Synthetic vs Native Antibodies

Group 6

Tiana Joseph, Neeha Chowdhury, and Kishan Ramrattan

Background/Hypothesis

Lyme disease is a tick-borne illness caused by spirochete *borrelia burgdorferi*, it produces, the ospC gene which is essential in producing an immune response because the antigen has a high affinity for binding. Therefore, we believe the cross-reactivity between the antigens and synthetic antibodies will be greater because the synthetic antibodies are better attuned to binding with the antigens. Antigens are molecules that cause antibodies to generate an immune response whenever foreign substances encounter the body. As for antibodies, these are molecules produced whenever a foreign response is received to help the antigen fight. Between the two types of antibodies used, there were native and synthetic ones. Native antibodies are antibodies that are formed within nature, while synthetic antibodies are antibodies formed through laboratory synthesis.

Hypothesis: The cross-reactivity between the antigens and synthetic antibodies will have a greater binding because the synthetic antibodies are better attuned to binding with the antigens.

Biological Question: Does the type of antibody influence cell binding?

Statistical Null Hypothesis: No effect (equal μ, no significant between the average optical density of the antibodies and antigens.

Methods

Dataset

	antigen	antibody	avg.od	Ab_class	class
1	O	anti-CT3	1.09321194	anti-CT3	synthetic
2	G	anti-CT3	0.98492037	anti-CT3	synthetic
3	T	anti-RT	0.91278214	anti-RT	synthetic
4	I	anti-CT3	0.87286921	anti-CT3	synthetic
5	K	anti-CT3	0.85383671	anti-CT3	synthetic
6	F	anti-RT	0.79537736	anti-RT	synthetic
7	I	anti-D	0.79209701	anti-Nat	native
8	B	anti-CT3	0.78333197	anti-CT3	synthetic
9	H	anti-RT	0.76514804	anti-RT	synthetic
10	T	anti-CT3	0.74418726	anti-CT3	synthetic
11	B	anti-CS	0.74362810	anti-CS	synthetic
12	D	anti-CT3	0.71934395	anti-CT3	synthetic
13	I	anti-CT1	0.71336870	anti-CT1	synthetic
14	A	anti-C	0.71168836	anti-Nat	native
15	F	anti-CT3	0.70427691	anti-CT3	synthetic
16	I	anti-RT	0.69416183	anti-RT	synthetic

Showing 1 to 17 of 164 entries, 5 total columns

This image is a portion of the data set used

The dataset contains the cross-reactivity values between the antigens and antibodies. There are five (5) columns and one hundred and sixty-four (164) rows. The columns are the antigen which is the ospC antigen. The antibody is the antibodies generated against the ospC antigens. The avg.od or average optical density contains the cross-reactivity values between the antigen and antibody pairing. The Ab-class or antibody class contains the class of antibodies, and the class contains the types of antibodies. All the columns contain characters except the avg.od which

contains numbers.

Data Visualization

The boxplot was used to compare the different types of antibodies and show the distribution of their average optical density based on the individual class of antibodies. Additionally, we used the violin plot to view the density of the data points for the avg.od grouped on the Ab_class. The two statistical methods used were the ANOVA test and the t-test. The ANOVA test was used to determine the statistical difference between the class of antibodies and the antibodies generated against the ospC antigens. We also used the t-test to determine the significance between the two groups of antigens (native and synthetic).

Statistical Methods

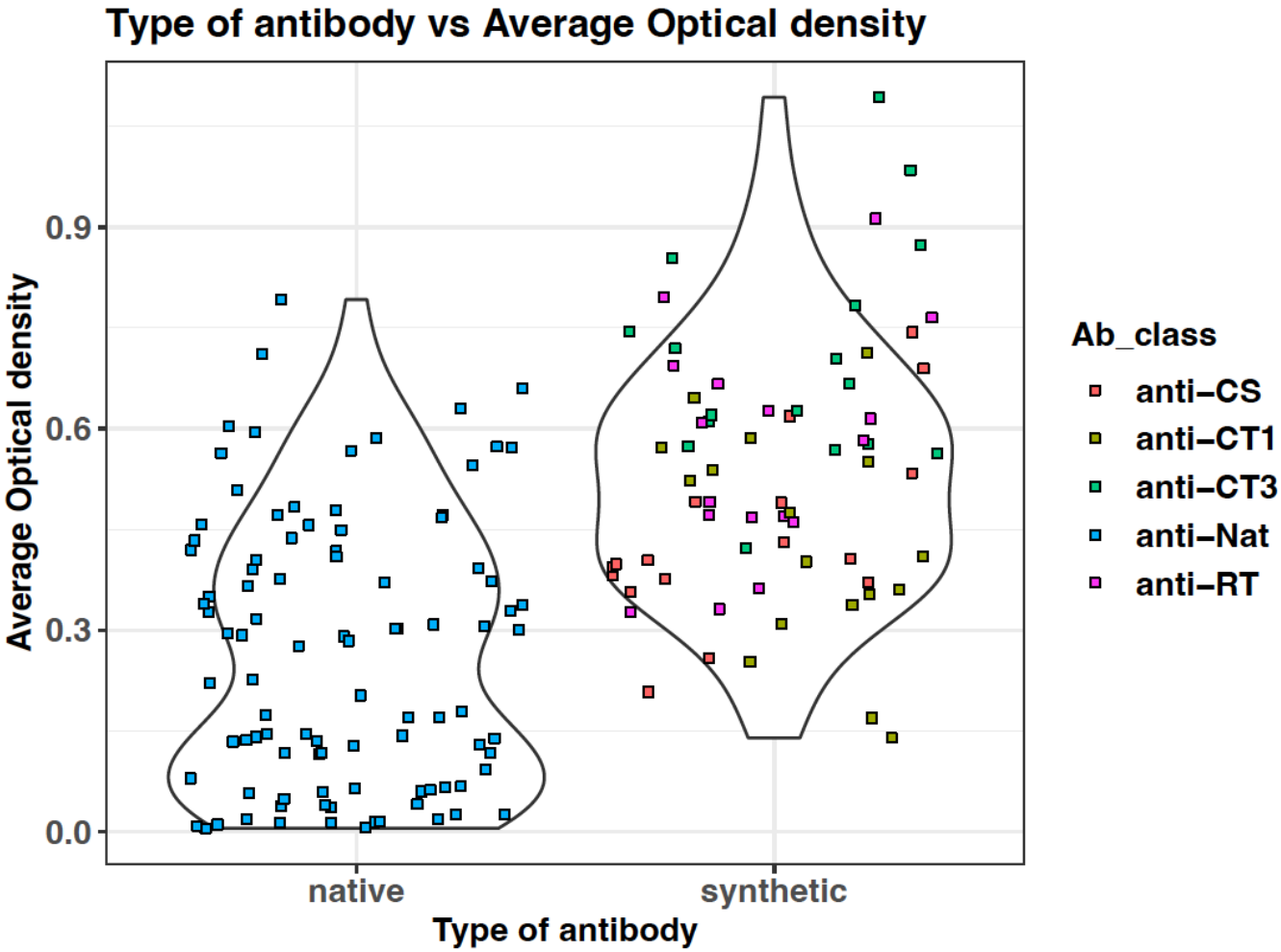
```
# reads the csv file
group6 = read_csv("group6_dataset.csv", show_col_types = FALSE)

# gives a brief summary of the structure of the dataset
glimpse(group6)

#select the distinct rows from a data set and deletes repeated information

#selects the types of antibody and antibodies generated against the ospC antigens

group6 %>% select(class,antibody) %>% distinct()


# x-axis is the class, and the y-axis is the average optical density

#geom_jitter was used to add the dots.

#theme_bw added a black and white background.

#fill and color were used to add color to the dot
```

```r
# lab was used to add titles to the axis and

#theme was used to edit the size and font of the text.


#violin plot
ggplot(group6,aes(x= class, y= avg.od)) + geom_violin() +
geom_jitter(aes(fill = Ab_class),shape=22) +theme_bw()  + labs(title =
"Type of antibody vs Average Optical density", y="Average Optical
density", x = "Type of antibody") + theme(text =
element_text(face="bold", size=12), axis.text =
element_text(face="bold", size=12), axis.title =
element_text(face="bold", size=12),
legend.text=element_text(face="bold", size=12),
legend.title=element_text(face="bold", size=12))

# The violin plot helps us to see where the data is more highly
concentrated.

#boxplot
ggplot(group6, aes(x= antibody, y= avg.od )) +
geom_boxplot(aes(fill=class), outlier.shape = NA, alpha= 0.3) +
geom_jitter(aes(fill=Ab_class),shape=21)  + theme_bw() + labs(title =
" Antibodies generated against ospC antigens vs Average Optical
Density", y="Average Optical density", x = "Types of Antibodies ") +
theme(text = element_text(face="bold", size=10), axis.text =
element_text(face="bold", size=10), axis.title =
element_text(face="bold", size=12), legend.text=
element_text(size=10), legend.title=element_blank(),
legend.position="bottom")

# The boxplot helps to visualize the data including high and low extremities.
This allows us to see where the data is spread out with specificity.


#test between the different types of antibodies
t.test(data=group6, avg.od ~ class)

#Anova for the class of antibody

model_Ab_class <- lm(data=group6, avg.od ~ Ab_class)

summary(model_Ab_class)

#Anova for antibody generated against the ospC antigen.

model_antibody <- lm(data=group6, avg.od ~ antibody)
summary(model_antibody)
```

Results

Figures and Captions

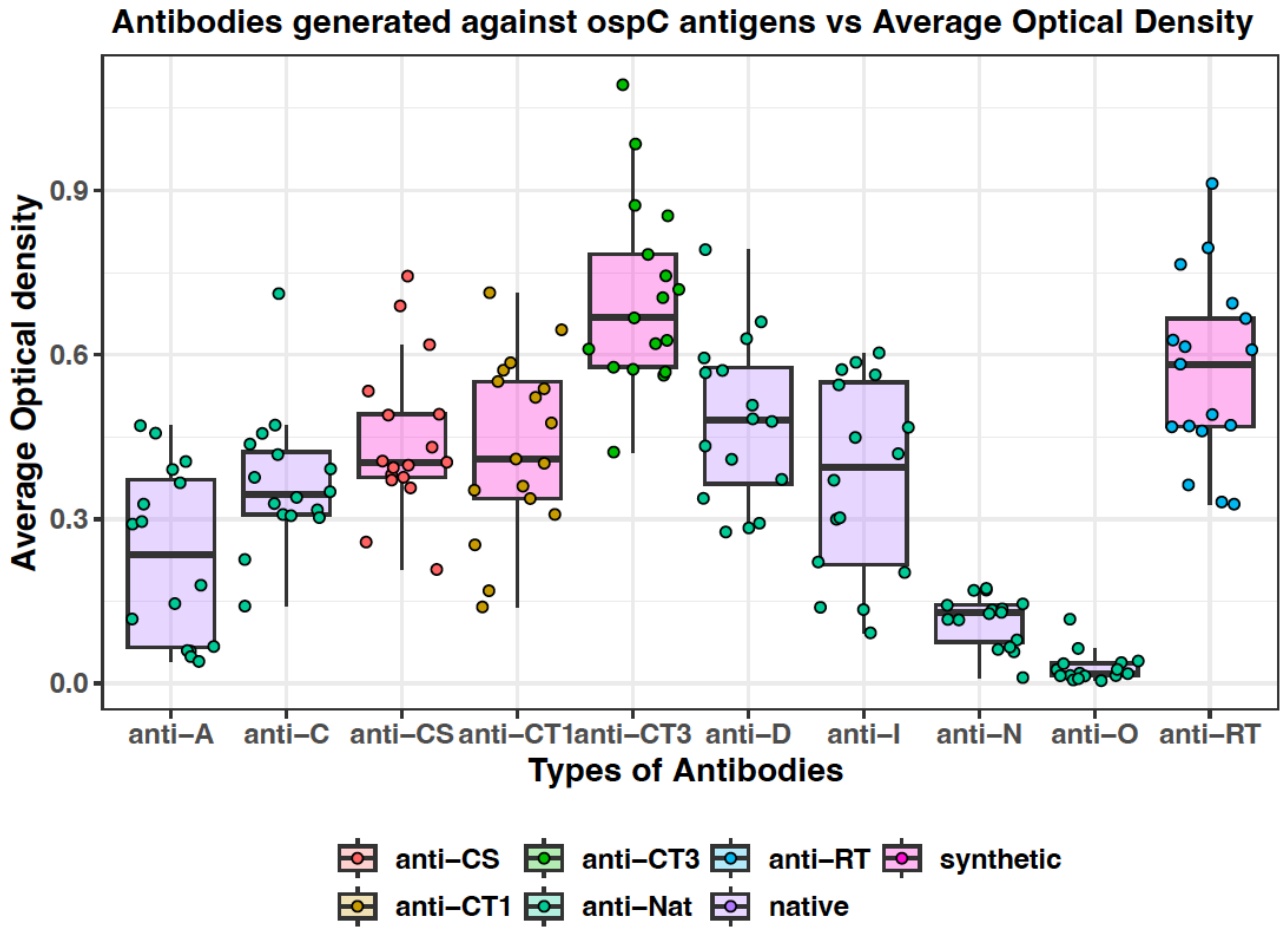
The violin plot shows the density of the cross-reactivity values against the types of antibodies and then colored by the class of antibodies.



The violin plot shows the concentration of the average optical density. The data points are distinguishable by color, and the blue points are all native antibodies, while the synthetic antibodies have a variation in the colors for the data points which highlight each antibody class. The density is in the center of the plot for the synthetic antibodies, and it is concentrated at a

higher point as opposed to the native.

The boxplot shows a comparison between the different types and classes of antibodies in relation to the cross-reactivity values.



The boxplot here shows all the values with more emphasis on the upper and lower extremities. The boxplot has been color-coded to distinguish different types of antibodies. The purple boxes are for native antibodies, and the pink is for synthetic antibodies. All the teal-colored points on the graph show one class of native antibodies, while the synthetic antibodies have a variation in color, which shows different classes of antibodies. Anti-CT3 is synthetic and has the highest average optical density, and anti-O is native and has the lowest. Most of the synthetic datasets

have a higher optical density as opposed to the native ones.

Statistical analysis

Two sample t-test

```
        Welch Two Sample t-test

data:  avg.od by class
t = -8.7205, df = 148.18, p-value = 5.207e-15
alternative hypothesis: true difference in means between group native and group synthetic is not equal to 0
95 percent confidence interval:
 -0.3322514 -0.2094901
sample estimates:
   mean in group native mean in group synthetic
              0.2662295              0.5371003
```

The t-test indicated that the synthetic antibody has the highest mean therefore it is expected that the synthetic type of antibody will have a greater binding. Since the p-value was 5.207e-15 and was less than 0.05, we rejected the null hypothesis and there is a significant difference between the two types of antibodies (native and synthetic).

One-Way ANOVA: Statistics-estimate of the class of antibody binding effects

```
Call:
lm(formula = avg.od ~ Ab_class, data = group6)

Residuals:
    Min      1Q  Median      3Q     Max
-0.2920 -0.1371 -0.0257  0.1296  0.5259

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.44422    0.04504   9.862  < 2e-16 ***
Ab_classanti-CT1 -0.01270    0.06370  -0.199 0.842225
Ab_classanti-CT3  0.26078    0.06370   4.094 6.74e-05 ***
Ab_classanti-Nat -0.17799    0.04887  -3.642 0.000365 ***
Ab_classanti-RT   0.12344    0.06370   1.938 0.054425 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1857 on 159 degrees of freedom
Multiple R-squared:  0.4062,    Adjusted R-squared:  0.3913
F-statistic: 27.19 on 4 and 159 DF,  p-value: < 2.2e-16
```

The ANOVA test indicated that the p-value is equal to 2x10^-16 with the synthetic antibody anti-CT3 and the native antibody anti-Nat both having a strong significance difference between the avg.od. However, anti-CT1 and anti-RT which are synthetic antibody has p-values are 0.842225 and 0.054425 which are more than or equal to 0.05 therefore there is no significant difference between them and avg.od. Additionally, the negative t-values show there is a significance in the p-values for anti-Nat, indicating a lower avg.od therefore it has a lower bind.

One-Way ANOVA: Statistics-estimate of the antibodies generated against the ospC antigens binding effects

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.29204 -0.08519 -0.00683  0.08707  0.38821

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.23255    0.03555   6.542 8.50e-10 ***
antibodyanti-C   0.13507    0.05027   2.687  0.00801 **
antibodyanti-CS  0.21167    0.04953   4.274 3.36e-05 ***
antibodyanti-CT1 0.19897    0.04953   4.017 9.18e-05 ***
antibodyanti-CT3 0.47245    0.04953   9.539  < 2e-16 ***
antibodyanti-D   0.24804    0.05027   4.934 2.07e-06 ***
antibodyanti-I   0.14059    0.05027   2.797  0.00582 **
antibodyanti-N  -0.11773    0.05027  -2.342  0.02047 *
antibodyanti-O  -0.20389    0.05027  -4.056 7.92e-05 ***
antibodyanti-RT  0.33511    0.04953   6.766 2.60e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1422 on 154 degrees of freedom
Multiple R-squared:  0.6629,    Adjusted R-squared:  0.6432
```

The ANOVA test indicated that the p-value is equal to 8.50 x 10^10 with all the antibodies having a strong significant difference between the types of antibodies and avg.od. However, native antibodies anti-C and anti-N don’t have a strong significance compared to the other antibodies. Additionally, anti-N and anti-O are also native antibodies, and they have values that are negative hence, the significance in p indicates there is a lower avg. od and a lower binding between the anti-N, and anti-O.

Conclusion

The boxplot showed that anti-CT3 has the highest binding while anti-O has the lowest binding against the ospC antigens. The violin plot demonstrated that the synthetic antibodies have many classes of antibodies, and they are concentrated at a higher point than the native antibodies. The t-test also indicated the synthetic antibody has a higher mean than the native antibodies. The p-value was 5.207e-15 which is less than 0.05, therefore we reject the null hypothesis, and this shows that there is a significant difference between the native and the synthetic antibodies. The first ANOVA test for the class of antibodies revealed that the synthetic antibody anti-CT3 and the native antibody anti-Nat both have a strong significance difference between the avg.od. However, it also indicated that the p-values for the synthetic antibody anti-CT1 and anti-RT showed no significant difference between avg.od. Therefore, we fail to reject our null hypothesis. However, the negative t-values and the significance of the p-values for anti-Nat, indicate a lower avg.od, therefore it has a lower bind. This conveys, that the native antibody has a lower binding against the native ospC antigens.

The second ANOVA test indicated there is a strong significance between the types of antibodies and the avg.od since the p-value was less than 0.05. Therefore, we reject the null hypothesis and this shows that there is a significant difference between the native and the synthetic antibodies. However, it also showed that anti-N and anti-O, native antibodies generated against the ospC antigens have a lower binding due to their negative t-values and significance p-value. Overall, our hypothesis was inconclusive because although there is a significant difference between avg.od of the native antibody and some of the synthetic antibodies. For the synthetic antibodies anti-CT1 and anti-RT, there is no significance between the avg.od. This suggests there

isn't enough evidence that synthetic antibodies will have a greater binding. Also, the class of antibody anti-Nat, indicates it has a lower bind, but only anti-N and anti-O antibodies generated against the ospC antigens have a lower bind. Therefore, the other native antibodies should have a high binding against the ospC antigens.