

**Inversions + Gene Expression:** I calculated the coefficient of variance (CV) for expression in a few different categories in Table 1 (inverted vs non-inverted, significant and non-significant inversions, and inverted and non-inverted within the ATCC strain). CV is higher in the non-inverted regions compared to inverted regions both overall and within the ATCC strain. Comparing the CV between expression in significant inversions (inversion alignment blocks that had a significant difference between the inverted and non-inverted sequences) and expression in non-significant inversions, we see that the significant inversions have a higher CV. So, when inversions cause a significant change in expression within that block, there is more variation? And looking at more long range impacts, inversions have less variation in expression than non-inverted regions?

I also did a a Fligner-Killeen test to test for homogeneity of variances (I think this is just variance and not CV). This is a non-parametric test. These tests say that the variation between each of the above tested groups is significantly different (Table 2). I found two tests that determine if two CVs are equal between different groups (Table 3). Based on these tests, it appears as though the CVs between inverted and non-inverted regions does not differ (in all blocks or when only considering the ATCC genome). There is a significant difference in CVs between the significant inverted alignment blocks (blocks that had a significant difference in expression between inverted and non-inverted sequences within the block) and the non-significant inverted alignment blocks. So, only some inversions have an impact on expression, and when they do, the CVs in these blocks is different from non-significant inverted alignment blocks. **What do you think of these results?**

**If you think I should be doing any other tests to look into how the variance in expression changes between inverted and non-inverted regions please let me know!**

I attempted to make some figures to try and visualize these results but I am not happy with either of them. Please ignore the aesthetics because these will be changed. Figure 1 shows the distribution of expression values in the inverted alignment blocks and the non-inverted alignment blocks. We run into the same issue here as in the substitutions paper with  $dN$ ,  $dS$ , and  $\omega$ . When put on a log scale, expression values of 0 are not included in the violin plot. Without the log scale, the graphs are unreadable. Figure 2 plots the mean values for expression in the inverted and non-inverted alignment blocks as well as the 95% confidence interval. I was planing on making whichever graph we choose for each of the various categories in Tables 1 and 2, perhaps putting all graphs into one figure. **Do you like either of these graphs? Can you think of a better way to display the difference in variance?**

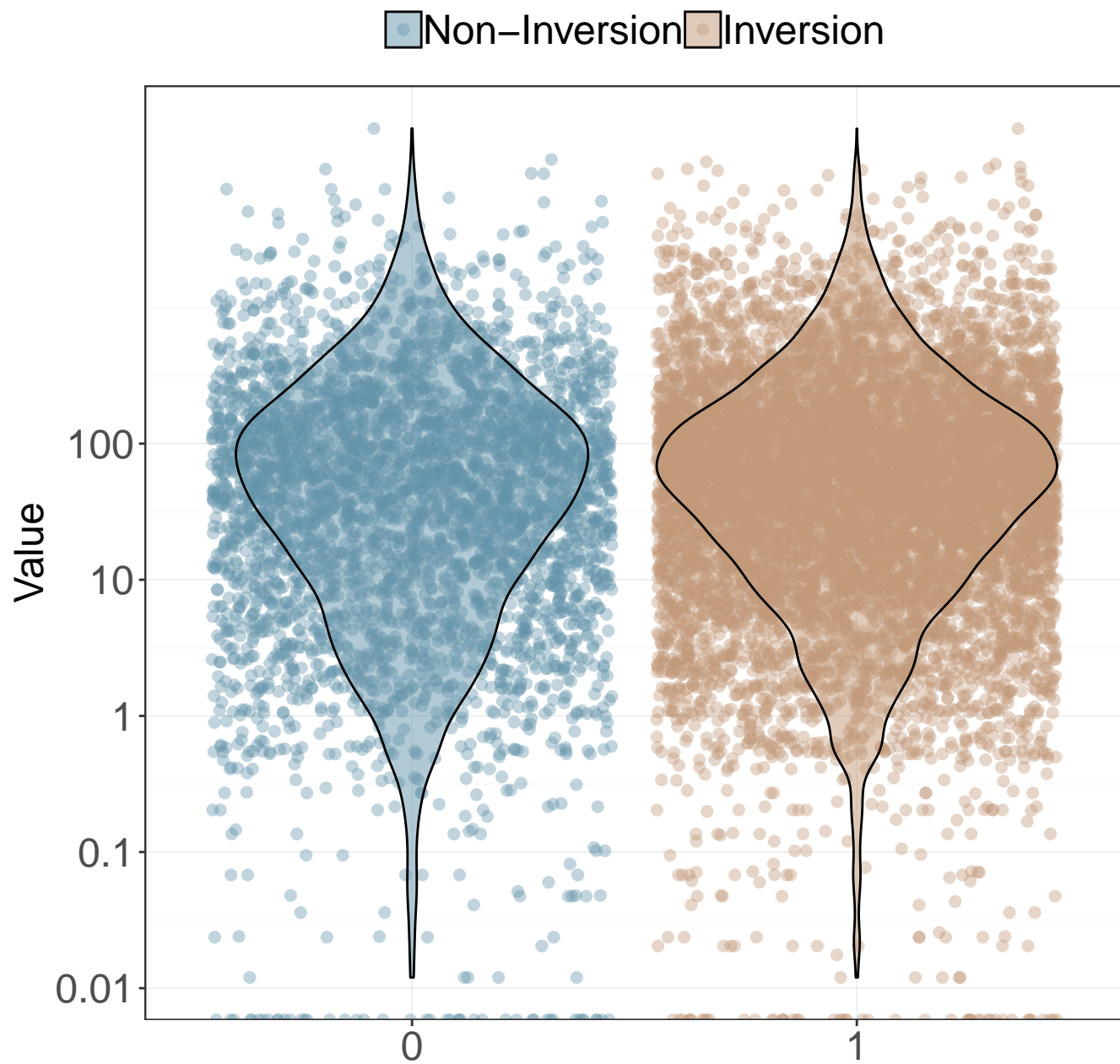


Figure 1: Violin plots of distribution of expression values between inverted and non-inverted alignment blocks.

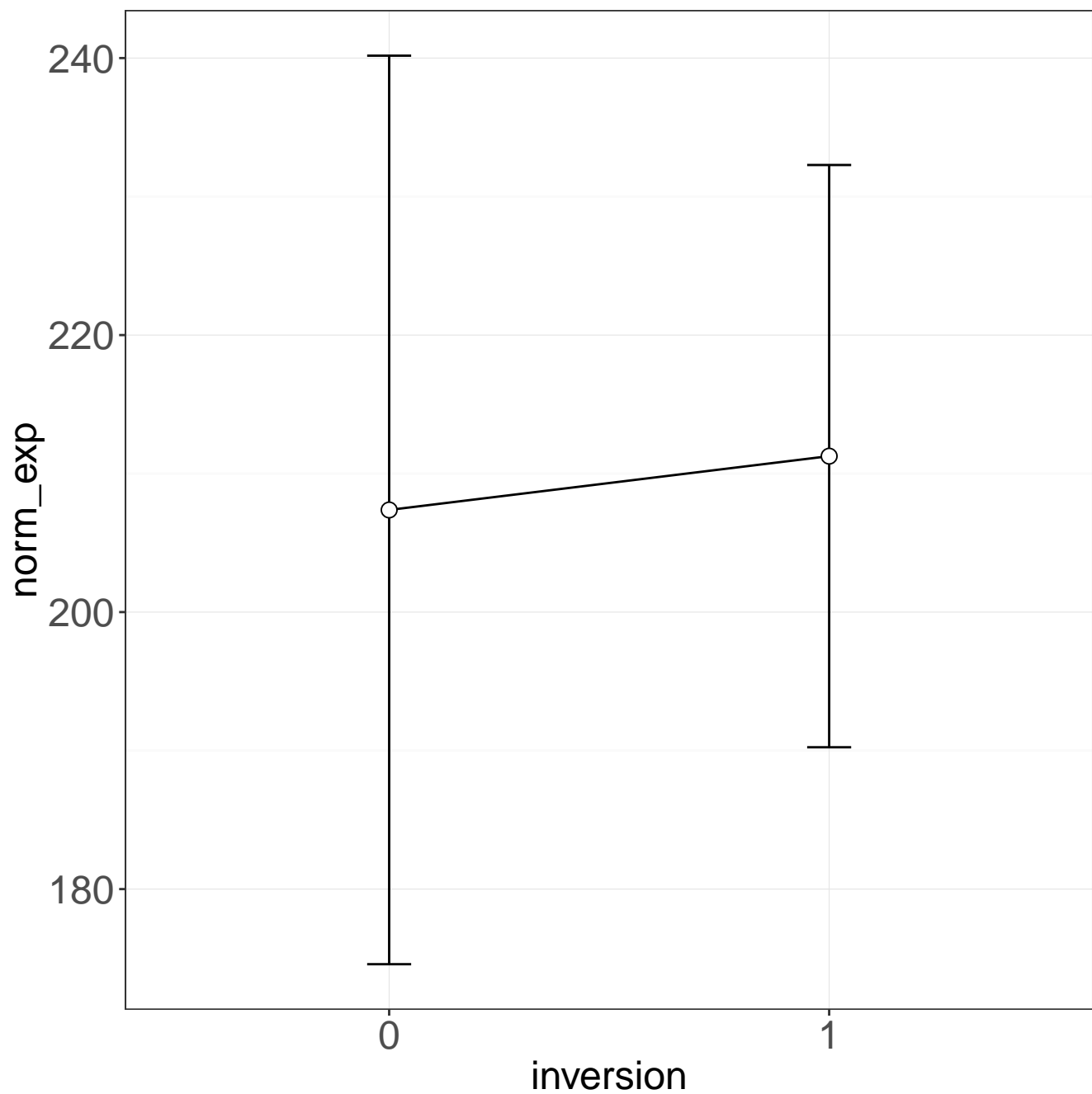


Figure 2: Plots of mean expression values between inverted and non-inverted alignment blocks with 95% confidence intervals.

Group	Inversions	Non-Inversions
All Blocks	3.26	3.43
Only ATCC genes	3.24	3.78
Group	Significant Inversions	Non-Significant Inversions
Significant Inversions	4.39	3.08

Table 1: Coefficient of variance in gene expression between different groups. “All Blocks” indicates all identified alignment blocks. “Only ATCC genes” indicates all ATCC genes that are both inverted and non-inverted. “Significant Inversions” indicates all inverted blocks that had a significant difference in gene expression between the inverted and non-inverted sequences. The coefficient variance in this group was calculated for the inversions that were significant inversions and non-significant inversions.

Group	chi-squared
All Blocks	6.005*
Only ATCC genes	6.000*
Significant Inversions	13.904***

Table 2: Fligner-Killeen test for homogeneity of variances in gene expression between different groups. “All Blocks” indicates all identified alignment blocks. “Only ATCC genes” indicates all ATCC genes that are both inverted and non-inverted. “Significant Inversions” indicates all inverted blocks that had a significant difference in gene expression between the inverted and non-inverted sequences. The coefficient variance in this group was calculated for the inversions that were significant inversions and non-significant inversions. All results are marked with significance codes as followed:  $< 0.001 = \text{***}$ ,  $0.001 < 0.01 = \text{**}$ ,  $0.01 < 0.05 = \text{*}$ ,  $> 0.05 = \text{NS}$ .

Group	Test Statistic	
	Asymptotic	M-SLRT
All Blocks	NS	NS
Only ATCC genes	NS	NS
Significant Inversions	8.738**	13.600***

Table 3: Tests for equality of coefficient of variances in gene expression. The “Asymptotic” test refers to the [cite Feltz and Miller 1996 here](#) asymptotic test. The “M-SLRT” test refers to the Modified Signed-Likelihood Ratio Test (M-SLRT) from [cite Krishnamoorthy and Lee 2014](#). “All Blocks” indicates all identified alignment blocks. “Only ATCC genes” indicates all ATCC genes that are both inverted and non-inverted. “Significant Inversions” indicates all inverted blocks that had a significant difference in gene expression between the inverted and non-inverted sequences. The coefficient variance in this group was calculated for the inversions that were significant inversions and non-significant inversions. All results are marked with significance codes as followed:  $< 0.001 = \text{***}$ ,  $0.001 < 0.01 = \text{**}$ ,  $0.01 < 0.05 = \text{*}$ ,  $> 0.05 = \text{NS}$ .