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

[How to hide message from API?]

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.

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.

Data Summaries

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

The dataset has 2465 data points and 20 rhesus monkeys. We first present our exploratory data analysis and summaries.

```
##
## 6104 6105 6107 6117 6118 6119 6125 6132 6160 6193 6199 6200 6201 6202 6203 6204
##   35  228  239  243    7   55  216  251  183  117   48  191   73   78  238  156
## 6205 6209 6210 6214
##    5   46   50    6
```

In the current analysis, each row represents one antibody and its measurements. While it is possible to treat the 20 rhesus macaques as the observational units, the analysis will become quite complex. We would have to use information from gene segments of the heavy and light chains of the antibodies to classify each antibody. There are five such segments, and each has 6 to 11 subtypes. (In total, there can be $7 * 7 * 6 * 11 * 6 = 1.9404 \times 10^4$ combinations of gene segments.) Thus, we decided to follow the convention of vaccine studies and treat each antibody as independent.

The main goal of the study is to test 1) whether the Treg suppression treatments can increase the diversity of antibodies and 2) enhance the effectiveness of vaccines. Our predictors are time points, which indicates time and the number of vaccines given up to that point, treatment or drug, and Isotype. We first take a look at these variables.

There are four time points; one before any procedure was done, and three after vaccine shots were administered to the macaques. In the treatment groups, groups 1-3 represent different doses of drug 1, groups 4-6 represent different doses of drug 2, and group 7 represents the control group. Later we'll look at the effect made by different drugs first and then different doses.

```
##
##    group 1 group 2 group 3 group 4 group 5 group 6 group 7
## 0     129       0       0      90       0       0      54
## 1     190      60      96     105     297     131     125
## 2     141     110       0     148      77       0     347
## 3     122       0       0     101       0       0     142
```

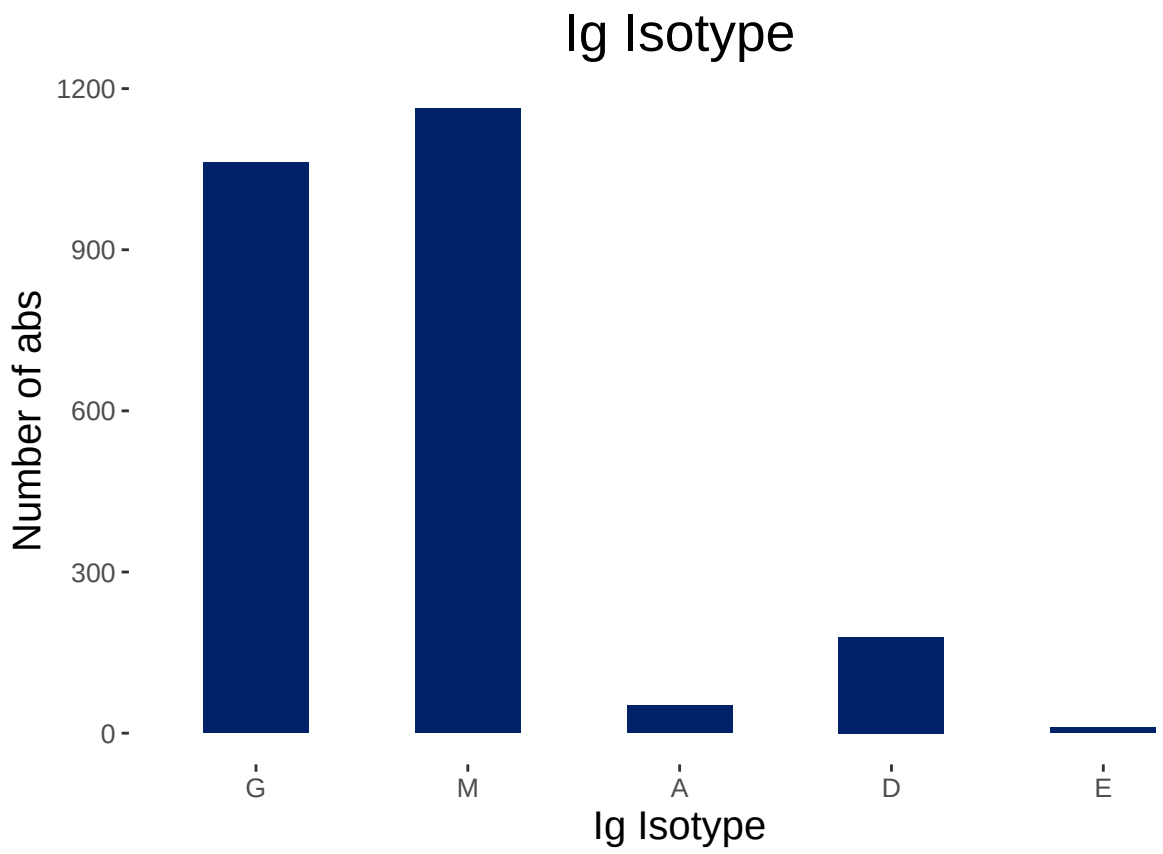
```
##
##    group 1 group 2 group 3 group 4 group 5 group 6 group 7
## 1     582     170      96       0       0       0       0
## 2       0       0       0     444     374     131       0
## 3       0       0       0       0       0       0     668
```

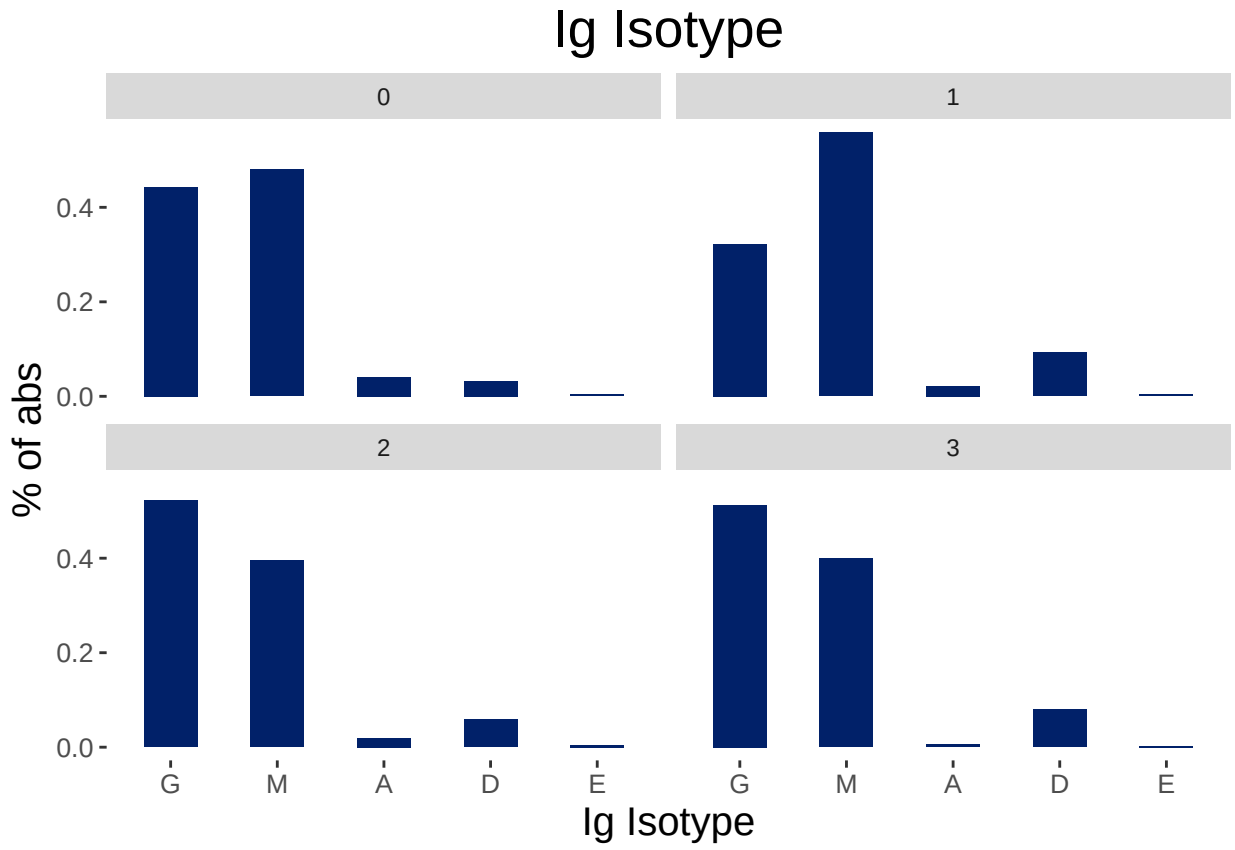
```
##
##    group 1 group 2 group 3 group 4 group 5 group 6 group 7
## 1     582     170      96       0       0       0       0
## 2       0       0       0     444     374     131       0
## 3       0       0       0       0       0       0     668
```

Each marcaque only received one kind of treatment.

```
##
##       group 1 group 2 group 3 group 4 group 5 group 6 group 7
## 6104        0       0      35       0       0       0       0
## 6105        0       0       0     228       0       0       0
## 6107        0       0       0       0       0       0     239
## 6117      243       0       0       0       0       0       0
## 6118        0       7       0       0       0       0       0
## 6119        0       0      55       0       0       0       0
## 6125        0       0       0     216       0       0       0
## 6132        0       0       0       0     251       0       0
## 6160      183       0       0       0       0       0       0
## 6193        0     117       0       0       0       0       0
## 6199        0       0       0       0       0      48       0
## 6200        0       0       0       0       0       0     191
## 6201        0       0       0       0      73       0       0
## 6202        0       0       0       0       0      78       0
## 6203        0       0       0       0       0       0     238
## 6204      156       0       0       0       0       0       0
## 6205        0       0       0       0       0       5       0
## 6209        0      46       0       0       0       0       0
## 6210        0       0       0       0      50       0       0
## 6214        0       0       6       0       0       0       0
```

Next, we'll take a look at the variable `Isotype`. There are 5 kinds of heavy chain for antibodies: IgG, IgA, IgM, IgE, IgD. The two most important kinds are IgG and IgM. IgM mostly occurs in the acute stage of infection, and IgM appears later in blood with better neutralizing potentials.


Ig Isotype
1200
900
600
300
0
Number of abs
G
M
A
D
E
Ig Isotype

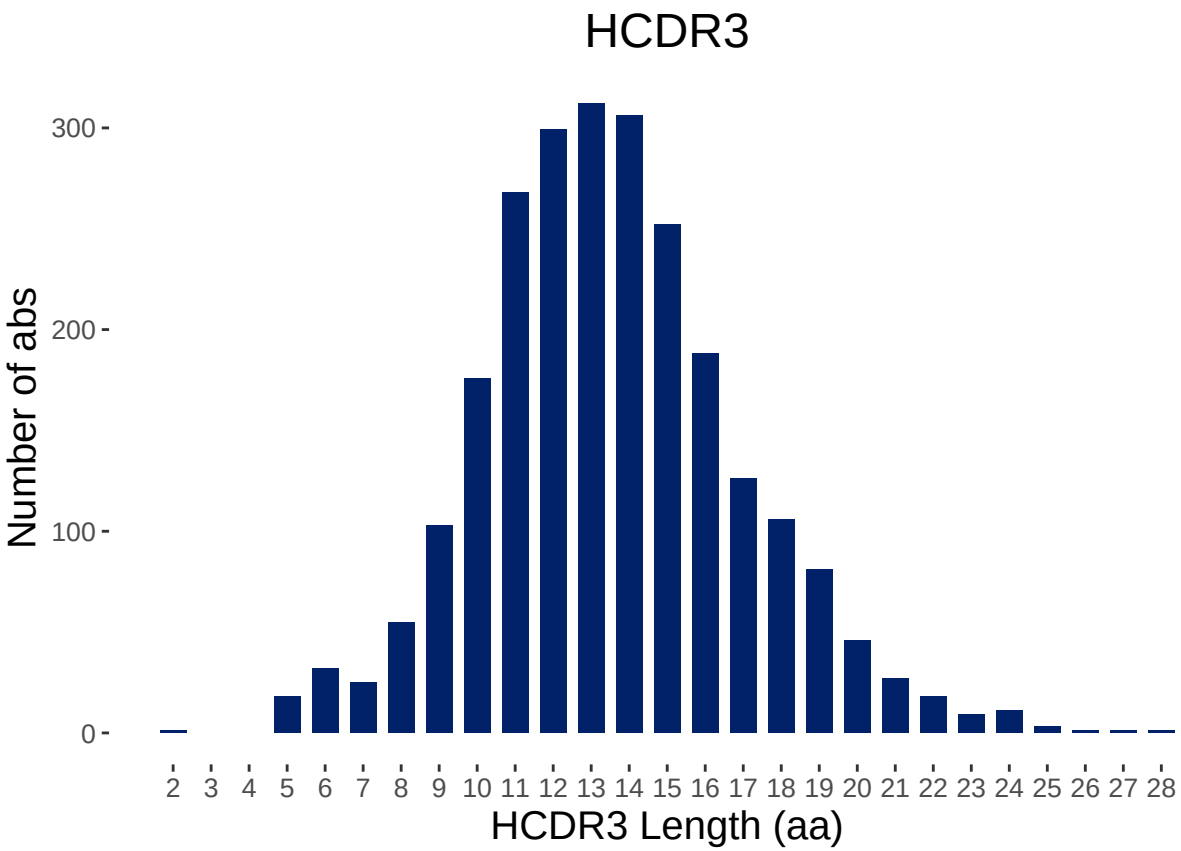


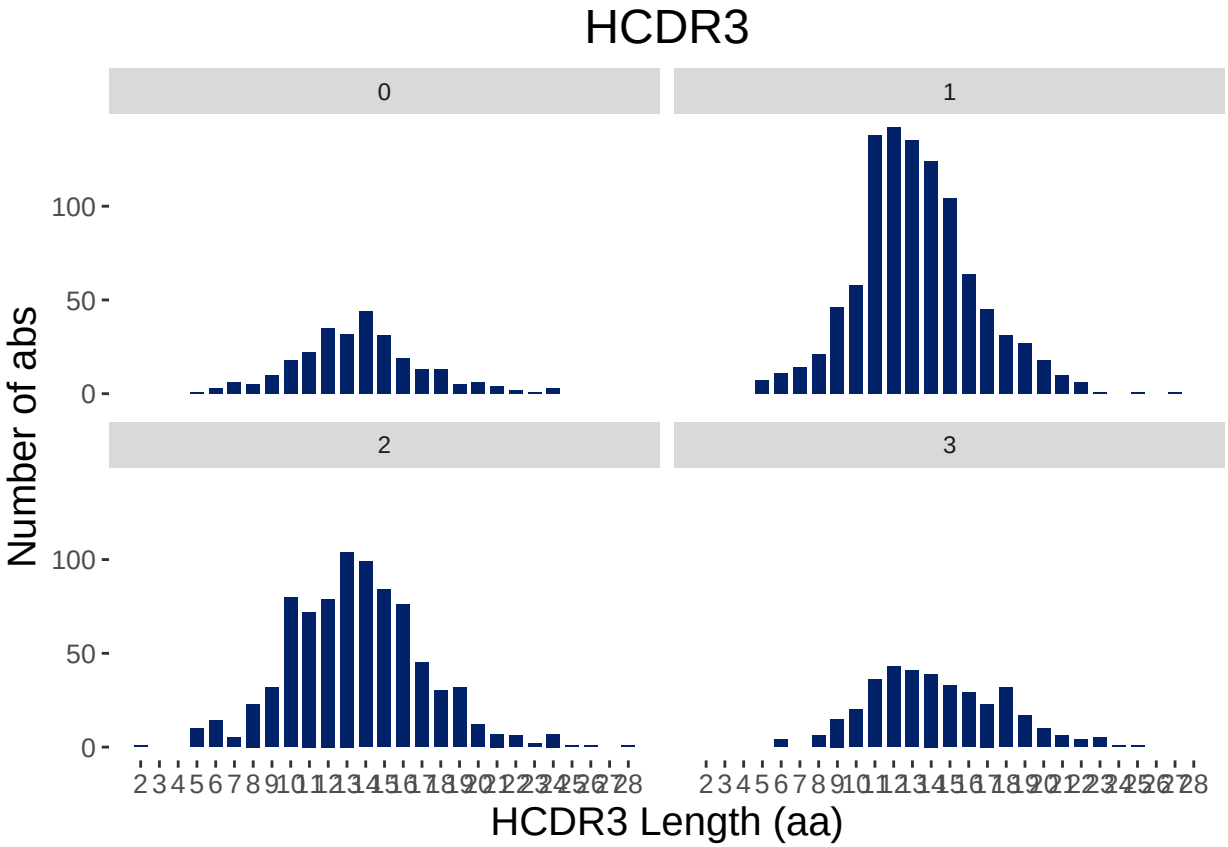

```
##   Isotype Ab # Ab %
## 1       A   51  2.1
## 2       D  179  7.3
## 3       E   10  0.4
## 4       G 1062 43.1
## 5       M 1163 47.2
```

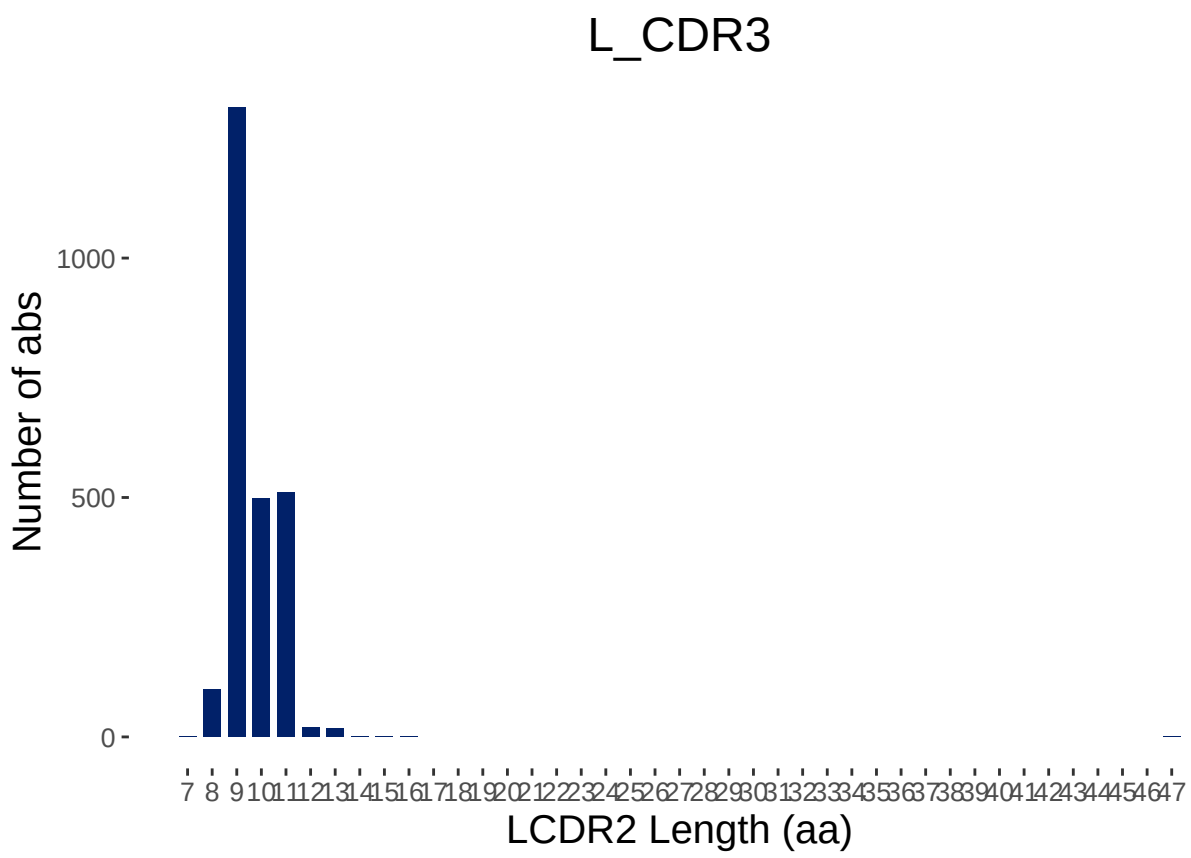
As expected, we see that IgG and IgM occupy the biggest proportion of all antibodies in all time points. We'll use the variable `Isotype` as a grouping covariate later.

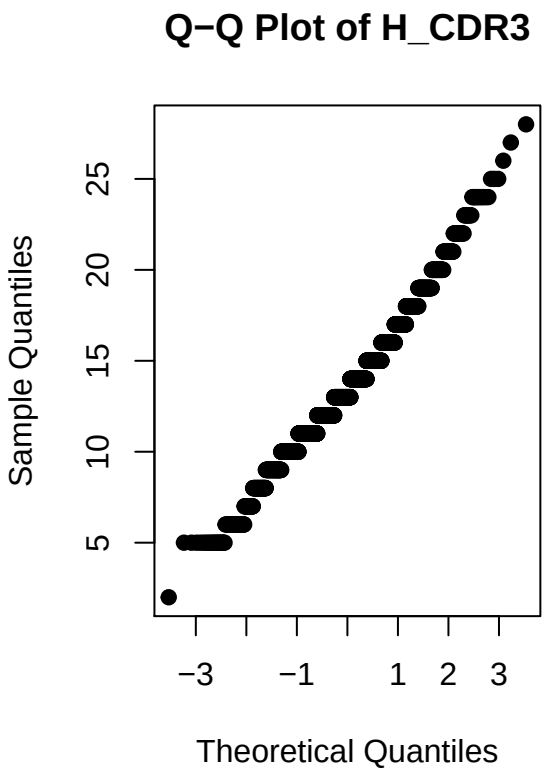
Next we'll examine our responses: `H_CDR3`, `HMuFreq`, `L_CDR3`, `LMuFreq`, `Binding`, and `Reactivity`. In each antibody, there are two sets of heavy chain and light chain, all of which forming a Y-shape immunoglobulin. Thus many of the variables start with H or L, indicating which chain the information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

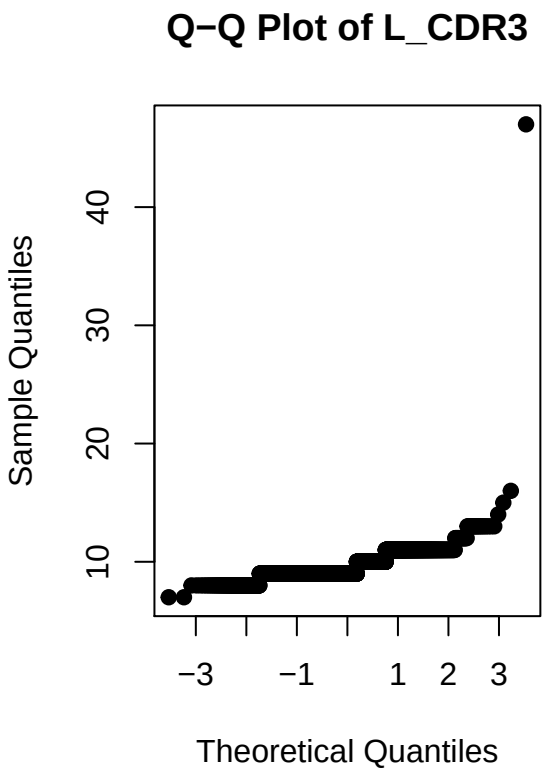
`H_CDR3` and `L_CDR3` indicates the length of the third complementarity-determining region on the variable heavy chain and light chain. The longer they are, the more potential there is to produce diverse antibodies. [Kan, could you check to see if this is correct?] In other words, we want the values 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. For the Q-Q plot for `L_CDR3`, we can see that there is one outlier. Without it, the distribution is likely normal. (We'll get to this soon.)


HCDR3
Number of abs
300
200
100
0
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
HCDR3 Length (aa)

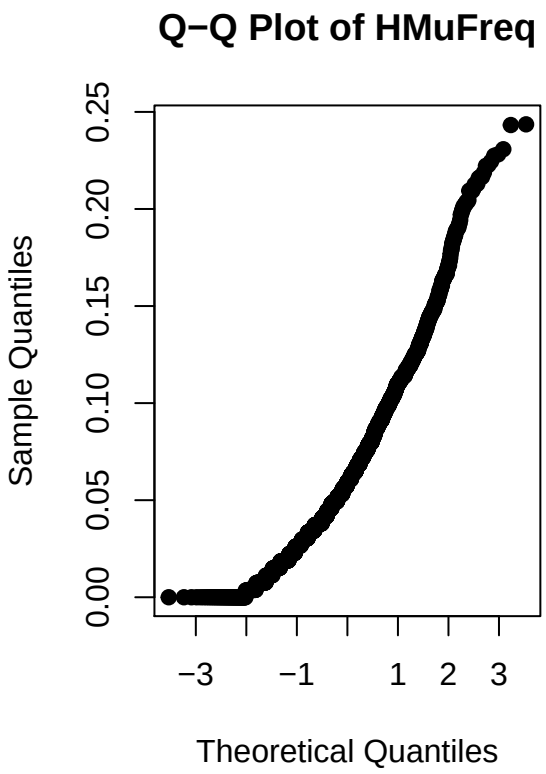

HCDR3
0
1
2
3
Number of abs
100
50
0
HCDR3 Length (aa)

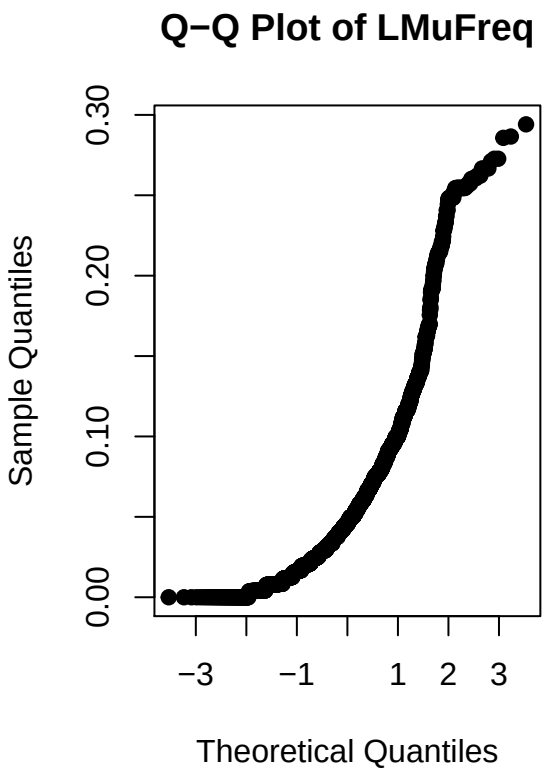

L_CDR3
Number of abs
1000
500
0
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
LCDR2 Length (aa)



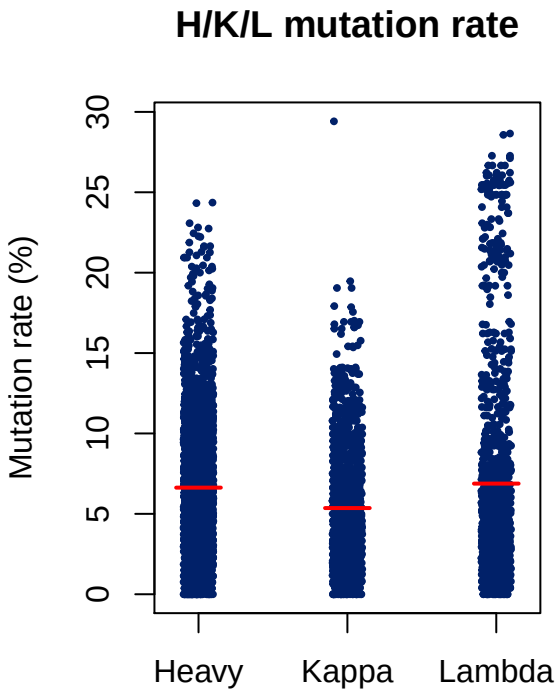


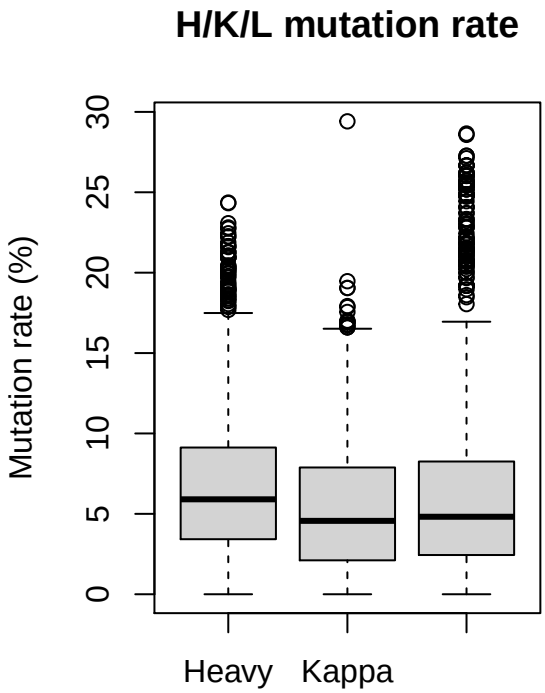
`HMuFreq` and `LMuFreq` are calculated by dividing `H_Substitution` by `H_VBase` for heavy chain and similarly for light chain. These two variables show how much the antibodies mutate. For the purpose of the study, the higher the mutation rate is, the better. Below we see some comparison of mutation 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?]

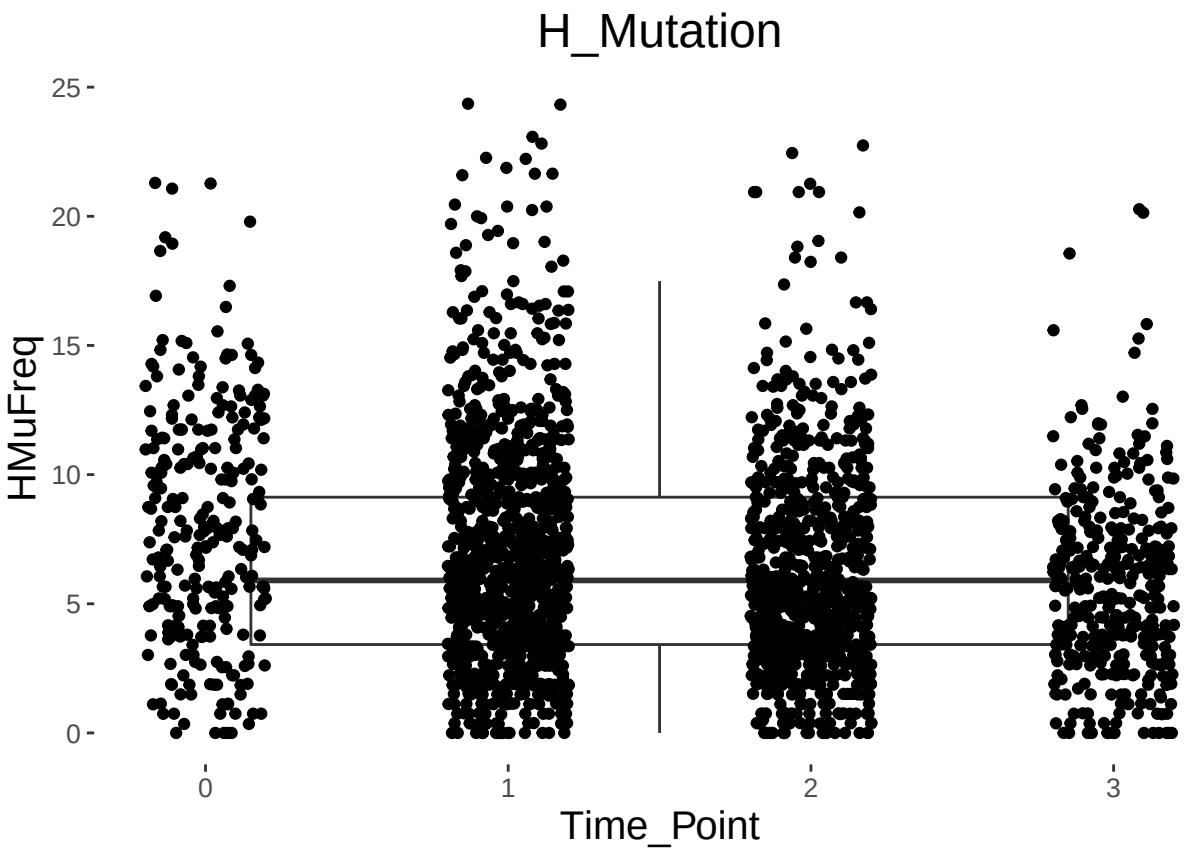




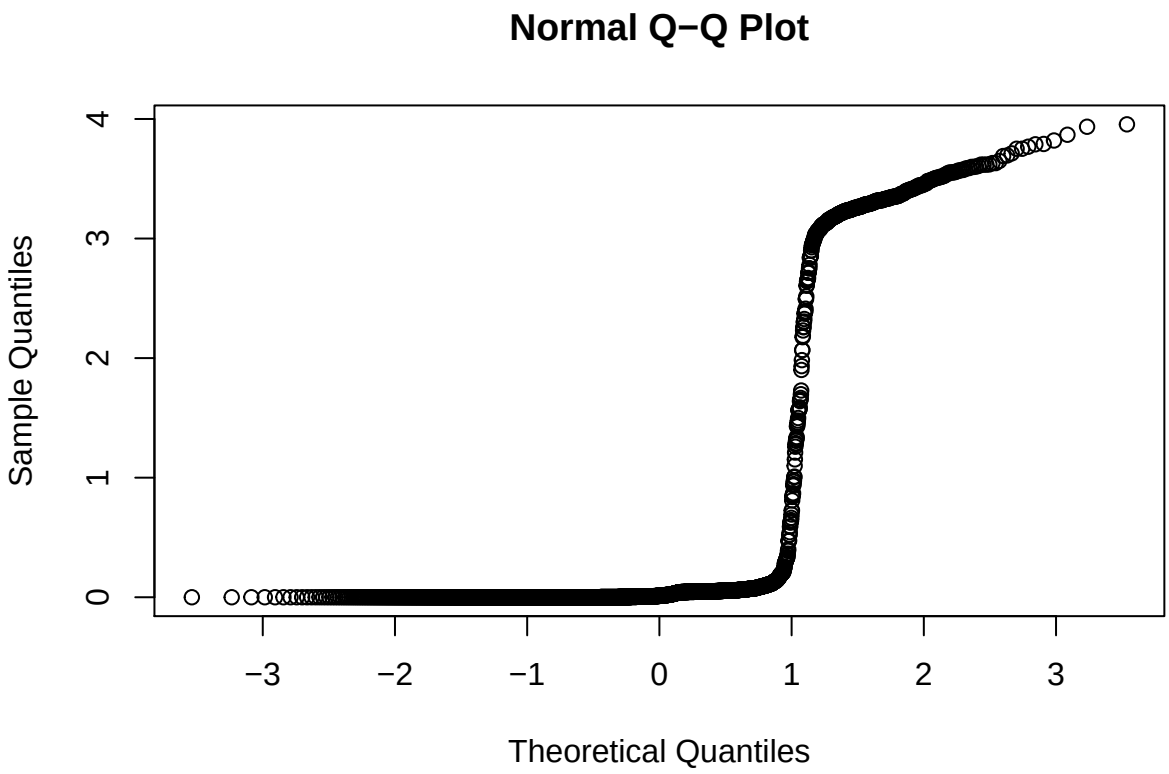

```
##         H_Mutation% K_Mutation% L_Mutation%
## Min.           0.00        0.00        0.00
## 1st Qu.        3.42        2.11        2.44
## Median         5.90        4.56        4.82
## Mean           6.63        5.36        6.88
## 3rd Qu.        9.13        7.88        8.25
## Max.          24.36       29.41       28.65
```


H/K/L mutation rate
Mutation rate (%)
0
5
10
15
20
25
30
Heavy
Kappa
Lambda


H/K/L mutation rate
Mutation rate (%)
0
5
10
15
20
25
30
Heavy
Kappa

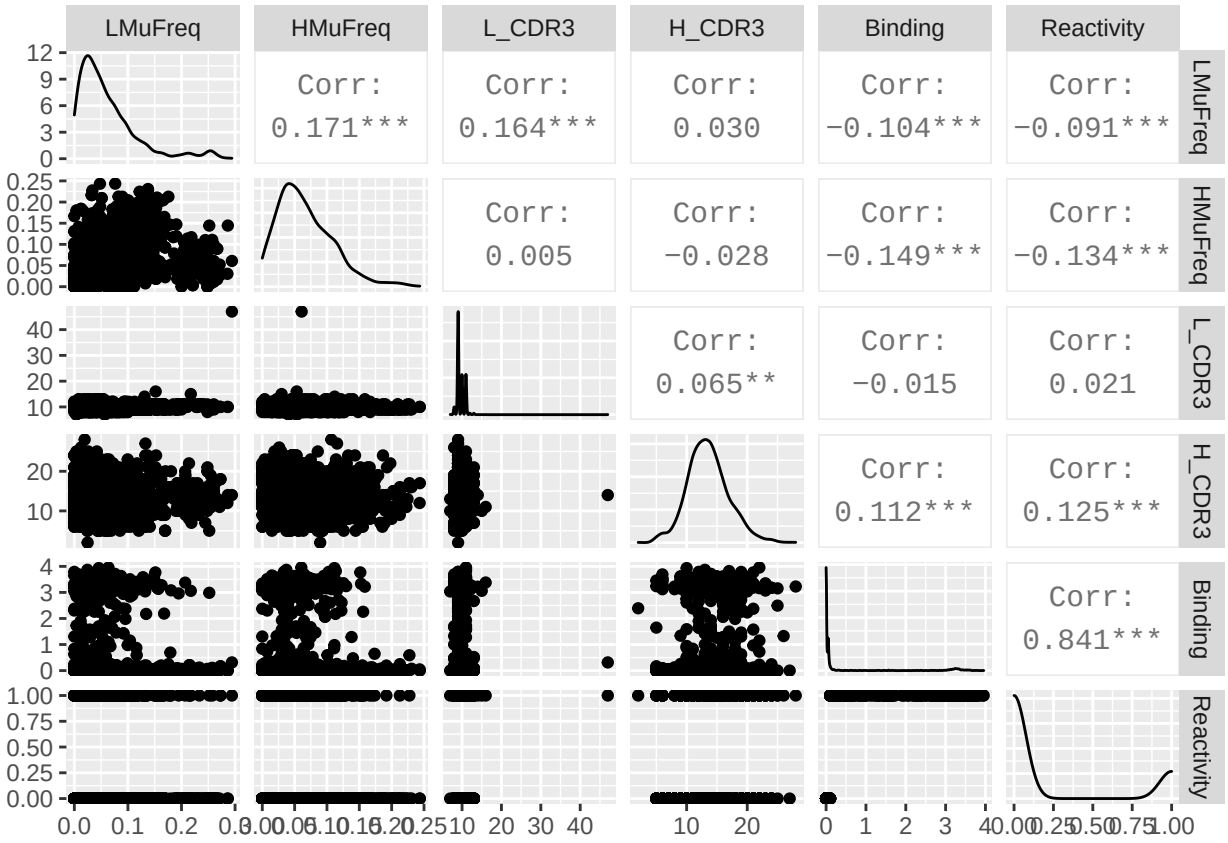


Lastly, `Binding` indicates the rate of neutralizing, meaning how much the antibodies bind with the virus and thus make the virus ineffective. This is the most important measure of the study. `Reactivity` turns `Binding` into a binary variable; `Binding` rate above 0.1 is considered reactive. In the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

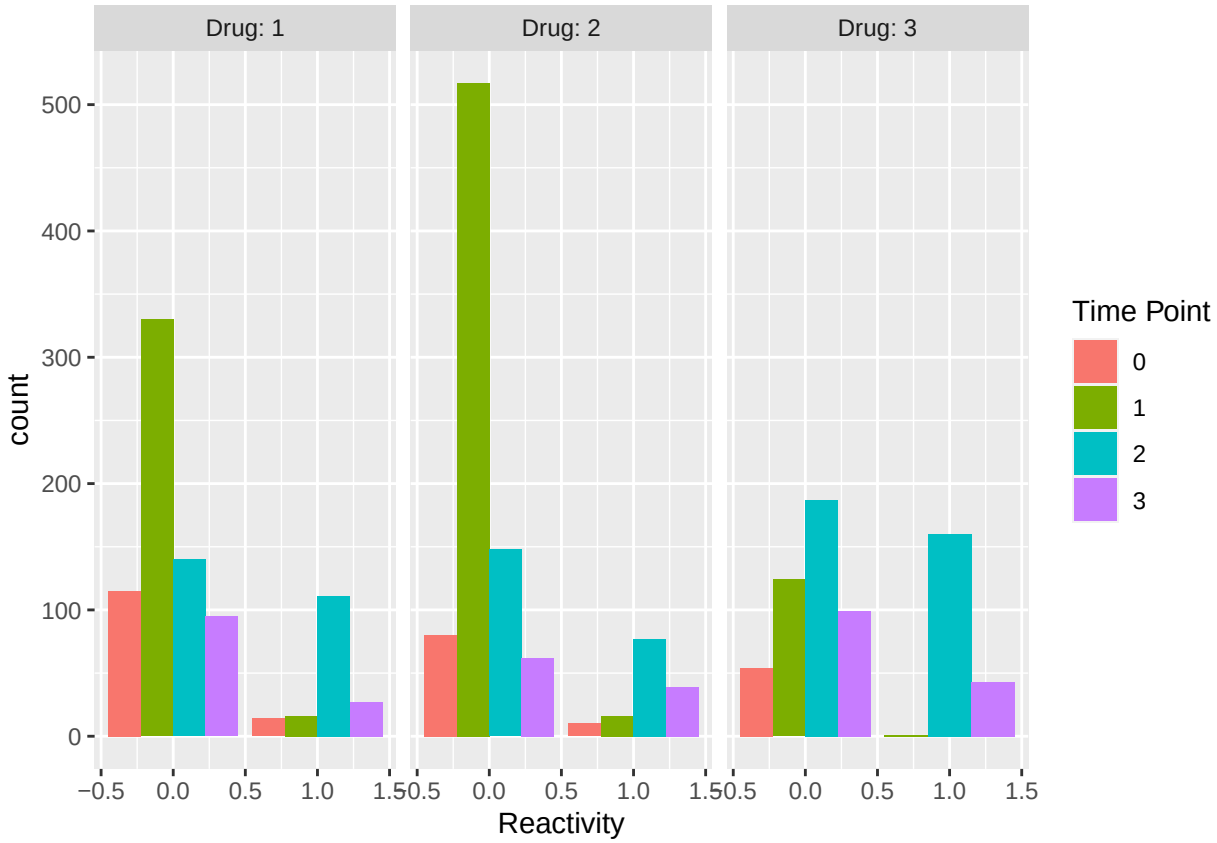


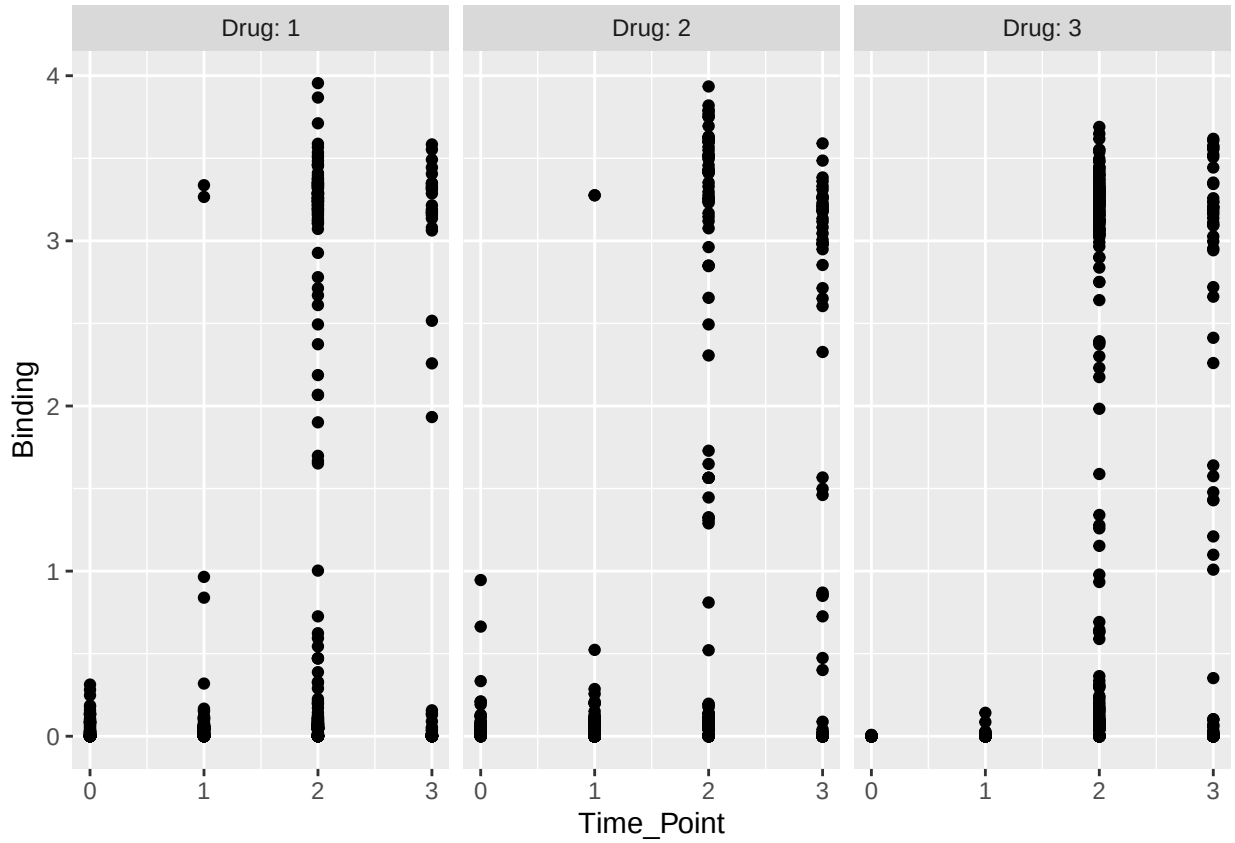

```
##
##    0    1
## 1951  514
```

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?]



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

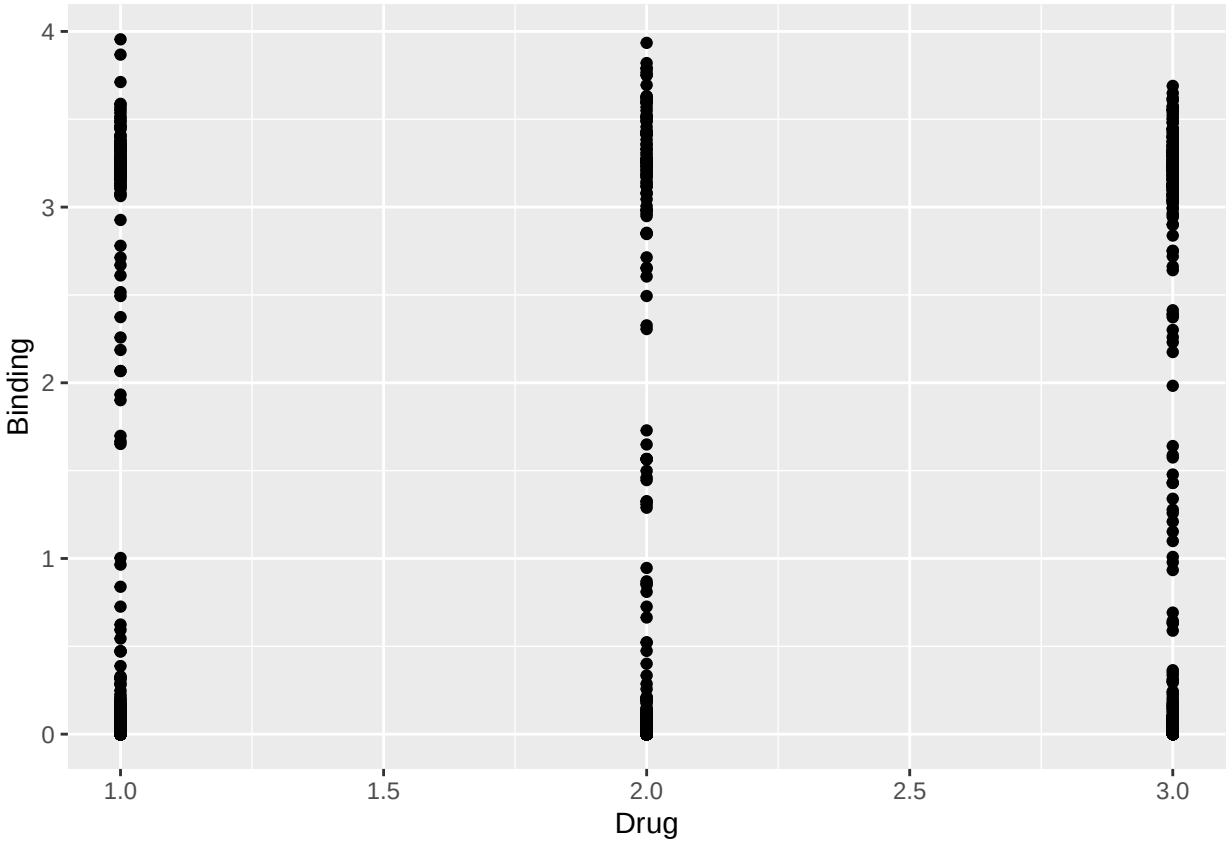


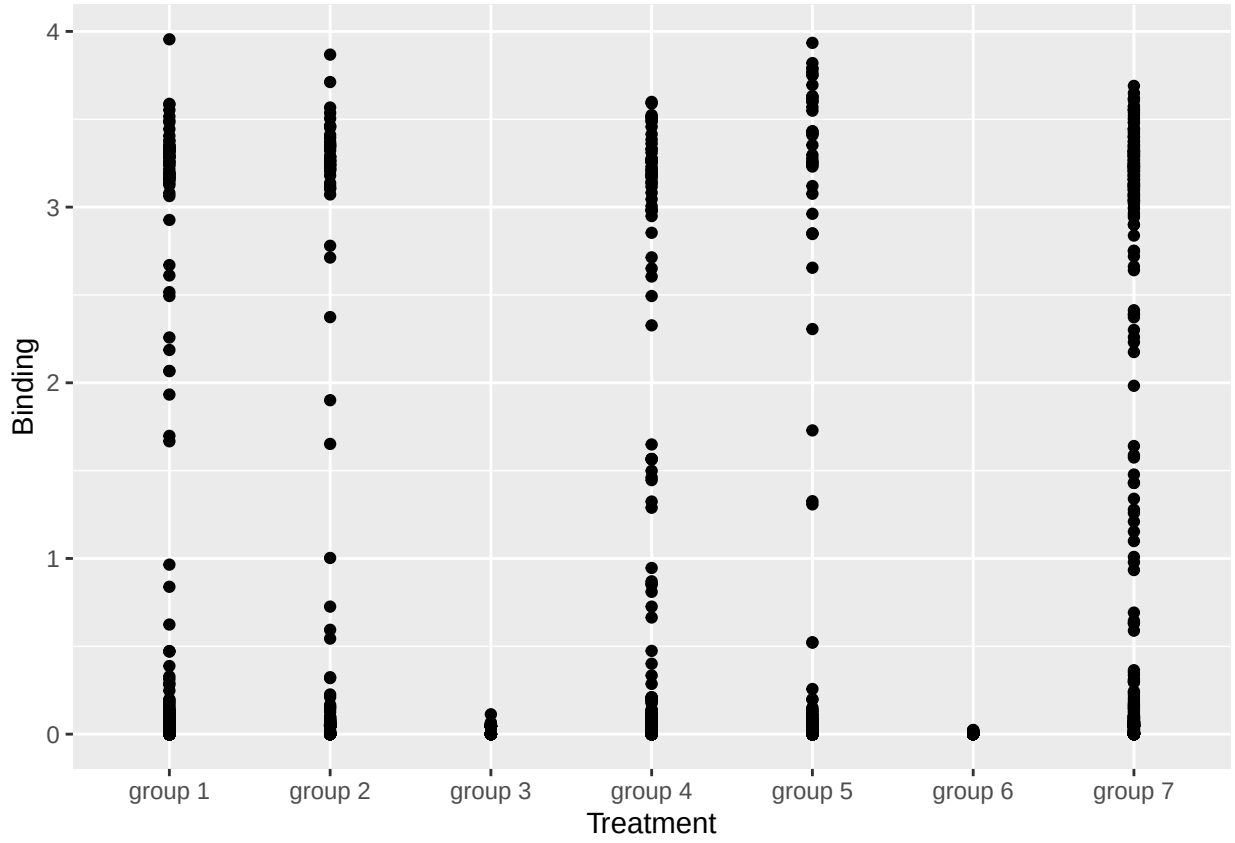

```
##
##       0   1
##   1 680 168
##   2 807 142
##   3 464 204
```

```
##
##       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
```

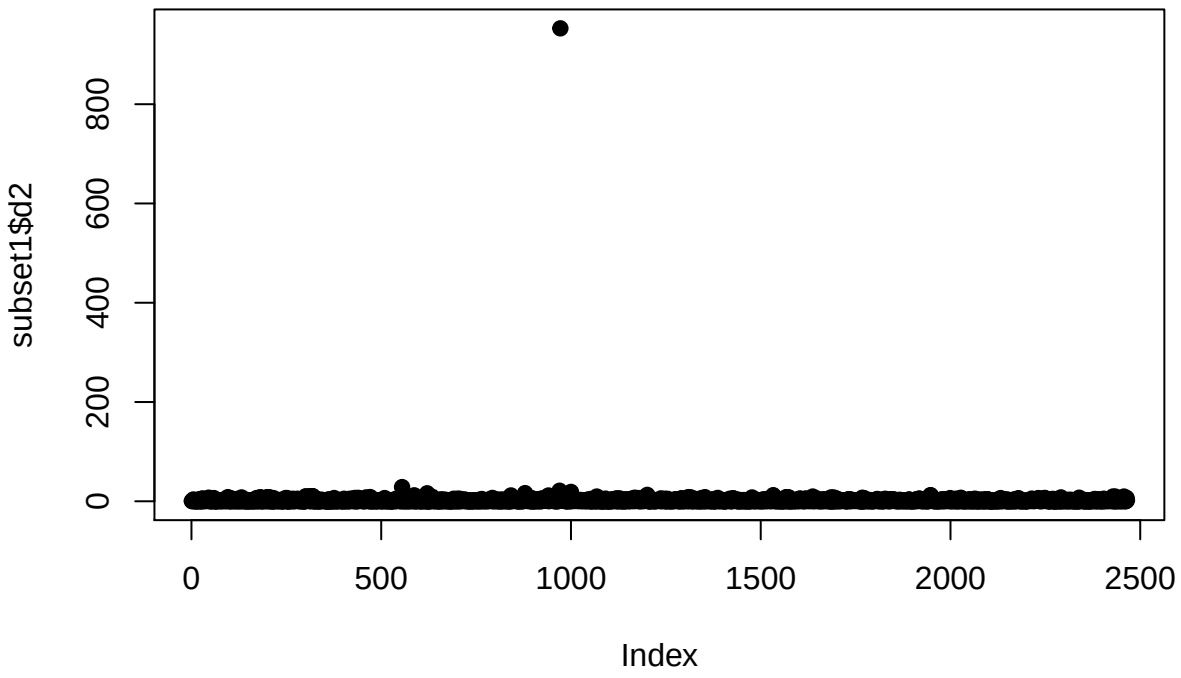

Binding
4
3
2
1
0
1.0
1.5
2.0
2.5
3.0
Drug

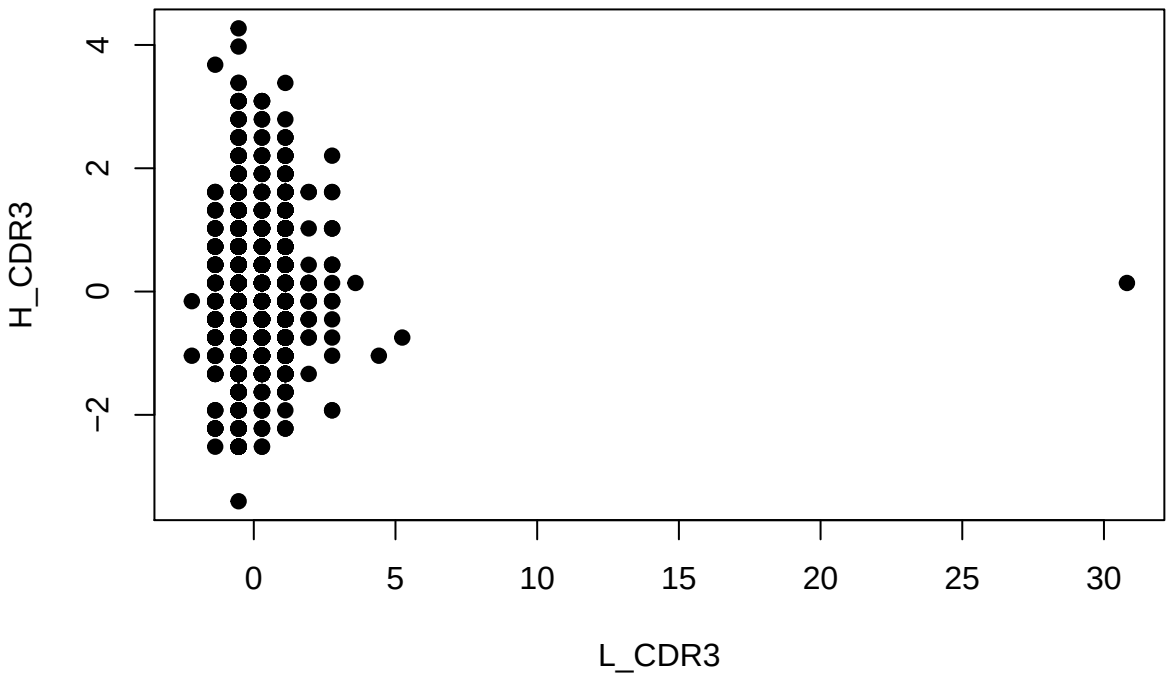


Outlier detection

Before we go into analyses, notice may have outlier in LCDR3 variable.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.00    9.00    9.00    9.65   10.00   47.00
```


subset1$d2
0
200
400
600
800
0
500
1000
1500
2000
2500
Index




```
## # A tibble: 1 x 4
##   L_CDR3 H_CDR3    d2 Z[,"L_CDR3"] [,"H_CDR3"] [,"d2"]
##    <dbl>  <dbl> <dbl>        <dbl>       <dbl>   <dbl>
## 1     47     14  953.         30.8       0.139    49.4
```

```
## [1] 972
```

Row 972 from `Data2` is in fact an outlier, as shown in the summary and plots above. The value for `L_CDR3` is quite unlikely. Since we can't go back to the original data, we remove the data point and will use the new dataset `Data3`.

```
## # A tibble: 1 x 19
##   MonkeyID  Drug Treatment Time_Point Isotype H_VBase H_Substitutions
##      <dbl> <dbl> <chr>          <dbl> <chr>     <dbl>           <dbl>
## 1     6107     3 group 7            2 G           263              16
## # ... with 12 more variables: H_Insertions <dbl>, H_Deletions <dbl>,
## #   HMuFreq <dbl>, H_CDR3 <dbl>, L_VBase <dbl>, L_Substitutions <dbl>,
## #   L_Insertions <dbl>, L_Deletions <dbl>, LMuFreq <dbl>, L_CDR3 <dbl>,
## #   Binding <dbl>, Reactivity <dbl>
```

Data Analysis

Multivariate Data Analysis

[Use the caret package to do training and prediction?]

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.

We will use `Reactivity` separately when we use logistic regression. Although `Time_Point` is also one of the predictors, we will use it for longitudinal analysis in the next section.

First, we use manova to test effects. Since we have a large same 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.

```
##             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
```

We next use multiple linear regression.

```
## Response Data3$L_CDR3 :
##
## Call:
## lm(formula = `Data3$L_CDR3` ~ drug * it)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5872 -0.6634 -0.5696  0.4304  6.3000
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.583333   0.193972  49.406   <2e-16 ***
## drug2       -0.011905   0.283946  -0.042   0.9666
## drug3        0.750000   0.433735   1.729   0.0839 .
## itD          0.100877   0.231230   0.436   0.6627
## itE          0.416667   0.408198   1.021   0.3075
## itG          0.116667   0.200512   0.582   0.5607
## itM          0.080057   0.199609   0.401   0.6884
## drug2:itD   -0.178972   0.329405  -0.543   0.5870
## drug3:itD   -0.944849   0.472420  -2.000   0.0456 *
## drug2:itE   -1.321429   0.714582  -1.849   0.0645 .
## drug3:itE          NA         NA      NA       NA
## drug2:itG   -0.002729   0.292246  -0.009   0.9925
## drug3:itG   -0.862752   0.440155  -1.960   0.0501 .
## drug2:itM   -0.028759   0.291370  -0.099   0.9214
## drug3:itM   -0.843770   0.439548  -1.920   0.0550 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9503 on 2447 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.006476,   Adjusted R-squared:  0.001198
## F-statistic: 1.227 on 13 and 2447 DF,  p-value: 0.2527
##
##
## Response Data3$LMuFreq :
##
## Call:
## lm(formula = `Data3$LMuFreq` ~ drug * it)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06859 -0.03658 -0.01446  0.02018  0.22777
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0393431  0.0110147   3.572 0.000361 ***
## drug2        0.0009367  0.0161238   0.058 0.953677
## drug3       -0.0069529  0.0246295  -0.282 0.777738
## itD          0.0164648  0.0131304   1.254 0.209981
## itE          0.0058906  0.0231795   0.254 0.799418
## itG          0.0251809  0.0113861   2.212 0.027089 *
## itM          0.0216787  0.0113348   1.913 0.055918 .
## drug2:itD    0.0013393  0.0187052   0.072 0.942928
## drug3:itD    0.0197308  0.0268263   0.736 0.462105
## drug2:itE    0.0007513  0.0405774   0.019 0.985229
## drug3:itE           NA         NA      NA       NA
## drug2:itG    0.0017932  0.0165951   0.108 0.913962
## drug3:itG   -0.0043914  0.0249941  -0.176 0.860545
## drug2:itM   -0.0040176  0.0165454  -0.243 0.808164
## drug3:itM    0.0096142  0.0249596   0.385 0.700130
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05396 on 2447 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.01073,    Adjusted R-squared:  0.005478
## F-statistic: 2.042 on 13 and 2447 DF,  p-value: 0.0147
##
##
## Response Data3$H_CDR3 :
##
## Call:
## lm(formula = `Data3$H_CDR3` ~ drug * it)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5829 -2.0000 -0.0933  2.0023 13.7416
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.7917     0.6790  21.783   <2e-16 ***
## drug2        -2.2679     0.9940  -2.282   0.0226 *
## drug3        -3.2917     1.5184  -2.168   0.0303 *
## itD          -1.1776     0.8095  -1.455   0.1458
## itE           2.2083     1.4290   1.545   0.1224
## itG          -0.2088     0.7019  -0.297   0.7661
## itM          -1.6639     0.6988  -2.381   0.0173 *
## drug2:itD     1.7472     1.1531   1.515   0.1299
## drug3:itD     2.4436     1.6538   1.478   0.1397
## drug2:itE    -2.0655     2.5015  -0.826   0.4091
## drug3:itE         NA         NA      NA       NA
## drug2:itG     1.5118     1.0231   1.478   0.1396
```

```
## drug3:itG     2.9672     1.5408   1.926   0.0543 .
## drug2:itM     2.1378     1.0200   2.096   0.0362 *
## drug3:itM     2.8823     1.5387   1.873   0.0612 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.327 on 2447 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.04222,    Adjusted R-squared:  0.03713
## F-statistic: 8.298 on 13 and 2447 DF,  p-value: < 2.2e-16
##
##
## Response Data3$HMuFreq :
##
## Call:
## lm(formula = `Data3$HMuFreq` ~ drug * it)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.091172 -0.028366 -0.006062  0.024649  0.185109
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.063852   0.008280   7.712 1.79e-14 ***
## drug2       -0.010070   0.012121  -0.831   0.4062
## drug3        0.015765   0.018514   0.852   0.3946
## itD         -0.004201   0.009870  -0.426   0.6704
## itE         -0.021395   0.017424  -1.228   0.2196
## itG          0.015267   0.008559   1.784   0.0746 .
## itM         -0.005718   0.008521  -0.671   0.5022
## drug2:itD    0.008037   0.014061   0.572   0.5677
## drug3:itD   -0.025158   0.020166  -1.248   0.2123
## drug2:itE    0.045493   0.030503   1.491   0.1360
## drug3:itE          NA         NA      NA       NA
## drug2:itG    0.022123   0.012475   1.773   0.0763 .
## drug3:itG   -0.036875   0.018788  -1.963   0.0498 *
## drug2:itM    0.011428   0.012437   0.919   0.3583
## drug3:itM   -0.019511   0.018763  -1.040   0.2985
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04056 on 2447 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.1012, Adjusted R-squared:  0.09639
## F-statistic: 21.18 on 13 and 2447 DF,  p-value: < 2.2e-16
##
##
## Response Data3$Binding :
##
## Call:
## lm(formula = `Data3$Binding` ~ drug * it)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -1.7208 -0.6290 -0.0707 -0.0147  3.5273
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5331     0.1989   2.681   0.0074 **
## drug2        -0.4841     0.2911  -1.663   0.0965 .
## drug3         0.4137     0.4447   0.930   0.3523
## itD          -0.4404     0.2371  -1.858   0.0633 .
## itE          -0.4693     0.4185  -1.121   0.2623
## itG           0.3817     0.2056   1.857   0.0634 .
## itM          -0.4274     0.2047  -2.089   0.0369 *
## drug2:itD     0.4918     0.3377   1.456   0.1455
## drug3:itD    -0.4800     0.4844  -0.991   0.3218
## drug2:itE     0.8512     0.7326   1.162   0.2454
## drug3:itE         NA         NA      NA       NA
## drug2:itG     0.2412     0.2996   0.805   0.4210
## drug3:itG     0.3922     0.4513   0.869   0.3849
## drug2:itM     0.4501     0.2987   1.507   0.1320
## drug3:itM    -0.4600     0.4507  -1.021   0.3075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9743 on 2447 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.2467, Adjusted R-squared:  0.2427
## F-statistic: 61.64 on 13 and 2447 DF,  p-value: < 2.2e-16
```

Finally, we use logistic regression.
[All the p-values are large. I am wondering if we should take out the variable `Reactivity` and just focus on `Binding`.]

```
##
## Call:
## lm(formula = Data3$Reactivity ~ drug * it)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60403 -0.25854 -0.06818 -0.05380  0.95238
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.20833    0.07463   2.792  0.00529 **
## drug2       -0.16071    0.10925  -1.471  0.14140
## drug3        0.12500    0.16688   0.749  0.45390
## itD         -0.12061    0.08897  -1.356  0.17531
## itE         -0.06548    0.15705  -0.417  0.67679
## itG          0.12594    0.07713   1.633  0.10261
## itM         -0.11251    0.07680  -1.465  0.14306
## drug2:itD    0.12633    0.12674   0.997  0.31898
## drug3:itD   -0.12761    0.18176  -0.702  0.48270
## drug2:itE    0.35119    0.27494   1.277  0.20160
## drug3:itE         NA         NA      NA       NA
## drug2:itG    0.08497    0.11243   0.756  0.44984
## drug3:itG    0.14475    0.16934   0.855  0.39276
## drug2:itM    0.13307    0.11210   1.187  0.23533
```

```
## drug3:itM   -0.16703    0.16912  -0.988  0.32343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3656 on 2450 degrees of freedom
## Multiple R-squared:  0.1937, Adjusted R-squared:  0.1894
## F-statistic: 45.28 on 13 and 2450 DF,  p-value: < 2.2e-16
```

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 it is adjusted based on the number of pairs and variables being compared.)

```
## [1] "L_CDR3  pairwise CI's"
##  contrast          estimate     SE   df lower.CL upper.CL
##  group 1 - group 2  0.08842 0.0829 2454  -0.2016    0.378
##  group 1 - group 3 -0.03487 0.1047 2454  -0.4012    0.331
##  group 1 - group 4  0.11097 0.0599 2454  -0.0987    0.321
##  group 1 - group 5 -0.01158 0.0630 2454  -0.2321    0.209
##  group 1 - group 6  0.08361 0.0919 2454  -0.2380    0.405
##  group 1 - group 7  0.11559 0.0540 2454  -0.0732    0.304
##  group 2 - group 3 -0.12328 0.1213 2454  -0.5477    0.301
##  group 2 - group 4  0.02255 0.0857 2454  -0.2773    0.322
##  group 2 - group 5 -0.10000 0.0879 2454  -0.4075    0.208
##  group 2 - group 6 -0.00480 0.1105 2454  -0.3913    0.382
##  group 2 - group 7  0.02717 0.0816 2454  -0.2584    0.313
##  group 3 - group 4  0.14583 0.1070 2454  -0.2283    0.520
##  group 3 - group 5  0.02328 0.1087 2454  -0.3571    0.404
##  group 3 - group 6  0.11848 0.1277 2454  -0.3282    0.565
##  group 3 - group 7  0.15046 0.1037 2454  -0.2124    0.513
##  group 4 - group 5 -0.12255 0.0667 2454  -0.3559    0.111
##  group 4 - group 6 -0.02735 0.0945 2454  -0.3579    0.303
##  group 4 - group 7  0.00462 0.0582 2454  -0.1990    0.208
##  group 5 - group 6  0.09520 0.0965 2454  -0.2423    0.433
##  group 5 - group 7  0.12717 0.0614 2454  -0.0876    0.342
##  group 6 - group 7  0.03198 0.0908 2454  -0.2857    0.350
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "LMuFreq  pairwise CI's"
##  contrast           estimate      SE   df lower.CL upper.CL
##  group 1 - group 2  0.007328 0.00472 2454 -0.00918  0.02384
##  group 1 - group 3 -0.006001 0.00596 2454 -0.02686  0.01486
##  group 1 - group 4  0.002551 0.00341 2454 -0.00939  0.01449
##  group 1 - group 5 -0.000993 0.00359 2454 -0.01355  0.01156
##  group 1 - group 6 -0.001342 0.00523 2454 -0.01966  0.01697
##  group 1 - group 7  0.003112 0.00307 2454 -0.00764  0.01386
##  group 2 - group 3 -0.013330 0.00691 2454 -0.03750  0.01084
##  group 2 - group 4 -0.004778 0.00488 2454 -0.02185  0.01230
##  group 2 - group 5 -0.008322 0.00500 2454 -0.02583  0.00919
##  group 2 - group 6 -0.008670 0.00629 2454 -0.03068  0.01334
##  group 2 - group 7 -0.004216 0.00465 2454 -0.02048  0.01205
##  group 3 - group 4  0.008552 0.00609 2454 -0.01275  0.02986
```

```
##  group 3 - group 5  0.005008 0.00619 2454 -0.01665  0.02667
##  group 3 - group 6  0.004659 0.00727 2454 -0.02077  0.03009
##  group 3 - group 7  0.009114 0.00591 2454 -0.01155  0.02978
##  group 4 - group 5 -0.003544 0.00380 2454 -0.01683  0.00974
##  group 4 - group 6 -0.003893 0.00538 2454 -0.02271  0.01493
##  group 4 - group 7  0.000562 0.00331 2454 -0.01103  0.01216
##  group 5 - group 6 -0.000349 0.00549 2454 -0.01957  0.01887
##  group 5 - group 7  0.004106 0.00350 2454 -0.00812  0.01633
##  group 6 - group 7  0.004454 0.00517 2454 -0.01364  0.02254
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "H_CDR3  pairwise CI's"
##  contrast          estimate    SE   df lower.CL upper.CL
##  group 1 - group 2   0.5757 0.295 2454 -0.45488    1.606
##  group 1 - group 3   0.9285 0.372 2454 -0.37339    2.230
##  group 1 - group 4   0.9693 0.213 2454  0.22406    1.715
##  group 1 - group 5   0.3928 0.224 2454 -0.39095    1.177
##  group 1 - group 6   0.7509 0.327 2454 -0.39208    1.894
##  group 1 - group 7   0.6651 0.192 2454 -0.00594    1.336
##  group 2 - group 3   0.3528 0.431 2454 -1.15543    1.861
##  group 2 - group 4   0.3936 0.305 2454 -0.67188    1.459
##  group 2 - group 5  -0.1829 0.312 2454 -1.27567    0.910
##  group 2 - group 6   0.1753 0.393 2454 -1.19820    1.549
##  group 2 - group 7   0.0894 0.290 2454 -0.92557    1.104
##  group 3 - group 4   0.0408 0.380 2454 -1.28891    1.371
##  group 3 - group 5  -0.5357 0.386 2454 -1.88738    0.816
##  group 3 - group 6  -0.1776 0.454 2454 -1.76477    1.410
##  group 3 - group 7  -0.2634 0.369 2454 -1.55299    1.026
##  group 4 - group 5  -0.5765 0.237 2454 -1.40570    0.253
##  group 4 - group 6  -0.2184 0.336 2454 -1.39301    0.956
##  group 4 - group 7  -0.3042 0.207 2454 -1.02780    0.419
##  group 5 - group 6   0.3581 0.343 2454 -0.84126    1.558
##  group 5 - group 7   0.2723 0.218 2454 -0.49084    1.035
##  group 6 - group 7  -0.0858 0.323 2454 -1.21483    1.043
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "HMuFreq  pairwise CI's"
##  contrast           estimate      SE   df lower.CL  upper.CL
##  group 1 - group 2  0.001782 0.00367 2454 -0.01105  0.014613
##  group 1 - group 3  0.014768 0.00463 2454 -0.00144  0.030977
##  group 1 - group 4 -0.000904 0.00265 2454 -0.01018  0.008375
##  group 1 - group 5 -0.005563 0.00279 2454 -0.01532  0.004195
##  group 1 - group 6 -0.009779 0.00407 2454 -0.02401  0.004452
##  group 1 - group 7  0.013055 0.00239 2454  0.00470  0.021410
##  group 2 - group 3  0.012986 0.00537 2454 -0.00579  0.031765
##  group 2 - group 4 -0.002686 0.00379 2454 -0.01595  0.010580
##  group 2 - group 5 -0.007345 0.00389 2454 -0.02095  0.006261
##  group 2 - group 6 -0.011561 0.00489 2454 -0.02866  0.005539
##  group 2 - group 7  0.011273 0.00361 2454 -0.00136  0.023910
##  group 3 - group 4 -0.015672 0.00473 2454 -0.03223  0.000884
```

```
##  group 3 - group 5 -0.020331 0.00481 2454 -0.03716 -0.003502
##  group 3 - group 6 -0.024547 0.00565 2454 -0.04431 -0.004785
##  group 3 - group 7 -0.001713 0.00459 2454 -0.01777  0.014343
##  group 4 - group 5 -0.004659 0.00295 2454 -0.01498  0.005665
##  group 4 - group 6 -0.008875 0.00418 2454 -0.02350  0.005750
##  group 4 - group 7  0.013959 0.00258 2454  0.00495  0.022968
##  group 5 - group 6 -0.004216 0.00427 2454 -0.01915  0.010717
##  group 5 - group 7  0.018618 0.00272 2454  0.00912  0.028120
##  group 6 - group 7  0.022834 0.00402 2454  0.00878  0.036891
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "Binding  pairwise CI's"
##  contrast          estimate     SE   df lower.CL upper.CL
##  group 1 - group 2  -0.5394 0.0951 2454  -0.8721  -0.2066
##  group 1 - group 3   0.3560 0.1202 2454  -0.0644   0.7763
##  group 1 - group 4  -0.0201 0.0688 2454  -0.2608   0.2205
##  group 1 - group 5   0.0164 0.0723 2454  -0.2366   0.2695
##  group 1 - group 6   0.3819 0.1055 2454   0.0129   0.7510
##  group 1 - group 7  -0.4236 0.0619 2454  -0.6403  -0.2070
##  group 2 - group 3   0.8953 0.1392 2454   0.4083   1.3823
##  group 2 - group 4   0.5192 0.0983 2454   0.1752   0.8633
##  group 2 - group 5   0.5558 0.1009 2454   0.2029   0.9086
##  group 2 - group 6   0.9213 0.1268 2454   0.4778   1.3648
##  group 2 - group 7   0.1157 0.0937 2454  -0.2120   0.4434
##  group 3 - group 4  -0.3761 0.1227 2454  -0.8055   0.0532
##  group 3 - group 5  -0.3395 0.1248 2454  -0.7760   0.0969
##  group 3 - group 6   0.0259 0.1465 2454  -0.4866   0.5384
##  group 3 - group 7  -0.7796 0.1190 2454  -1.1960  -0.3632
##  group 4 - group 5   0.0366 0.0765 2454  -0.2311   0.3043
##  group 4 - group 6   0.4021 0.1084 2454   0.0228   0.7813
##  group 4 - group 7  -0.4035 0.0668 2454  -0.6371  -0.1699
##  group 5 - group 6   0.3655 0.1107 2454  -0.0218   0.7528
##  group 5 - group 7  -0.4401 0.0704 2454  -0.6865  -0.1937
##  group 6 - group 7  -0.8056 0.1042 2454  -1.1701  -0.4410
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [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
```

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.

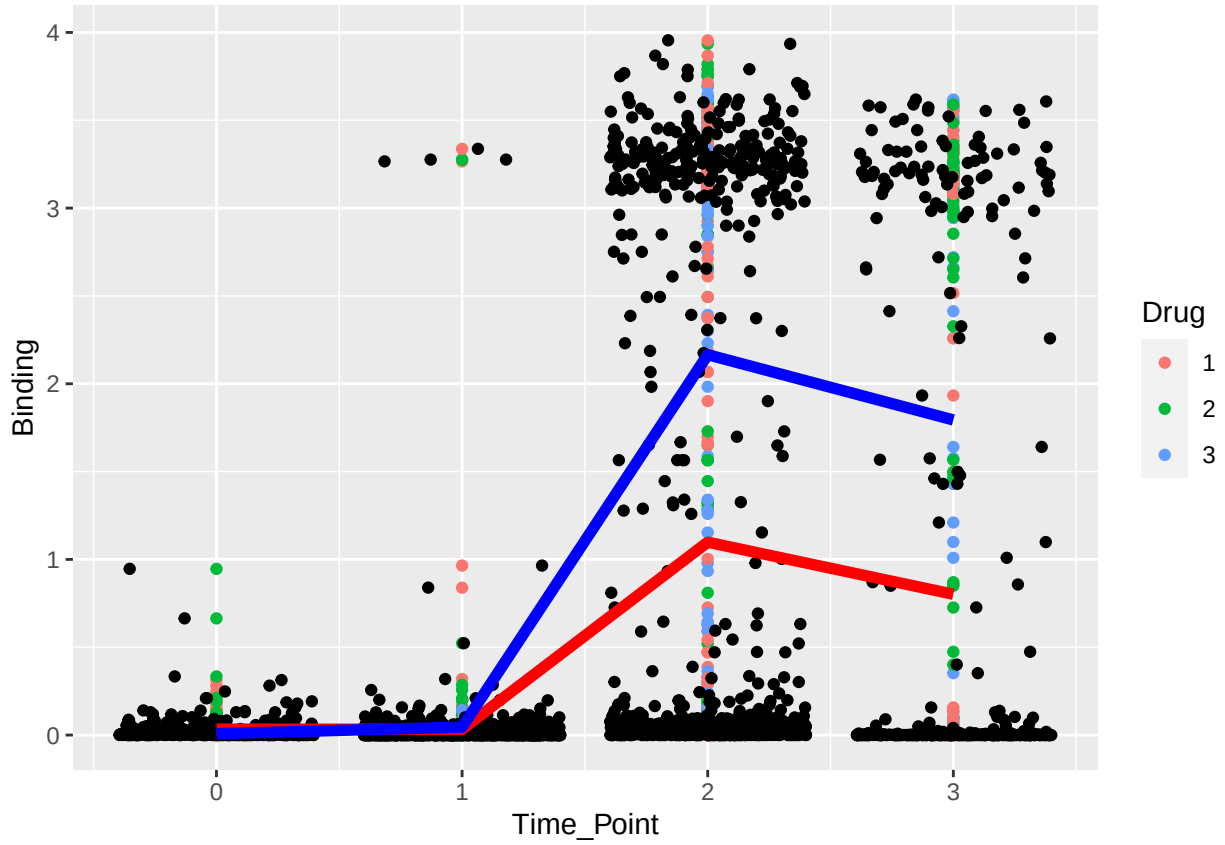
We could conclude that the drugs/treatment groups do increase mutation rate; however, they do not increase binding rate. That is to say, although the treatments do help increase the diversity of antibodies, they are not specific to the HIV antigens and thus do not increase binding.

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 bindign 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.

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

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

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

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

[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