



INTEGRATIVE ANALYSIS OF NEXT GENERATION SEQUENCING DATA

ABSCHLUSSPRÄSENTATION ZUM SOFTWAREPROJEKT
„FUNCTIONAL GENOMICS“

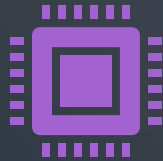
Gruppe 1:

Maximilian Otto, Lucas Rieckert, Kevin Zidane

STRUCTURE:



1. Introduction



2. Methods



3. Results




4. Conclusion

INTRODUCTION


- Main goals:
 - Identifying correlation between histone modification profiles (HMP) and gene expression
 - Create a prediction model for the expression of genes
- Data: ChIP-Seq, ATAC-Seq, RNA-Seq
- Samples:
 - Tissue types: limb, forebrain and lung tissue
 - Development states: cells from embryos, 14.5 days post conception (E14.5d)
cells from completely developed mice, postnatal (P0)
- Considered marks: H3K4me3, H3K4me1, H3K27ac, H3K36me3,
H3K9me3, H3K27me3

METHODS

- Data collection and quality verification
 - ENCODE Database 
 - FastQC
- Univariate analyses
 - Mapping
 - Peak calling
 - Correlation between histone marks
 - Characterization of H3K27ac-peaks

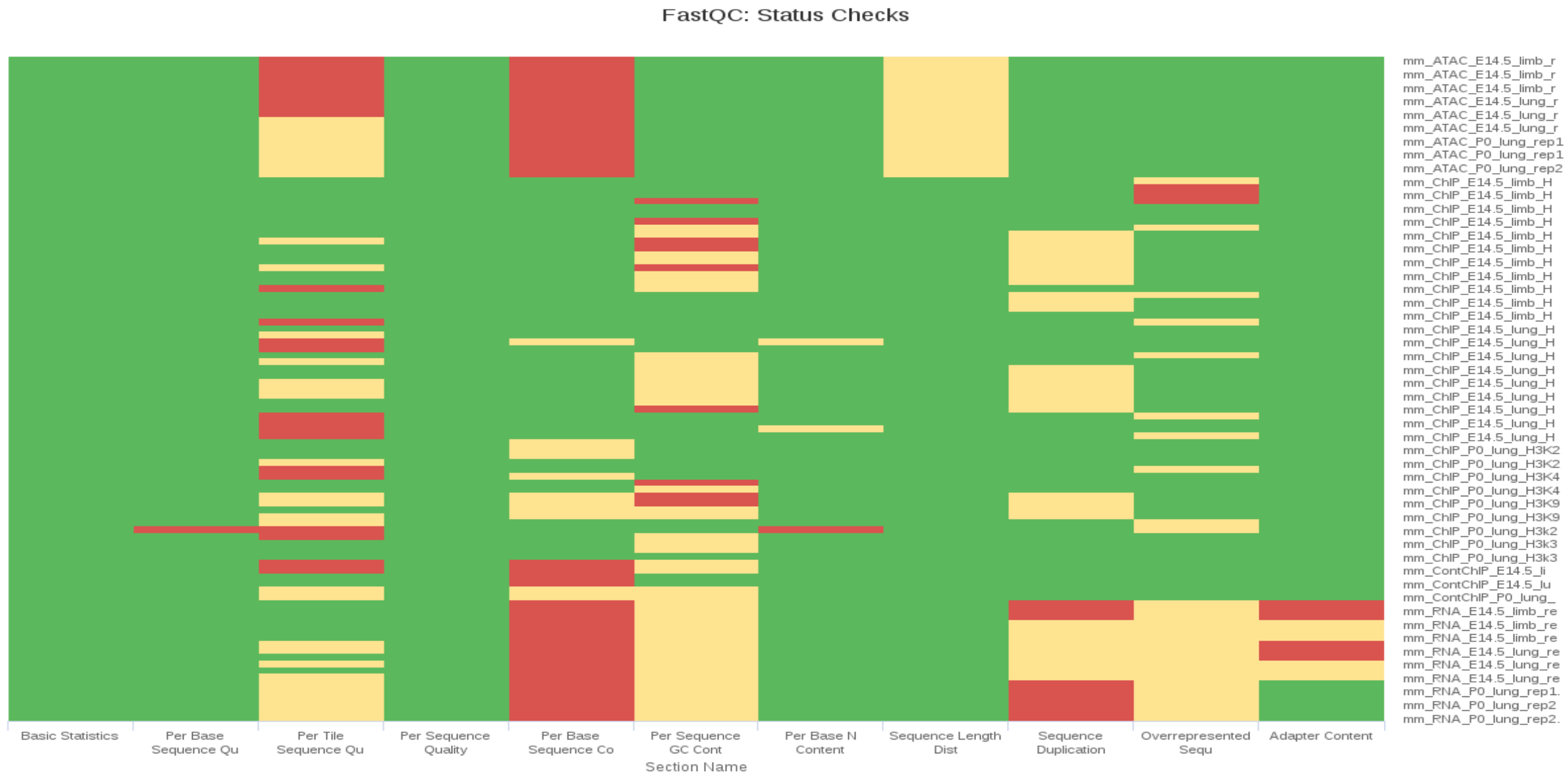


METHODS

- Multivariate analyses with ChromHMM
 - Differential gene expression analyses
 - Multivariate statistical prediction model
- 

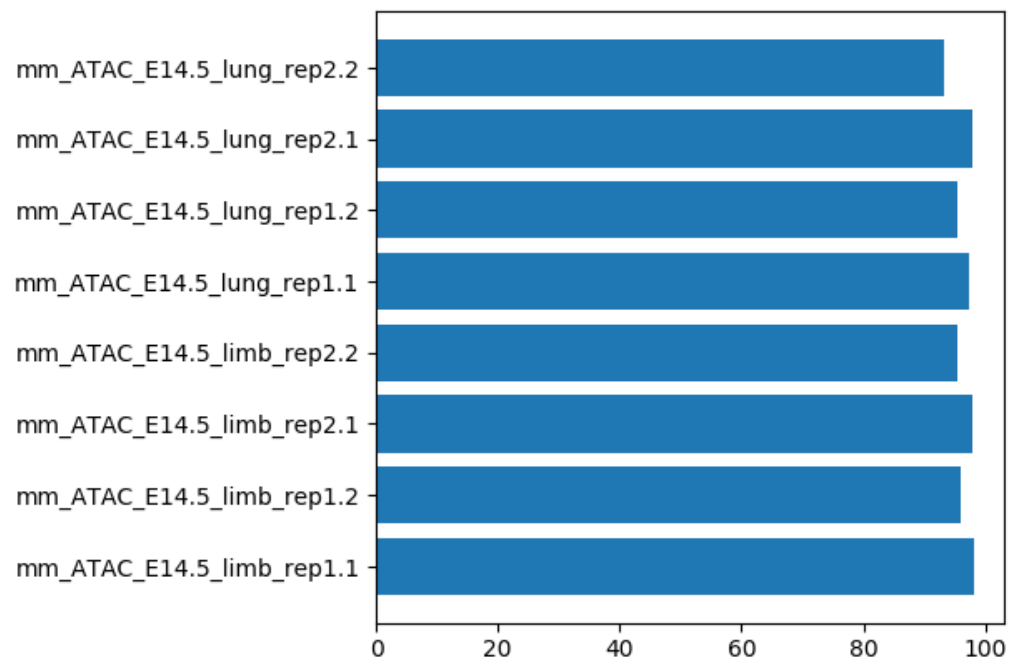


RESULTS: DATA QUALITY

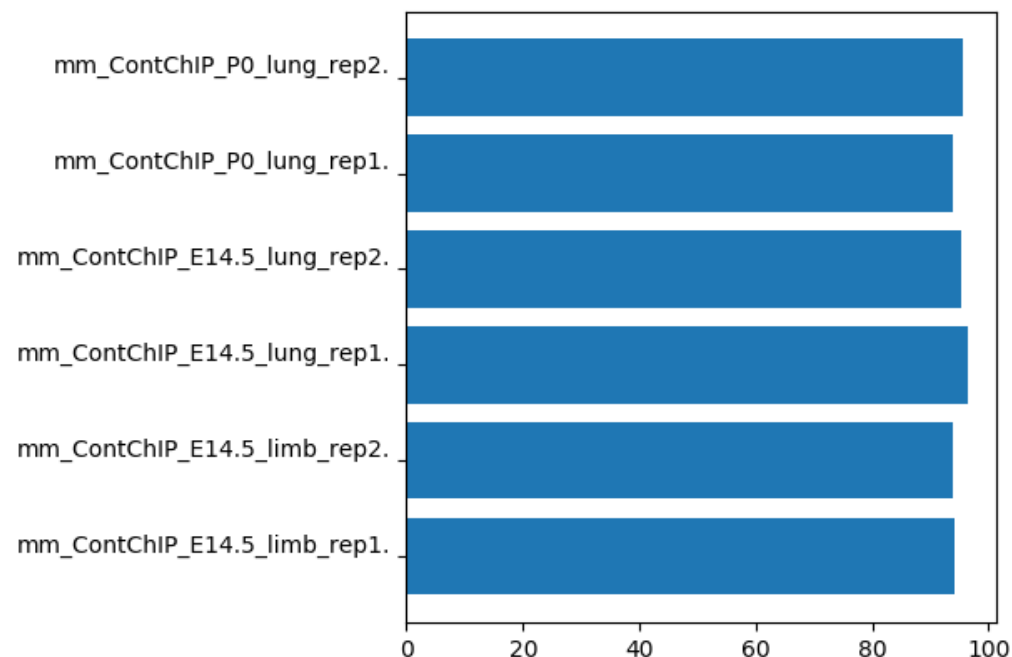


RESULTS: UNIVARIATE ANALYSIS

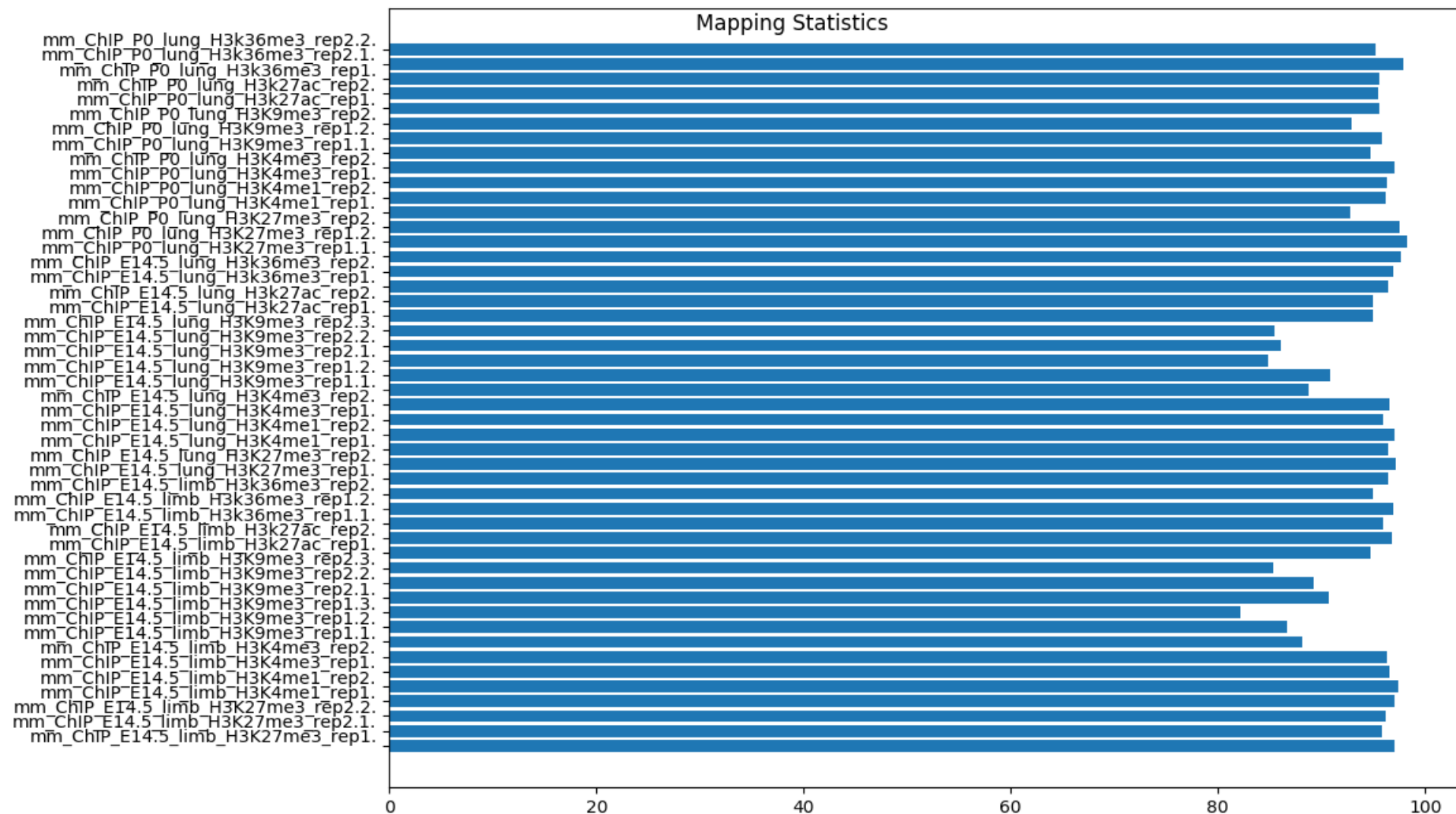
Mapping Statistics



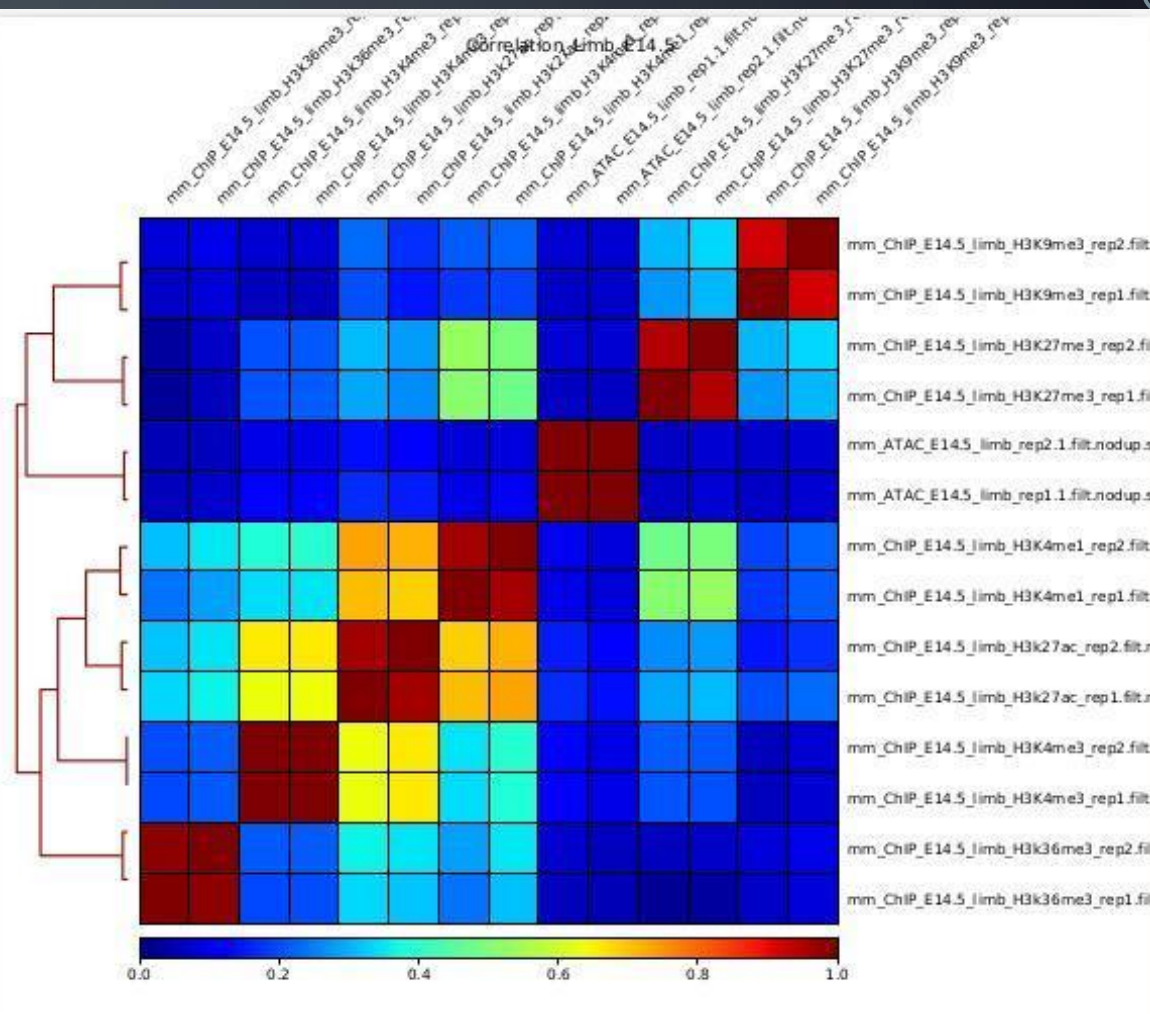
Mapping Statistics



RESULTS: UNIVARIATE ANALYSIS

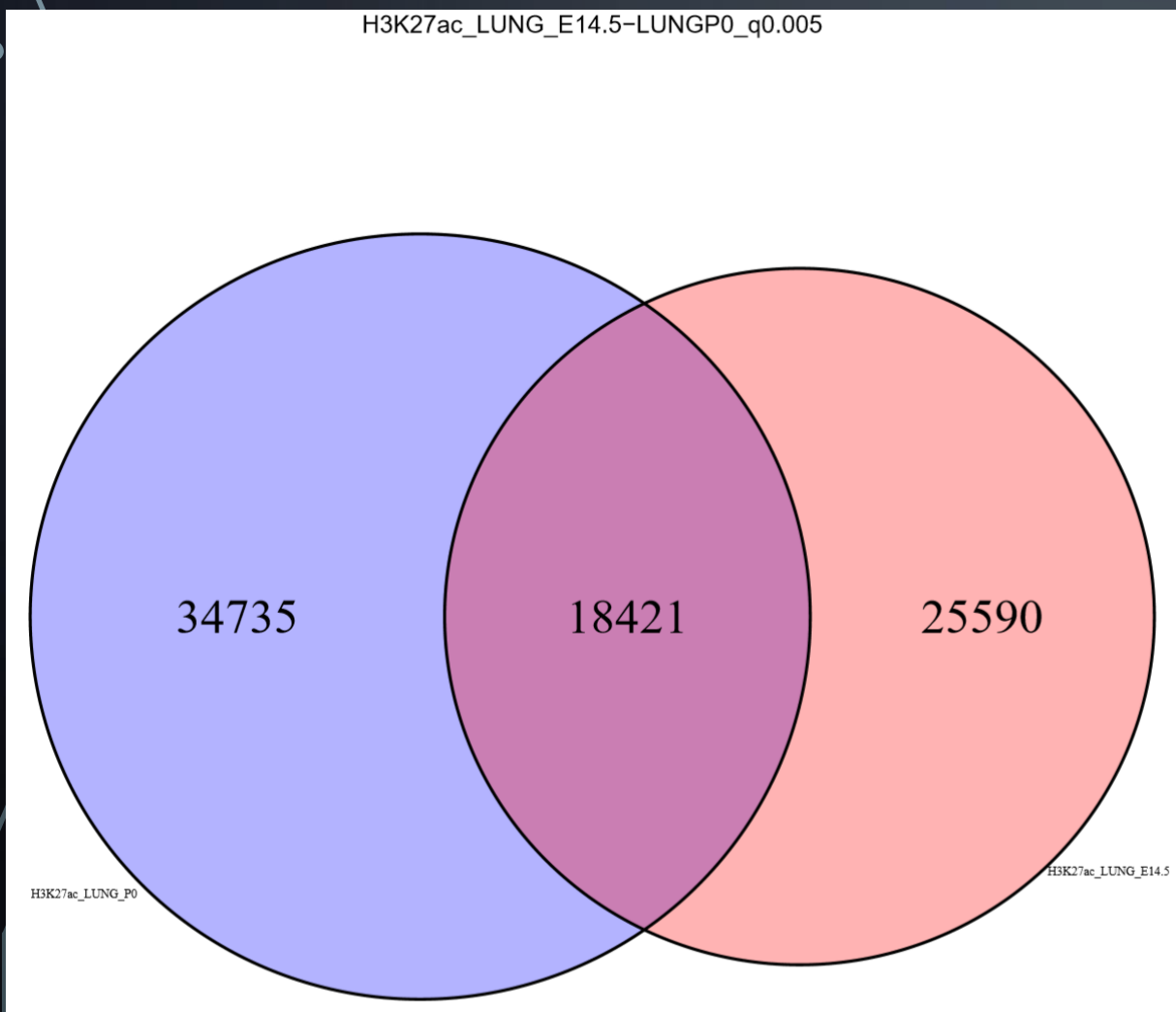


Correlation between histone marks:



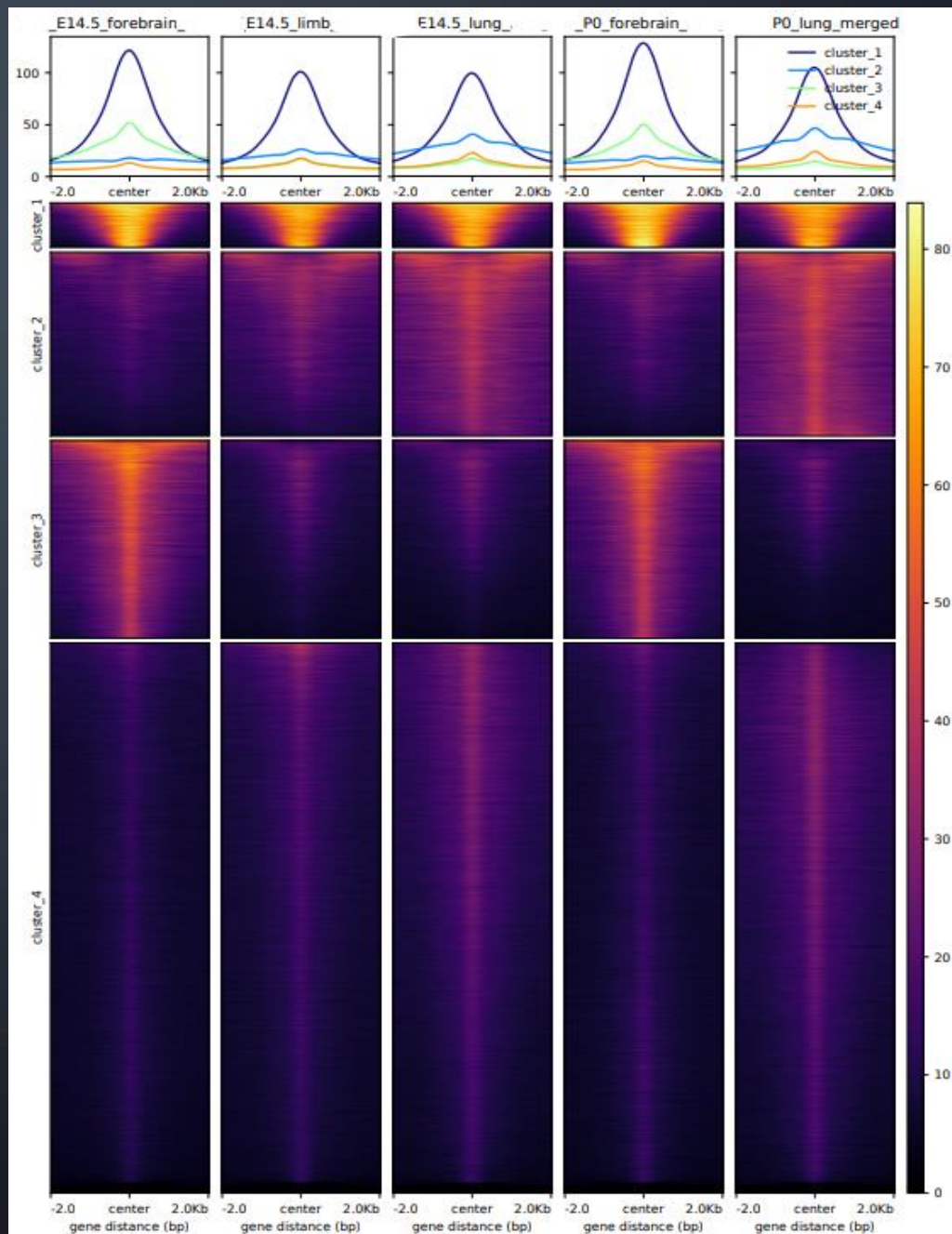
RESULTS: UNIVARIATE ANALYSIS

Surveying of the amount of common peaks:

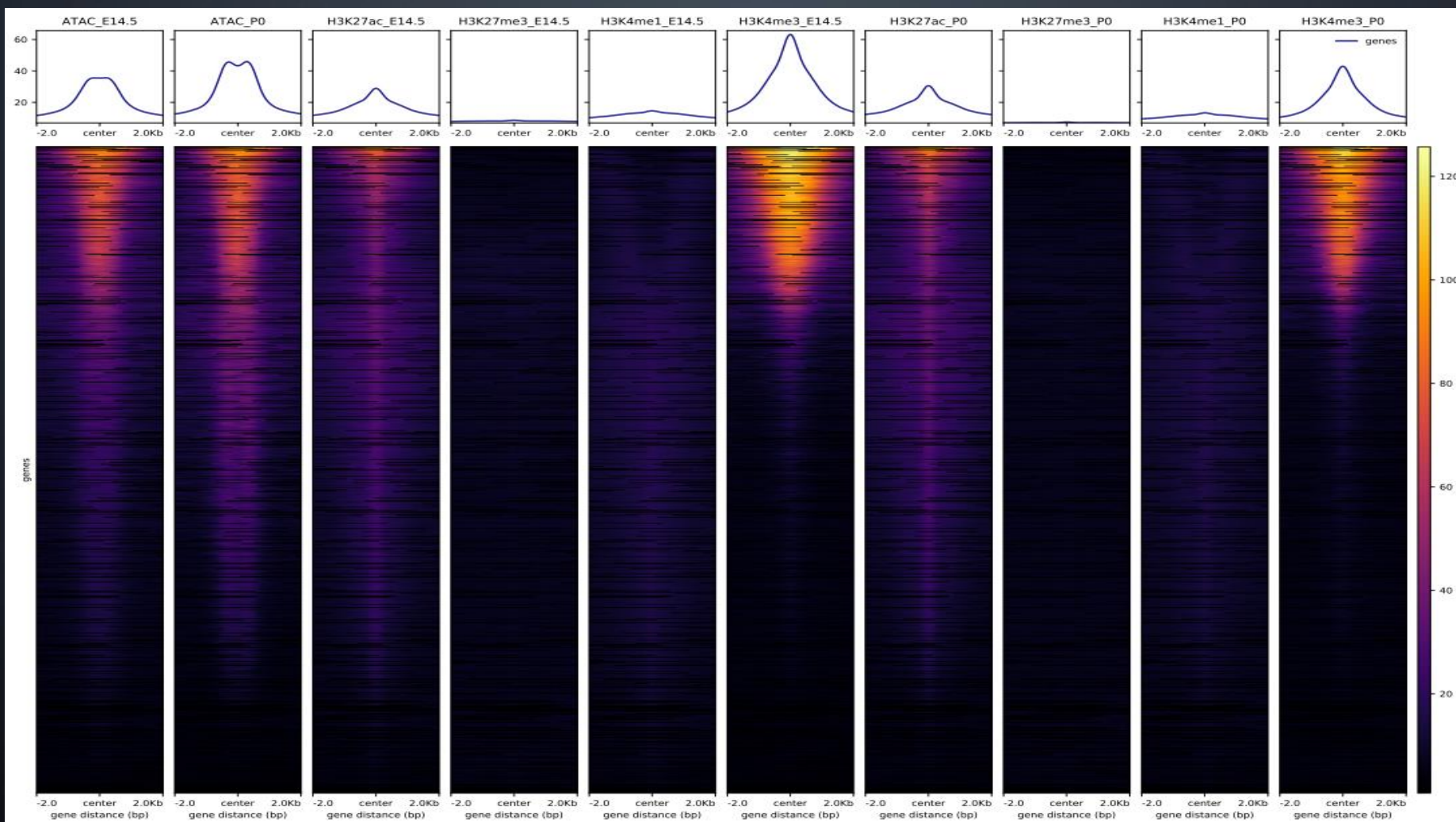


RESULTS: UNIVARIATE ANALYSIS

Clustered heatmap of read coverage in common H3K27ac peaks of limb, lung and forebrain at E14.5 and P0 apart from limb in P0. We can observe a cluster for forebrain and lung, but less for limb.



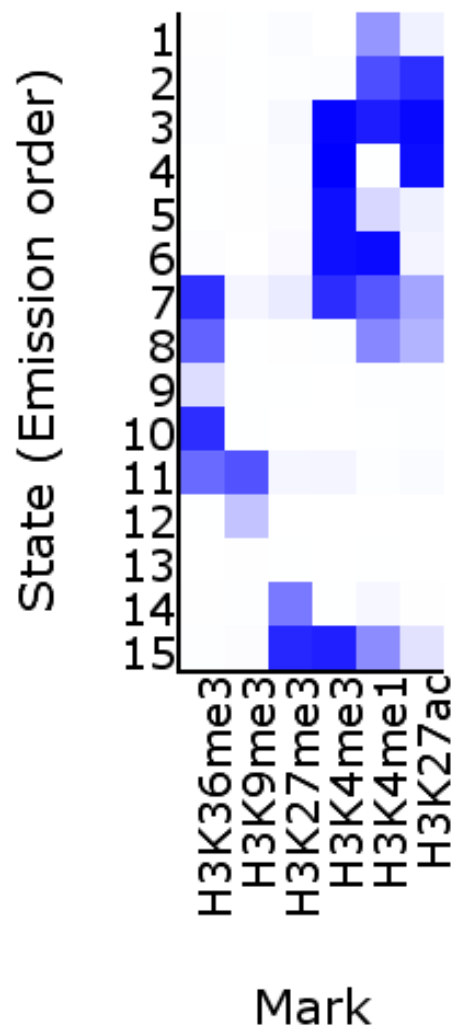
RESULTS: UNIVARIATE ANALYSIS



Heatmap of read coverage of ATAC-SEQ, H3K27ac, H3K27me3, H3K4me1 and H3K4me3 in lung at regions of H3K27ac-peaks at E14.5 and P0.

RESULTS: UNIVARIATE ANALYSIS

Emission Parameters



ChromHMM's obtained states:

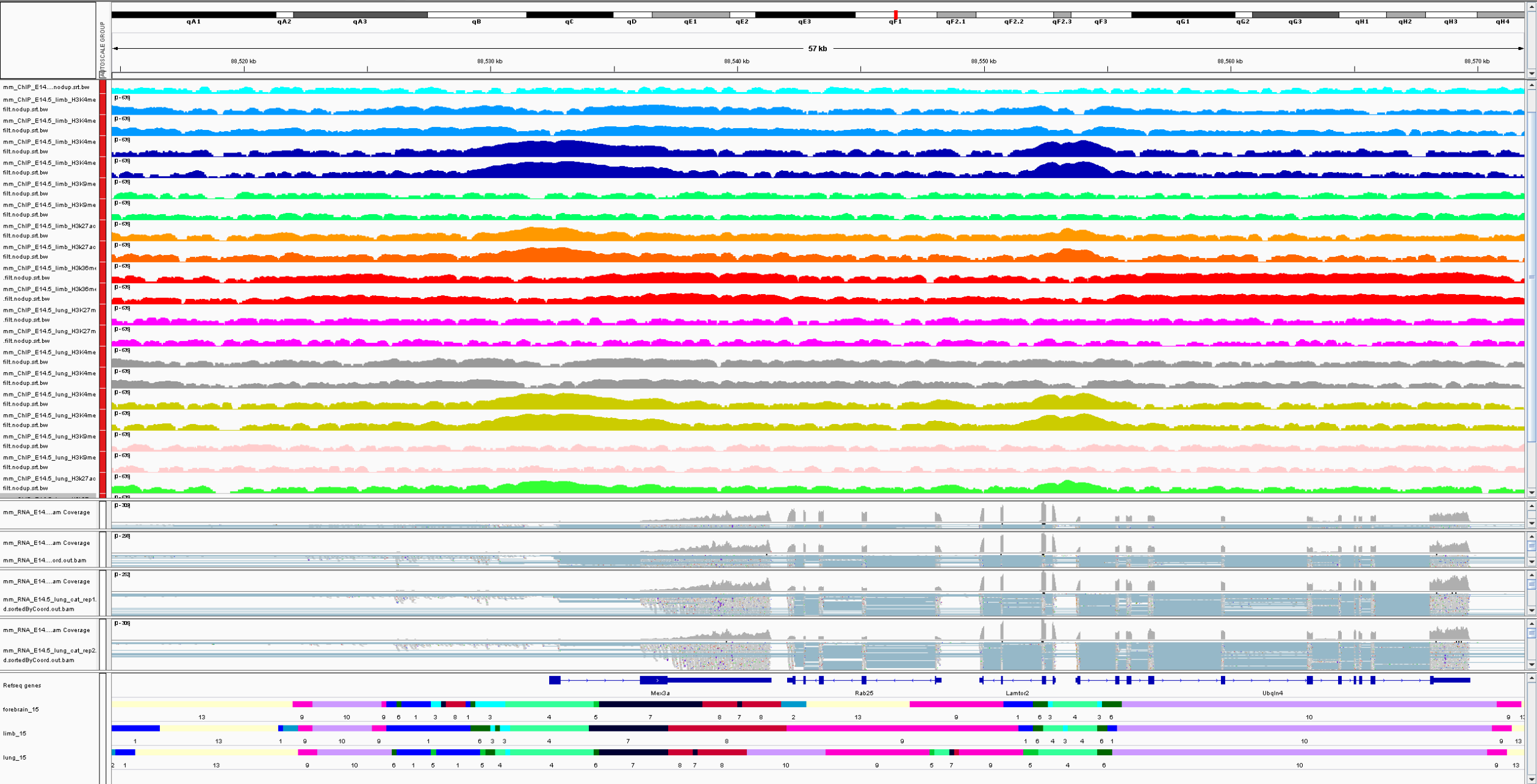
- 1 = weak TSS dist. - Enhancer
- 2 = poised, TSS-dist. - Enhancer
- 3 = Strong, TSS-dist - Enhancer
- 4 = active TSS
- 5 = Poised, TSS-prox - Enhancer
- 6 = weak inactive - Promoter
- 7 = Strong, TSS-prox - Enhancer, Transk. for 5'-3'
- 8 = Initiation - Transkription
- 9 = Strong - Transkription
- 10 = H3K36me3 associated, but unknown functionality
- 11 = unknown functionality
- 12 = H3K36me3 associated - Heterochromatin
- 13 = no Signal
- 14 = Polycomb associated - Heterochromatin
- 15 = Bivalent - Promoter

Mouse (mm10)

chr3

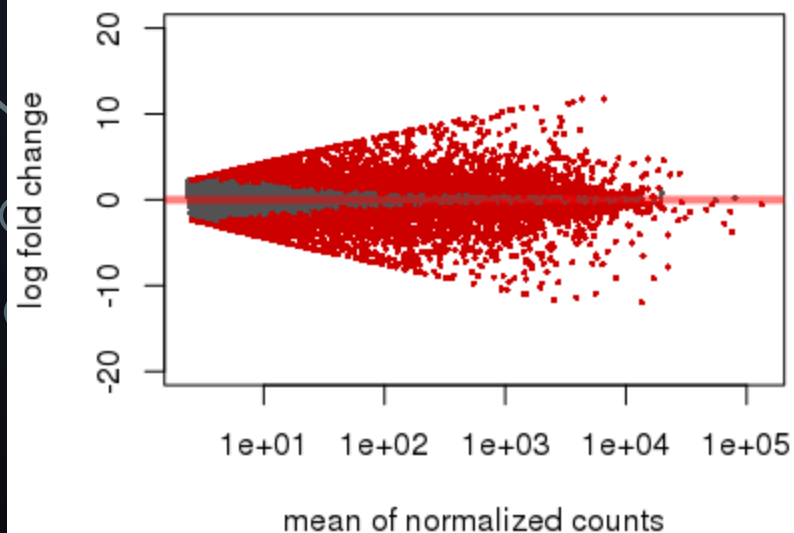
chr3:88,514,637-88,572,083

Go

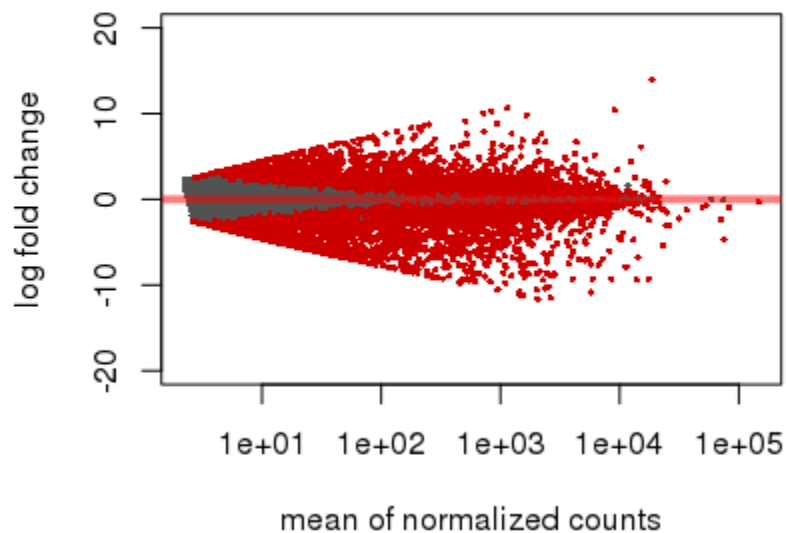


RESULTS: DIFFERENTIAL GENE ANALYSIS

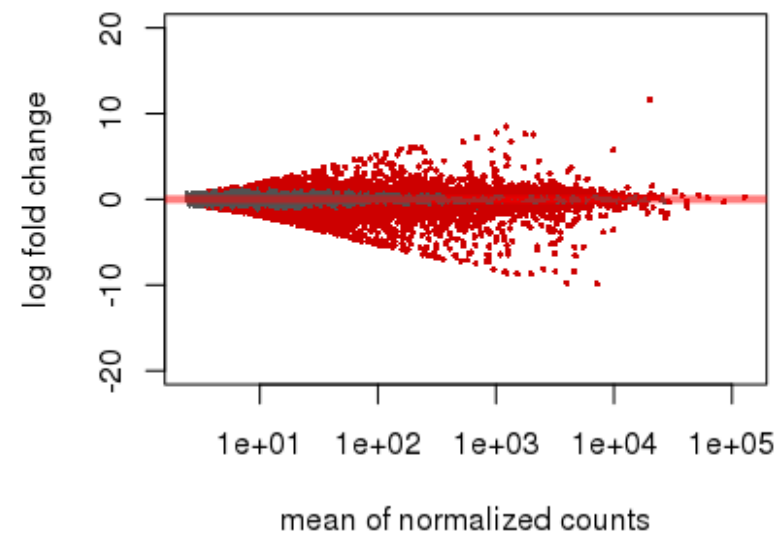
forebrain vs limb (normal)



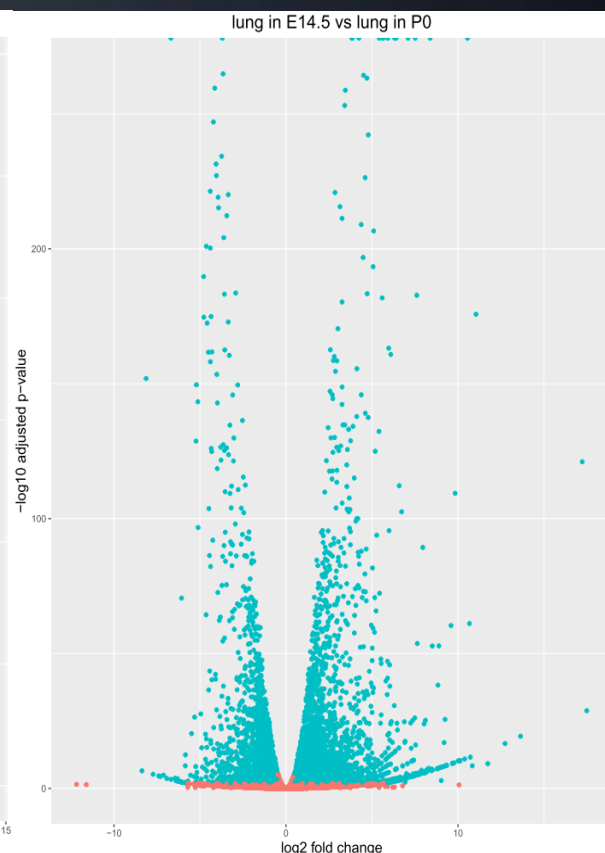
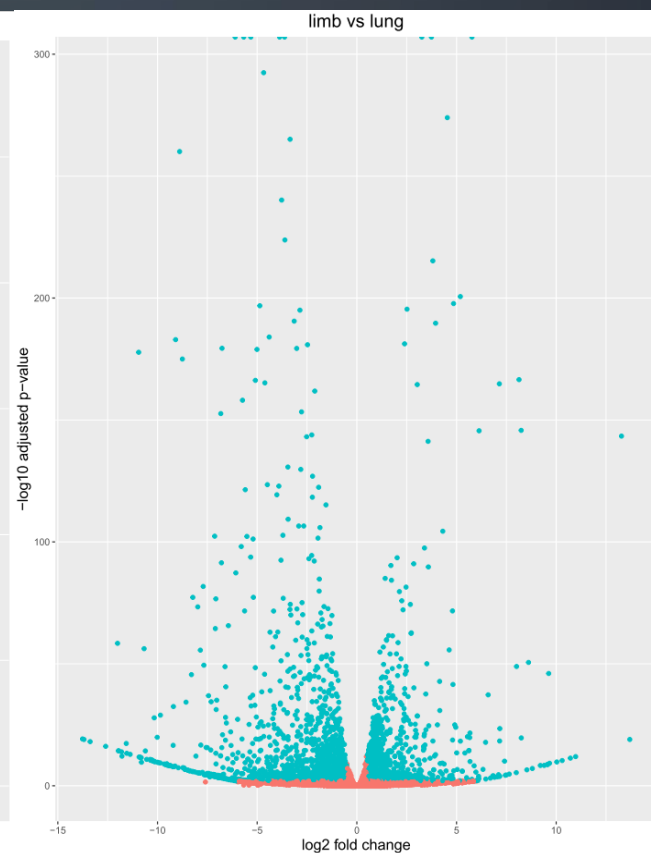
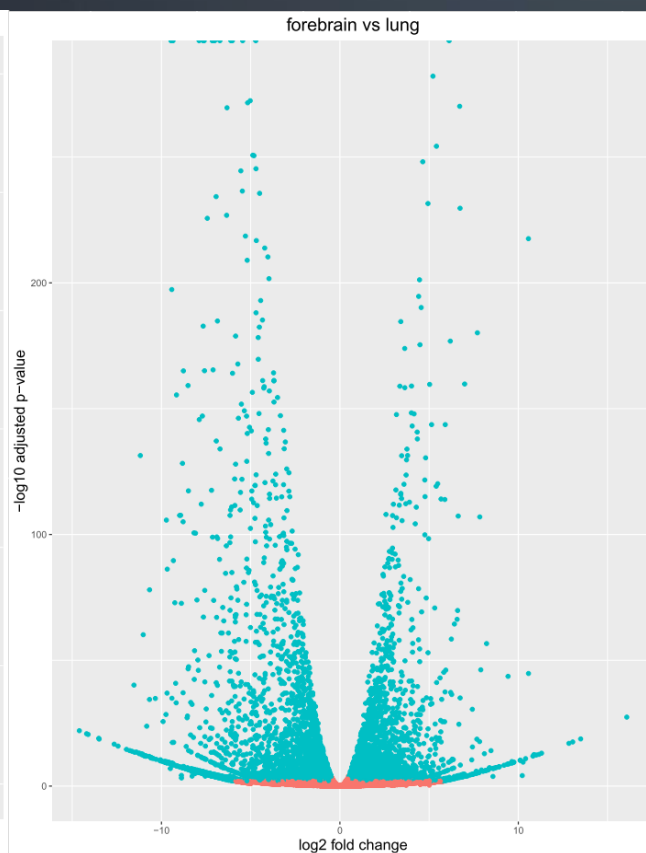
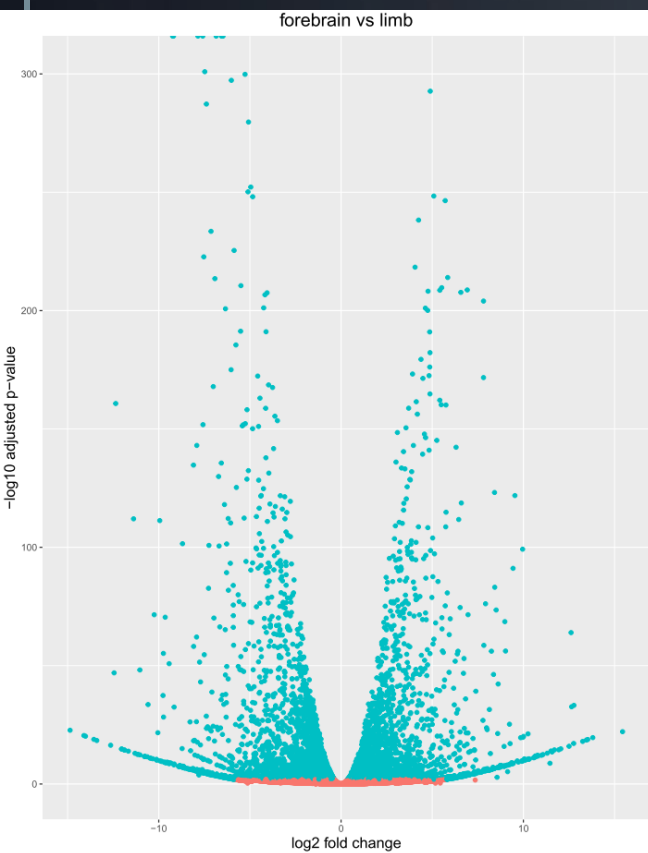
forebrain vs lung (normal)



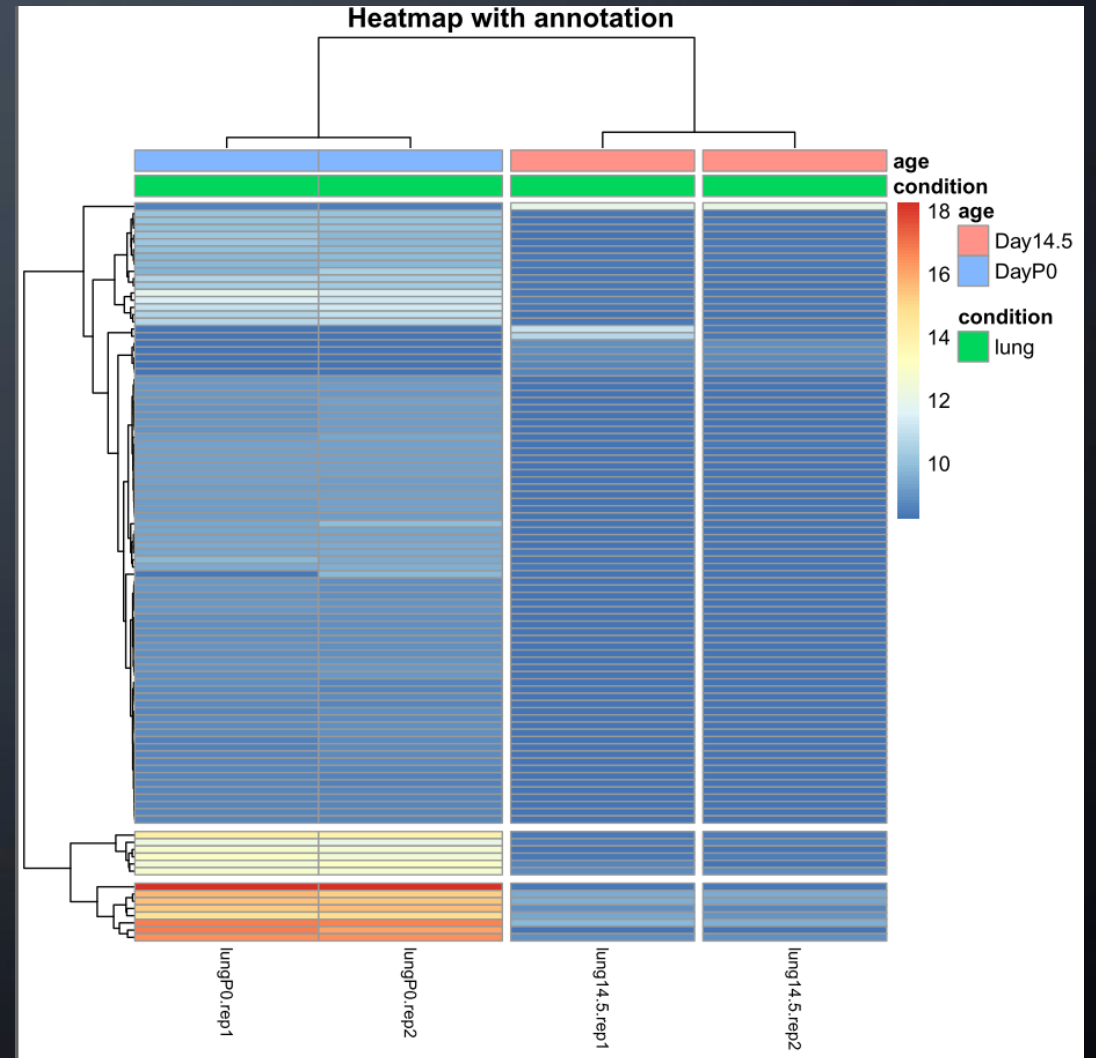
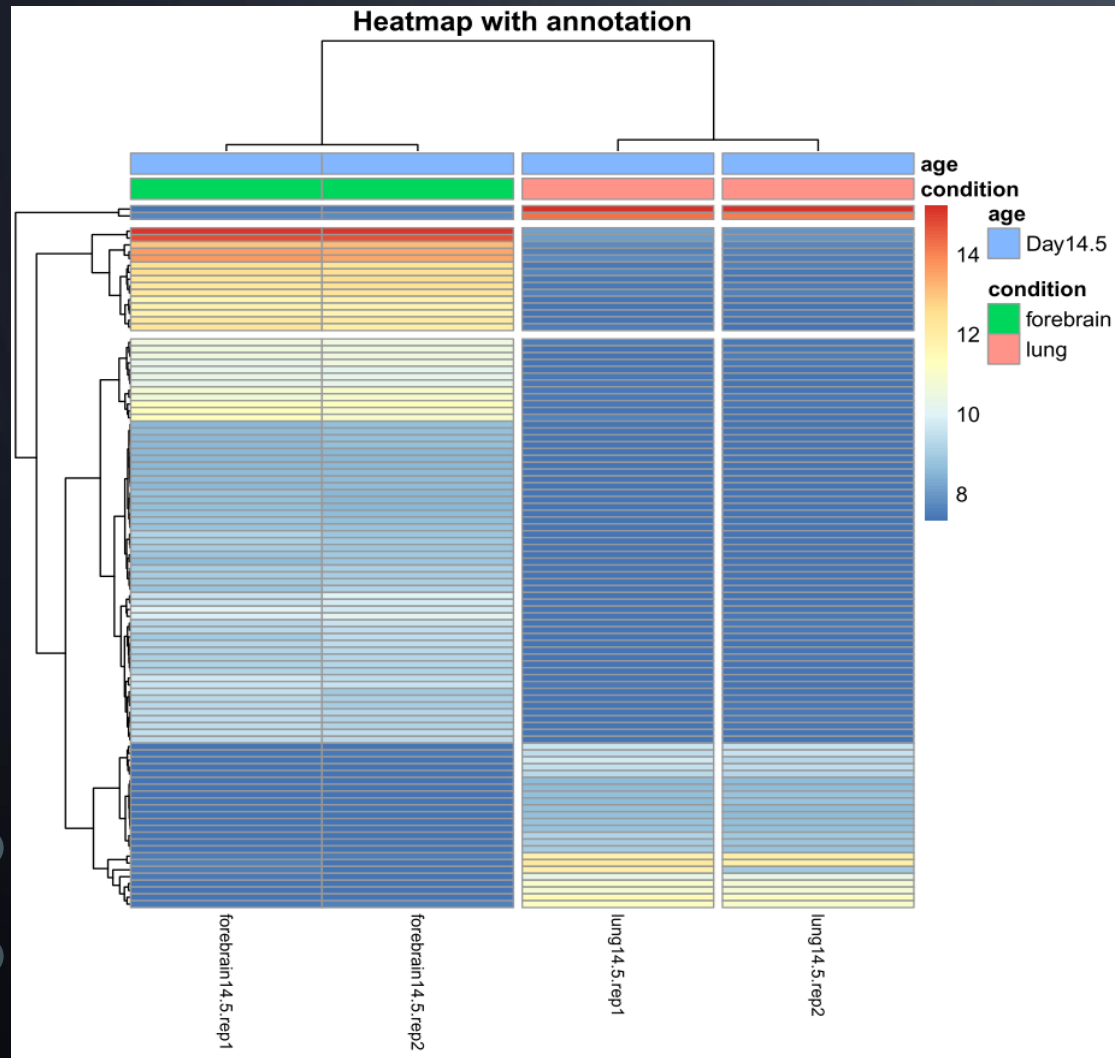
limb vs lung (normal)



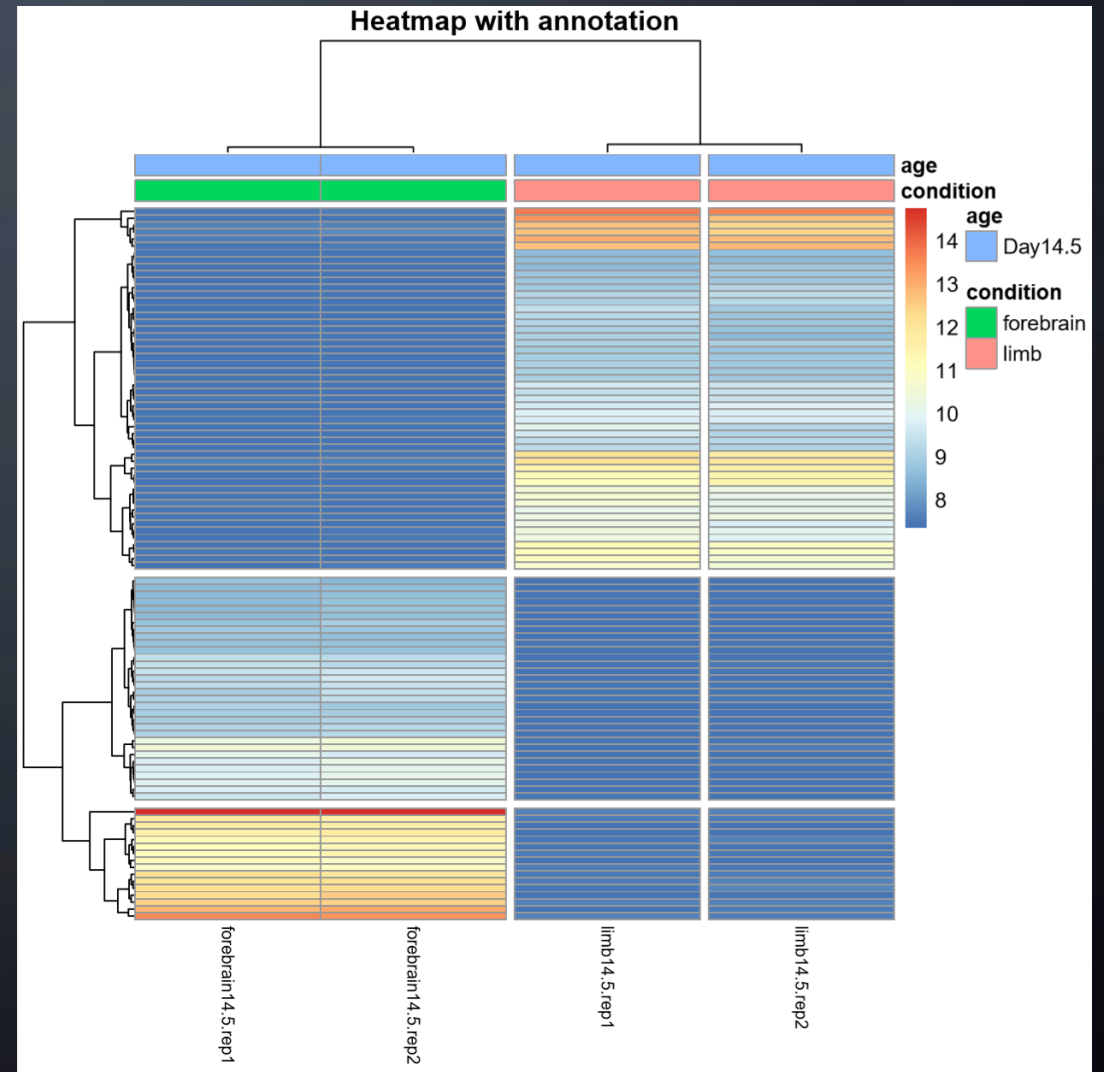
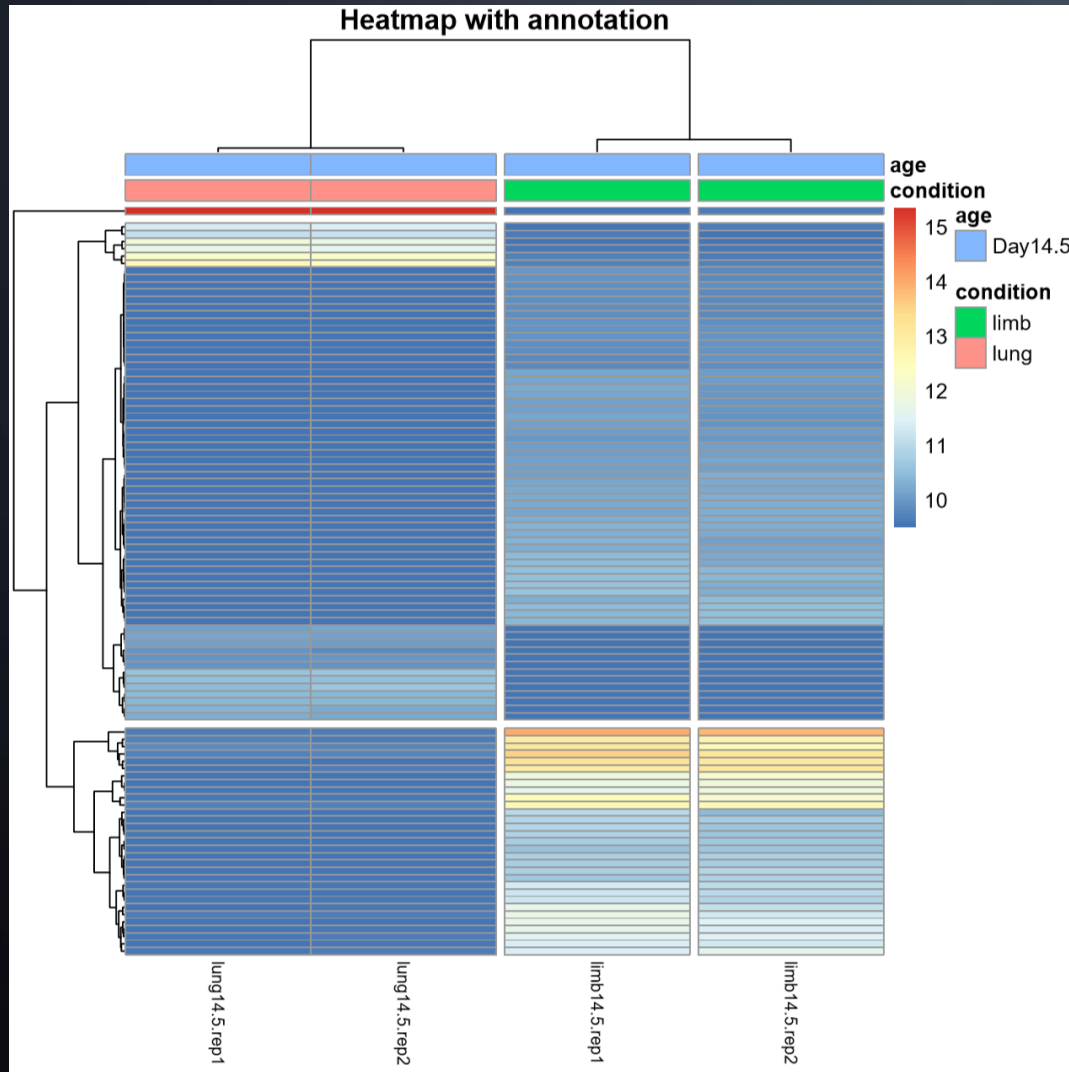
RESULTS: DIFFERENTIAL GENE ANALYSIS



RESULTS: DIFFERENTIAL GENE ANALYSIS



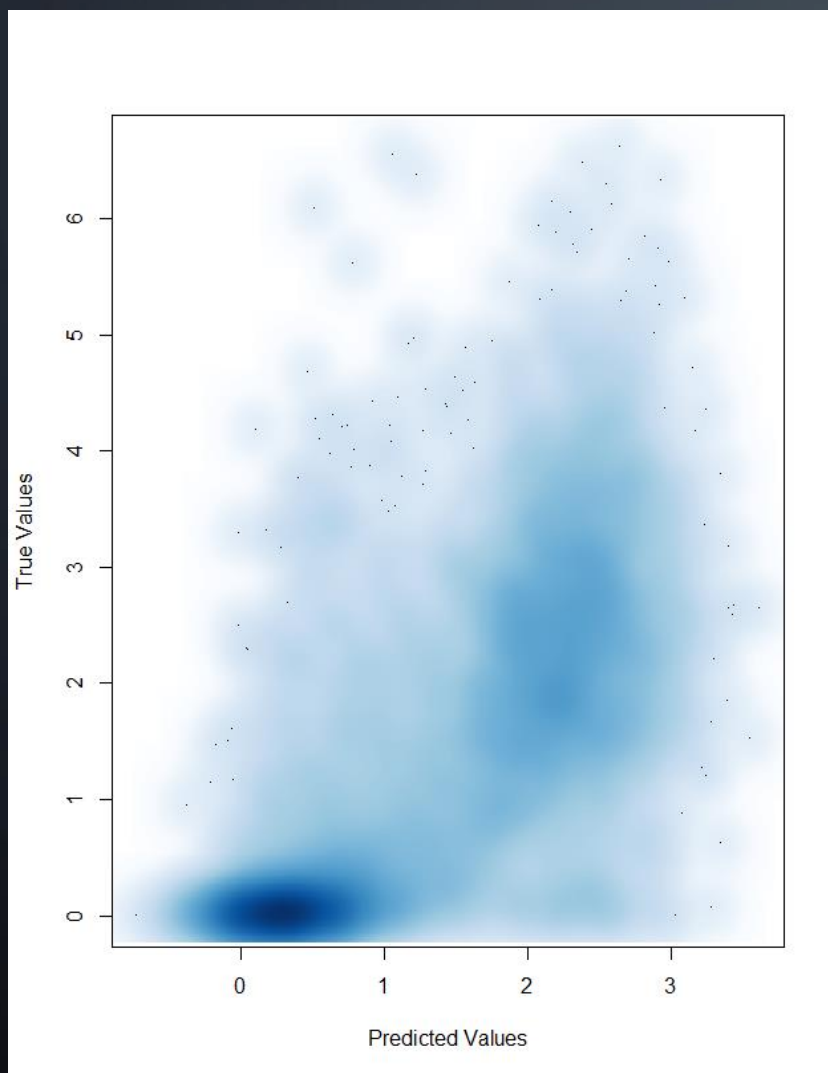
RESULTS: DIFFERENTIAL GENE ANALYSIS



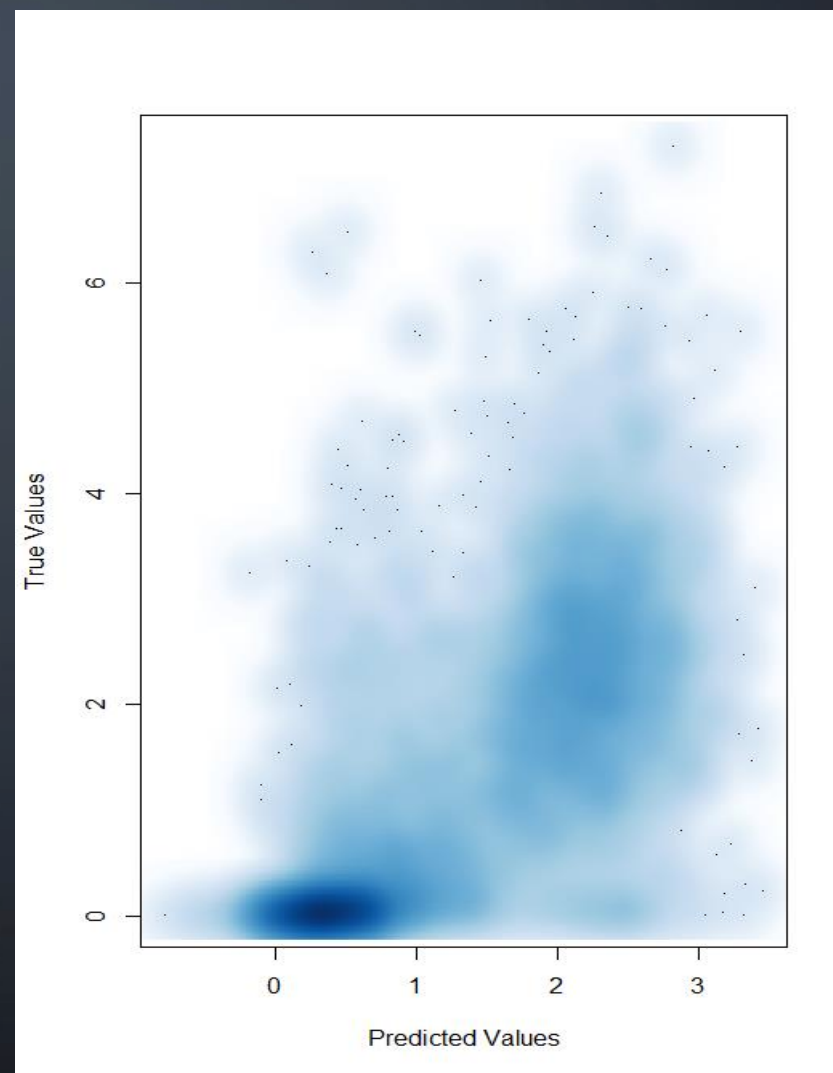
RESULTS: PREDICTION MODEL

- Prediction model for the limb tissue data
 - Trained on 75% of the data and tested on the other 25%
- Beta-coefficients:
 - H3K4me3: 0.864
 - H3K4me1: 0.126
 - H3K27ac: 0.416
 - H3K36me3: 0.019
 - H3K9me3: -0.344
 - H3K27me3: -0.358
- R²-value: 0.4549
- Adjusted R²-value: 0.4547
- Correlation between test dataset and the RPKM values: 0,683

RESULTS: PREDICTION MODEL



Lung



Limb

CONCLUSION

- Coherence between functionality of chromatin states and their responsible histone modifications
- Some genes are only distinctly necessary for different tissues at different development states
- Prediction model: correlation coefficient (between HMP and expressed genes) = 0.7
- General expression of genes, depending on HMPs, could be predictable over different tissues
- Further questions:
 - Including other marks in analysis may enhance prediction accuracy