Final Project – **Antibody Response Induced by HIV Vaccines and T-cell Suppression Treatments in Rhesus Macaques** – Second Draft

Kan Luo, Shih-Ni Prim, Frederick Davey, Rizwana Rehman

10/25/2020

Contents

Notes/questions

- The resulting dataset from this section is called `Data2`, and one outlier will be removed in a later section, which results in a final dataset `Data3`. So `Data3` is used for analysis.
- Questions for Kan marked as [Kan. . .] below.
- Comments for future addition and revisions in [. . .]
- Instead of looking at average of Binding, look at average of reactivity to see the percentage of reactive. Perhaps use natural log to transform Binding.
- Who can figure out how to put all figures and tables after all text and before supplmental materials?
- Dig deeper into analysis results?

- Not sure what sections the professor wants

Abstract

Introduction

A dominant vaccine development strategy is to induce neutralizing antibodies by immunizing humans with the virus' glycoproteins. However, HIV vaccines that adopted this strategy mostly failed due to the fact that HIV is an RNA virus, which mutates rapidly to escape the inhibition of neutralizing antibodies. By the time the body generates neutralizing antibodies against the glycoproteins of some HIV strains, the RNA virus has already mutated. Thus, the existing neutralizing antibody fails to recognize, bind with, and neutralize the HIV virus. One possible solution is to increase the number of potential neutralizing antibodies that will cycle in the body by releasing a variety of antibodies after glycoprotein immunization.

Our dataset includes measurements of antibodies measured in 20 rhesus macaques after they were given the same HIV vaccine at three different time points and one of three randomly selected anti-Treg treatments. Blood samples were collected two weeks after vaccine dosing, and antibodies were isolated from those samples. A different number of antibodies were collected from each blood sample, limited by assay yield. Each observation contains information about the antibody isolated post the glycoprotein immunization. In the current report, we test if the different Treg treatments and number of vaccine injections cause changes in the antibody characteristics and if the changes are related to immunization/treatment timepoints.

Specifically, we evaluate:

1. **Do treatments, time points, or isotypes have effects on the mutation frequency or the amino acid count in the third complementarity determining region (CDR3)?** This will be evaluated with a three-way MANOVA, followed by further pair-wise analysis of any values found significantly different.
2. **Do the paired antibody heavy chain and light chain have different mutation rate, CDR3, or binding values?** This will be evaluated with paired comparisons by antibody.
3. **How does the binding strength of the antibodies develop in response to the number of vaccine dosages by treatment?** This will be evaluated with a longitudinal analysis by test subject.

Methodologies

During the experiment, 20 rhesus macaques were given glycoprotein immunization and supplemental antibody doses, as well as one of three treatments (two experimental regulatory T-cell suppression treatments and one control). Regulatory T (Treg) cells prevent autoimmune diseases and suppress allergic reactions by inhibiting adaptive antibody immune response in the germinal center. Theoretically, this adaptive response lowers the effectiveness of vaccines. Thus the experiment used T-cell suppression treatments to investigate the effect on immunization. These drugs are widely used in post transplant immunosuppression treatment to prevent rejection.

For the analysis of mutation frequency and CDR3 count, each antibody within the same treatment is treated as an independent observation. While we might expect different variance within subject vs between subjects, the number of potential antibodies observed is much higher than the number sampled in a blood draw. A human antibody is formed by a heavy chain and light chain. For heavy chain, human has about 51 V-gene segments, 25 D-gene segments and 6 J-gene segment. For light chain (kappa and lambda), there are 71 V-gene segments and 9 J-gene segments[ref.5]. Any heavy chain V-D-J combination and light chain V-J combinations can randomly happen in germline center. Theoretically, there can be $51 * 25 * 6 * 71 * 9 = 4.88835 \times 10^6$ combinations of gene segments. Considering the frequently happened mutation, insertion, class switching, each individual can have over **10 billion** different antibodies. Thus, we decided to follow the convention of vaccine studies and treat each antibody as independent. [Kan – can you identify a reference article or journal

here that uses this convention? We don't need to quote it / change the answer, it's just defending our claim that it's a standard practice]

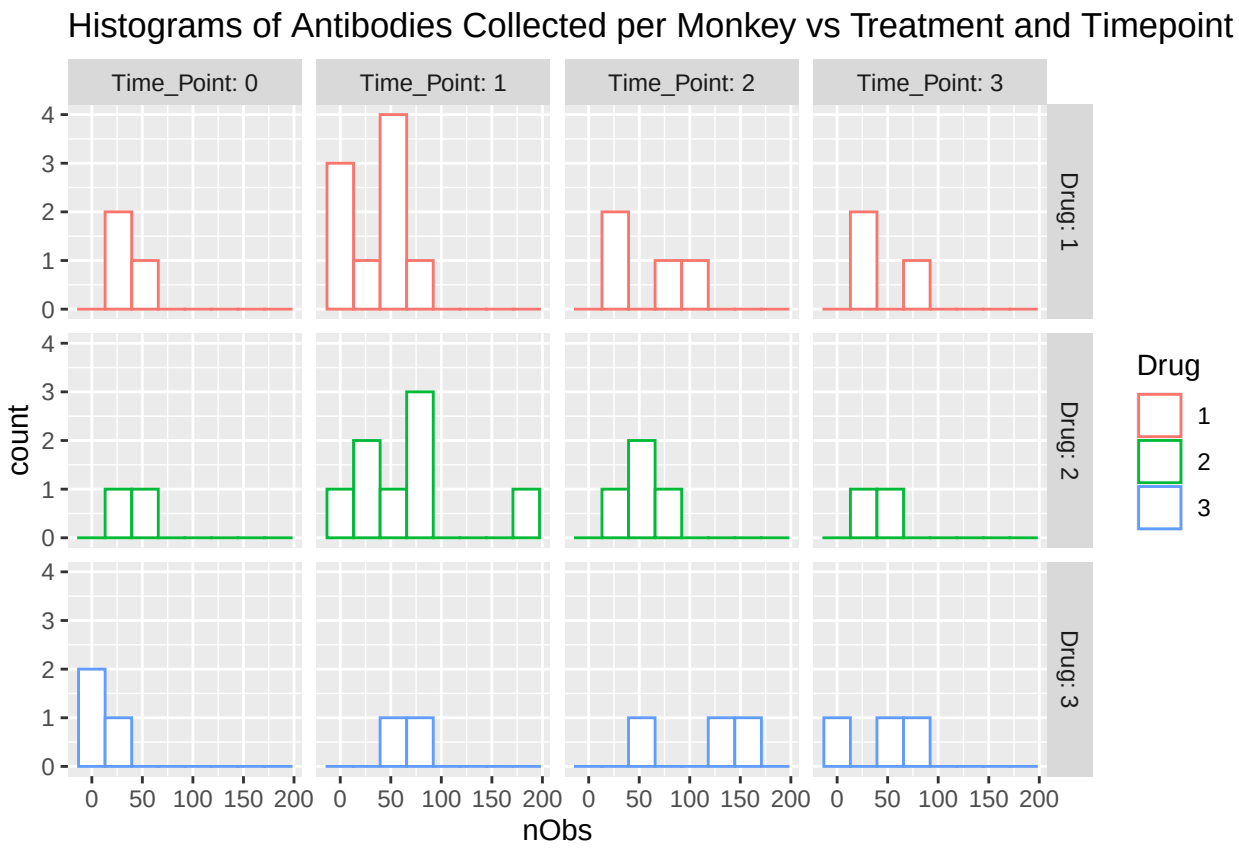
To evaluate the mutation frequency and CDR count vs Treatments, Timepoints, and Isotypes MANOVA, we will be running the [xxx] function from the [xxx] package [package citation]

For the longitudinal analysis of binding strength vs number of vaccine doses, we will be using the gls function from the nlme package[7].

Data Summaries

[Feel free to condense any tables or figures or make them look better.]

The dataset has 2465 antibodies collected from 20 rhesus monkeys. We first present our exploratory data analysis and summaries. First an overview of the number of antibodies collected per treatment at each timepoint



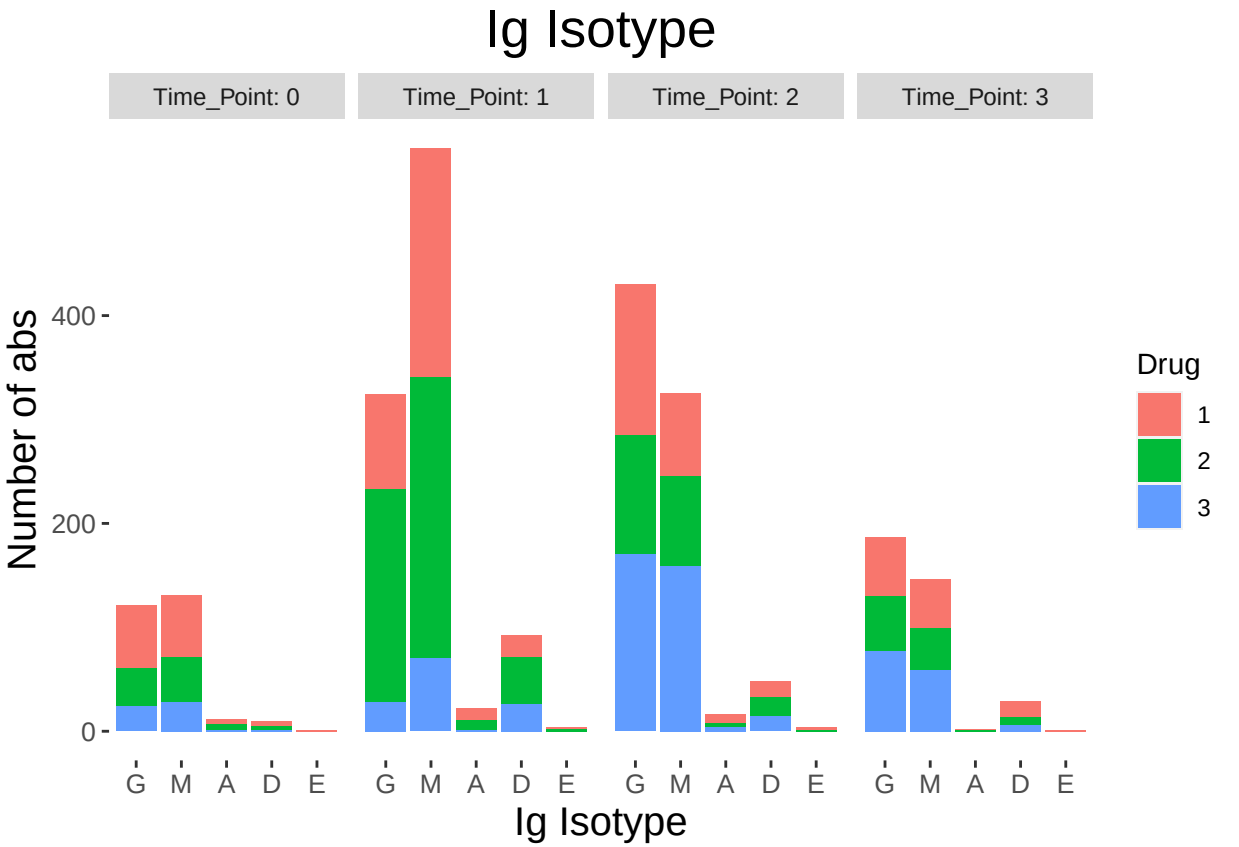

```
##     TimePoint
## Drug   0   1   2   3
##    1 129 346 251 122
##    2  90 533 225 101
##    3  54 125 347 142
```

```
## , , Isotype = A
##
##     TimePoint
## Drug   0   1   2   3
##    1   4  11   8   1
```

```
##    2   6  10   4   1
##    3   1   1   4   0
##
## , , Isotype = D
##
##     TimePoint
## Drug   0   1   2   3
##    1   4  22  15  16
##    2   4  45  19   7
##    3   1  26  14   6
##
## , , Isotype = E
##
##     TimePoint
## Drug   0   1   2   3
##    1   1   2   3   1
##    2   0   2   1   0
##    3   0   0   0   0
##
## , , Isotype = G
##
##     TimePoint
## Drug   0   1   2   3
##    1  60  91 145  57
##    2  37 205 115  53
##    3  24  28 170  77
##
## , , Isotype = M
##
##     TimePoint
## Drug   0   1   2   3
##    1  60 220  80  47
##    2  43 271  86  40
##    3  28  70 159  59
```

There are four time points; time zero was collected before any procedure was done. Times 1,2, and 3 were collected two weeks after an initial and two booster vaccine shots were administered to the macaques. In the treatment groups, groups 1-3 represent different doses of anti T-Reg drug 1, groups 4-6 represent different doses of anti-Treg drug 2, and group 7 represents the control group [We have 20 monkeys, do we need to mention groups at all? Can we just stick to drugs 1,2, and control group 3?].

Next, we'll take a look at the variable `Isotype`. There are 5 kinds of immunoglobulin isotypes: IgG, IgA, IgM, IgE, IgD[ref.6]. The two most important kinds are IgG and IgM. IgM occurs in the acute stage of infection and perform an role of primary response. The secondary response IgG appears later in serum with higher binding affinity, and neutralizing potentials against toxins and virus. IgA mostly found in mucosal tissues such as Nasal mucosa. Non-dominant IgD and IgE are typically lower than 1% in blood.



As expected, we see that IgG and IgM occupy the biggest proportion of all antibodies in all time points. Before immunization (time point 0), there are similar weight of IgG and IgM found in blood. After the 1st immunization (time point 1), primary immune response results an increase of IgM, followed with IgG increase at later time point 2 and 3. We'll use the variable `Isotype` as a grouping covariate later.

For the response variables, we will begin with an outlier check, then review their distributions across treatment and timepoints.

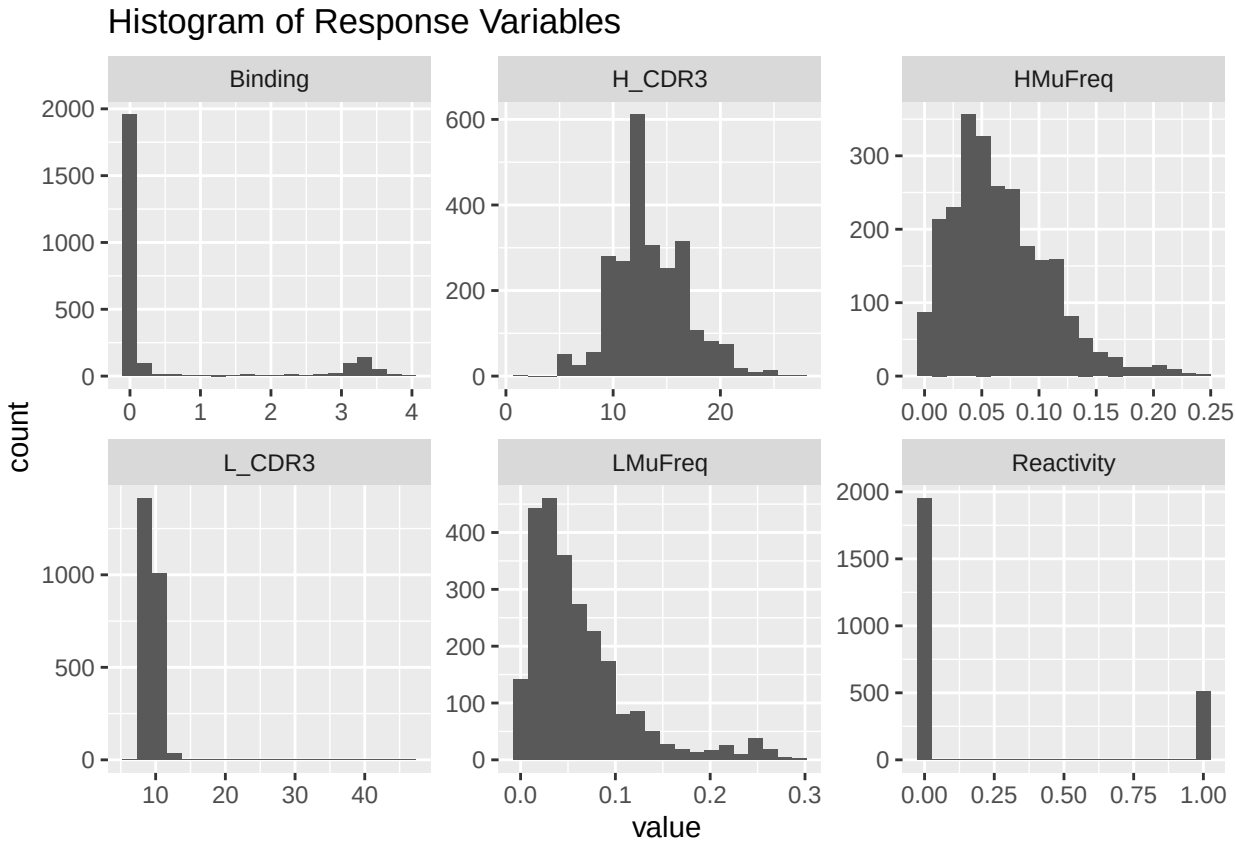


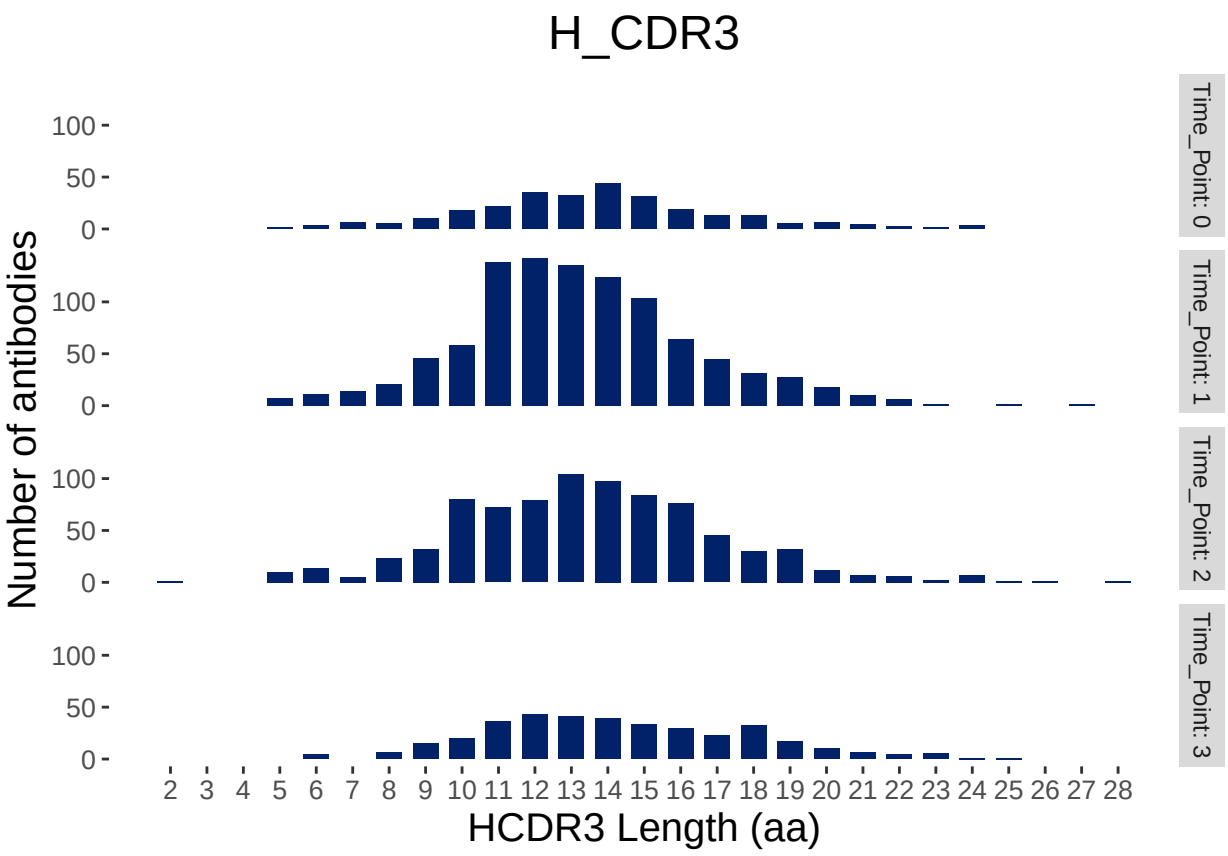
Table 1: Normal Standardized L_CDR3 Statistics

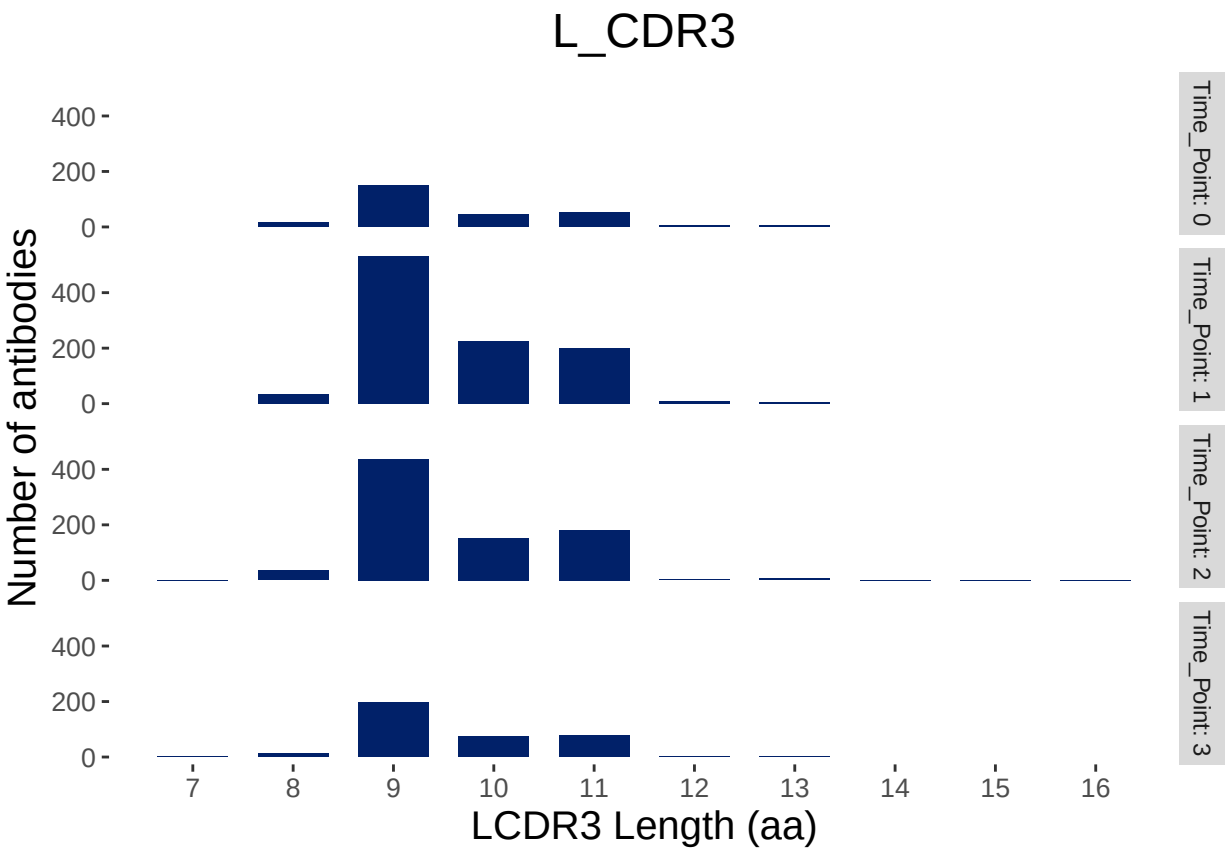
V1	Min. :-2.1860	1st Qu.:-0.5361	Median :-0.5361	Mean : 0.0000	3rd Qu.: 0.2888	Max. :30.8110

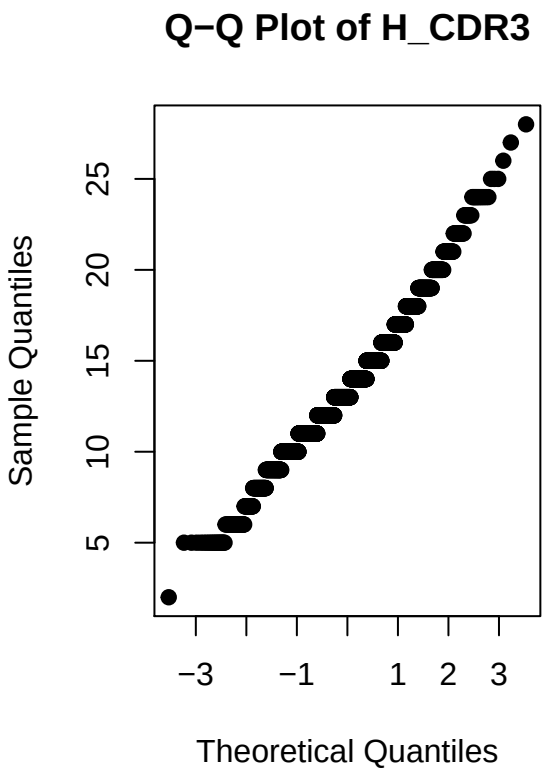
L_CDR3 appears to have an outlier several standard. This is most likely caused by a laboratory error, so we will exclude it and continue the analysis with the remaining data.

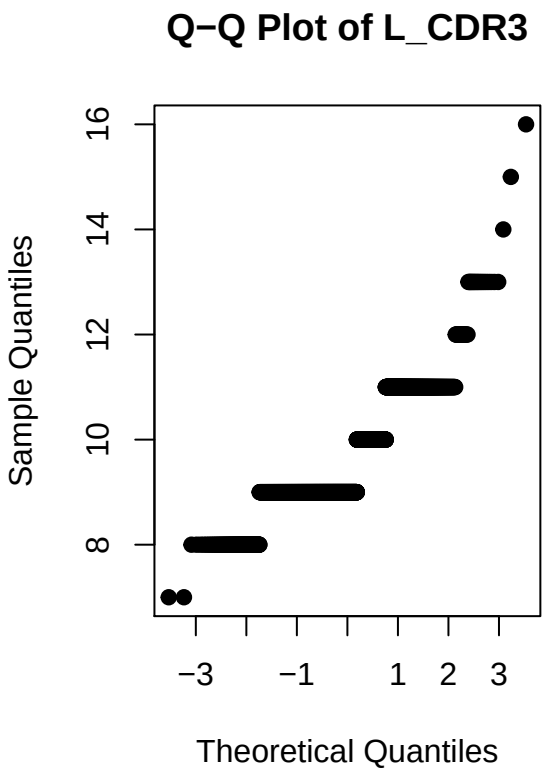
Next we'll examine our responses: `H_CDR3`, `HMuFreq`, `L_CDR3`, `LMuFreq`, `Binding`, and `Reactivity`. In each antibody, there are two sets of heavy chain and light chain, all of which forming a Y-shape immunoglobulin. Thus many of the variables start with H or L, indicating which chain the information comes from.

`H_CDR3` and `L_CDR3` indicates the length of the third complementarity-determining region on the variable heavy chain and light chain. The longer they are, the more potential there is to produce diverse antibodies. [Kan, could you check to see if this is correct?] In other words, we want the values to be higher. Below we see that the distributions are roughly normal with the center around 13 for `H_CDR3`, with all data points, and slightly centers for different time points.

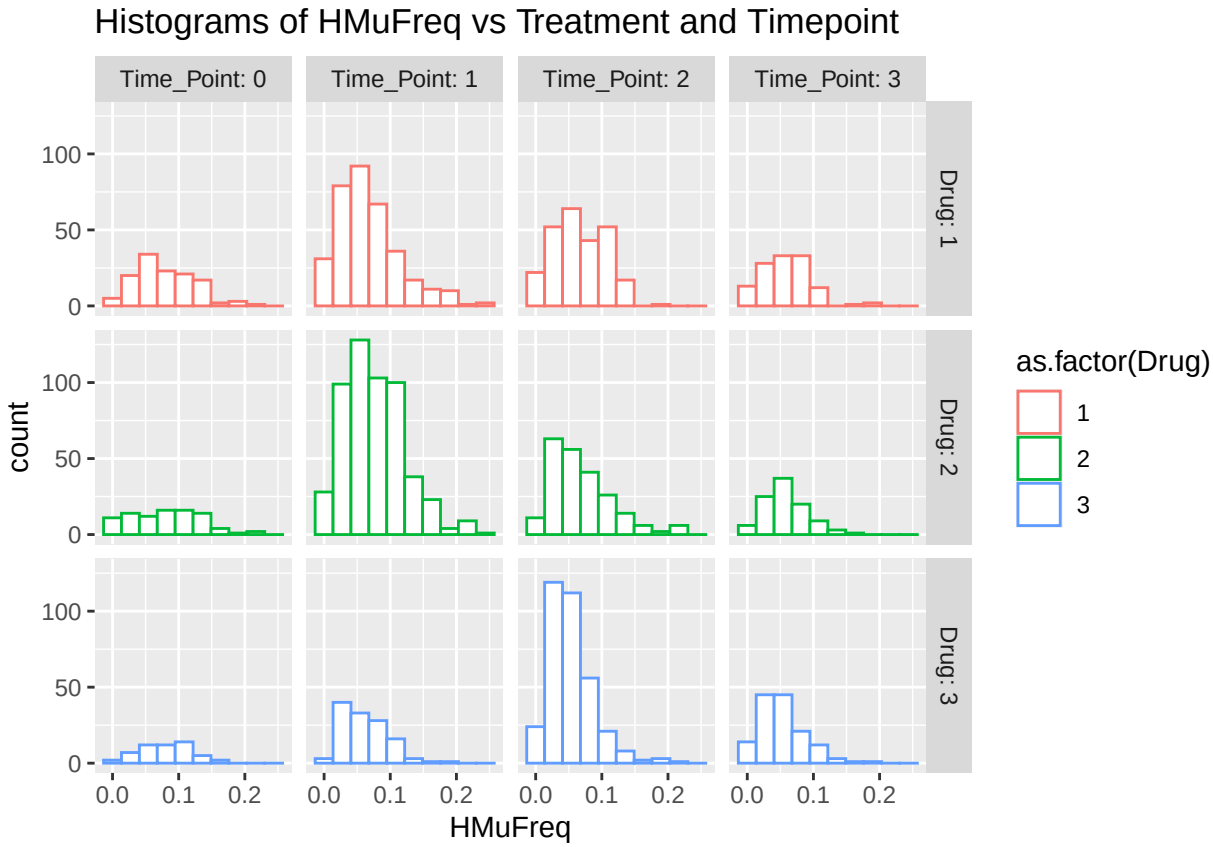

H_CDR3
Number of antibodies
100
50
0
HCDR3 Length (aa)
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Time_Point: 0
Time_Point: 1
Time_Point: 2
Time_Point: 3

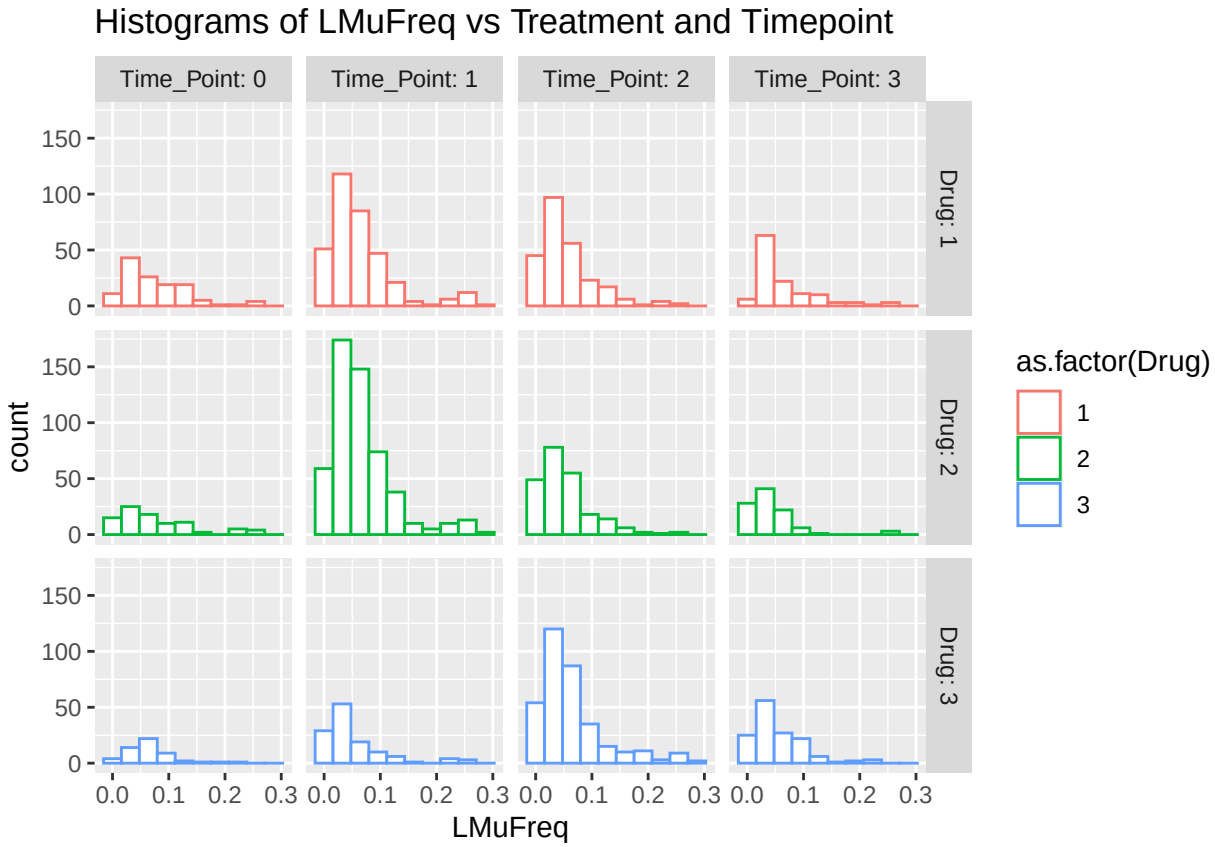

L_CDR3
Number of antibodies
400
200
0
LCDR3 Length (aa)
7
8
9
10
11
12
13
14
15
16
Time_Point: 0
Time_Point: 1
Time_Point: 2
Time_Point: 3

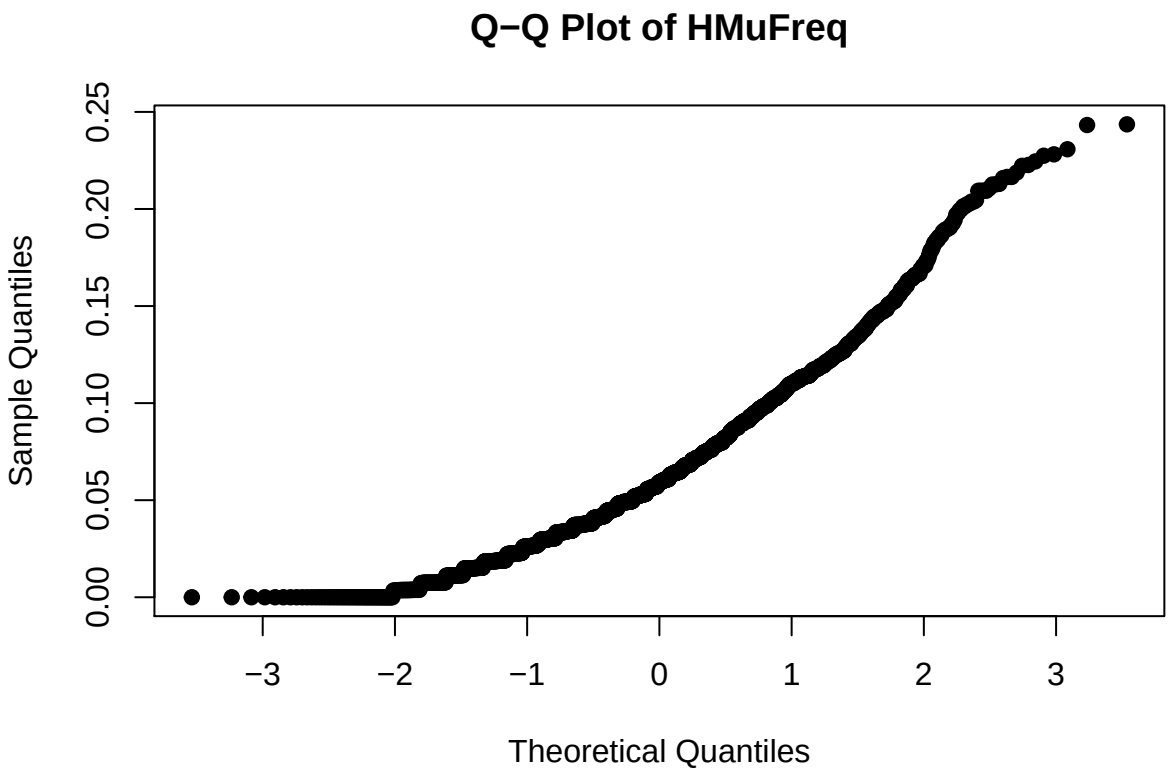


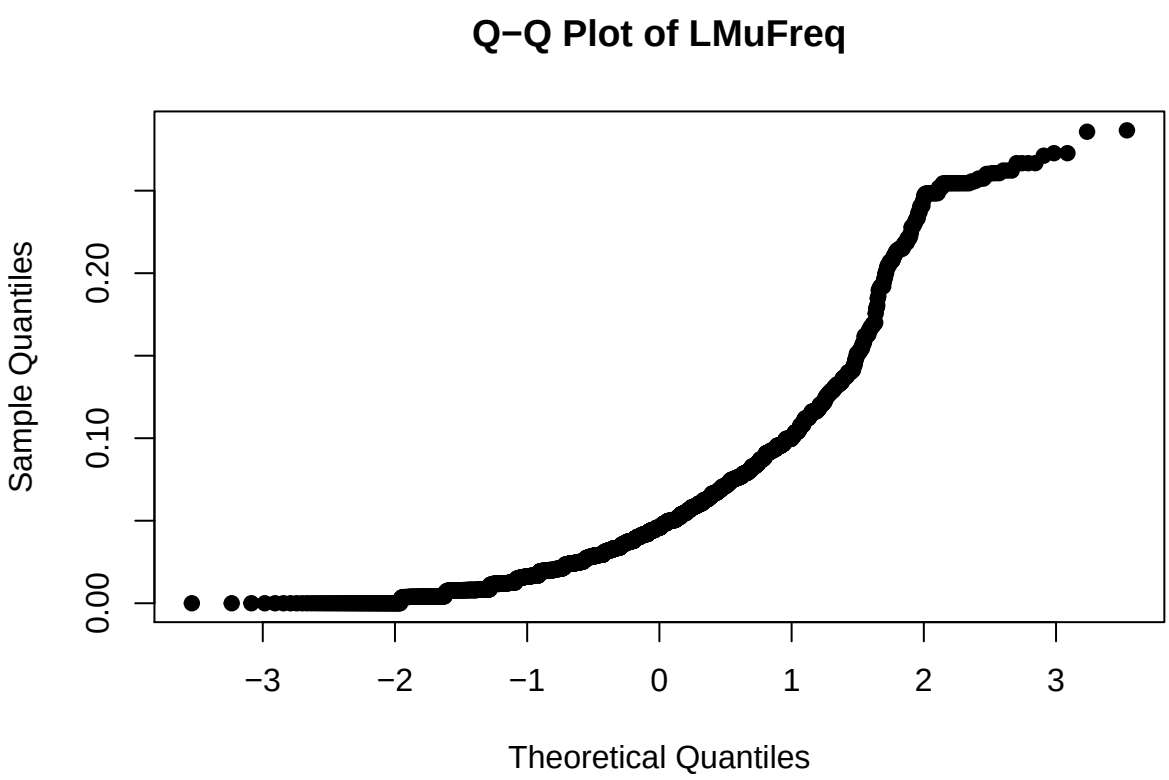


`HMuFreq` and `LMuFreq` are calculated by dividing `H_Substitution` by `H_VBase` for heavy chain and similarly for light chain. These two variables show how much the antibodies mutate. A higher mutation rate is usually indicative of better virus neutralization. Below we see some comparison of mutation rate between heavy chain and light chain. (Kappa and Lambda are two kinds of light chain.) [Kan, is there a reason to split up light chain into Kappa and Lambda? Could we simply plot heavy chain vs. light chain?]

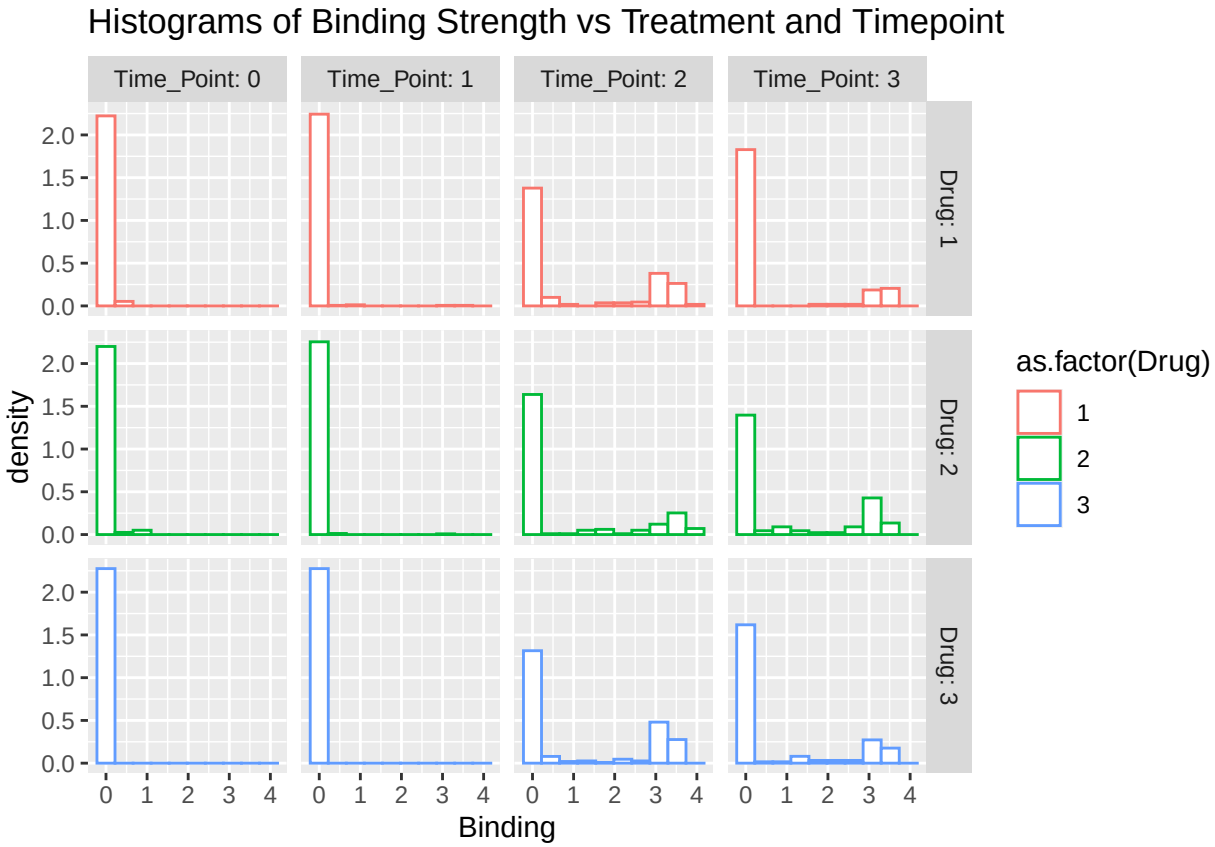

Histograms of HMuFreq vs Treatment and Timepoint
Time_Point: 0
Time_Point: 1
Time_Point: 2
Time_Point: 3
Drug: 1
Drug: 2
Drug: 3
count
100
50
0
0.0
0.1
0.2
HMuFreq
as.factor(Drug)
1
2
3

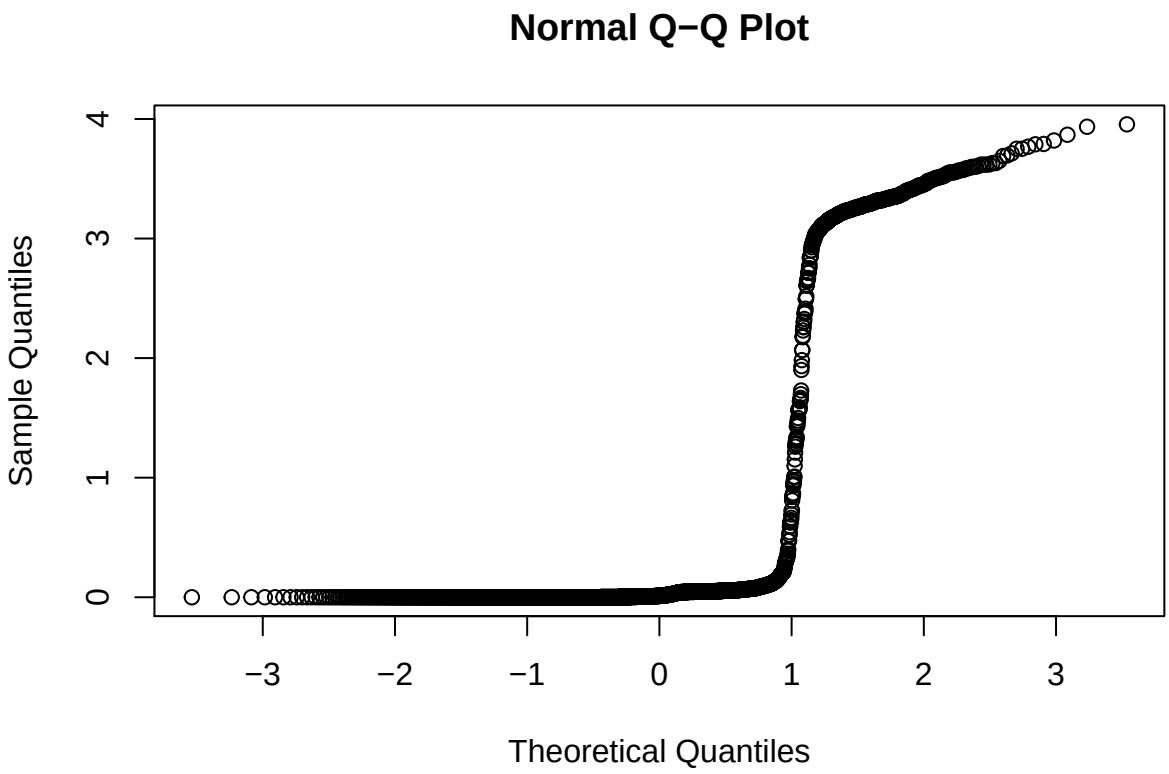

Histograms of LMuFreq vs Treatment and Timepoint
Time_Point: 0
Time_Point: 1
Time_Point: 2
Time_Point: 3
Drug: 1
Drug: 2
Drug: 3
count
LMuFreq
as.factor(Drug)
1
2
3


Q−Q Plot of HMuFreq
Sample Quantiles
0.00
0.05
0.10
0.15
0.20
0.25
−3
−2
−1
0
1
2
3
Theoretical Quantiles

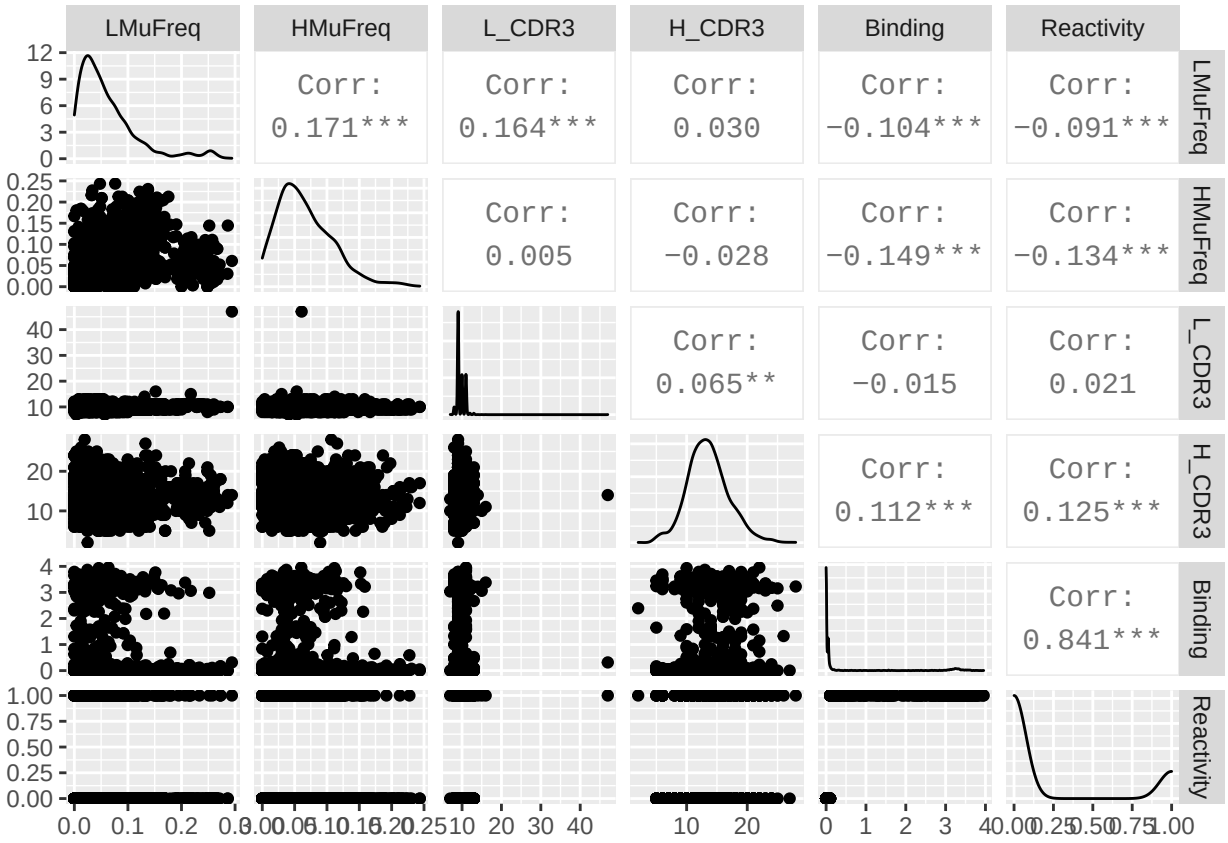


Lastly, `Binding` indicates the rate of neutralizing, meaning how much the antibodies bind with the virus and thus make the virus ineffective. This is the most important measure of the study. `Reactivity` turns `Binding` into a binary variable; `Binding` rate above 0.1 is considered reactive. In the Q-Q plot of `Binding`, we can see that it is not normally distributed. However, since our sample size is larger than 2000, we can use the Central Limit Theorem and assume normality.

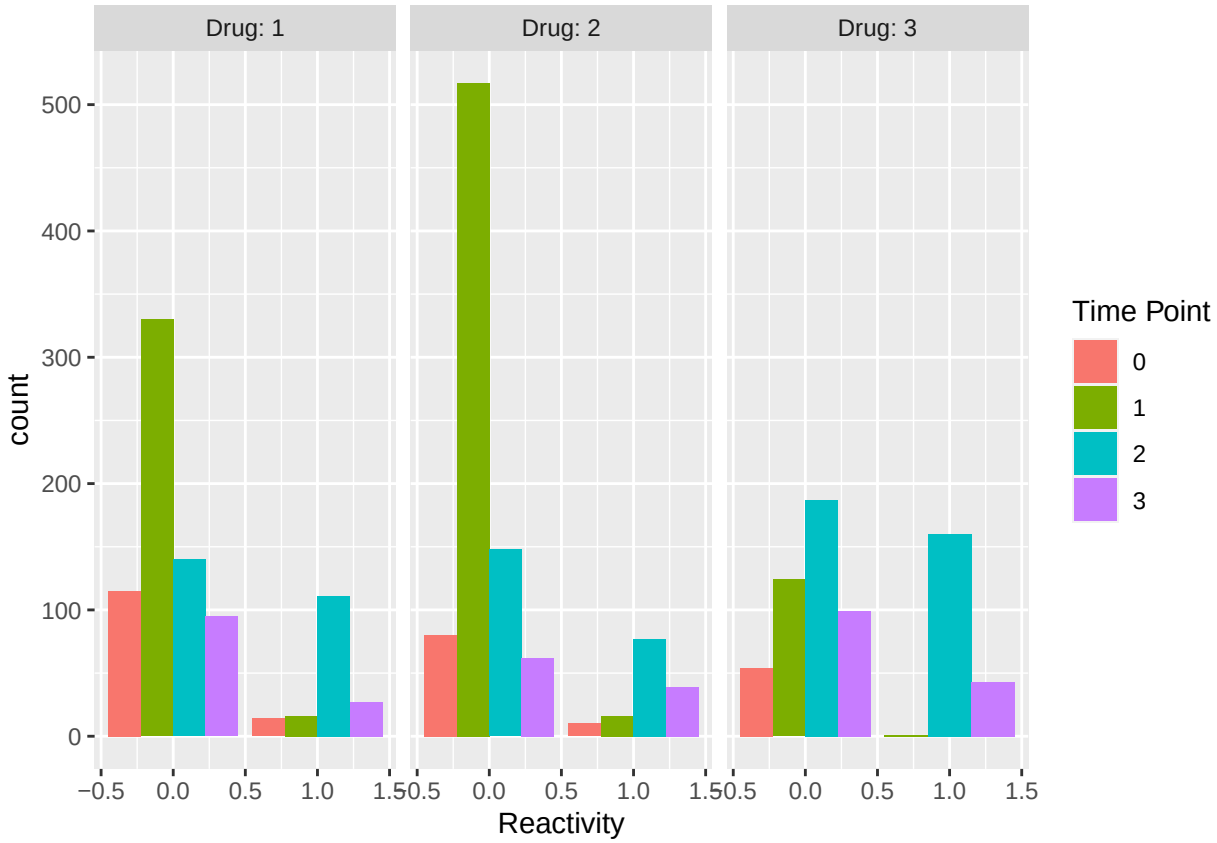

Histograms of Binding Strength vs Treatment and Timepoint
Time_Point: 0
Time_Point: 1
Time_Point: 2
Time_Point: 3
Drug: 1
Drug: 2
Drug: 3
density
Binding
as.factor(Drug)
1
2
3

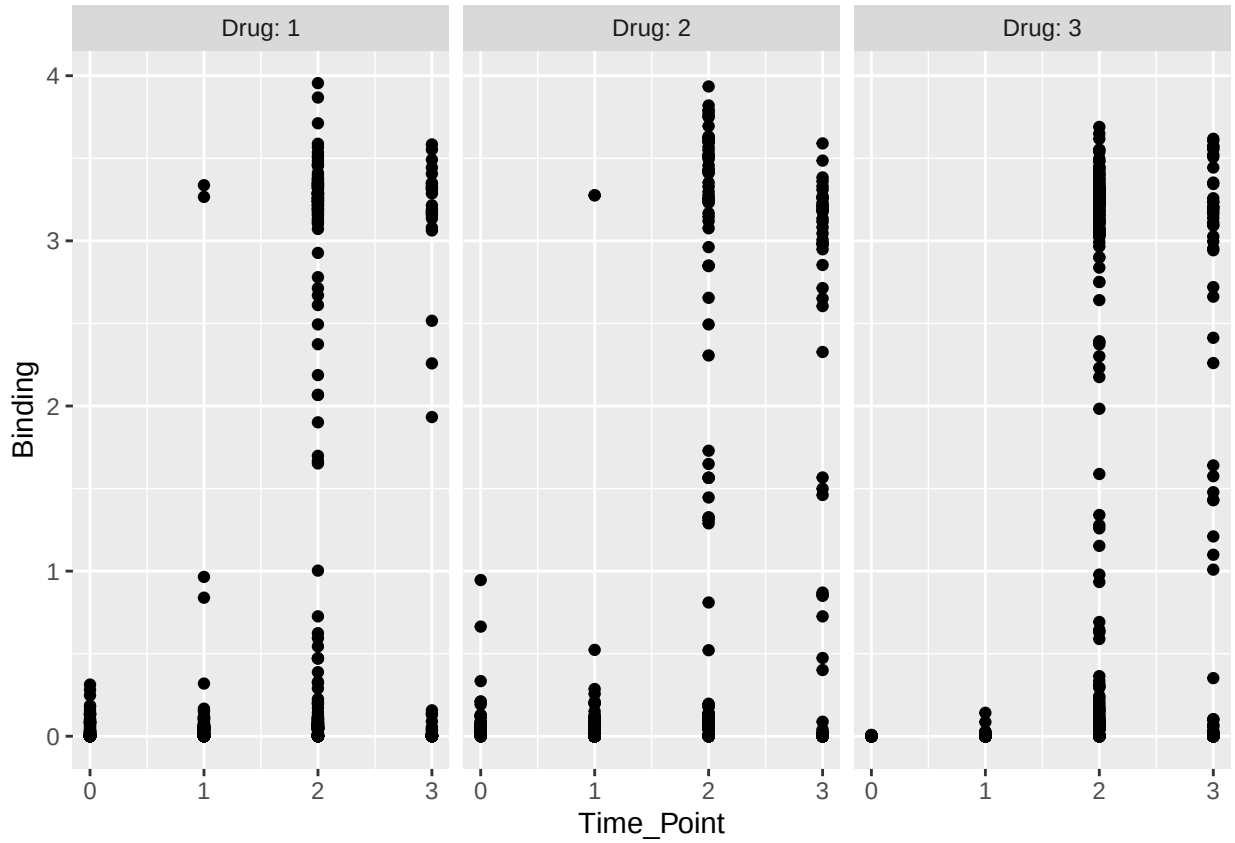


Let's take a look at these response variables and check whether they might be correlated. In the plot below, we can see that none of the response variables are highly correlated, except for `Binding` and `Reactivity`, which is expected, because `Reactivity` is a binary variable derived from `Binding`. We will only choose one of these two variables in each analysis based on the type of analysis. [What else to point out here?]



* [We should trim this, move to the data analysis section if desired] Now we use some plots to see whether the response variables might be different for different time points, treatment groups or drugs, and grouping covariate (`Isotype`).


Drug: 1
Drug: 2
Drug: 3
count
500
400
300
200
100
0
−0.5
0.0
0.5
1.0
1.5
Reactivity
Time Point
0
1
2
3

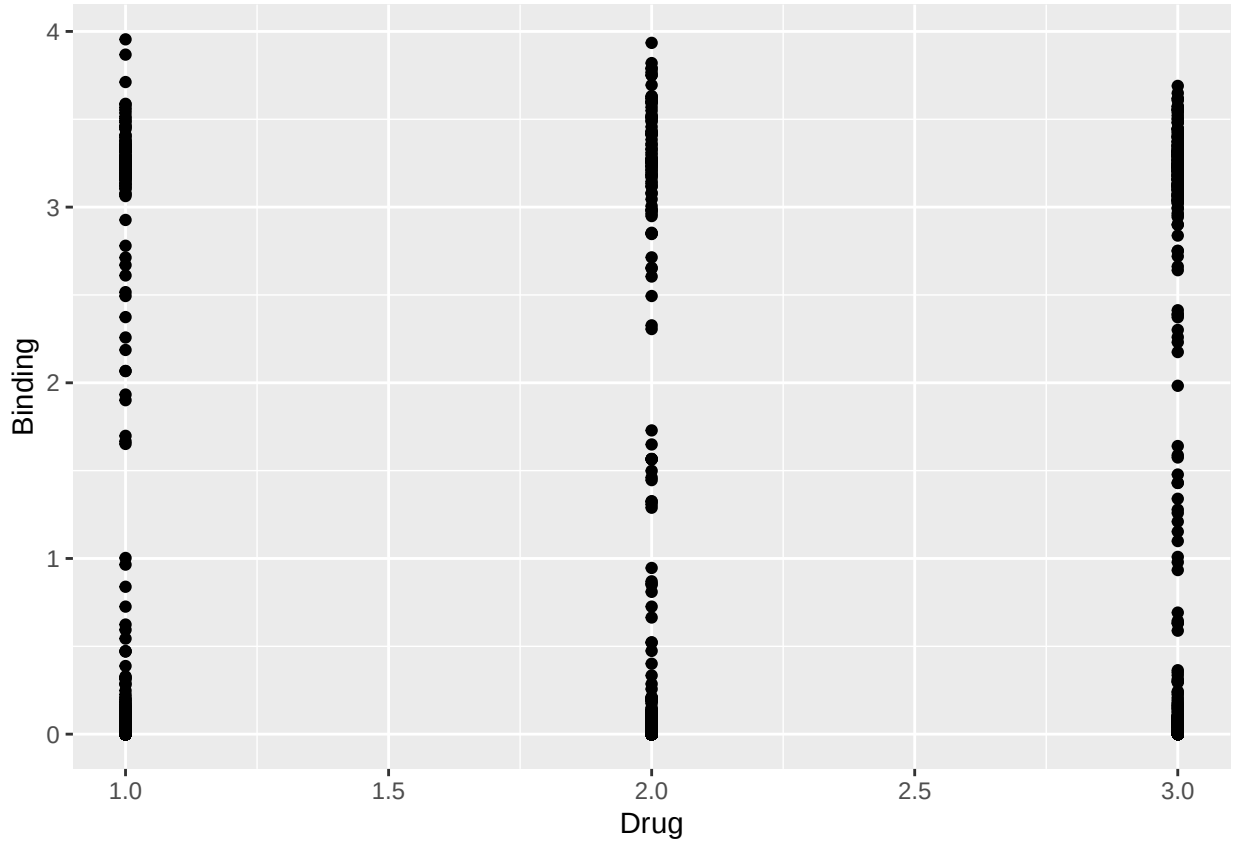



```
##
##       0   1
##   1 680 168
##   2 807 142
##   3 464 204
```

```
##
##       0   1
##   0 249  24
##   1 971  33
##   2 475 348
##   3 256 109
```

`Binding` or `Reactivity` do seem to be affected by various predictors, and the boxplots for `Binding` do appear quite different.

```
## # A tibble: 3 x 6
##    Drug avgLMuFreq avgHMuFreq avgBinding varBinding avgReact
##   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>    <dbl>
## 1     1     0.0616    NA          0.450      1.14     0.198
## 2     2     0.0616     0.0730     0.334      0.864    0.150
## 3     3     0.0594     0.0559     0.807      1.81     0.305
```



Data Analysis

Multivariate Data Analysis

Now we want to test whether predictors `Drug` and `Isotype` have effects on the five responses: `H_CDR3`, `HMuFreq`, `L_CDR3`, `LMuFreq`, and `Binding`. We choose `Binding` here, because all the variables are continuous.

[I don't think this is an accurate sample size. Some of the isotypes only have 1-10 samples in some of the time points. I think we should trim to IgG and IgM, can possibly include IgD]

First, we use manova to test effects. Since we have a large sample size (n = 2464), we can assume normality. In the output, we can see that the main effects of `Drug` and `Isotype` and the interaction effects are all significant.

#[How are we managing controls? / Controlling for time 0 data?]

```
##             Df   Pillai approx F num Df den Df    Pr(>F)
## drug         2 0.058434   14.711     10   4888 < 2.2e-16 ***
## it           4 0.300048   39.672     20   9784 < 2.2e-16 ***
## drug:it      7 0.069958    4.960     35  12235 < 2.2e-16 ***
## Residuals 2447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pairwise comparison

To know more details about which groups have different means, we use pairwise comparisons for each treatment group, drug, and Isotype. (We set α as 0.05, and use a Bonferroni correction.)

```
## [1] "L_CDR3  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  1 - 2      0.0451 0.0450 2458  -0.0870    0.177
##  1 - 3      0.1018 0.0492 2458  -0.0429    0.246
##  2 - 3      0.0567 0.0480 2458  -0.0844    0.198
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "LMuFreq  pairwise CI's"
##  contrast  estimate      SE   df lower.CL upper.CL
##  1 - 2    -0.000176 0.00256 2458 -0.00770  0.00735
##  1 - 3     0.002320 0.00280 2458 -0.00592  0.01056
##  2 - 3     0.002496 0.00273 2458 -0.00554  0.01053
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "H_CDR3  pairwise CI's"
##  contrast estimate    SE   df lower.CL upper.CL
##  1 - 2      0.4907 0.160 2458   0.0204    0.961
##  1 - 3      0.4438 0.175 2458  -0.0712    0.959
##  2 - 3     -0.0469 0.171 2458  -0.5492    0.456
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "HMuFreq  pairwise CI's"
##  contrast estimate      SE   df lower.CL  upper.CL
##  1 - 2      -0.006 0.00199 2458 -0.01186 -0.000145
##  1 - 3       0.011 0.00218 2458  0.00461  0.017432
##  2 - 3       0.017 0.00213 2458  0.01076  0.023276
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "Binding  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  1 - 2       0.118 0.0522 2458  -0.0355    0.271
##  1 - 3      -0.356 0.0571 2458  -0.5235   -0.188
##  2 - 3      -0.473 0.0557 2458  -0.6372   -0.310
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
```

```
## [1] "L_CDR3  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  A - D      0.1136 0.1510 2456  -0.3838    0.611
##  A - E      0.0667 0.3289 2456  -1.0170    1.150
```

```
##  A - G      0.0041 0.1364 2456  -0.4451    0.453
##  A - M      0.0441 0.1361 2456  -0.4042    0.492
##  D - E     -0.0469 0.3091 2456  -1.0651    0.971
##  D - G     -0.1095 0.0769 2456  -0.3627    0.144
##  D - M     -0.0695 0.0764 2456  -0.3210    0.182
##  E - G     -0.0626 0.3022 2456  -1.0581    0.933
##  E - M     -0.0225 0.3021 2456  -1.0177    0.973
##  G - M      0.0400 0.0404 2456  -0.0931    0.173
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "LMuFreq  pairwise CI's"
##  contrast  estimate      SE   df lower.CL upper.CL
##  A - D    -0.021206 0.00858 2456 -0.04947  0.00705
##  A - E    -0.006829 0.01869 2456 -0.06840  0.05475
##  A - G    -0.023476 0.00775 2456 -0.04900  0.00205
##  A - M    -0.021669 0.00773 2456 -0.04714  0.00380
##  D - E     0.014376 0.01756 2456 -0.04348  0.07223
##  D - G    -0.002270 0.00437 2456 -0.01666  0.01212
##  D - M    -0.000463 0.00434 2456 -0.01476  0.01383
##  E - G    -0.016647 0.01717 2456 -0.07321  0.03992
##  E - M    -0.014839 0.01716 2456 -0.07138  0.04170
##  G - M     0.001807 0.00230 2456 -0.00576  0.00937
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "H_CDR3  pairwise CI's"
##  contrast estimate    SE   df lower.CL upper.CL
##  A - D       0.297 0.530 2456   -1.448    2.043
##  A - E      -2.229 1.154 2456   -6.032    1.574
##  A - G      -0.728 0.479 2456   -2.304    0.849
##  A - M       0.503 0.478 2456   -1.070    2.076
##  D - E      -2.527 1.085 2456   -6.100    1.046
##  D - G      -1.025 0.270 2456   -1.914   -0.137
##  D - M       0.206 0.268 2456   -0.677    1.089
##  E - G       1.502 1.060 2456   -1.992    4.995
##  E - M       2.733 1.060 2456   -0.760    6.225
##  G - M       1.231 0.142 2456    0.764    1.698
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "HMuFreq  pairwise CI's"
##  contrast estimate      SE   df lower.CL upper.CL
##  A - D     0.00523 0.00659 2456  -0.0165  0.02693
##  A - E     0.00848 0.01435 2456  -0.0388  0.05576
##  A - G    -0.01628 0.00595 2456  -0.0359  0.00332
##  A - M     0.00393 0.00594 2456  -0.0156  0.02349
##  D - E     0.00325 0.01349 2456  -0.0412  0.04768
##  D - G    -0.02151 0.00335 2456  -0.0326 -0.01046
##  D - M    -0.00130 0.00333 2456  -0.0123  0.00968
##  E - G    -0.02476 0.01319 2456  -0.0682  0.01868
```

```
##  E - M    -0.00455 0.01318 2456  -0.0480  0.03888
##  G - M     0.02021 0.00176 2456   0.0144  0.02602
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "Binding  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  A - D     0.30395 0.1610 2456   -0.227    0.834
##  A - E     0.20847 0.3508 2456   -0.947    1.364
##  A - G    -0.66527 0.1454 2456   -1.144   -0.186
##  A - M     0.30220 0.1451 2456   -0.176    0.780
##  D - E    -0.09548 0.3296 2456   -1.181    0.990
##  D - G    -0.96922 0.0820 2456   -1.239   -0.699
##  D - M    -0.00175 0.0814 2456   -0.270    0.267
##  E - G    -0.87374 0.3223 2456   -1.936    0.188
##  E - M     0.09373 0.3222 2456   -0.968    1.155
##  G - M     0.96747 0.0431 2456    0.825    1.109
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
```

[We should have code build these tables, why / where are we fitting for groups? This muddies the conclusions]

Here are the pairs that have significant differences:

- Treatment
  - L_CDR3: none
  - LMuFreq: none
  - H_CDR3:
    - group 1 > group 4
  - HMuFreq:
    - group 1 > group 7
    - group 3 < group 5
    - group 3 < group 6
    - group 5 > group 7
    - group 6 > group 7
  - Binding:
    - group 1 < group 2
    - group 1 > group 6
    - group 2 > group 3
    - group 2 > group 4
    - group 2 > group 5
    - group 2 > group 6
    - group 3 < group 7
    - group 4 > group 6
    - group 4 < group 7
    - group 5 < group 7
    - group 6 < group 7
- Drug
  - L_CDR3: none

  - LMuFreq: none
  - H_CDR3:
    * 1 > 2
  - HMuFreq:
    * 1 < 2
    * 1 > 3
    * 2 > 3
  - Binding:
    * 1 < 3
    * 2 < 3
- Isotype
  - L_CDR3: none
  - LMuFreq: none
  - H_CDR3:
    * IgD < IgG
  - HMuFreq:
    * IgD < IgG
    * IgG > IgM
  - Binding:
    * IgA < IgG
    * IgD < IgG
    * IgG > IgM

In short, `L_CDR3` and `LMuFreq` do not have significnat paired differences.

For `H_CDR3`, treatment group 1 (drug 1) is higher than treatment group 7 (control), and IgG has a longer H_CDR3 length than IgD.

For `HMuFreq`, drug 2 has the highest mean, followed by drug 1 and control. More specifically, treatment groups 5 and 6 (two doses in drug 2) have the highest mutation rates. IgG has higher mutation rate than IgD.

For `Binding`, drug 3 has the highest mean, but drug 1 and drug 2 do not have significant differences. IgG has higher binding rate than IgA, IgD, and IgM.

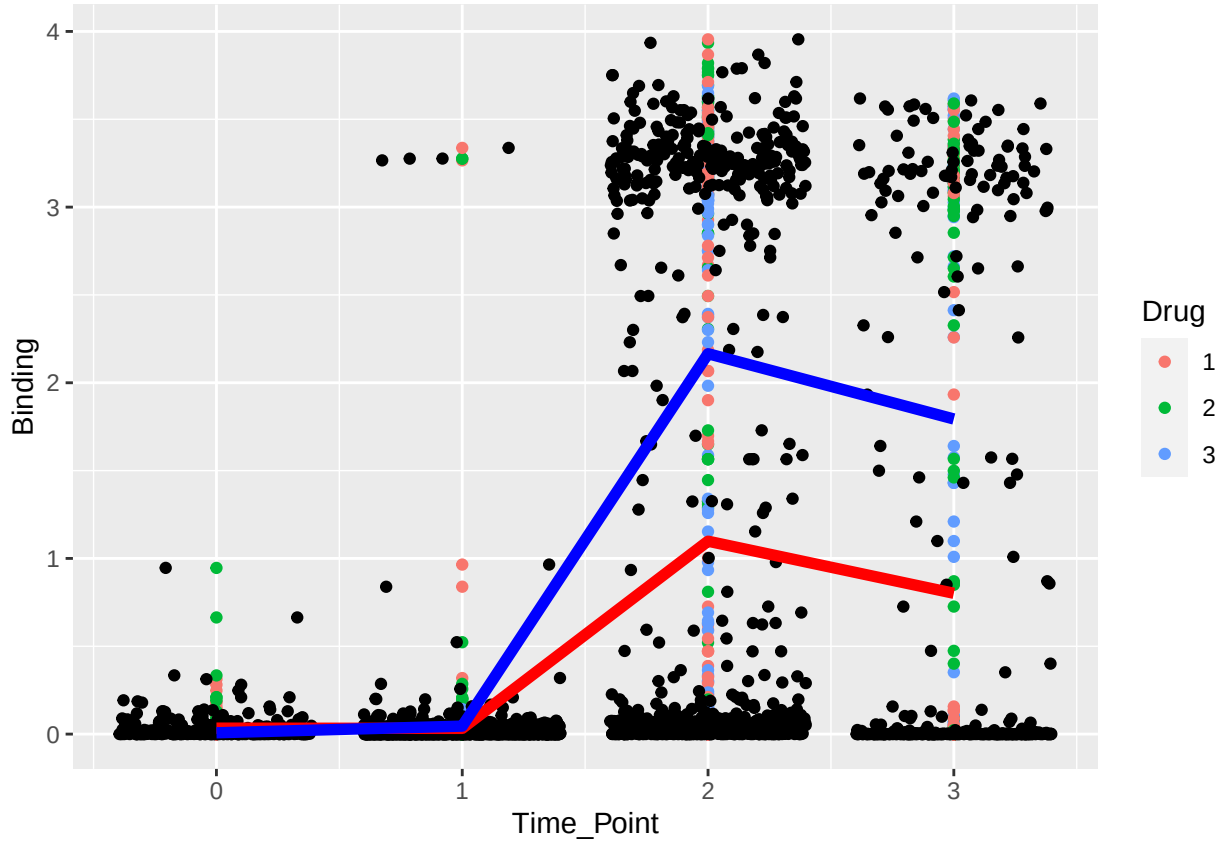
We could conclude that the drugs/treatment groups do increase mutation rate; however, they do not increase binding rate. That is to say, although the treatments do help increase the diversity of antibodies, they are not specific to the HIV antigens and thus do not increase binding.

Longitudinal Data Analysis

[This section probably needs some more revisions, since we still have three more lectures.]

First we don't consider treatments but only plot the mean trend over time. The plot shows that binding does vary over time. The red line shows the mean trend over time, and the blue line shows the variance over time. The variance does not seem equal over time, so we use unequal variance over time for the covariance structure.

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + e_{ij}$$



Here we use `Binding` as the response, `Time_Point` as the time factor, and `Drug` as the covariates. Random effect for both intercept and slope. Now we want to add one covariate: `Drug`. We use two indicator variables: `D1` and `D2`, where

$$D1 = \begin{cases} 1 & \text{if Drug} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$D2 = \begin{cases} 1 & \text{if Drug} = 2 \\ 0 & \text{otherwise} \end{cases}$$

Assuming that the random effects are the same for each drug, our full model is:

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + D1_i(\beta_2 + \beta_3 Time_{ij}) + D2_i(\beta_4 + \beta_5 Time_{ij}) + b_{0i} + b_{1i} Time_{ij} + e_{ij}$$

$$\mathbf{b_i} \sim N\left(0, \mathbf{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix}\right)$$

Drug 1: $Y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_2 + \beta_3 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$
Drug 2: $Y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_4 + \beta_5 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$
Drug 3: $Y_{ij} = \beta_0 + \beta_1 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$

```
## Linear mixed-effects model fit by REML
##  Data: dataLDA
##        AIC      BIC    logLik
```

```
##   3673.651 3749.144 -1823.826
##
## Random effects:
##  Formula: ~Time_Point | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.6893734 (Intr)
## Time_Point  0.6524155 -0.999
## Residual    0.2169511
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | Time_Point
##  Parameter estimates:
##         1         0         2         3
## 1.0000000 0.3827704 7.1563614 6.5754859
## Fixed effects: binding ~ Time_Point + D1 + D1:Time_Point + D2 + D2:Time_Point
##                    Value Std.Error   DF    t-value p-value
## (Intercept)   -0.0432994 0.3982221 2441 -0.1087318  0.9134
## Time_Point     0.1772970 0.3772249 2441  0.4700034  0.6384
## D1            -0.3162408 0.5043030   17 -0.6270850  0.5389
## D2            -0.8725346 0.5123761   17 -1.7029181  0.1068
## Time_Point:D1  0.2407466 0.4811398 2441  0.5003672  0.6169
## Time_Point:D2  0.7867662 0.4891206 2441  1.6085323  0.1078
##  Correlation:
##               (Intr) Tm_Pnt D1     D2     T_P:D1
## Time_Point    -0.998
## D1            -0.790  0.788
## D2            -0.777  0.775  0.614
## Time_Point:D1  0.782 -0.784 -0.998 -0.608
## Time_Point:D2  0.769 -0.771 -0.608 -0.998  0.605
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.18636862 -0.29361355 -0.10681754  0.02325697 15.04537744
##
## Number of Observations: 2464
## Number of Groups: 20
```

The p-values for `Drug` and the interaction of `Drug` and `Time_Point` are large. So we try another model with `Time_Point` as the only predictor. [This is skipping the part where we fit only main effect (not interaction) with Drug]

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$$

$$\underbrace{\begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{im_i} \end{bmatrix}}_{\mathbf{Y_i}} = \underbrace{\begin{bmatrix} 1 & Time_{i1} \\ \vdots & \vdots & \vdots \\ 1 & Time_{im_i} \end{bmatrix}}_{\mathbf{X_i}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} 1 & Time_{i1} \\ \vdots & \vdots \\ 1 & Time_{im_i} \end{bmatrix}}_{\mathbf{Z_i}} \underbrace{\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}}_{\mathbf{b_i}} + \underbrace{\begin{bmatrix} e_{i1} \\ \vdots \\ e_{im_i} \end{bmatrix}}_{\mathbf{e_i}}$$

$$\mathbf{b_i} \sim N\left(0, \mathbf{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix}\right)$$

$$\mathbf{e}_{ij} \sim N(0, \mathbf{R}_i = \sigma^2 I_{mi})$$

[need to consider whether time point 2 is the optimal point]

```
## Linear mixed-effects model fit by REML
##  Data: dataLDA
##        AIC     BIC    logLik
##   3661.551 3713.83 -1821.776
##
## Random effects:
##  Formula: ~Time_Point | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.6628601 (Intr)
## Time_Point  0.6255252 -0.998
## Residual    0.2163048
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | Time_Point
##  Parameter estimates:
##         1         0         2         3
## 1.0000000 0.3842252 7.1971216 6.6106513
## Fixed effects: binding ~ Time_Point
##                  Value Std.Error   DF   t-value p-value
## (Intercept) -0.5031390 0.1871486 2443 -2.688447  0.0072
## Time_Point   0.5695081 0.1798267 2443  3.166983  0.0016
##  Correlation:
##            (Intr)
## Time_Point -0.998
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.15655984 -0.26620653 -0.11153392  0.02881313 15.06729096
##
## Number of Observations: 2464
## Number of Groups: 20
```

This simpler model has lower AIC and BIC, as shown below. So we prefer the model with `Time_Point` as the predictor and, with the low p-values of the slope of `Time_Point`, conclude that the binding rates vary over time. In other words, the number of HIV vaccines given do affect the binding rate, but the drugs given do not have significant effects.

```
##      df      AIC df.1      BIC
## lda  13 3673.651   13 3749.144
## lda2  9 3661.551    9 3713.830
```

Discussion

In this study we used both multivariate and longitudinal data analysis to examine the effects of HIV vaccines and Treg suppression treatments. Although the study provides evidence to support the concept of using immunosuppressing treatments to increase diversity, the added diversity does not seem to improve the binding rate. In other words, this study does not provide evidence to show that the added treatments can enhance the effects of HIV vaccines.

List of variables

- Treatment: Treatment A is the mock control, and treatment B and C are two different kinds of Treg inhibitor treatments.
- Time Points: 0 represents before immunization; 1 represents 2 weeks post 1st immunization; 2 represents 2 weeks post 2nd immunization; and 3 represents 2 weeks post 3rd immunization, respectively.
- Isotype: the category of antibody type
- H_ID and L_ID: heavy chain and light chain IDs for the particular observation
- H_V, H_D and H_J: the gene segments used in heavy chain VDJ recombination in that antibody. The same applies to L_V and L_J
- H_VBase: the number of nucleotide of the heavy chain variable region
- H_Substitutions, H_Insertions, H_Deletions: the number of relative nucleotide mutations.
- HMuFreq: calculated by H_Substitutions / H_VBase
- H_CDR3: the number of amino acid of the heavy chain's third complementarity determining region
- Binding: affinity of antibodies against a selected HIV glycoprotein. The larger value indicates stronger binding

Reference

The dataset, which can be found here, was provided by Kan Luo, as he was one of authors for the following four publications that used the dataset:

1. Luo K, Liao HX, Zhang R, et al. Tissue memory B cell repertoire analysis after ALVAC/AIDSVAX B/E gp120 immunization of rhesus macaques. *JCI Insight*. 2016;1(20):e88522. Published 2016 Dec 8. doi:10.1172/jci.insight.88522
2. Bradley, T., Kuraoka, M., Yeh, C.-H., Tian, M., Chen, H., Cain, D. W., . . . Haynes, B. F. (2020). Immune checkpoint modulation enhances HIV-1 antibody induction. *Nature Communications*, 11(1), 948. doi:10.1038/s41467-020-14670-w
3. Easterhoff, D., Pollara, J., Luo, K., Tolbert, W. D., Young, B., Mielke, D., . . . Ferrari, G. (2020). Boosting with AIDSVAX B/E Enhances Env Constant Region 1 and 2 Antibody-Dependent Cellular Cytotoxicity Breadth and Potency. *Journal of Virology*, 94(4), e01120-01119. doi:10.1128/jvi.01120-19
4. Wiehe, K., Easterhoff, D., Luo, K., Nicely, N. I., Bradley, T., Jaeger, F. H., Dennison, S. M., Zhang, R., Lloyd, K. E., Stolarchuk, C., Parks, R., Sutherland, L. L., Scearce, R. M., Morris, L., Kaewkungwal, J., Nitayaphan, S., Pitisuttithum, P., Rerks-Ngarm, S., Sinangil, F., Phogat, S., ... Haynes, B. F. (2014). Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. *Immunity*, 41(6), 909–918. https://doi.org/10.1016/j.immuni.2014.11.014

5. Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Muller W, Bontrop R, Lemaitre M, Malik A, Barbie V, Chaume D. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* 1999;27:209–212. doi: 10.1093/nar/27.1.209.

6. Jenny M Woof , Dennis R Burton,Human antibody-Fc receptor interactions illuminated by crystal structures.*Nat Rev Immunol.* 2004 Feb;4(2):89-99. doi: 10.1038/nri1266.

7. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-150, <URL: https://CRAN.R-project.org/package=nlme>.