* chip - 1A - replicate similarity heatmap - reads.pdf
  + Summary: Chip-seq data sets show high reproducibility. Additionally, the later two timepoints are much more similar to each other than to the initial timepoint, indicating that more variability in Groucho binding is contained in the first transition (1.5-4 hr -> 4-6.5 hr) than the second transition (4-6.5 -> 6.5 - 9 hr). (Groucho binding is more dynamic at earlier timepoints)
  + Method
    - Generated using ChIPQC (A Bioconductor package developed for the analysis and quality control of chip-seq data)
    - Calculates a correlation statistic based on binning normalized reads by peak. Differences in peak intensity between replicates and treatments will result in a lower test statistic.
* chip - 1B - replicate similarity heatmap - peaks.png
  + Summary: Chip-seq data sets show high reproducibilty.
  + Method
    - Generated using BEDTools (A command-line utility useful for the programmatic analysis and manipulation of genomic region data)
    - The correlation coefficient (Jaccard statistic) is calculated from only binding regions (as determined by the peak-calling software used, MACS2), ignoring peak intensity. A greater percentage of overlap between peaks will result in a higher correlation.
  + Conclusions
    - The heatmap pattern is slightly different from chip - 1A. Probably indicative of a different pattern of changes in peak position versus changes in peak intensity.
* chip - 1C - replicate similarity PCA.png
  + Summary: Principal component analysis of chip-seq replicates.
  + Method
    - Principal component analysis of each replicate, generated from the intensity (normalized reads) of overlapping peaks in each data set
* chip - 1D - replicate similarity Venn - peak overlap.pdf
  + Summary: Venn diagram of the number of overlapping peaks in each replicate for all three timepoints and Input
  + Method
    - Minimum overlap of 1 bp between replicates
* chip - 2 - Associated feature Venn.pdf
  + Summary: Overlap of potential Groucho targets identified in each dataset
  + Method
    - Generated a consensus peak set for each timepoint (peaks present in both replicate)
    - Identified closest feature to each peak
    - Removed duplicate features (genes with mulitple associated peaks are only included once)
* chip - 3A - Overlapping peak Venn.pdf
  + Summary: Overlap of peaks between datasets (potential Groucho binding sites)
  + Method
    - Generated a consensus peak set for each timepoint (peaks present in both replicates)
    - Removed peaks overlapping peaks identified in Input (presumably artifactual)
    - Calculated whether each peak overlaps consensus peaks in all other datasets
  + Conclusions
    - Interestingly, there are no common peaks to the timepoint 1 and 3 datasets that are not also identified in timepoint 2. This indicates there are no identified Groucho binding sites that are bound at early stages, lose Groucho binding, and then gain it again.
    - There are a significant number of peaks that remain bound through all timepoints, however. I refer to these as constitutive binding sites, though it's impossible to say if Groucho is binding at these sites in all cell types, as the chip-seq data is an embryo-wide average.
* chip - 3B - Overlapping peak venn with input.png
  + Summary: Similar to the previous Venn diagram, but includes peaks identified in the Input samples
  + Method
  + Conclusions
    - Most peaks overlapping with Input are present in all three Gro datasets, leading me to believe they are false-positives
* chip - 4A - Nearest feature dotchart.pdf
  + Summary: Dotchart of Groucho binding region localizaiton relative to the nearest gene
  + Method
    - Generated using ChIPpeakAnno (A Bioconductor package designed fo the annotation of chip-seq binding data)
  + Conclusions
* chip - 4B - Nearest feature piechart.pdf
  + Summary: Identical data to the dotchart above. Pie charts are generally considered problematic as humans are genreally incapable of making accurate comparisons between the areas of irregular shapes. So the dotchart is probably preferable. I don't really have a strong opinion on this however. Ane pie charts are very popular.
* chip - 5 - Genomic Feature Enrichment.png
  + Summary: Heatmap of enrichment of various types of features for reads. Yellow indicates enrichment, blue indicates depletion (below average levels if distribution of reads was random)
  + Method
    - Generated with ChIPQC
  + Conclusions
    - Strong enrichment in 5' UTRs, though this may be due to these areas of the genome being more accessible in actively transcribed genes, so chip-seq naturally generates above average levels of reads from these areas
    - Slight enrichment of reads in promoter regions, with lower levels/depletion furthur upstream of genes
    - Slight enrichment in introns
* chip - 6 - Average peak width boxplot.pdf
  + Summary: A boxplot indicating average peak width in each sample and input. Shows an increasing trend in binding site size through development, with an average peak size of 500 - 750 bp
  + Method
  + Conclusions
    - Slight increase in peak size during devleopment.
      * Difficult to say if this is biologically relevant or a product of slight variability in library preparation
    - Should probably remove input from the graph as it's not very informative