

Supplementary Methods

Bayesian models of IFN-I resistance of HIV-1 plasma isolates from longitudinally sampled participants

To create a simple model of the temporal dynamics of type I interferon (IFN-I) resistance, IFN α 2 and IFN β IC₅₀ values were each modeled using a Bayesian change point hierarchical model. The model is based on a segmented regression of the log IC₅₀ making the following simplifying assumptions:

- Each participant has a level of resistance at the acute infection stage drawn from separate population-level distributions for typical, non- or fast progressors.
- Each participant has a drop (or rise) in IFN-I resistance from acute levels drawn from separate population-level distributions for typical, non- or fast progressors.
- Each participant has a time to nadir drawn from separate population-level distributions for typical, non- or fast progressors.
- Resistance changes linearly from onset of symptoms to time of nadir.
- The nadir of IFN-I resistance represents a changepoint in the data. Following this point, changes in IFN-I resistance are modeled as a linear function of CD4+ T cell count changes away from the count found at nadir.
- Where data is not present, CD4+ T cell counts are assumed to be linearly interpolated between adjacent observations.

The log IC₅₀ observation from each viral isolate i was modeled as a normal distribution $IC50_i \sim \text{Normal}(\mu_i, \sigma)$ with mean μ_i where:

$$\mu_i = \begin{cases} \alpha_{\text{person}_i} + \delta_{\text{person}_i} \frac{\text{time}_i}{s_{\text{person}_i}} & \text{if } \text{time}_i < s_{\text{person}_i} \\ \alpha_{\text{person}_i} + \delta_{\text{person}_i} + \beta_{\text{person}_i} (\text{CD4}_{\text{person}_i, \text{time}_i} - \text{CD4}_{\text{person}_i, s_{\text{person}_i}}) & \text{if } \text{time}_i \geq s_{\text{person}_i} \end{cases}$$

where the parameters α_j represent the level of IFN-I resistance at symptom onset, δ_j represents the change from symptom onset to nadir and s_j represents the time of nadir in person j . Study participant data is represented by time_i corresponding to the time since onset of symptoms and person_i recording the participant from which isolate i was collected, $\text{CD4}_{j,k}$ containing the estimated CD4+ T cell count for person j at time k and progression_j is the disease progression type (fast/non/typical) for participant j . The hierarchical probabilities

for these parameters were:

$$\begin{aligned}
\sigma &\sim \text{Gamma}(1, 0.1) \\
\alpha_j &\sim \begin{cases} \text{Normal}(\theta_{\alpha, \text{typical}}, \tau_{\alpha}) & \text{if progression}_j = \text{typical} \\ \text{Normal}(\theta_{\alpha, \text{typical}} + \theta_{\alpha, \text{fast}}, \tau_{\alpha}) & \text{if progression}_j = \text{fast} \\ \text{Normal}(\theta_{\alpha, \text{typical}} + \theta_{\alpha, \text{non}}, \tau_{\alpha}) & \text{if progression}_j = \text{non} \end{cases} \\
\delta_j &\sim \begin{cases} \text{Normal}(\theta_{\delta, \text{typical}}, \tau_{\delta}) & \text{if progression}_j = \text{typical} \\ \text{Normal}(\theta_{\delta, \text{typical}} + \theta_{\delta, \text{fast}}, \tau_{\delta}) & \text{if progression}_j = \text{fast} \\ \text{Normal}(\theta_{\delta, \text{typical}} + \theta_{\delta, \text{non}}, \tau_{\delta}) & \text{if progression}_j = \text{non} \end{cases} \\
s_j &\sim \begin{cases} \text{NegativeBinomial}(\theta_{s, \text{typical}}, \tau_s) & \text{if progression}_j = \text{typical} \\ \text{NegativeBinomial}(\theta_{s, \text{typical}} \exp(\theta_{s, \text{fast}}), \tau_s) & \text{if progression}_j = \text{fast} \\ \text{NegativeBinomial}(\theta_{s, \text{typical}} \exp(\theta_{s, \text{non}}), \tau_s) & \text{if progression}_j = \text{non} \end{cases} \\
\beta_j &\sim \begin{cases} \text{Normal}(\theta_{\beta, \text{typical}}, \tau_{\beta}) & \text{if progression}_j = \text{typical} \\ \text{Normal}(\theta_{\beta, \text{fast}}, \tau_{\beta}) & \text{if progression}_j = \text{fast} \\ \text{Normal}(\theta_{\beta, \text{non}}, \tau_{\beta}) & \text{if progression}_j = \text{non} \end{cases}
\end{aligned}$$

where j indicates each participant and $\text{NegativeBinomial}(x, y)$ represents a negative binomial distribution parameterized such that the expected value is x and the variance is $x + \frac{x^2}{y}$. All hyperparameters were given prior probabilities of $\theta_x \sim \text{Normal}(0, 10)$ for parameters representing the means of a distribution and $\tau_x \sim \text{Gamma}(1, 0.1)$ for parameters representing standard deviations other than $\theta_{\alpha, \text{typical}}$ and $\theta_{s, \text{typical}}$ which were given a flat prior and $\tau_s \sim \text{Cauchy}(0, 10)$.

For computational efficiency, the nadir time parameter s was discretized to weekly intervals, assumed to fall within 1 to 150 weeks after symptom onset and marginalized out of the joint probability:

$$p(\text{IC50}, \dots) = p(\dots) \prod_{i=1}^n \sum_{s=1}^{150} \text{Normal}(\text{IC50}_i | \mu_{i,s}, \sigma) \text{NegativeBinomial}(s | \theta_s, \tau_s)$$

where \dots represents all parameters other than s and $\mu_{i,s}$ is defined the same as μ_i :

$$\mu_{i,s} = \begin{cases} \alpha_{\text{person}_i} + \frac{\text{time}_i}{s} \delta_{\text{person}_i} & \text{if time}_i < s \\ \alpha_{\text{person}_i} + \delta_{\text{person}_i} + \beta_{\text{person}_i} (\text{CD4}_{\text{person}_i, \text{time}_i} - \text{CD4}_{\text{person}_i, s}) & \text{if time}_i \geq s \end{cases}$$

Posterior probabilities were estimated with 50 Markov chain Monte Carlo chains of 5000 iterations each using Stan (91).

Bayesian models of IFN-I resistance of outgrowth and rebound HIV-1 isolates

To compare the IFN-I resistance of viral isolates derived from plasma samples collected during acute, chronic and rebound infections, as well as from viably frozen PBMCs collected during ART suppression (QVOA), IFN α 2 and IFN β IC₅₀ values were modeled using a Bayesian hierarchical model. The model is based on the assumptions that:

- Isolates found at acute infection form a base level of IFN-I resistance for a given person. Resistances in virus isolated from chronic, ART suppressed and rebound infection for this person are modelled as changes from this initial level.
- The mean IC₅₀ level within each person for acute isolates and the change from acute levels for chronic, QVOA and rebound isolates are drawn from a population-level distribution for that type.
- QVOA isolates are separated into two populations; a “pre” group composed of QVOA viruses isolated from study participants prior to or in the absence of treatment interruption (ATI) and a “post” group of QVOA viruses isolated from participants following ATI and reinitiation of ART.
- In both QVOA populations, the viruses can include some proportion of rebound-like isolates. This mixture is modeled in both pre- and post-treatment so that differences in mixture proportion between the two populations can be assessed.
- Variation in the potency of INF-I used to experimentally determine IC₅₀ values may shift the inferred resistance for isolates tested in other studies. This effect is modeled as a multiplicative shift in IC₅₀ for all isolates measured outside this study (acute recipient and chronic donor isolates from ref. 50).
- Isolates from participants who received exogenous IFN α 2 during treatment interruption may display altered interferon resistance. This effect is modeled as a multiplicative shift in IC₅₀ for all rebound isolates from such participants (participants 004, 030, and 044 from ref. 38).

The log IC₅₀ observation from each viral isolate i from acute, chronic and rebound isolates was modeled as a normal distribution:

$$\text{IC50}_i \sim \text{Normal}(\mu_{\text{type}_i, \text{person}_i}, \sigma_{\text{type}_j})$$

with the mean resistance for isolate type j from person k :

$$\mu_{j,k} \sim \begin{cases} \text{Normal}(\alpha_k + \beta_{\text{batch}} \text{batch}_k, \psi_j) & \text{if } j = \text{acute} \\ \text{Normal}(\alpha_k + \beta_{j,k} + \beta_{\text{batch}} \text{batch}_k + \beta_{\text{ifn}} \text{IFN}_k, \psi_j) & \text{if } j = \text{rebound} \\ \text{Normal}(\alpha_k + \beta_{j,k} + \beta_{\text{batch}} \text{batch}_k, \psi_j) & \text{otherwise} \end{cases}$$

where type_i indicates whether isolate i was isolated during acute, chronic, QVOA or rebound infection from participant person_i , batch_k indicates when isolates from person k were tested in another study and IFN_k indicates when person k was treated with exogenous IFN α 2 prior to and during treatment interruption. Parameters are included for the mean resistance level during acute infection for each person α_k , standard deviation of isolates of type j within a person σ_j , standard deviation of mean resistance for type j isolates among people ψ_j , change from acute levels in isolates of type j in a given participant $\beta_{j,k}$, the effects of exogenous IFN treatment β_{ifn} and batch to batch variation in IFN in isolates assayed in previous studies β_{batch} .

For QVOA isolates, the IC_{50} was modeled as a mixture of two populations such that:

$$p(\text{IC50}_i | \mu_{\text{qvoa}, \text{person}_i}, \sigma_{\text{qvoa}}, \mu_{\text{rebound}, \text{person}_i}, \sigma_{\text{rebound}}, \phi_{\text{prePost}_i}) = \\ \phi_{\text{prePost}_i} \text{Normal}(\text{IC50}_i | \mu_{\text{rebound}, \text{person}_i}, \sigma_{\text{rebound}}) \\ + (1 - \phi_{\text{prePost}_i}) \text{Normal}(\text{IC50}_i | \mu_{\text{qvoa}, \text{person}_i}, \sigma_{\text{qvoa}})$$

where prePost_i indicates whether isolate i was isolated pre- or post-ATI and ϕ_{pre} and ϕ_{post} represent the proportion of rebound-like virus present in pre- and post-ATI QVOA isolates.

The hierarchical parameter priors were modeled as:

$$\begin{aligned} \sigma_j &\sim \text{Gamma}(1, 0.1) \\ \psi_j &\sim \text{Gamma}(1, 0.1) \\ \phi_{\text{pre}} &\sim \text{Uniform}(0, 1) \\ \phi_{\text{post}} &\sim \text{Uniform}(0, 1) \end{aligned}$$

where j indicates the isolate type (acute, chronic, QVOA, rebound). All α_k , $\beta_{j,k}$ were given flat priors and $\beta_{\text{ifn}} \sim \text{Normal}(0, 10)$ and $\beta_{\text{batch}} \sim \text{Normal}(0, 10)$.

Posterior probabilities were estimated with 50 Markov chain Monte Carlo chains of 5,000 iterations each using Stan (91).