language Julia. We ran the computations over 25 million iterations, a sufficient number to achieve convergence of the algorithm, and considered a burn-in length of 15 million iterations. For each estimation (one for each population), we ran the algorithm twice with different random seeds to check that we got consistent results. The computations were monitored by looking at the acceptance rates and the values taken by the MCMC chains.

Setting the MH algorithms using CMAES

The CMA-ES (Covariance Matrix Adaptation - Evolution Strategy) algorithm is a stochastic method for optimization that gives good results in a wide range of problems, including problems that are non-linear, non-separable and in high dimension [18].

This algorithm searches for the maximum of a function over generations.

At each generation, a sample of λ points (i.e. parameter vectors) is generated, according to a multidimensional normal distribution whose mean and covariance matrix are computed from the selected points of the previous generation. Among these λ offspring, we select μ (those which give the highest values for the posterior). These are the ones we use for the next generation.

This continues until we reach the maximum number of generations n_g or if the estimated value for the maximum no longer changes sufficiently over the generations. The basic equation for sampling the individuals, for generation number g=0,1,2,…,n_g reads [19]:

$$\theta_k^{(g+1)} \sim m^{(g)} + \sigma^{(g)} \mathcal{N}\left(\mathbf{0}, C^{(g)}\right) \quad \text{for } k = 1, \ldots, \lambda$$

with $m^{(g)}$ the mean value of the search distribution, $\sigma^{(g)}$ the step-size and $C^{(g)}$ the covariance matrix at generation g.

We used this algorithm to find the maximum a posteriori (MAP) of the distribution defined in eq.7, for each patient independently. We used it as starting point for our MCMC algorithm.

Moreover, learning the covariance matrix in the CMA-ES is analogous to learning the inverse Hessian matrix in a quasi-Newton method [19].

Thus, we chose for the covariance matrix of the Metropolis proposal the one learned by the CMA-ES method and set $\Sigma = C^{(ng)}$.

Uncertainty propagation and comparison of Quantities of Interest (QoI)

With the MCMC algorithm, we could estimate the posterior distributions of the parameters for the 3 subpopulations ($JAK2^{V617F}$, $CALR^{m}$ or MPL^{m}) and for each patient. From the latter, we could propagate the uncertainties on the model outputs, using a Monte-Carlo method (by sampling 1 million parameter vectors in the MCMC chain). That is to say that we computed the dynamics of the heterozygous and homozygous clonal fractions among HSC, progenitors and mature cells along with their uncertainties. We can see the resulting dynamics, along with a 95% credibility interval, in figures S6-8. For most patients, there is a good fit between the inferred dynamics and the data for progenitors and mature cells. As expected when dealing with large, heterogeneous cohorts, some of the patient dynamics are not adequately described by the model. This is the case for patient 31 (homozygous immature clonal fraction dynamics), patient 23 and patient 15. Yet, overall, the model was able to capture the main dynamics of the clonal fractions of mutated cells under treatment, which gave us confidence in the insights we draw from it.

Interestingly, there were some patients (for example 2 or 6) that seemed to present some oscillations in their experimental clonal fractions over time. This pattern cannot be captured by the model as it is. We suspect that these oscillations resulted from dosage variation. We aim to investigate this hypothesis in a further work by adding some regulation mechanism or taking into account the pharmacodynamics. At this stage, however, the current model was rich enough to describe the most pertinent features of the observed dynamics. A more complex model would not be appropriate for our research questions since it would result in more parameters to estimate and increase the risk of overfitting.

From the inferred dynamics of mutated HSC during IFNα treatment, we can compute some quantities of interest (QoI). The first QoI is the response factor, *R*, defined as the ratio of the inferred HSC clonal fraction (for homozygous or heterozygous cells, or VAF) at a given time (t=3,000 days) compared to its initial value.

We calculated heterozygous, homozygous and global HSC molecular response factors from the model with uncertainty propagation. We thus obtained the distribution of *R*, from which we could compute quantiles.

$R>1$ indicates that there is no response or an adverse response, as the VAF among HSC is higher at t=3,000 days than initially. If $R<1$, there is a response to the therapy. The lower this value, the better the response. If *R* reaches approximately zero, it means that there is a complete remission (CMR).

This response factor can be computed not only at time t=3,000 days, but also as a function of time. It will start at 1 when the therapy begins, and evolve over time. Its evolution can be plotted. If decreasing below 0.5, we call it partial molecular remission (PMR). The time when this value is reached is called PMR time. Of course, for longer times, a complete remission might be reached, but comparing the PMR times between responders, and sub-populations of responders, gives us insights into how quickly patients respond to treatment.

Both these QoI can be used to make comparisons between patients and, more interestingly, between groups of patients. Thus, we can compare groups according to their driver mutations, or the average IFNα dosage they received in the first 450 days of therapy, or even compare the heterozygous response to the homozygous one. We use the Mann-Whitney U test to examine whether subgroups differ in terms of the QoI. When we want to separate a population according to a continuous quantity, notably to test the dose effect, we divide it according to the median value of this quantity, so that we get two groups of the same size, which can be termed as low-dose and high-dose patients. However, since these categories are determined by an empirical criteria and not a clinical one, we also test the nullity of the coefficient in the linear regression of the log of the response against the dose (Student Test), a test which does not rely on the categorization of patients.

MODEL PARAMETERS

Several assumptions

The model presented above has potentially many patient-specific parameters. An excessive number of parameters risks overfitting and non-identifiability. To address this issue, we made additional assumptions concerning the parameters and their values.

Since it is particularly challenging to infer the parameter values of HSC when no observations of these cells are available, we chose to assume that the parameter α was independent of cell type (*wt*, *het* or *hom*) and unaffected by IFNα. We took $\alpha=1/30$, $\gamma=1/300$ (that is, one order of magnitude lower than α), and $\chi=0.1$ (meaning that the wild-type active HSC represent 10% of the pool of HSC). To our knowledge, for the *in vivo* human hematopoiesis, there are no exact values for these parameters that can be found in the biological literature. These 3 parameters values, describing the wild-type HSC dynamics, were fixed according to their likely orders of magnitude. We conducted a sensitivity analysis and verified that these choices did not strongly influence the model output nor the quality of the fit. Note that the parameters we infer are relative to these fixed parameter values and do not correspond to true absolute values.

We assumed that wild-type HSC are not much influenced by IFNα, that is $\gamma^*=\gamma$ and $\beta^*=\beta$. We considered that, before the start of the treatment, the transitions between quiescent and active

compartments were similar for mutated heterozygous and homozygous cells, that is $\gamma_{het}=\gamma_{hom}=\gamma$ and $\beta_{het}=\beta_{hom}=\beta$. With our model, we wanted to study the effect of IFNα on the quiescent exit of heterozygous mutated cells. This could be done by considering that the treatment affects γ_{het}, β_{het}, or both parameters. This last choice would result in a non-identifiability problem. We chose to consider that the treatment affects γ_{het}, therefore we estimated γ^*_{het} and set $\beta^*_{het}=\beta$. The same goes for homozygous cells.

For $CALR^m$ patients, the previous hypotheses concerning γ, β and χ were not required, since we did not consider inactive $CALR^m$ cells.

For hematopoietic progenitors and mature cells, we had more biological knowledge that could help us set some values. First, we considered that δ_i and δ_m did not depend on the zygosity or on IFNα. We set $\delta_m=1$ per day (knowing that wild-type granulocytes have a lifespan of the order of magnitude of the day) and $\delta_i=1/6$ per day (assuming that the order of magnitude of the lifespan of progenitors should be about some days).

We assumed that progenitor cells proliferate in the same way no matter the cell type, that is $\kappa^*_{i,het}=\kappa^*_{i,hom}=\kappa_i^*$ and $\kappa_{i,hom}=\kappa_{i,het}=\kappa_i$. This assumption allowed inferring the initial conditions η_{het} and η_{hom} from data of progenitor cells. Finally, we assumed that mutated cells may proliferate faster than wild-type cells at the latest stages of hematopoiesis, but considered no difference between heterozygous or homozygous cells (since the data we had for granulocytes did not make the distinction between the zygosity): thus, we take $k_{m,het}=k_{m,hom}$. Finally, we considered that the therapy affects equally all mature cells: $k^*_{m,hom}=k^*_m=k^*_{m,het}$.

For each patient, we ended up with five parameters (3 for $CALR^m$ patients) and two initial conditions at the stem cell level (both for homozygous and heterozygous clones) that we needed to infer from data. These parameters include Δ^*_{het} and Δ^*_{hom} related to the differentiation of the mutated heterozygous and homozygous HSC into progenitors under IFNα, and γ^*_{het} and γ^*_{hom} to model the effect of the treatment on the quiescent exit of the cells. These two last parameters did not appear in the simplified model for $CALR^m$. We also estimated $k_{m,het}$ which models the proliferation advantage of mutated cells at the latest stages of hematopoiesis. Finally, we had to estimate the unknown initial conditions η_{het} and η_{hom} that correspond to the proportions of mutated heterozygous and homozygous HSC (over the whole wild-type HSC pool).

Table 2 summarizes the parameters of the model and our assumptions.

Population approach and prior distributions

Several sets of parameter values were inferred: One for $JAK2^{V617F}$ patients, one for $CALR^m$ patients, and one (thus non-hierarchical) for the single MPL^m patient. These are the three populations we

considered, based on the driver mutation. As previously explained, using a population approach, we considered that patients belonging to the same population must have somewhat similar parameter values. Technically, we took the population's effect into account through hyperparameters, as previously explained.

For $JAK2^{V617F}$ patients, we introduced hyperparameters for all parameters except the initial conditions η_{het} and η_{hom}. For both of them, we considered uniform prior distributions over the range [0,3]. For the other parameters, conditionally on the hyperparameters, the prior distributions were given in eq. 5. They were distributed according to truncated Gaussian distributions, with mean τ_i and variance σ_i^2 for the i^{th} parameter. For the variance of the i^{th} parameter, we chose an improper prior distribution: an inverse-gamma (0,0) law. The hyperparameters corresponding to the means of the Gaussian distributions followed *a priori* a uniform distribution, over the same ranges used to truncate the Gaussian distribution. The ranges were [1,20], [1/300, 0.1] and [-1,1] for $k_{m,het}$, y^*_{het} (and y^*_{hom}) and Δ^*_{het} (and Δ^*_{hom}) respectively.

For $CALR^m$ patients, the same goes for $k_{m,het}$, η_{het} and η_{hom}. But for the parameters modeling the balance of divisions a mutated HSC can undergo under treatment, we adapted the hierarchical framework as follows. We considered that, for a given patient *j*, $\Delta^*_{het,j}|\Delta_{het} \sim \mathcal{N}_{[-1,1]}(\Delta_{het}, 0.1^2)$ and $\Delta^*_{hom,j}|\Delta_{het} \sim \mathcal{N}_{[-1,1]}(\Delta_{het}, 0.1^2)$, with $\Delta_{het}=\Delta_{hom}$ a parameter shared by the $CALR^m$ population. This hierarchical model means that, *a priori*, the treatment had neither a positive effect nor an adverse effect, and that the mutated clonal fractions continued their expansion according to the value of Δ_{het} before treatment (uniformly distributed over [0,0.3]).

VALIDATION

Validation of the Statistical estimation method

Models with too many parameters are ill-posed when different parameters values conduct to the same model output. Our model simplifications and the resulting reduction of the parameter space dimension circumvent this problem. To demonstrate it and test the ability of our statistical method to infer parameter values from noisy data, we simulated data from 30 virtual $JAK2^{V617F}$ patients. We sampled their parameters from arbitrary population distributions, computed the corresponding dynamics using our model, and generated noisy data for progenitors and mature cells. Not all virtual patients had the same number of data to reproduce the heterogeneity of the original cohort.

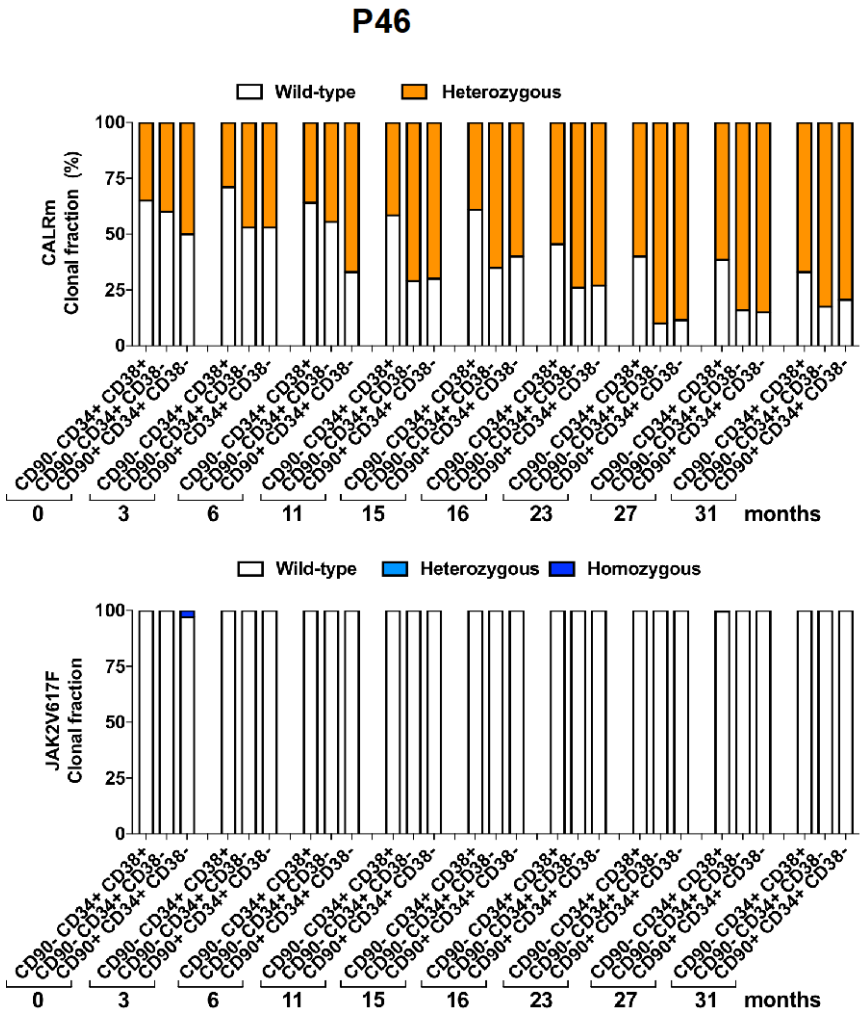
Then, we used our hierarchical Bayesian estimation method to retrieve the true parameter values from our noisy virtual data. We found (results not shown) that we could estimate with accuracy the dynamics of the mutated HSC clonal fractions and the long-term response factor (R-factor) even for

virtual patients with few data points. This result justifies using the R factor as a key indicator for comparing molecular responses between patients and studying how HSC are targeted differently according to zygosity and IFNα dosages. We also got good results for the parameter estimation, at the individual level for most of our virtual patients, but especially at the population level. This latter result gave us confidence in the conclusions we draw about the mechanism of action of IFNα on heterozygous *vs* homozygous HSC.

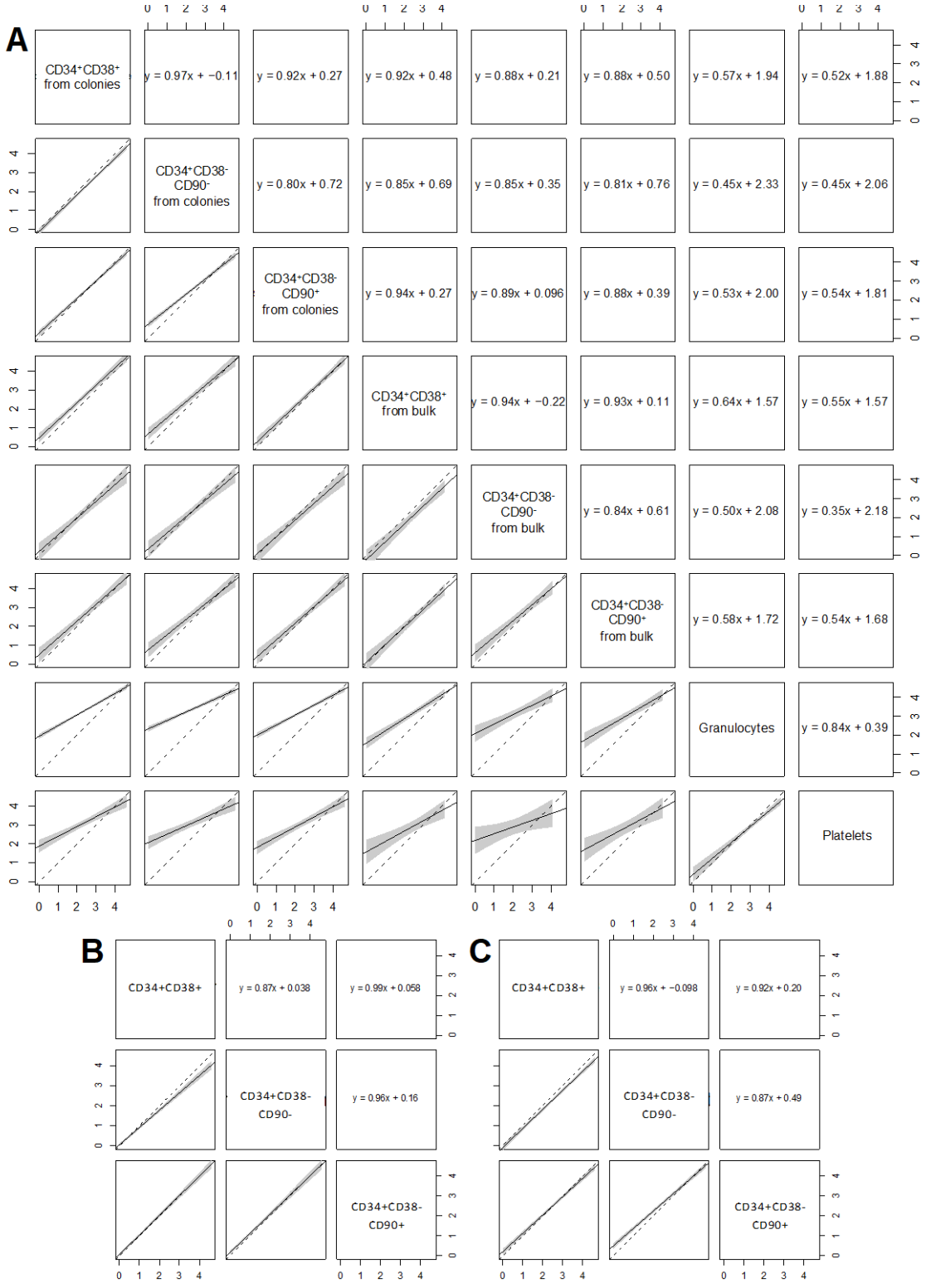
Validation of the predictive capacity of the model

With our mathematical model, we aim to infer the long-term response of patients to IFNα. To validate the model's predictive capacity, we need to test it with patients and data that have not been used for the calibration. For that purpose, we temporarily excluded two $JAK2^{V617F}$ patients from our cohort and calibrated the model on the remaining patients. Excluded patients are patients #12 and #22. We chose them as test patients because we have many observations for them within the first 300 days of therapy. Furthermore, they both exhibit an initial increase of their VAF in mature cells followed by a decrease, making it more challenging to obtain accurate predictions. When calibrating the model on the cohort without these two test patients, we learn a population effect through our hyperparameters. Then, we make the parameter estimation for our two test patients. The previously learned population effect is now used as prior knowledge. For inferring the long-term dynamics of the test patients, we also have to use some early observations. We progressively consider more and more data points and quantitatively evaluate the quality of the predictions by computing the Mean Absolute Error (MAE) between our predictions and our observations. The results are shown in Supplemental Figure S10. For both patients, already with two observations, we have rather good long-term predictions (median value) but with large 95% credibility intervals. By adding progressively more observations for the parameter estimation procedure, the predictive error decreases, and the credibility interval becomes smaller. These results show that we can accurately predict the long-term response of IFNα for $JAK2^{V617F}$ patients and give us confidence in the conclusions we draw from the model and its calibration.

Supplemental figures



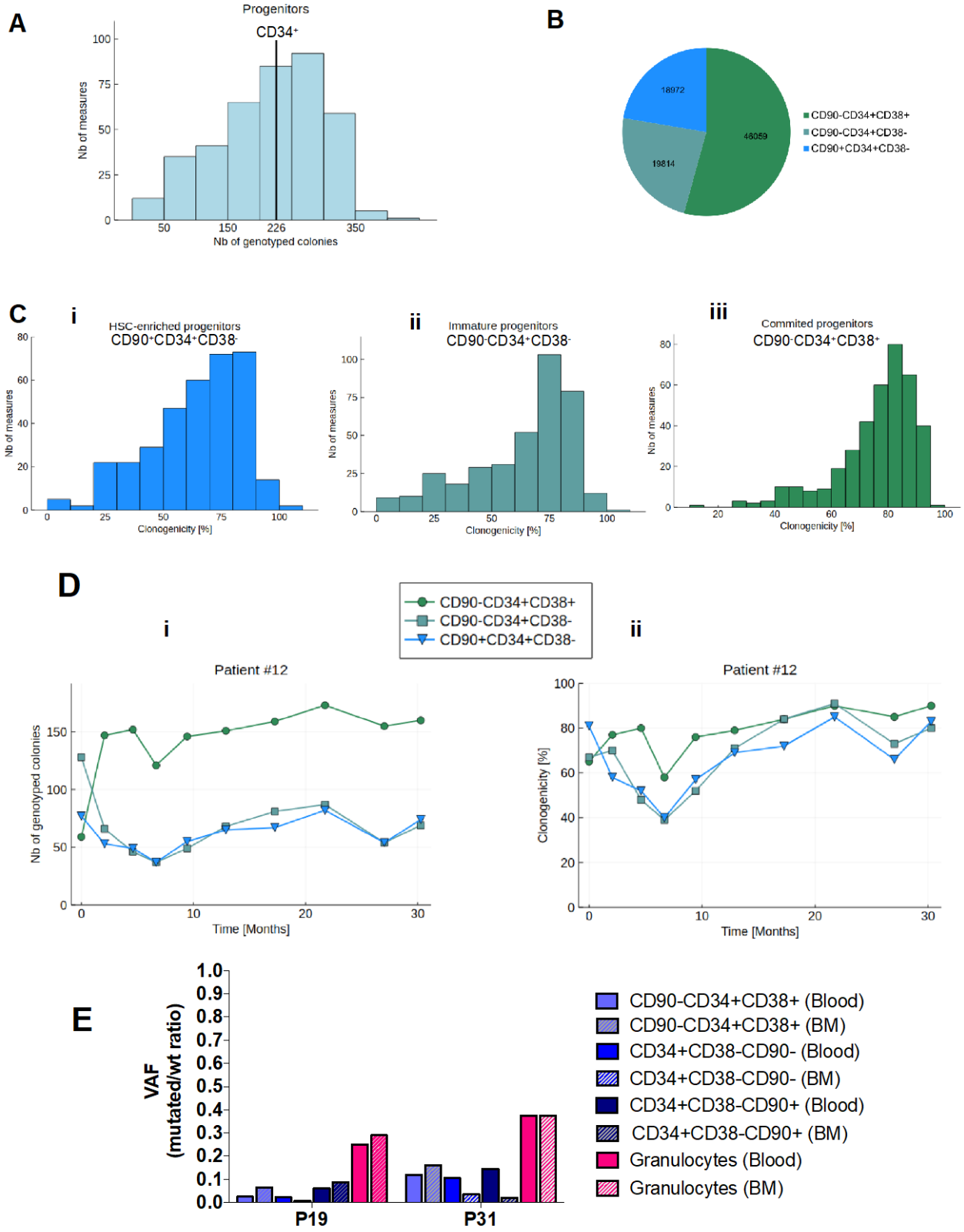
Supplemental figure S1: Clonal architecture of the *CALR*^{m} and *JAK2*^{V617F} progenitors assessed in P46 patient during 31 months. The *JAK2*^{V617F} homozygous subclone was only present at very low clonal fraction in only one progenitor subtypes before IFNα treatment.



Supplemental figure S2: Correlations between clonal fraction and allele burden measurements in different cells.

(A) Correlations between measured allele burdens in different immature cell types (top three rows) and mature cell types (bottom two rows). Each point corresponds to a pair of measurements in the same patient at the same time after applying the transformation log(1 + p), where p is the VAF between 0 and 100. Dashed lines correspond to the first bisector y = x. Solid lines are linear regression lines and shaded bands are 95% confidence intervals for the regression lines. Two sets of measurements are similar if the regression line is close to the first bisector (y = x).

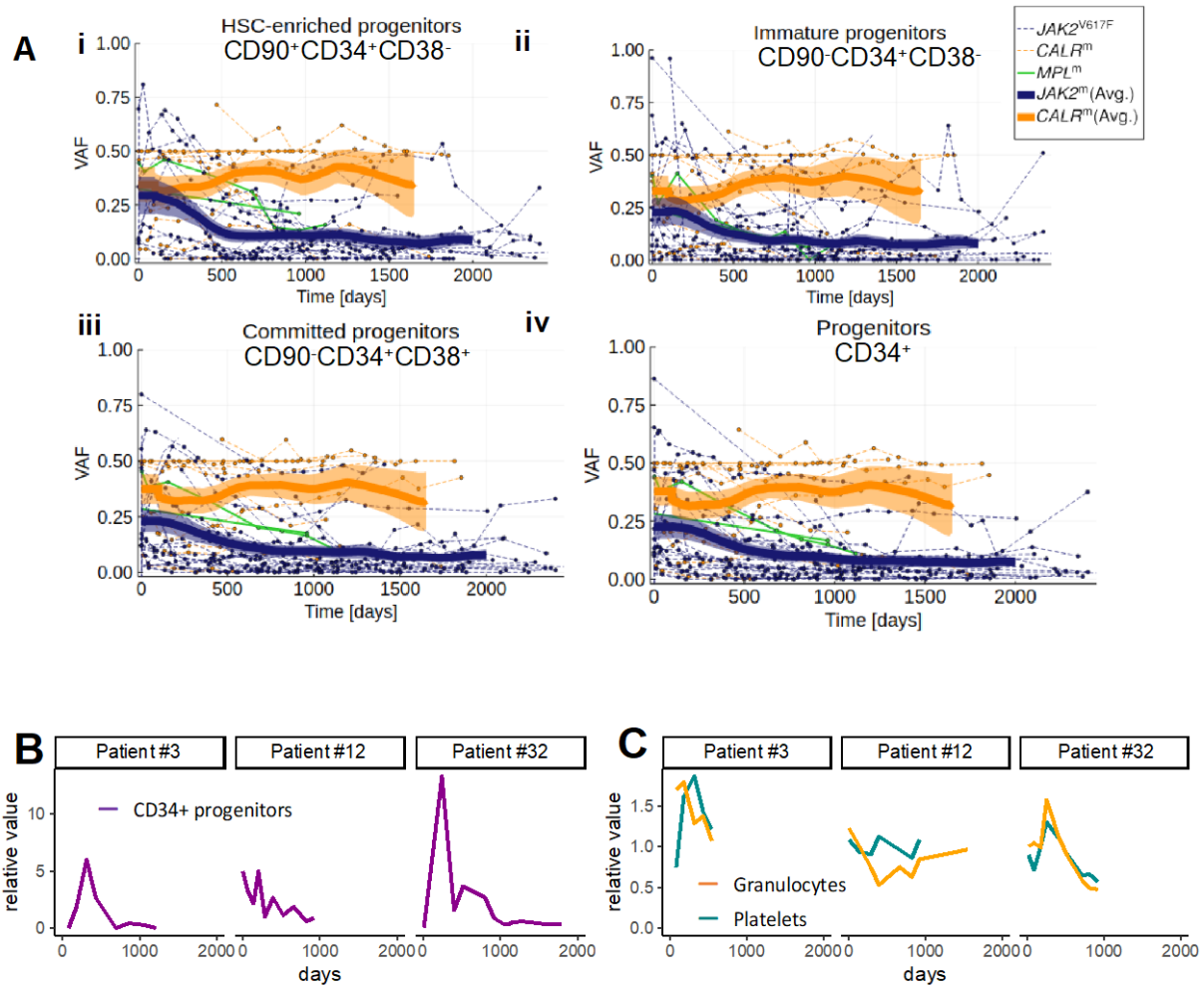
(B-C) Correlations between measured homozygous (B) and heterozygous (C) clonal fractions in different immature cell compartments. Analyses and aesthetics are the same as in panel A.



Supplemental figure S3: Primary data obtained from single cell experiments during the follow-up.

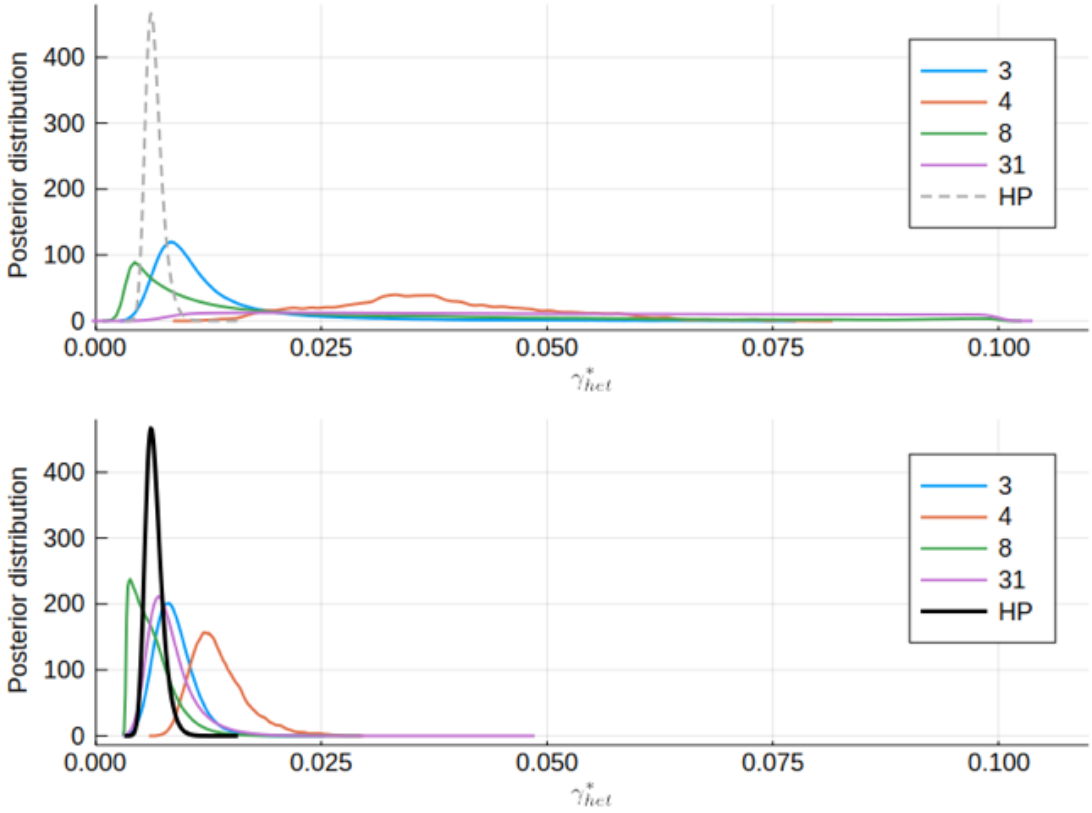
(A) Distribution of the 395 timepoint measures in function of the number of genotyped progenitor colonies. The median number of genotyped colonies was 226 per timepoint. (B) Absolute numbers of $CD90^{+}CD34^{+}CD38^{-}$, $CD90^{-}CD34^{+}CD38^{-}$ and $CD90^{-}CD34^{+}CD38^{+}$ progenitor colonies genotyped for the 48 patients. (C) Distribution of the 395 timepoint measures in function of the clonogenicity.

The median clonogenicities were 67%, 71% and 79% for CD90^{+}CD34^{+}CD38^{-}, CD90^{-}CD34^{+}CD38^{-} and CD90^{-}CD34^{+}CD38^{+}, respectively. (D) Number of genotyped colonies overtime (i) and of clonogenicity (ii) overtime in patient P12. (E) VAF measures in different type of progenitors and granulocytes in the blood and bone marrow of 2 patients (P19 and P31) at a given timepoint.



Supplemental figure S4: Proportion of mutated hematopoietic progenitors and kinetics of total number of progenitors, platelets and granulocytes in the blood induced by IFNα

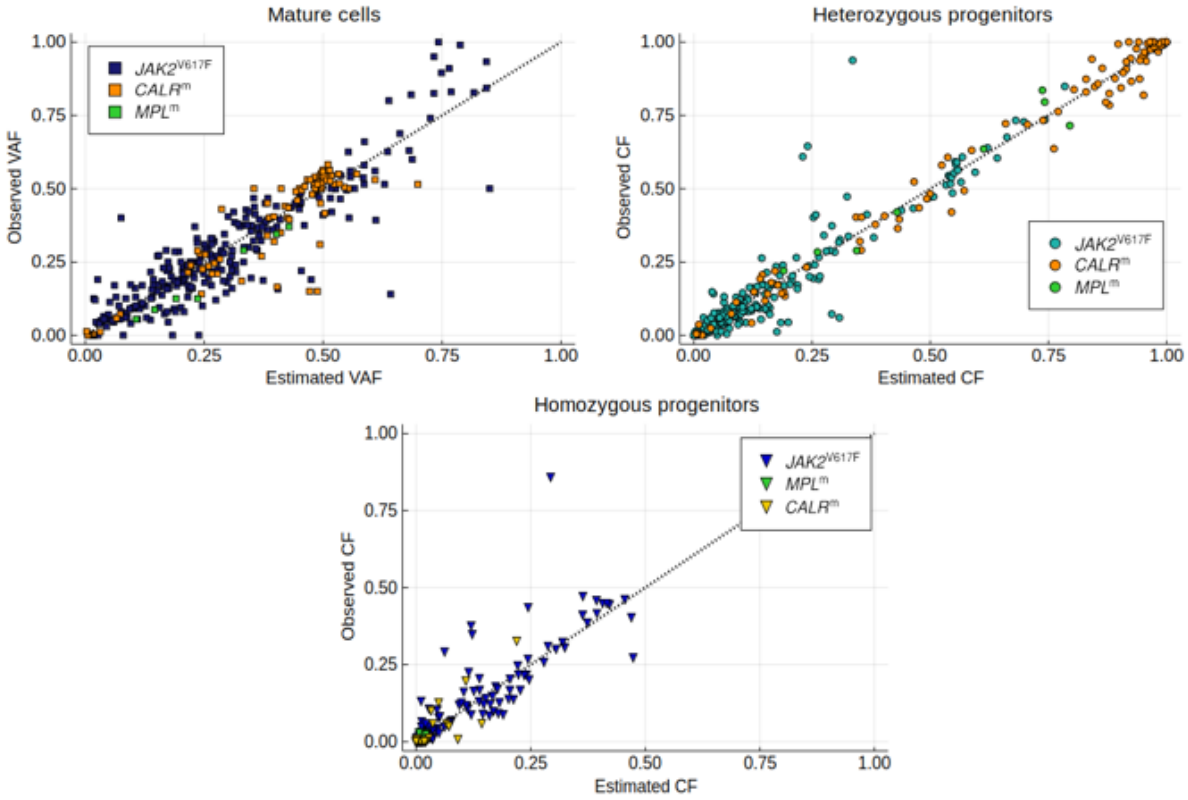
(A) Effect of IFNα in different hematopoietic compartments during the clinical survey. Graph lines indicate the VAF measured or calculated in (i) HSC-enriched immature progenitors (CD90^{+}CD34^{+}CD38^{-}), (ii) immature progenitors (CD90^{-}CD34^{+}CD38^{-}), (iii) committed progenitors (CD90^{-}CD34^{+}CD38^{+}) and (iv) pooled progenitors (CD34^{+}). Thin lines represent data from each of the 48 patients whose MPN harboring *JAK2*^{V617F} (dotted blue), *CALR*^{m} (dotted orange) or *MPL*^{m} (green). Thick curves are smoothed averages (floating averages over ±100 days) from the 32 *JAK2*^{V617F} (blue) or 14 *CALR*^{m} (orange) case data. (B and C) Cell counts for CD34^{+} progenitors (B) and mature cells (C) in the peripheral blood of three patients, relative to the mean across these three patients.



Supplemental figure S5: Comparison between a population-based approach and a standard Bayesian estimation

Study of the effect of the hierarchical Bayesian estimation (bottom) and a standard one (top) with the example of the parameter γ^*_{het}. The top graph is an estimation of individual parameters from four patients. Patient #4 had only 3 data points, so for him γ^*_{het} was estimated with a high variance. Patient #31 had a lot of data points but with a very flat dynamics for its mutated heterozygous progenitors so, given the data, his parameter γ^*_{het} could potentially have very different values with equal probabilities without impacting the quality of the fits. Patients #3 and #8 had more characteristic dynamics, their parameters γ^*_{het} were therefore estimated with high accuracy: values far from the *maximum a posteriori* would result in bad fits.

The bottom graph shows results of the hierarchical estimation based on the population-based approach, we did not estimate each patient's parameters independently but used a hierarchical framework. We considered *a priori* that each patient parameter vector is a sample from the same distribution, describing both the inter-individual variability (with the probability distribution) and the similarity in the population (the fact that the distribution is common for all individuals in the population). Technically, we introduced a population parameter called hyper-parameter (HP) that was estimated. This figure shows clearly the effect of HP on all patients, their posterior distributions (i.e. the estimation of the parameter γ^*_{het} with uncertainties) is more concentrated. This method also reduces the variance of the parameters.

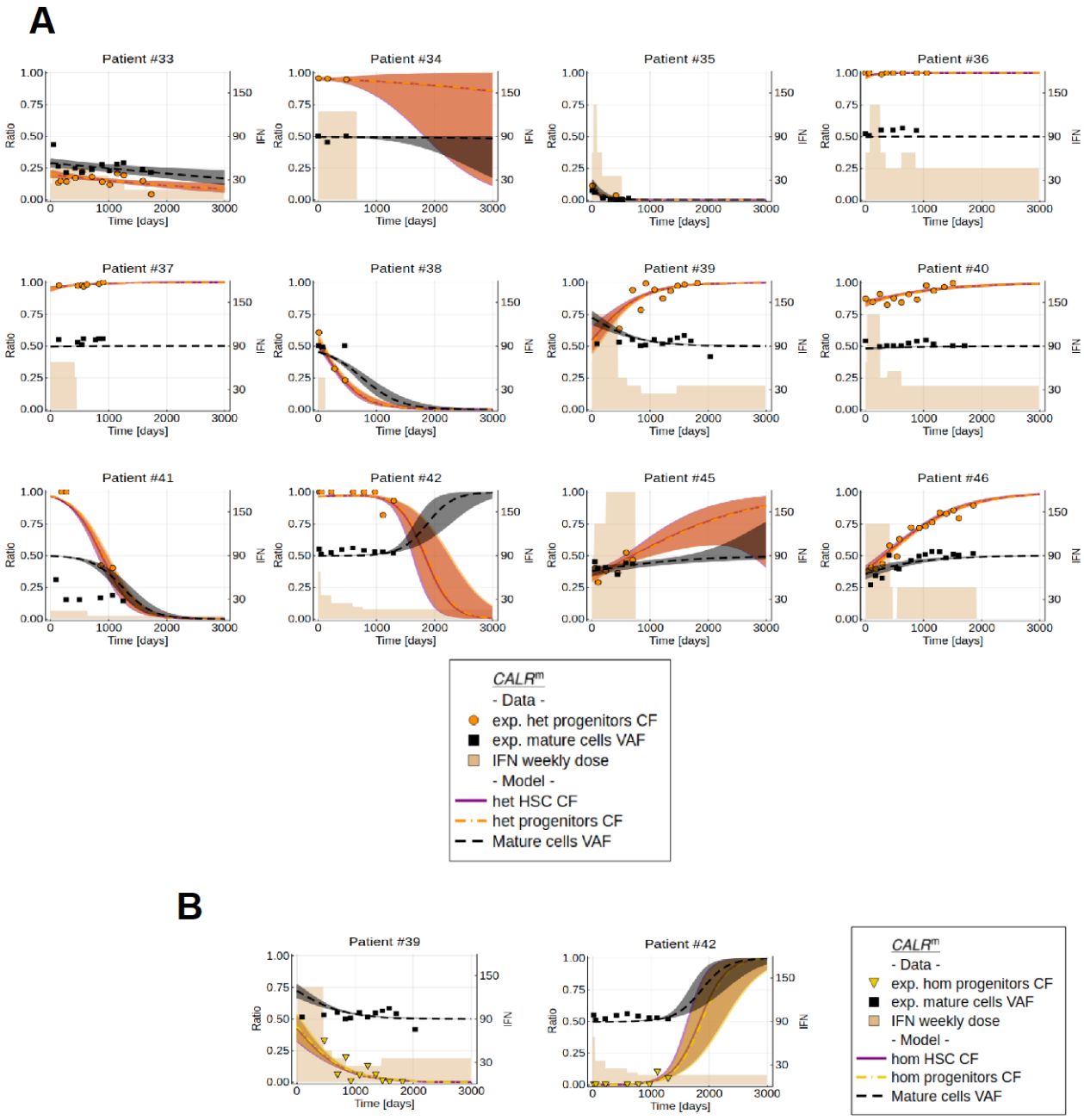


Supplemental figure S6: Comparison between observed and inferred values

Comparisons between observed (y-axis) and inferred (x-axis, median) values of CF or VAF among progenitors and mature cells. The linear dotted line represents the exact fits, points away from this line are not in agreement with the model. This figure syntheses the observations and corresponding estimations from the inferred dynamics of all patients after the start of the treatment.

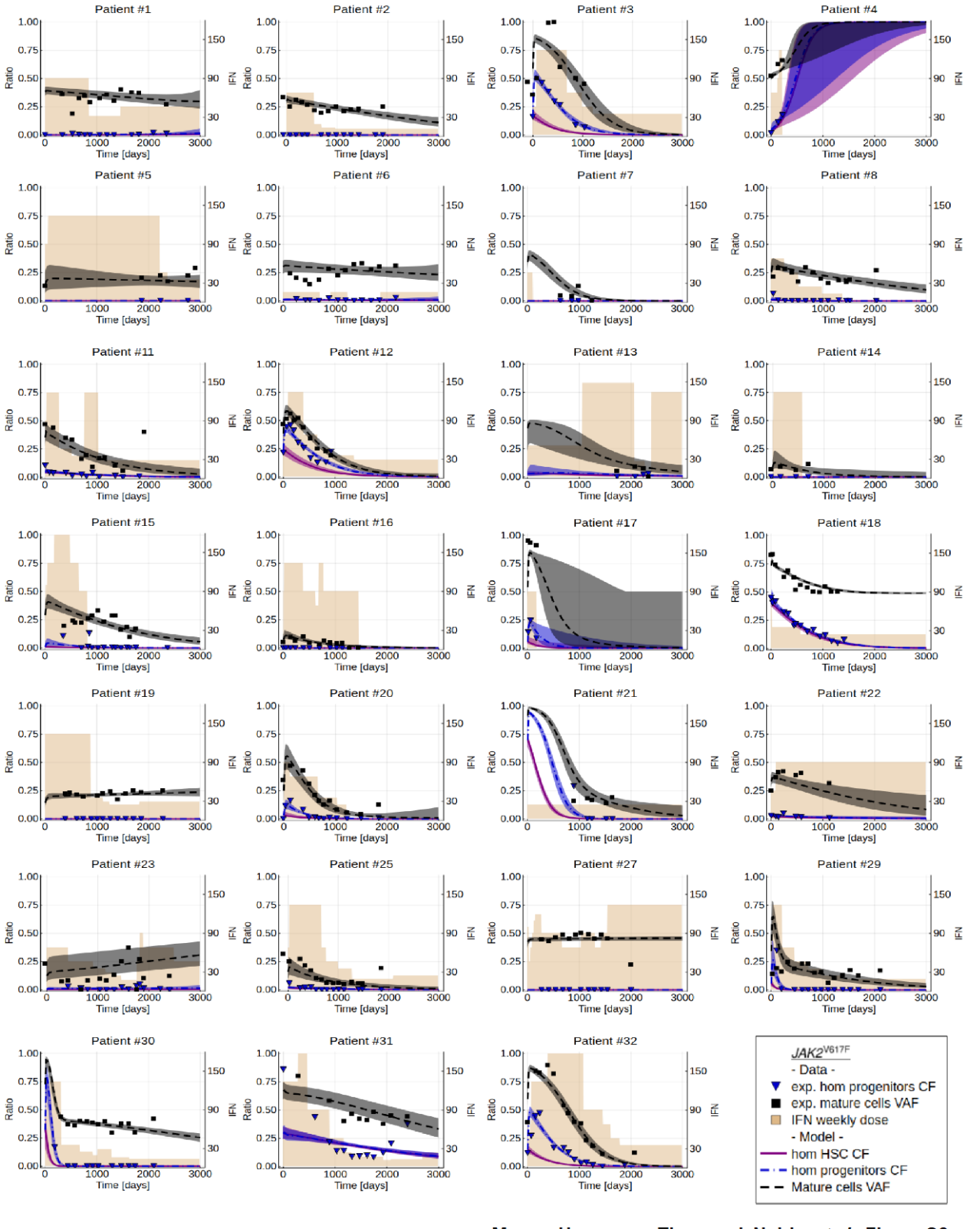
For mature cells, we have 98, 264 and 7 datapoints for *CALR*^{m}, *JAK2*^{V617F} and *MPL*^{m} cases, respectively. Most of the squares are localized near the dotted line, indicating that the model describes correctly the mature VAF dynamics. For high experimental VAF in mature cells, the model underestimates the theoretical VAF.

For progenitors, we have 96, 258 and 6 datapoints for *CALR*^{m}, *JAK2*^{V617F} and *MPL*^{m} cases, respectively. There is a very good fit between observed and estimated heterozygous CF for most of the data points, with values ranging from 0 to 1. It is harder to evaluate the quality of the estimations for homozygous cells because most of the experimental points have a negligible value.



Supplemental figure S7: Inferred *CALR*^{m} dynamics

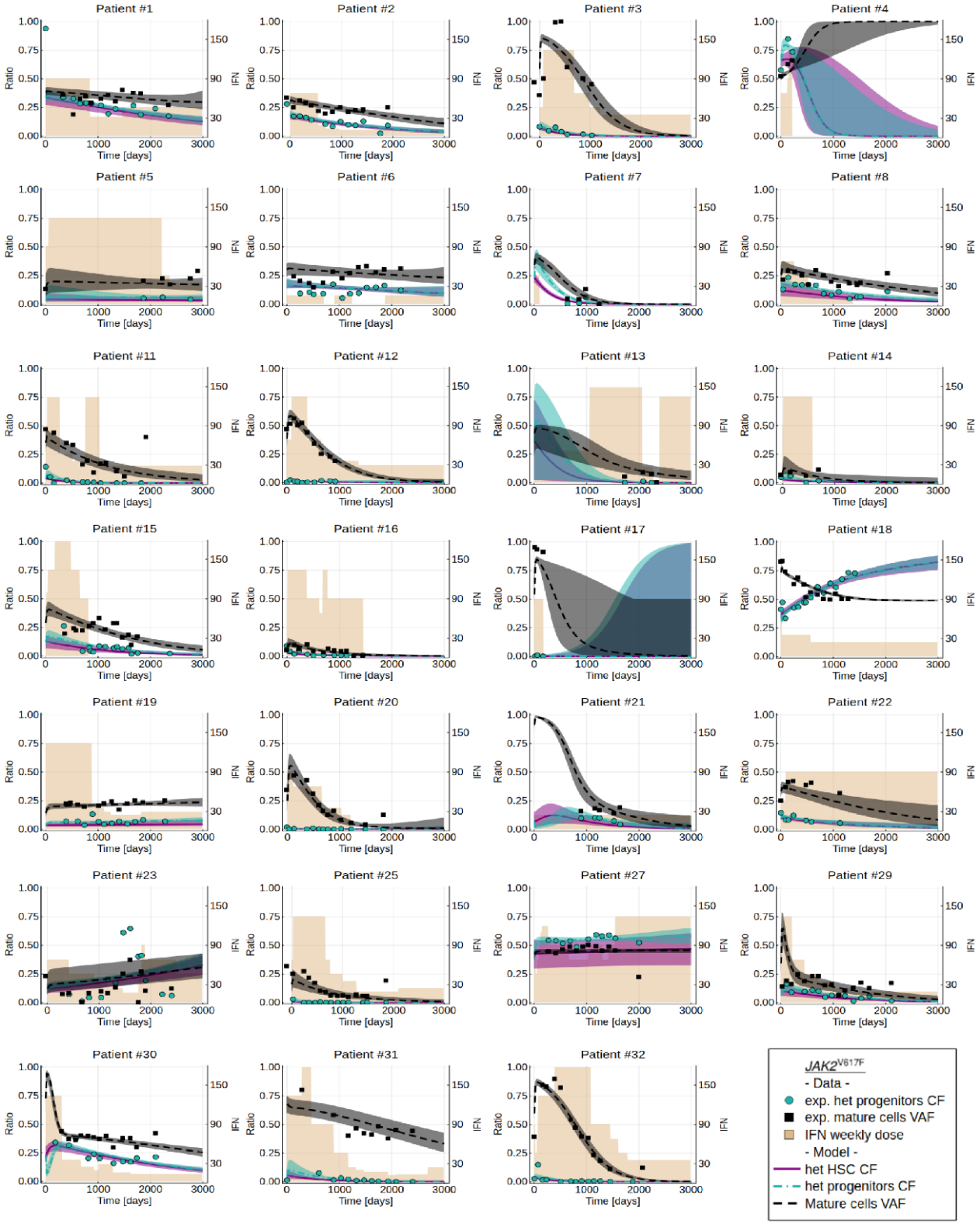
All the dynamics of inferred mutated progenitors and HSC (Clonal Fraction CF) and mutated mature cells (VAF) are presented for (A) heterozygous and (B) homozygous *CALR*^{m} cells for the 12 patients considered in the hierarchical Bayesian framework. These patients were considered because they had a sufficient number of data points (>2 progenitor time point measurements from the start of treatment). The dots, triangles and squares are the experimental data values. The curves were determined with the mathematical model (median values). The purple line is the inferred dynamics of mutated HSC. The shaded areas represent 95% confidence intervals. The shaded beige areas correspond to the dosage of IFNα received overtime. Overall, there are very good fits. Patient #41, who exhibited a low mature VAF (~0.2-0.25) within the first months of therapy that was not compatible – according to our model – with a clonal fraction among progenitors almost equal to 1, was the only *CALR*^{m} case that had discrepancy between the data and the model.



Supplemental figure S8: Inferred homozygous *JAK2*^{V617F} dynamics

All the dynamics of inferred homozygous mutated progenitors (CF), mature cells (VAF) and HSC (CF) are presented for 27 *JAK2*^{V617F} cases for whom we made the parameter estimation using our hierarchical Bayesian framework. These cases were considered because they had a sufficient number

of data points (>2 progenitor measures from the start of treatment). Only a few cases had homozygous clones. Triangles and squares are experimental data values. The curves were determined with the model (median values). The purple line is the inferred dynamic of mutated HSC. The shaded areas represent 95% credibility intervals. The shaded beige areas correspond to the dosage of IFNα received overtime. For most of the cases, there is a good fit between the experimental data and the inferred curves. In particular, our model reproduced and explained the increase followed with the decrease of the homozygous progenitor CF observed within the first months of therapy. For patient #31 only, the model transiently failed to describe the experimental homozygous CF dynamics, with first a decrease followed by a plateau and finally an increase.

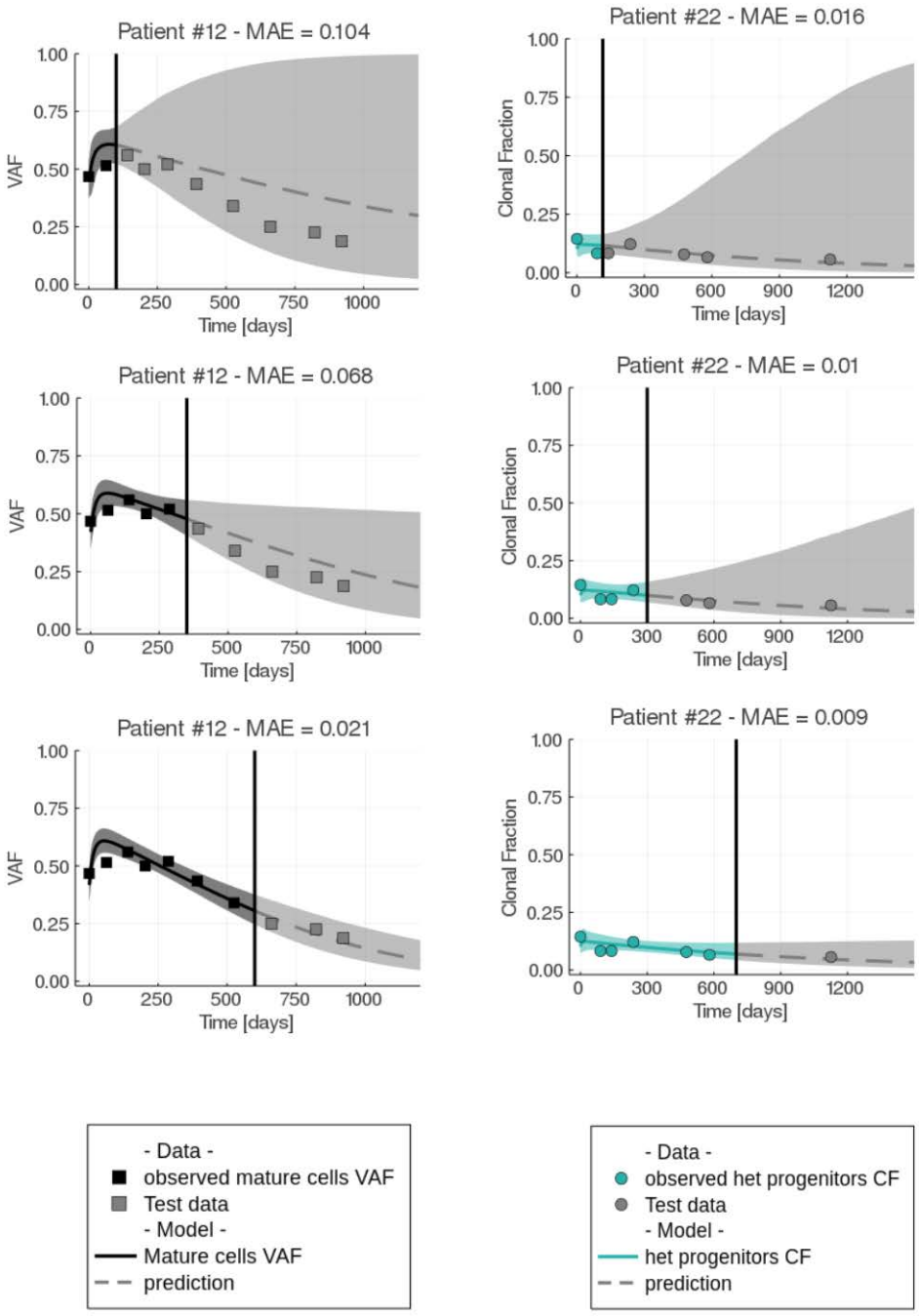


Mosca, Hermange, Tisserand, Noble, *et al.*, Figure S9

Supplemental figure S9: Inferred heterozygous *JAK2*^{V617F} dynamics

All the dynamics of inferred heterozygous mutated progenitors (CF), mature cells (VAF) and HSC (CF) are presented for the 27 *JAK2*^{V617F} cases. The dynamics for each case were obtained through the same parameter estimation as in Supplemental Figure S8. Of note, our model estimates heterozygous and homozygous dynamics together, so the mature VAF presented in the figures is the same as in

Supplemental Figure S8. The dots and squares are the experimental data values. The curves were determined with the model (median values). The purple line is the inferred dynamic of mutated HSC. The shaded areas represent 95% credibility intervals. The shaded beige areas correspond to the dosage of IFNα received overtime. For most cases, there were good fits between our estimated dynamics and the experimental data, for both mature VAF and heterozygous CF. Only a few patients had experimental dynamics that could not be described by the model. This is the case for patient #23 (first left, row six), which exhibited a very late increase and then decrease of both its progenitor heterozygous CF and mature VAF. The patient #15 had similar dynamics. Interestingly, patients #2 or #6 had oscillatory dynamics. These phenomena cannot be captured by our model in its present state because we suspect that it relies on treatment dosage variations. As a perspective, it will be interesting to study some regulatory or pharmacodynamic effects. Note that not all progenitor data points shared the same uncertainty: some heterozygous CF were calculated from a smaller number of sampled and genotyped progenitors than for others, so they came with more uncertainty, which might further justify that the inferred curves did not pass through these points (more details in the supplementary online methods).



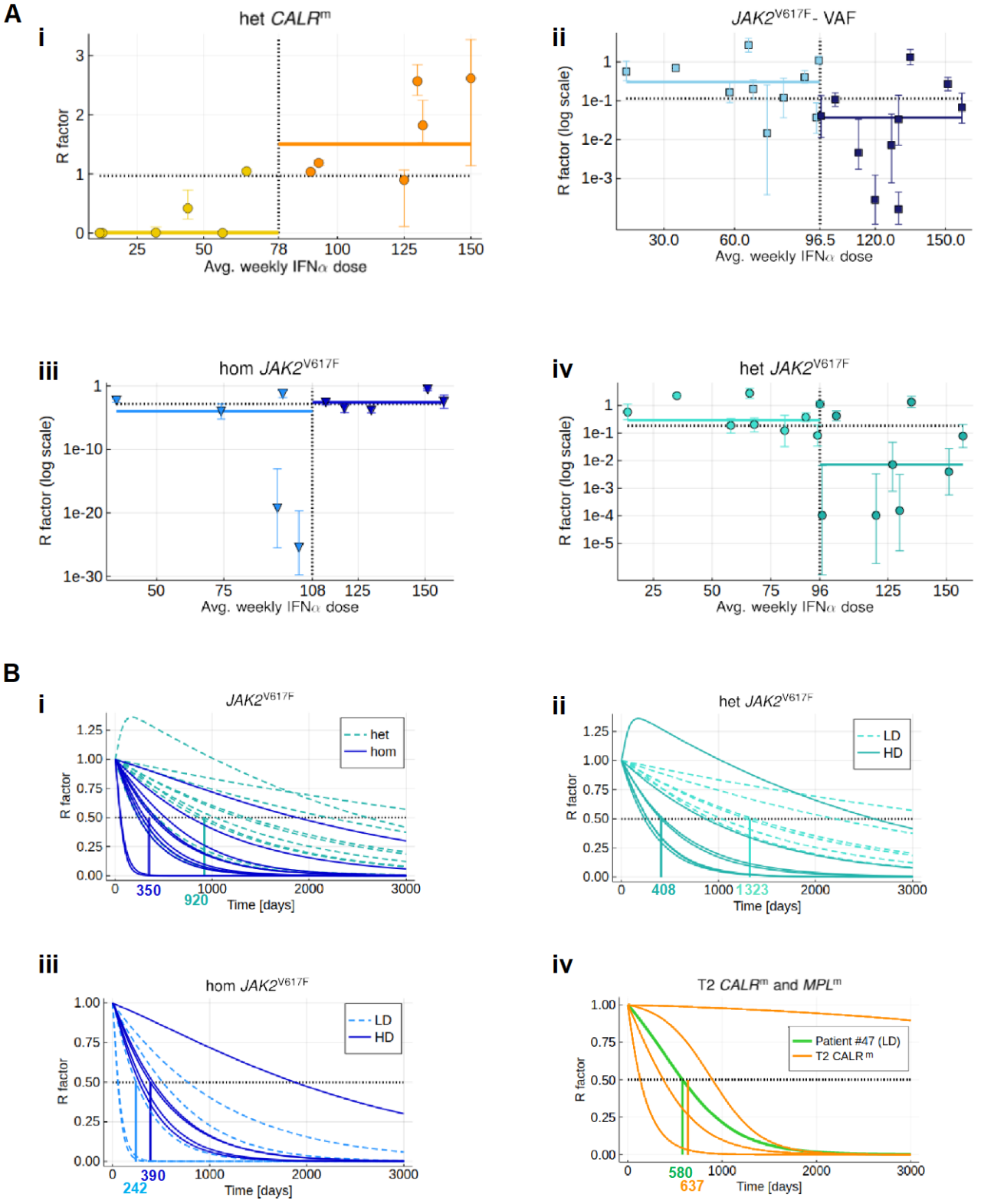
S10: Validation of the predictive capacity of the model

Comparison between model predictions and experimental VAF measurements in mature cells from patients #12 (A) or experimental CF measurements in heterozygous progenitor from patient #22 (B) according to the number of early observations used for inferring the long-term molecular response to IFNα.

Observations before the vertical line are used to predict the long-term response to the therapy.

The solid line represents the (median) value of the inferred VAF in mature cells (A) or CF in heterozygous progenitors (B), when the dashed line represents the (median) value of the predicted VAF in mature cells (A) or CF in heterozygous progenitors (B). Predictions are given with a 95% credibility interval. The predictions' quality are evaluated by computing the Mean Absolute Error (MAE) over the remaining observations (test data beyond the vertical line).

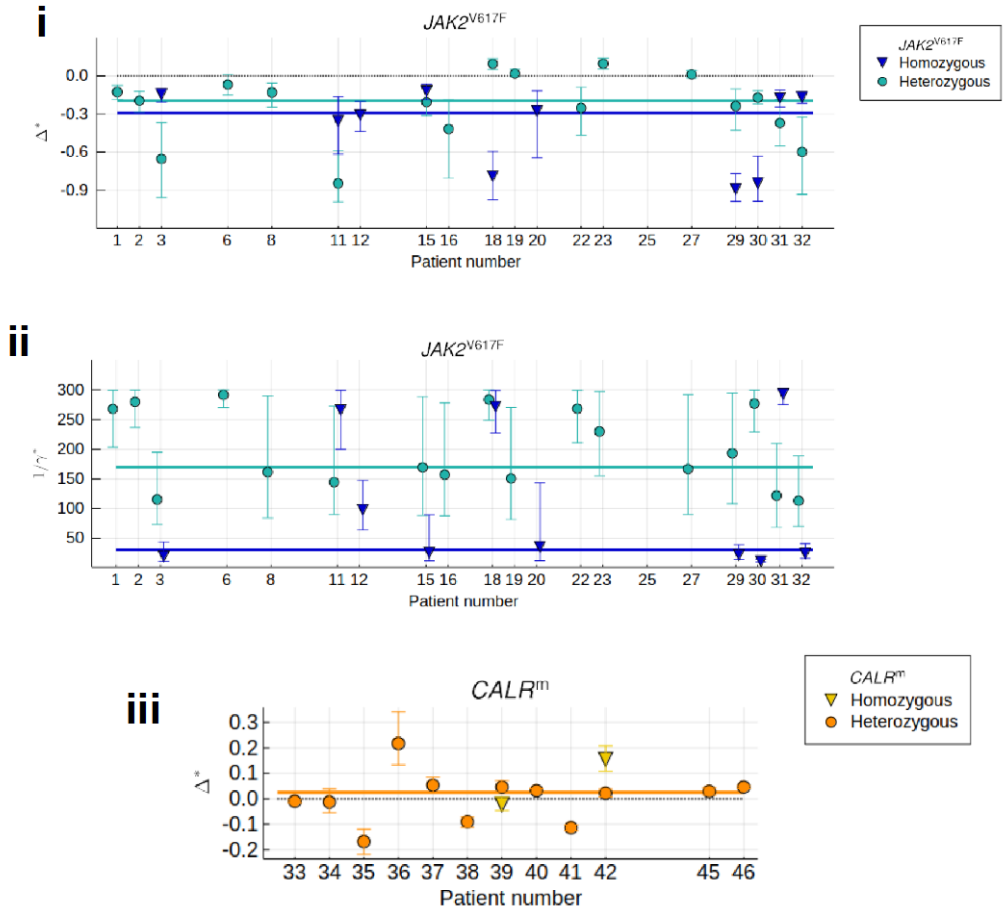
With as few as two observations, the prediction (mean value) was close to the withheld data. The error and uncertainty are decreasing as the number of data used increases for the predictions.



Supplemental figure S11: IFNα impacts molecular response differently according to the mutation type and zygosity in HSC and to the IFNα dosages

(A) Molecular response factor (R) was calculated at the end of the treatment (3,000 days) and plotted against the various weekly IFNα dosages (averaged over 450 days) for (i) heterozygous *CALR*^{m} HSC; (ii) global *JAK2*^{V617F} HSC; (iii) homozygous *JAK2*^{V617F} HSC and (iv) heterozygous *JAK2*^{V617F} HSC. Points represent individual median values. We observed a significant tendency to worse response (R>1) for heterozygous *CALR*^{m} HSC while continuously increasing the dosage (P<1e-4 when testing the nullity of the linear regression coefficient). On the contrary a trend for a better response (R→ 0)

for global *JAK2*^{V617F} HSC was observed when the dosage was increased (P=0.0747 when testing the nullity of the linear regression coefficient). There was no effect of the dosage but overall a very good response for homozygous *JAK2*^{V617F} HSC whereas there was a trend to have an improved response for heterozygous *JAK2*^{V617F} HSC while continuously increasing the dosage (P=0.0498 when testing the nullity of the linear regression coefficient). The error bars represent the 95% credibility interval resulting from the Bayesian estimation. The vertical dotted lines delimit the groups of patients into two subgroups of the same size, below or above the median dosage. The horizontal black dotted lines represent the median value of the R factor for all patients within the group and the horizontal colored solid lines represent median values of the R factor computed among patients of each dosage groups. (B) Molecular response factor (R factor) dynamics were calculated during the survey for (i) heterozygous *versus* homozygous *JAK2*^{V617F} HSC, (ii) heterozygous *JAK2*^{V617F} HSC in patients treated with high *versus* low IFNα dosages, (iii) homozygous *JAK2*^{V617F} HSC in patients treated with high *versus* low IFNα dosages, (iv) heterozygous *MPL*^{m} HSC and type 2 *CALR*^{m} HSC. Only responding patients are considered here. The median time to achieve a 50% decrease of R factor (horizontal black dash lines) was calculated for each patient (and clone). Homozygous *JAK2*^{V617F} HSC, with a median time to achieve a 50% decrease of R factor (PMR) of 350 days, were depleted faster than heterozygous ones (PMR=920 days). The dosage had a significant effect (Mann-Whitney U test, P=0.0221) on the PMR for heterozygous patients (408 days for LD *versus* 1323 days for HD). No significant difference was found for homozygous responders with a PMR equal to 242 and 390 days for LD and HD, respectively.



Supplemental figure S12: IFNα differentially impacts on *JAK2*^{V617F} and *CALR*^{m} HSC homeostasis

Estimated parameters (mean values and 95% credibility intervals) for each patient. (i) Δ^*_{hom} and Δ^*_{het} for *JAK2*^{V617F} homozygous and heterozygous HSC estimated for each patient. A tendency for homozygous cells to encounter more differentiated divisions than heterozygous cells under IFNα was observed. The green circles are the mean values for heterozygous HSC and the blue triangles are the mean values for homozygous HSC, with 95% credibility interval; (ii) inverse ratios of the parameters γ^*_{het} and γ^*_{hom}. This can be seen as the time spent by the mutated cells in the inactive compartment of our model (these times have to be considered relatively to the fixed value γ =1/300). The green circles are the mean values for heterozygous HSC and the blue triangles are the mean values for homozygoys HSC, with 95% credibility interval; (iii) Δ^*_{het} and Δ^*_{hom} for *CALR*^{m} HSC. The orange circles are the mean values for heterozygous HSC and the yellow triangles are the mean values for homozygous HSC, with 95% credibility interval.

Table S1

General information of all the studied patients. Table presents the patient number, gender, age at the beginning of the treatment, MPN type, disease driver and associated driver mutations, IFNα follow-up time, mean IFNα dosage received over the 450 first days of therapy, hematological responses, molecular responses in mature and different progenitor compartments at the start, the end or during the treatment, global molecular HSC response (in terms of VAF) inferred by the model (median value) and clinical observations. HR: hematological response, SE: side effect, CMR: complete molecular response when mutation was reduced to undetectable levels, PMR: partial molecular response when the mutant allele burden decrease was >50%, MMR: minor molecular response when the mutant allele burden decrease was 25-49%. NR: not responder, NA: not applicable and NE: not estimated.

Patients #	Gender	Age at the beginning of treatment	MPN type	Previous treatment	Mutation(s)	IFN follow-up (months)	Mean weekly IFN dose received over the 450 first days of therapy [µg/week]	Hematological response	Molecular response				Model-based global HSC response factor (R) at 3000 days	Observations
									Mature cells	immature cells				
										Common myeloid progenitors	Immature progenitors	Immature progenitors enriched in HSC		
1	F	25	PV	none	JAK2V617F	149	90	HR	CMR	PMR	PMR	PMR	0.41	
2	F	34	PV	phlebotomy	JAK2V617F	60	68	HR	CMR	PMR	CMR	CMR	0.2	
3	F	45	ET	hydroxyurea	JAK2V617F	50	120	HR	NR	PMR	MMR	PMR	2.8E-04	
4	F	56	PV	hydroxyurea	JAK2V617F	8	42	Stop SE	NA	NA	NA	NA	NA	Not followed, intolerant
5	M	35	PV	hydroxyurea	JAK2V617F	95	129	HR	NA	NA	NA	NA	NA	Followed during the course of treatment
6	F	66	ET	none	JAK2V617F	79	14	HR	NA	NA	NA	NA	0.57	Followed during the course of treatment
7	F	51	ET	none	JAK2V617F	36	11	Stop SE	NA	NA	NA	NA	NA	Not followed, intolerant
8	F	59	PV	hydroxyurea	JAK2V617F	46	58	HR	MMR	MMR	PMR	PMR	0.17	
9	F	56	PV	none	JAK2V617F	21	22	HR	NR	PMR	CMR	PMR	NE	
10	F	29	ET	none	JAK2V617F	8	68	HR	NR	NR	MMR	NR	NE	
11	F	52	PV	hydroxyurea	JAK2V617F	60	97	HR	NR	PMR	CMR	CMR	0.041	
12	F	63	PV	hydroxyurea	JAK2V617F	41	113	HR	MR	NR	NR	NR	4.6E-03	
13	M	48	ET	phlebotomy	JAK2V617F	77	50	HR	NA	NA	NA	NA	NA	Followed during the course of treatment
14	M	35	ET	none	JAK2V617F	19	133	Stop SE	NA	NA	NA	NA	NA	Not followed, intolerant
15	M	57	PV	hydroxyurea	JAK2V617F	30	157	NR	NA	NA	NA	NA	0.068	Followed during the course of treatment
16	M	31	PV	hydroxyurea	JAK2V617F	47	127	HR	CMR	PMR	CMR	CMR	7.19E-03	
17	M	71	PV	hydroxyurea	JAK2V617F	6	36	Stop SE	NA	NA	NA	NA	NA	Not followed, intolerant
18	F	57	PMF	none	JAK2V617F	46	35	HR	MMR	NR	PMR	MMR	0.7	
19	M	48	PV	hydroxyurea	JAK2V617F	75	135	HR	NA	NA	NA	NA	1.33	Followed during the course of treatment
20	M	45	PV	hydroxyurea	JAK2V617F	60	74	HR	PMR	NR	CMR	CMR	0.015	
21	F	39	PV	hydroxyurea	JAK2V617F	54	23	HR	NA	NA	NA	NA	NA	Followed during the course of treatment
22	M	52	PV	hydroxyurea	JAK2V617F	37	81	NR	NR	PMR	PMR	PMR	0.12	
23	F	53	PV	hydroxyurea	JAK2V617F	73	66	HR	NA	NA	NA	NA	2.73	Followed during the course of treatment
24	F	51	ET	hydroxyurea	JAK2V617F	24	56	HR	CMR	CMR	CMR	PMR	NE	
25	M	62	PV	hydroxyurea	JAK2V617F	61	130	HR	CMR	CMR	CMR	CMR	0.033	
26	F	56	PMF	hydroxyurea	JAK2V617F	3	90	HR	MMR	NE	NE	NE	NE	
27	M	69	ET	hydroxyurea	JAK2V617F	54	96	HR	NA	NA	NA	NA	1.11	Followed during the course of treatment
28	F	50	PV	hydroxyurea	JAK2V617F	5	34	Stop SE	NE	NE	NE	NE	NE	Not followed, intolerant
29	M	40	ET	hydroxyurea	JAK2V617F	70	95	HR	NA	NA	NA	NA	0.037	Followed during the course of treatment
30	F	62	ET	none	JAK2V617F	70	103	HR	NA	NA	NA	NA	0.11	Followed during the course of treatment
31	M	65	PV	phlebotomy	JAK2V617F	80	151	HR	PMR	PMR	MMR	PMR	0.27	
32	M	52	PV	hydroxyurea	JAK2V617F	61	130	HR	PMR	CMR	PMR	CMR	1.6E-04	
33	M	48	ET	none	CALRdel52	57	44	HR	PMR	PMR	MMR	PMR	0.42	
34	F	68	ET	hydroxyurea	CALRIns5	22	125	Stop SE	NA	NA	NA	NA	0.89	Not followed, Intolerant
35	F	44	ET	hydroxyurea	CALRdel16	17	57	HR	CMR	PMR	CMR	CMR	6.00E-08	
36	M	46	ET	hydroxyurea	CALRdel46	35	90	HR	NR	NR	NR	NR	1.03	
37	M	66	PMF	hydroxyurea	CALRdel52	16	66	Stop SE	NA	NA	NA	NA	1.04	Not followed, Intolerant
38	M	44	ET	hydroxyurea	CALRdel13	4	12	Stop SE	NA	NA	NA	NA	2.00E-04	Not followed, Intolerant
39	M	57	ET	hydroxyurea	CALRdel52	67	132	HR	NA	NA	NA	NA	0.72	Followed during the course of treatment
40	M	53	ET	hydroxyurea	CALRdel34	49	93	HR	NR	NR	NR	NR	1.19	
41	F	59	ET	none	CALRIns5	41	11	HR	NA	NA	NA	NA	3.00E-04	Followed during the course of treatment
42	M	62	PMF	hydroxyurea	CALRdel52	42	32	HR	NR	NR	NR	NR	2.06	
43	F	54	ET	hydroxyurea	CALRdel52	22	45	HR	NA	NA	NA	NA	NE	Followed during the course of treatment
44	M	34	ET	none	CALRins5	24	26	HR	NR	MMR	NR	NR	NE	
45	M	63	PV	hydroxyurea	CALRdel52 JAK2V617F MPLS505N	22	150	NR	NR	NR	NR	NR	2.61	
46	M	45	PMF	hydroxyurea	CALRdel46 JAK2V617F	61	130	HR	NR	NR	NR	NR	2.57	
47	F	66	ET	hydroxyurea	MPLW515K	37	64	HR	PMR	PMR	PMR	PMR	1.18E-03	
48	F	63	ET	none	MPLW515R	32	45	HR	MMR	MMR	PMR	MMR	NE	

Table S2

Parameter values from the model. Table shows all the parameters of the model (for *JAK2*^{V617F} and *MPL*^{m} patients). The distinctions between the three cell types (*wt*, *het* or *hom*) and between before or during treatment are made. We indicated in this table the values that are fixed, the relations between parameters resulting from our assumptions, the parameters that are not relevant (NR) because they will not affect the model output after normalization, and finally the 7 parameters that we estimate. For these latter, we indicate the order of magnitude found from our estimations.

	Before treatment and initial conditions		Under IFNα (t≥0)	
	Parameter	Value	Parameter	Value
WT	α	$1/30$	α^*	$=\alpha$
	Δ	0	Δ^*	$=\Delta$
	γ	$1/300$	γ^*	$=\gamma$
	β	$=\gamma(1-\chi)/\chi$	β^*	$=\beta$
	κ_i	NR	$k_i^* = \kappa_i^* / \kappa_i$	NR
	δ_i	$1/6$	δ_i^*	$=\delta_i$
	κ_m	NR	$k_m^* = \kappa_m^* / \kappa_m$	NR
	δ_m	1	δ_m^*	$=\delta_m$
	χ	$0{,}1$		
	N_{HSC}	NR		
Het	α_{het}	$=\alpha$	α^*_{het}	$=\alpha$
	Δ_{het}	0	Δ^*_{het}	Estimated ($\sim$ -0.2)
	γ_{het}	NR	γ^*_{het}	Estimated ($\sim 1/150$)
	β_{het}	NR	β^*_{het}	$=\beta$
	$k_{i,het} = \kappa_{i,het}/\kappa_i$	1	$k_{i,het}^* = \kappa_{i,het}^* / \kappa_{i,het}$	$= k_i^*$
	$\delta_{i,het}$	$=\delta_i$	$\delta_{i,het}^*$	$=\delta_i$
	$k_{m,het} = \kappa_{m,het}/\kappa_m$	Estimated (~ 7.5)	$k_{m,het}^* = \kappa_{m,het}^* / \kappa_{m,het}$	$= k_m^*$
	$\delta_{m,het}$	$=\delta_m$	$\delta_{m,het}^*$	$=\delta_m$
	χ_{het}	$=\chi$		
	η_{het}	Estimated (~ 0.3)		
Hom	α_{hom}	$=\alpha$	α^*_{hom}	$=\alpha$
	Δ_{hom}	0	Δ^*_{hom}	Estimated ($\sim$ -0.3)
	γ_{hom}	NR	γ^*_{hom}	Estimated ($\sim 1/50$)
	β_{hom}	NR	β^*_{hom}	$=\beta$
	$k_{i,hom} = \kappa_{i,hom}/\kappa_i$	1	$k_{i,hom}^* = \kappa_{i,hom}^* / \kappa_{i,hom}$	$= k_i^*$
	$\delta_{i,hom}$	$=\delta_i$	$\delta_{i,hom}^*$	$=\delta_i$
	$k_{m,hom} = \kappa_{m,hom}/\kappa_m$	$= k_{m,het}$	$k_{m,hom}^* = \kappa_{m,hom}^* / \kappa_{m,hom}$	$= k_m^*$
	$\delta_{m,hom}$	$=\delta_m$	$\delta_{m,hom}^*$	$=\delta_m$
	χ_{hom}	$=\chi$		
	η_{hom}	Estimated (~ 0.3)		

Suppl. REFERENCES

1. James C, Ugo V, Le Couedic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature*. 2005;434(7037):1144–8.
2. Chaligne R, James C, Tonetti C, et al. Evidence for MPL W515L/K mutations in hematopoietic stem cells in primitive myelofibrosis. *Blood*. 2007;110(10):3735–43.
3. Xu J, Koelle S, Guttorp P, et al. Statistical inference for partially observed branching processes with application to cell lineage tracking of in vivo hematopoiesis. *The Annals of Applied Statistics*. 2019;13(4):2091–2119.
4. Kimmel M. Stochasticity and determinism in models of hematopoiesis. *Advances in Experimental Medicine and Biology*. 2014;119–152.
5. Marciniak-Czochra A, Stiehl T, Ho A, Jäger W, Wagner W. Modeling of asymmetric cell division in hematopoietic stem cells|regulation of selfrenewal is essential for efficient repopulation. *Stem cells and development*. 2009;377–386.
6. Colijn C, Mackey M. A mathematical model of hematopoiesis-I. Periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*. 2005;237(2):117–132.
7. Foo J, Drummond MW, Clarkson B, Holyoake T, Michor F. Eradication of Chronic Myeloid Leukemia Stem Cells: A Novel Mathematical Model Predicts No Therapeutic Benefit of Adding G-CSF to Imatinib. *PLoS Comput Biol*. 2009;5(9):e1000503.
8. Roeder I, Horn M, Glauche I, et al. Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. *Nat Med*. 2006;12(10):1181–1184.
9. Michor F, Hughes TP, Iwasa Y, et al. Dynamics of chronic myeloid leukaemia. *Nature*. 2005;435(7046):1267–70.
10. El-Khoury M, Cabagnols X, Mosca M, et al. Different impact of calreticulin mutations on human hematopoiesis in myeloproliferative neoplasms. *Oncogene*. 2020;39(31):5323–5337.
11. Nam AS, Kim KT, Chaligne R, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature*. 2019;
12. Hirsch P, Mamez AC, Belhocine R, et al. Clonal history of a cord blood donor cell leukemia with prenatal somatic JAK2 V617F mutation. *Leukemia*. 2016;30(8):1756–9.
13. Dupont S, Masse A, James C, et al. The JAK2 617V>F mutation triggers erythropoietin hypersensitivity and terminal erythroid amplification in primary cells from patients with polycythemia vera. *Blood*. 2007;110(3):1013–21.
14. Catlin SN, Abkowitz JL, Guttorp P. Statistical Inference in a Two-Compartment Model for Hematopoiesis. *Biometrics*. 2001;57(2):546–553.
15. Llamosi A, Gonzalez-Vargas AM, Versari C, et al. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. *PLoS Comput Biol*. 2016;12(2):e1004706.
16. Gelman A, Carlin JB, Stern HS, et al. Bayesian Data Analysis. Chapman and Hall/CRC; 2013.
17. Andrieu C, Thoms J. A tutorial on adaptive MCMC. *Stat Comput*. 2008;18(4):343–373.
18. Hansen N. The CMA evolution strategy: a comparing review. Towards a new evolutionary computation. 2006.
19. Hansen N. The CMA Evolution Strategy: A Tutorial. *arXiv:1604.00772 [cs, stat]*. 2016;