Bayesian hierarchical regression models of viral properties

Summary

Each viral property, e.g. IFNα2 IC_{50} or Env/RT ratios, was modeled using a Bayesian hierarchical model. The model is based on a linear regression estimating the differences between donor plasma viruses and viruses from donor genital fluid or recipient plasma viruses along with the effects of HIV subtype and IFNα2- and IFNβ-selection. Unlike a normal linear regression, this model allows accounting for 1) nested measurements within transmission pairs, 2) multiple transmissions from a single donor, 3) heteroscedasticity among virus populations, 4) censored data where we only know that a measurement is less than a given value.

These hierarchical models are based on the assumption that observations within a single patient-fluid-treatment are independent and identically normally distributed with mean and variance drawn from common population-level distributions. Estimates of the population-level distributions can then be used to infer broader patterns in the data.

Methods

Data was first transformed as follows:

Variable	Transformation
Env/RT	log
Infectivity	log
Replicative capacity	log
IFNα2 IC_{50}	log
IFNβ IC_{50}	log
IFNα2 Vres	logit
IFNβ Vres	logit
p24 release	logit

The observation from each viral isolate i was then modeled as a normal distribution $N(\mu_i, \sigma_i^2)$

with mean μ_i:

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

and variance σ_i^2:

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

where pair_i indicates the pair identity of the i^{th} observation, donor_j is the estimated mean of untreated donor plasma viral isolates from pair j and $\mathbb{1}()$ is an indicator function that is 1 if True and 0 if False. The various 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 IFNα2- or IFNβ-selection on donor or recipient viruses. So for example, a donor plasma virus i from pair 2 would have mean $\mu_i = \text{donor}_2$ and an IFNα2-selected recipient virus from pair 3 (which happened to be clade B) would have mean:

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

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

In Autologous and Bnaber IC_{50}, observations less than or 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:

$$p(\text{IC}_{50} = 20) = \int_{-\infty}^{\log(20)} N(\mu_i, \sigma_i^2)$$

Vres measurements were calculated as the amount of p24 released under maximum IFN dose divided by the released p24 without IFN as measured by ELISA. The limit of detection on these measurement was 0.1 so concentrations ≤ 0.1 were measured as 0.1. To account for

this, the probability of these observations was considered to be:

$$p\left(\text{Vres} = \frac{0.1}{\text{Untreated p24}}\right) = \int_{-\infty}^{\text{logit}\left(\frac{0.1}{\text{Untreated p24}}\right)} N(\mu_i, \sigma_i^2)$$

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

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

and coefficients σ from population-level normal hyperpriors:

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

The effect hyperparameters $\mu_{\text{recipient}}$, μ_{genital}, μ_{clade}, $\mu_{\text{donorAlpha}}$, $\mu_{\text{donorBeta}}$, $\mu_{\text{recipientAlpha}}$ and $\mu_{\text{recipientBeta}}$ were all given a flat prior probability. The variance parameters σ_{donor}, $\sigma_{\text{recipient}}$, σ_{genital}, σ_{clade}, $\sigma_{\text{donorAlpha}}$, $\sigma_{\text{donorBeta}}$, $\sigma_{\text{recipientAlpha}}$, $\sigma_{\text{recipientBeta}}$, ϕ_{donor}, $\phi_{\text{donorAlpha}}$, $\phi_{\text{donorBeta}}$, $\phi_{\text{recipient}}$, ϕ_{genital}, θ_{donor}, $\theta_{\text{donorAlpha}}$, $\theta_{\text{donorBeta}}$, $\theta_{\text{recipient}}$ and θ_{genital} were given a prior of Gamma(1,2) reflecting prior knowledge that the standard deviation in these assays was unlikely to be greater than several logs.

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

Markov Chain Monte Carlo sampling of the posterior probability distributions of the models was implemented in Stan using the `R` package `rstan` and run in 50 chains with each having a 50,000 iteration burnin and 50,000 iterations of sampling every 25$^{\text{th}}$ iteration. Code is available at:
`https://github.com/sherrillmix/hivPair/blob/master/bayesIC50.R`.

Simple comparison

As an example of a comparison of the Bayesian estimates with a simpler analysis, we can look at the estimated change in IFNβ IC_{50} between untreated donor plasma viruses and IFNβ-selected donor plasma viruses (Figure 3). We observed $\log_{10}(IC_{50})$ in both untreated and IFNβ-selection viral isolates for 3 donors with averages:

Donor	Untreated	IFNβ-selected	Difference
CH148	-4.269	-2.831	1.438
CH492	-4.203	-2.717	1.487
CH596	-4.162	-2.820	1.343

The simplest estimate would be to take the average, 1.423, and the standard deviation, 0.0733, of the three differences and estimate the 95% confidence interval on the mean as:

$$1.423 \pm \frac{1.96 \times 0.0733}{\sqrt{3}} = 1.423 \pm 0.0829$$

Or equivalently an estimate that IFNβ-selected donor viruses have an IC_{50} 26.5× (95% confidence interval: 21.9–32.0×) higher than untreated isolates.

From the Bayesian model, we obtained estimates of 20.7× (95% credible interval: 11.0–36.2×) higher IC_{50}. So the Bayesian model is being more conservative with wider intervals in its estimation due to incorporating uncertainty in our estimates of untreated and IFNβ-selected IC_{50} for each donor.