Inversions and Gene Expression Paper Revisions:

Papers looking at inversions across the genome

Naseeb et al. 2016

- Engineered various strains of yeast with a particular inversion (46Kb-800Kb) per strain to look at positioning of inversions, fitness, and expression. Mostly looking at impact of inversions involving the centromere or not to see how those impact fitness. Found large differences in expression (transcription) in inverted strains but no change in fitness (phenotype). Changes in expression were across genome (not just in inverted regions).
- Used limma (predecessor to DESeq) to look do a differential gene expression analysis (whole transcriptome, global expression profile). + FDR info Compared each gene inverted gene to it's controlled non-inverted gene (same gene because it was the same species). Visualized these difference with a volcano plot.
- Looked at genes within inversion and the closes 10, 20, and 40 genes on either side of the inversion.
- Used microarray and real time PCR on a subset of genes to validate transcriptome data.
- PCA to show strength in expression changes with the position of the inversion (in relation to the centromere I think)
- enrichment of deferentially expressed genes with in inverted regions checked using binomial and chi-squared tests. Found that DE genes are scattered around the genome rather than concentrated in inversion or near breakpoints.

Cui et al. 2012

- identified naturally occurring inversion (not bioinformatically determined) in *Staphylococcus aureus* that is reversible and causes change in phenotype.
- thought to be a bet hedging move to have some cells in population with one phenotype and some cells with another
- looked at expression and phenotype changes through DNA microarray and phenotypic array
- I think used PCR to see if recA was over expressed. Does not really mention any serious bioinformatics for gene expression methods

Alokam et al. 2002

- Identified existing inversions (500Kb - 700Kb) between *E. coli* and *Salmonella*, one spanning terminus

- these inversions persisted because they did not impact replication balance and replichore halves
- deleterious inversions would have been lost and not detected
- does not really look at expression at all, except to mention that gene dosage likely plays a role in not detecting inversions

1 Resampling (Permutations)

Since it has been so long since I looked at this project I wanted to just confirm that this is the approach we thought was best before I continue.

We have various block of alignment with various lengths. These blocks can therefore have varying numbers of genes in them.

From my notes, I am not sure which of the following options we decided on:

1. For all the block sizes, I will resample (without replacement?) individual columns (one nucleotide in the alignment = one column) to fill up the total length of the block while also keeping the total number of genes in that block the same.
2. For all block sizes, I will resample (without replacement?) entire genes (consecutive columns), to fill up the total number of genes within each block length? But then the resampled block could end up being longer than the original because not all genes are the same length.