

Each variable of interest, e.g. IFN α IC₅₀ or Env:RT ratios, was modeled using a Bayesian hierarchical model. The model is based on a linear regression estimating the differences (β s) between donor plasma viruses and viruses from donor genital fluid or recipient viruses along with the effects of clade and IFN α - and IFN β -selection. Unlike a normal linear regression, this model allows added hierarchy to deal with linked viruses within pairs and includes variance parameters to deal with potential heteroscedasticity among virus populations.

Data was first transformed as follows:

| Variable | Transformation |
|---|----------------|
| Env/RT | log |
| Infectivity (RLU/pg RT) | log |
| Pooled donor Replicative capacity (day 7 p24) | identity |
| IFNbeta IC50 (pg/ml) | log |
| IFNa2 IC50 (pg/ml) | log |
| p24 release with IFNa2 | logit |
| p24 release without IFNa2 | logit |
| Autologous IC50 | log |

The i^{th} observation for each donor-recipient pair was then modeled as a normal distribution $N(\mu_i, \sigma_i^2)$ with mean μ_i :

$$\begin{aligned} \mu_i = & \text{base}_{\text{pair}_i} + \beta_{\text{recipient}, \text{pair}_i} \mathbb{1}(\text{recipient}_i) + \beta_{\text{genital}, \text{pair}_i} \mathbb{1}(\text{genital}_i) + \beta_{\text{clade}, \text{pair}_i} \mathbb{1}(\text{cladeB}_i) \mathbb{1}(\text{recipient}_i) \\ & + \beta_{\text{donorAlpha}, \text{pair}_i} \mathbb{1}(\text{donor}_i \ \& \ \text{alphaSelect}_i) + \beta_{\text{donorBeta}, \text{pair}_i} \mathbb{1}(\text{donor}_i \ \& \ \text{betaSelect}_i) \\ & + \beta_{\text{recipientAlpha}, \text{pair}_i} \mathbb{1}(\text{recipient}_i \ \& \ \text{alphaSelect}_i) + \beta_{\text{recipientBeta}, \text{pair}_i} \mathbb{1}(\text{recipient}_i \ \& \ \text{betaSelect}_i) \end{aligned}$$

and variance σ_i^2 :

$$\sigma_i^2 = \begin{cases} \sigma_{\text{genital}, \text{pair}_i}^2 & \text{if genital}_i \\ \sigma_{\text{recipient}, \text{pair}_i}^2 & \text{if recipient}_i \\ \sigma_{\text{donorAlpha}, \text{pair}_i}^2 & \text{if donor}_i \ \& \ \text{alphaSelect}_i \\ \sigma_{\text{donorBeta}, \text{pair}_i}^2 & \text{if donor}_i \ \& \ \text{betaSelect}_i \\ \sigma_{\text{donor}, \text{pair}_i}^2 & \text{otherwise} \end{cases}$$

where pair_i indicates the pair identity of the i^{th} observation, $\mathbb{1}()$ is an indicator function that is 1 if True and 0 if False. The β are coefficients modeling the change expected for viruses in recipients, in donor genital samples, in recipients from HIV clade B and the effects of IFN α - or IFN β -selection on donor and recipient viruses. So for example, a donor plasma virus i from pair 2 would have mean $\mu_i = \text{base}_2$ and an IFN α -selected recipient virus from pair 3 (which happened to be clade B) would have mean:

$$\mu_i = \text{base}_3 + \beta_{\text{recipient}, 3} + \beta_{\text{clade}, 3} + \beta_{\text{recipientAlpha}, 3}$$

For a trio where one donor transmitted viruses to two separate recipients, recipient parameters were estimated independently for each recipient.

In Autologous IC50, observations less than equal to 20 were censored at 20 (the maximum amount of plasma tolerated by the cells). To model this, the probability of these observations was considered to be: $p(\text{IC50} = 20) = \int_{-\infty}^{\log(20)} N(\mu_i, \sigma_i^2)$

The coefficients β for each pair j come from population-level normal hyperpriors:

$$\begin{aligned} \text{base}_j &\sim N(\mu_{\text{base}}, \sigma_{\text{base}}^2) \\ \beta_{\text{recipient},j} &\sim N(\mu_{\text{recipient}}, \sigma_{\text{recipient}}^2) \\ \beta_{\text{genital},j} &\sim N(\mu_{\text{genital}}, \sigma_{\text{genital}}^2) \\ \beta_{\text{clade},j} &\sim N(\mu_{\text{clade}}, \sigma_{\text{clade}}^2) \\ \beta_{\text{donorAlpha},j} &\sim N(\mu_{\text{donorAlpha}}, \sigma_{\text{donorAlpha}}^2) \\ \beta_{\text{donorBeta},j} &\sim N(\mu_{\text{donorBeta}}, \sigma_{\text{donorBeta}}^2) \\ \beta_{\text{recipientAlpha},j} &\sim N(\mu_{\text{recipientAlpha}}, \sigma_{\text{recipientAlpha}}^2) \\ \beta_{\text{recipientBeta},j} &\sim N(\mu_{\text{recipientBeta}}, \sigma_{\text{recipientBeta}}^2) \end{aligned}$$

and coefficients σ from population-level normal hyperpriors:

$$\begin{aligned} \sigma_{\text{donor},j} &\sim N(\theta_{\text{donor}}, \phi_{\text{donor}}^2) \\ \sigma_{\text{donorAlpha},j} &\sim N(\theta_{\text{donorAlpha}}, \phi_{\text{donorAlpha}}^2) \\ \sigma_{\text{donorBeta},j} &\sim N(\theta_{\text{donorBeta}}, \phi_{\text{donorBeta}}^2) \\ \sigma_{\text{recipient},j} &\sim N(\theta_{\text{recipient}}, \phi_{\text{recipient}}^2) \\ \sigma_{\text{genital},j} &\sim N(\theta_{\text{genital}}, \phi_{\text{genital}}^2) \end{aligned}$$

The effect hyperparameters $\mu_{\text{recipient}}, \mu_{\text{genital}}, \mu_{\text{clade}}, \mu_{\text{donorAlpha}}, \mu_{\text{donorBeta}}, \mu_{\text{recipientAlpha}}$ and $\mu_{\text{recipientBeta}}$ were all given a flat prior probability. The variance parameters $\sigma_{\text{base}}, \sigma_{\text{recipient}}, \sigma_{\text{genital}}, \sigma_{\text{clade}}, \sigma_{\text{donorAlpha}}, \sigma_{\text{donorBeta}}, \sigma_{\text{recipientAlpha}}, \sigma_{\text{recipientBeta}}, \phi_{\text{donor}}, \phi_{\text{donorAlpha}}, \phi_{\text{donorBeta}}, \phi_{\text{recipient}}, \phi_{\text{genital}}, \theta_{\text{donor}}, \theta_{\text{donorAlpha}}, \theta_{\text{donorBeta}}, \theta_{\text{recipient}}$ and θ_{genital} were given a prior of $\text{Gamma}(1,2)$ reflecting prior knowledge that the standard deviation in these assays was unlikely to be greater than several logs. For Replicative Capacity, which was modeled on a linear scale, the variance parameters were given a vague prior of $\text{Gamma}(1, \frac{1}{100})$.

Plots and statistics are based on the estimated posterior probabilities of the population-level effects $\mu_{\text{recipient}}, \mu_{\text{genital}}, \mu_{\text{clade}}, \mu_{\text{donorAlpha}}, \mu_{\text{donorBeta}}, \mu_{\text{recipientAlpha}}$ and $\mu_{\text{recipientBeta}}$.

The model and Markov Chain Monte Carlo sampling of the posterior probability distribution was implemented in Stan using the R package `rstan`. Code is available at: <https://github.com/sherrillmix/hivPair/blob/master/bayesIC50.R>.