

Author contributions:

This final report was written and created by both Hannah Boyd and Aureo Zanon. Both parties participated equally in coding, troubleshooting, background research, and data interpretation. Aureo Zanon was instrumental in developing our code. Hannah Boyd provided valuable insight into the biological relevance of our results. We would like to thank Neha Patwardhan, Colin Howe, and Olivia Rumble for their valuable advice on GO term analysis and ATAC seq processing.

I.) Cell line comparison I – CMP vs Ery

Background: This project looks at both RNA seq and ATAC seq analysis of various stages of blood cell differentiation. All cell lines were isolated from C57BL6J mice femur and tibia. Our first pairwise analysis compares common myeloid progenitor cells (CMPs) and erythroblasts (Ery).

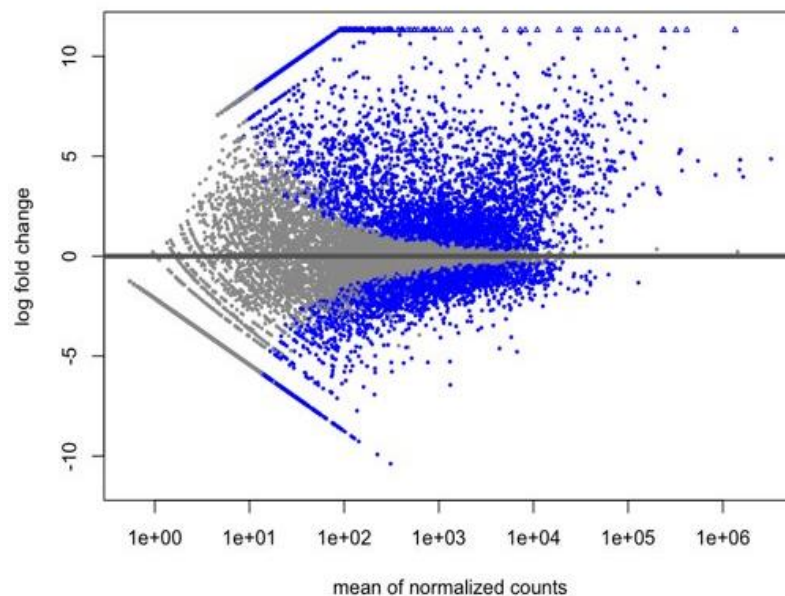
a. RNA seq

i. Comparison between DEseq2 and limma voom

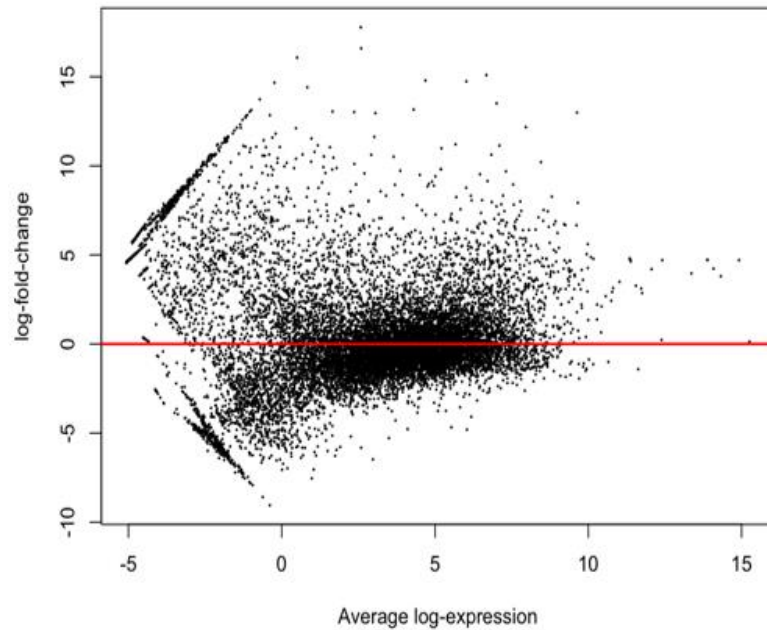
Although DEseq2 and limma voom process the same raw data from RNA sequencing it is common for these packages to produce different results. Many of these differences come from the analysis pipelines used in the generation of differential gene expression results. Such as DEseq2 uses a non-linear model, while limma voom uses a linear model.

Our analyses found that DEseq2 identified 4915 differentially expressed genes, whereas limma voom identified 3042 differentially expressed genes (see MA plots below). Based off these results we decided to use our DEseq2 results for the rest of our analyses since it was the most inclusive for measuring differentially expressed genes.

DEseq2 MA plot:

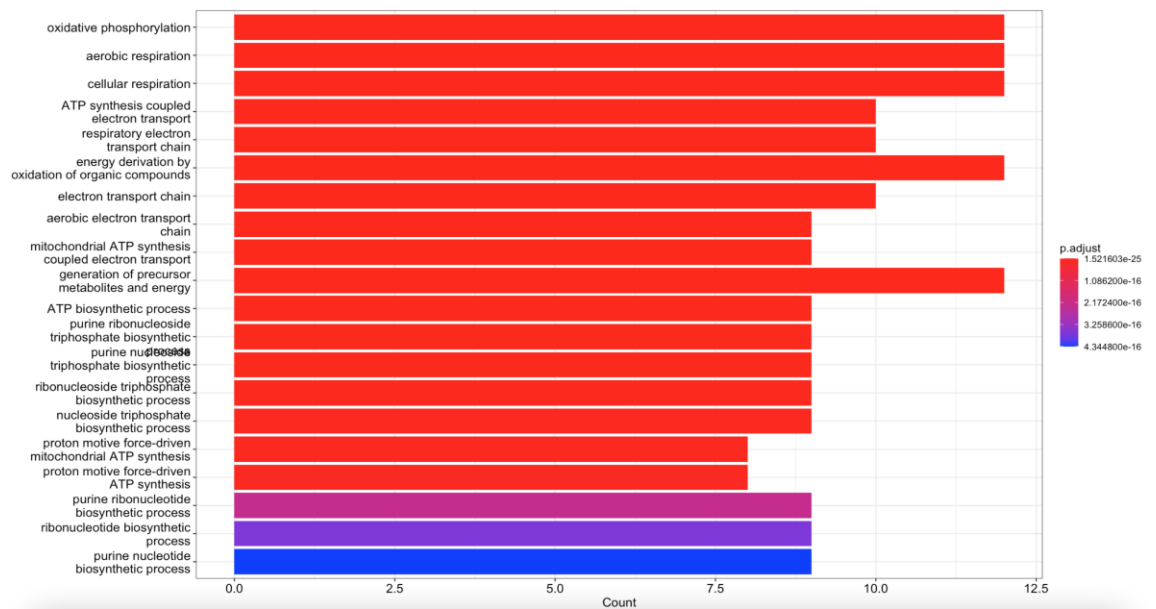


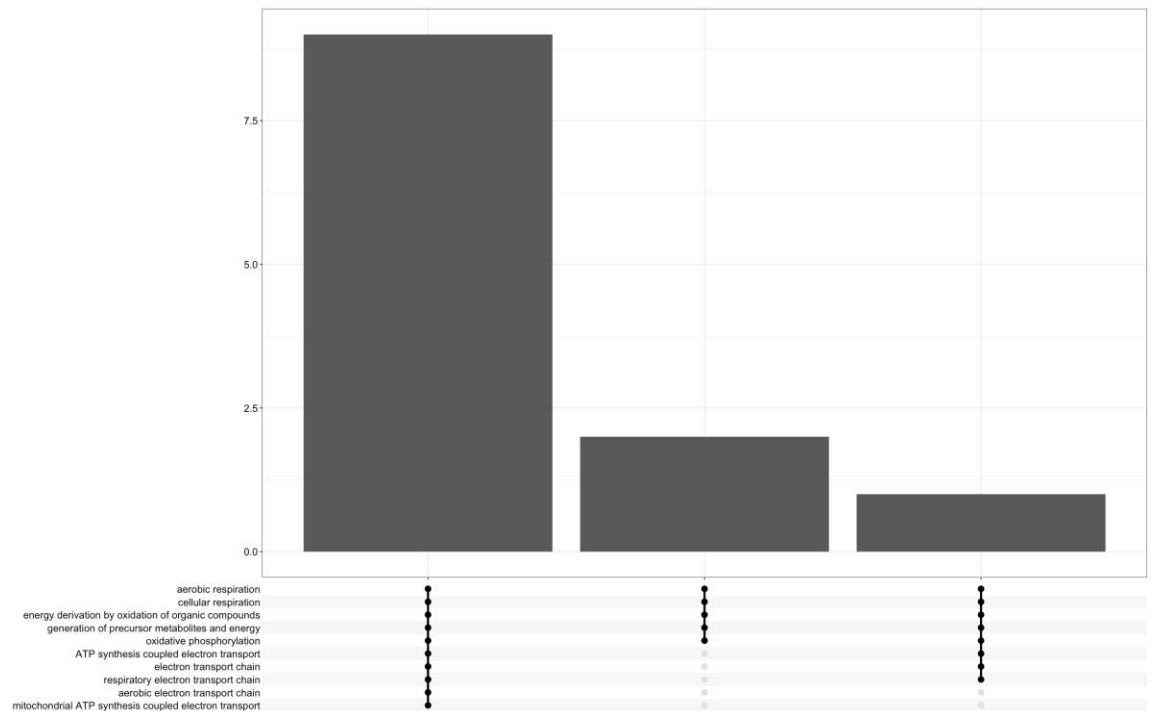
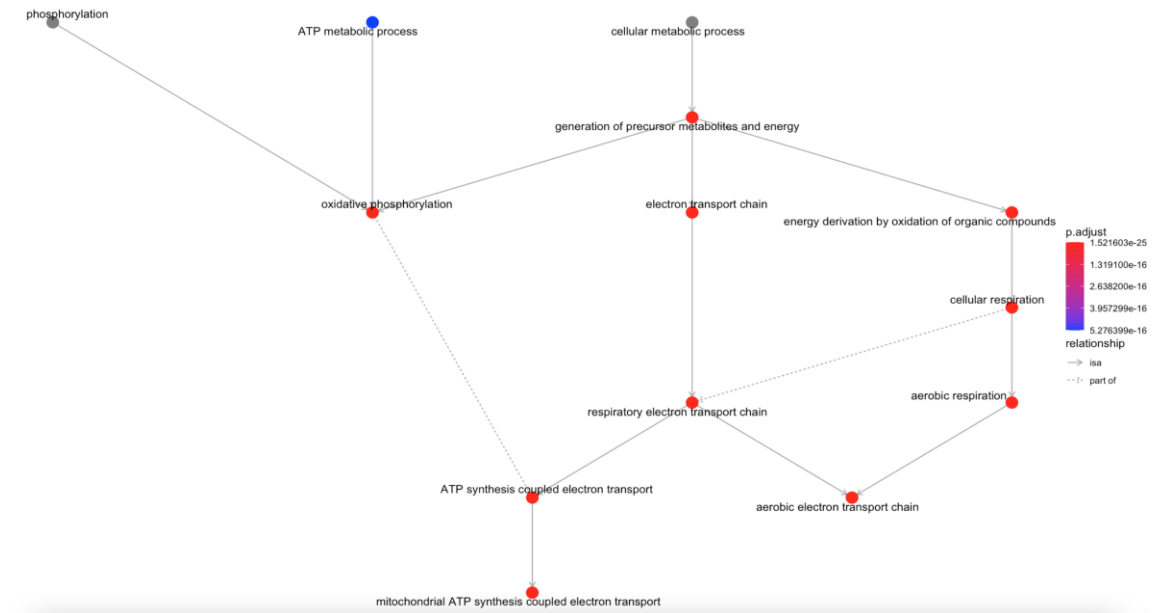
Limma voom MA plot:



ii. Functions of differentially expressed genes

Several genes were identified to be differentially expressed when comparing cells from the CMPs and Erys cell lines. The major ontologies of these genes fall into the categories listed below. Most of the identified ontologies are closely related to energy production and mitochondrial function. Cellular differentiation is an energy intensive process that requires ATP in order for molecular pathways to be activated and cytoskeletal restructuring to occur (Hu et al., 2016).

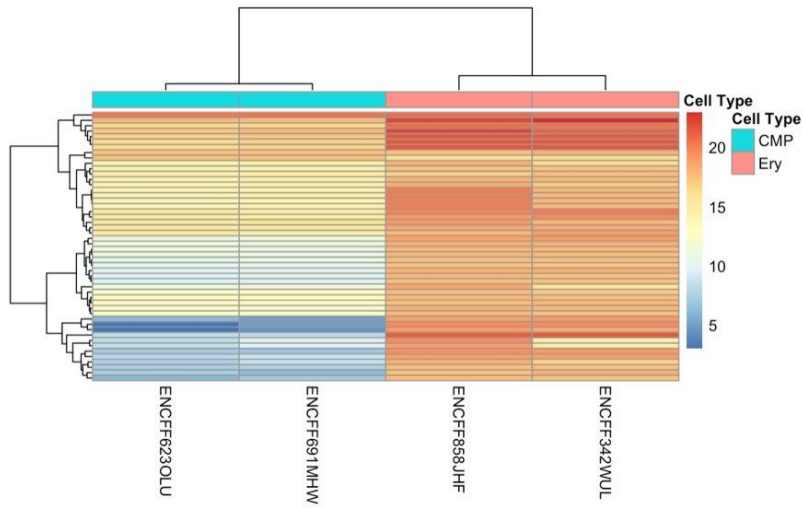




iii. Hierarchal tree using RNA seq data

Hierarchal trees are the most popular method for clustering RNA seq data. Hierarchal clustering is accomplished by having each data point as its own cluster and then combining these clusters into larger groups called clades. These clades allow us to see trends of expression for the genes identified by our DEseq2 analysis. Genes which have increased expression are coded in red, whereas genes that have decreased expression are coded in blue. We can see that in general Ery have higher levels of gene expression

compared to CMPs which have lower levels of gene expression (see hierarchical tree below).

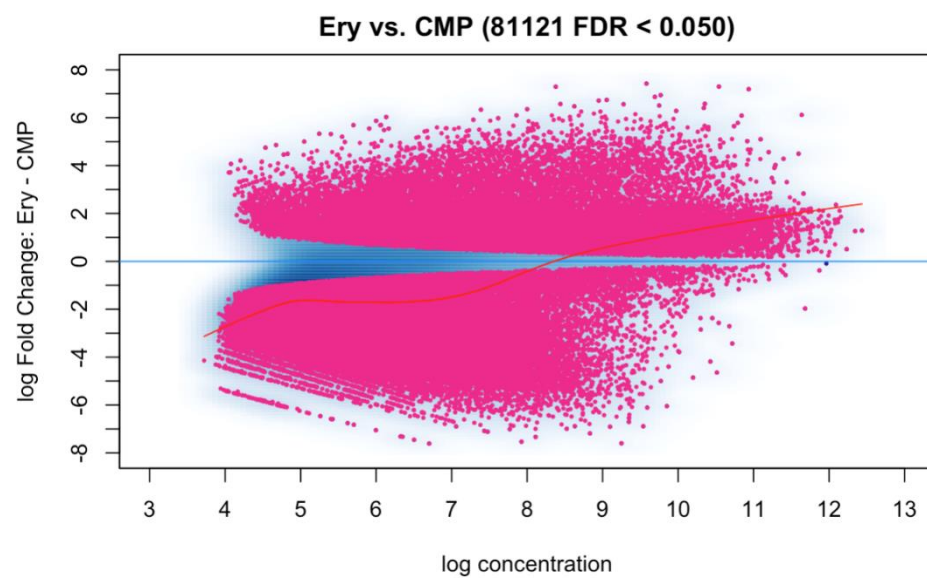
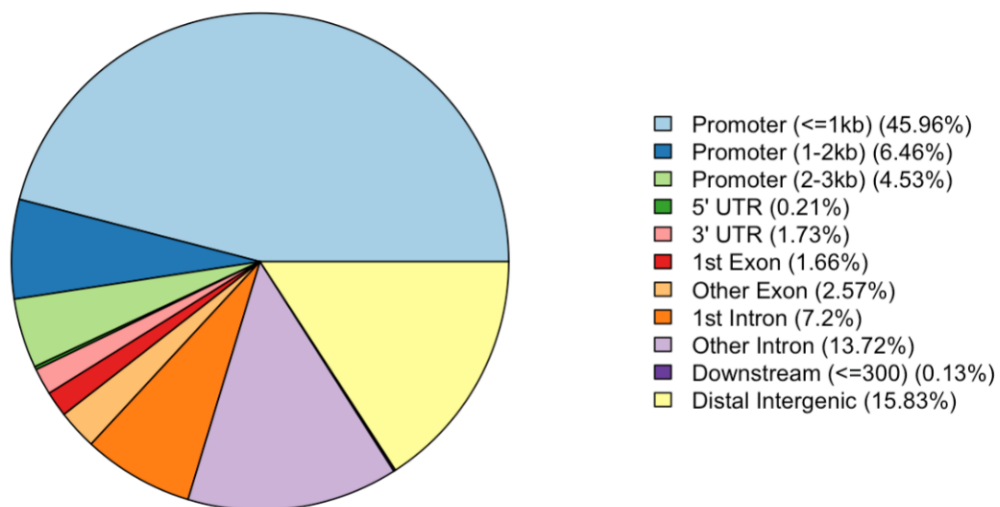


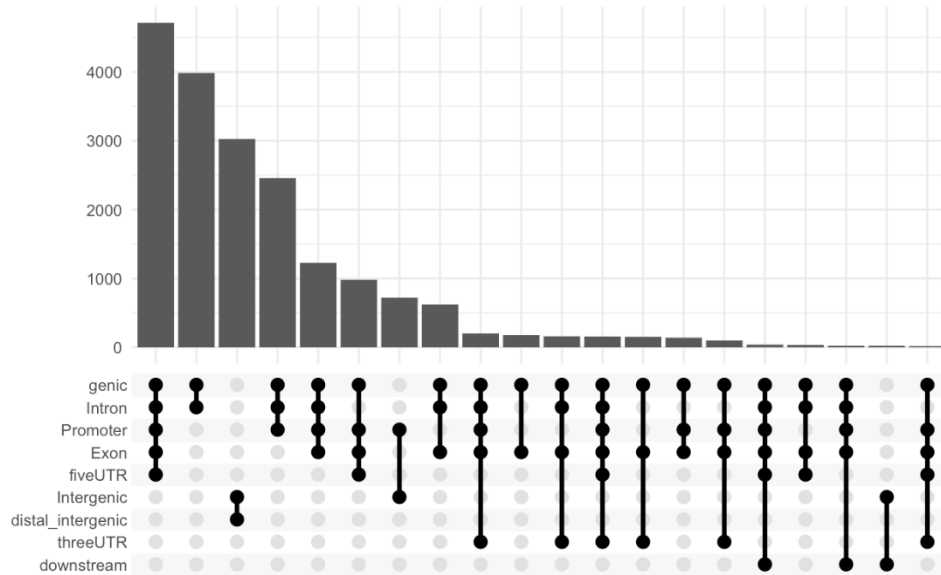
b. ATAC seq

i. Regions of differential chromatin patterns and gene expression implications

ATAC seq is a popular technique for measuring chromatin accessibility. Chromatin accessibility is a predictive measure for gene expression. For transcription factors to initiate transcription it is necessary for them to be able to access DNA. When chromatin is open (not condensed) it is much easier for gene expression machinery to begin transcribing a given gene. ATAC seq does not directly measure gene expression, but it does give us correlative information that if a given area is accessible it is more likely to be expressed. ATAC seq data is often combined with RNA sequencing (as was done in our analyses) or ChIP seq to give a more in-depth picture of a given genomic landscape.

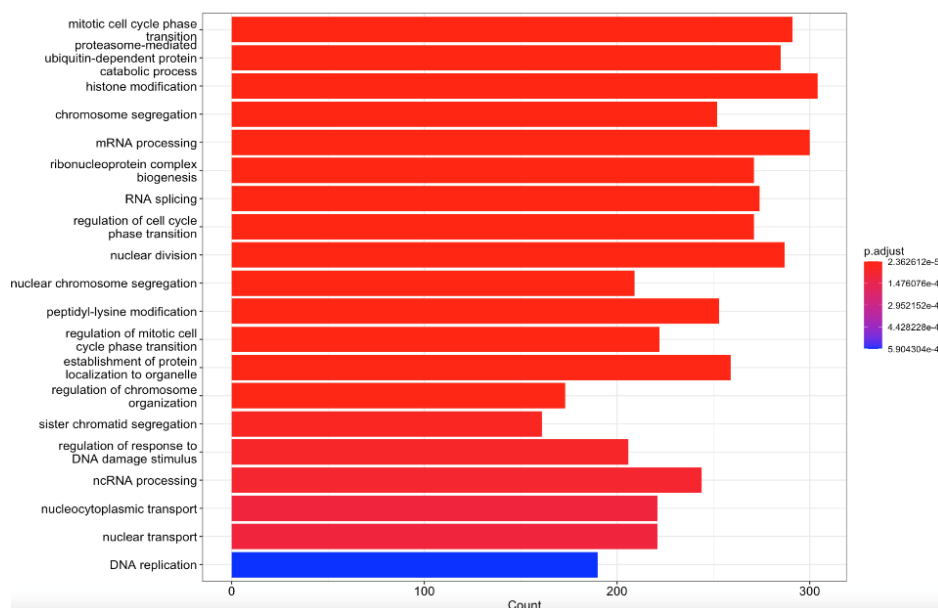
As visualized below in our peak annotation pie chart, more than half of the peaks fall into promoter regions. This suggests that most of the accessible regions in our comparison are at promoter sights. Promoter sights are a hot spot for gene regulation, especially transcription factor binding which in most cases causes gene expression of a given gene.

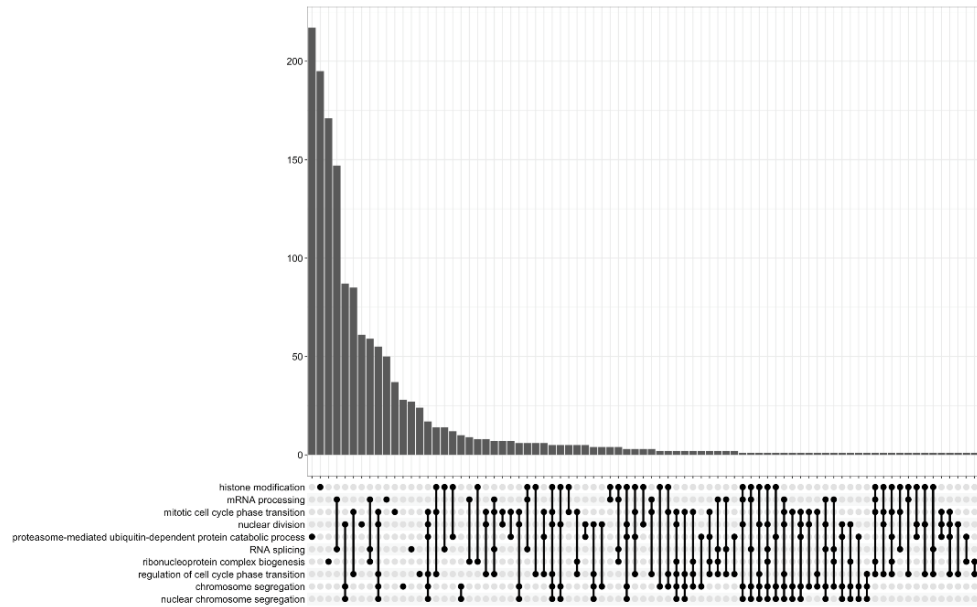
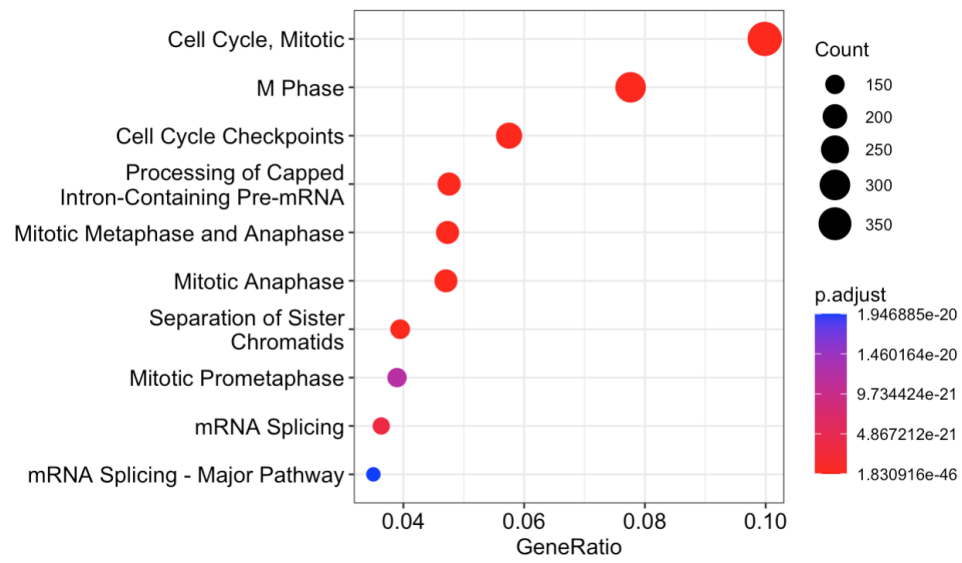


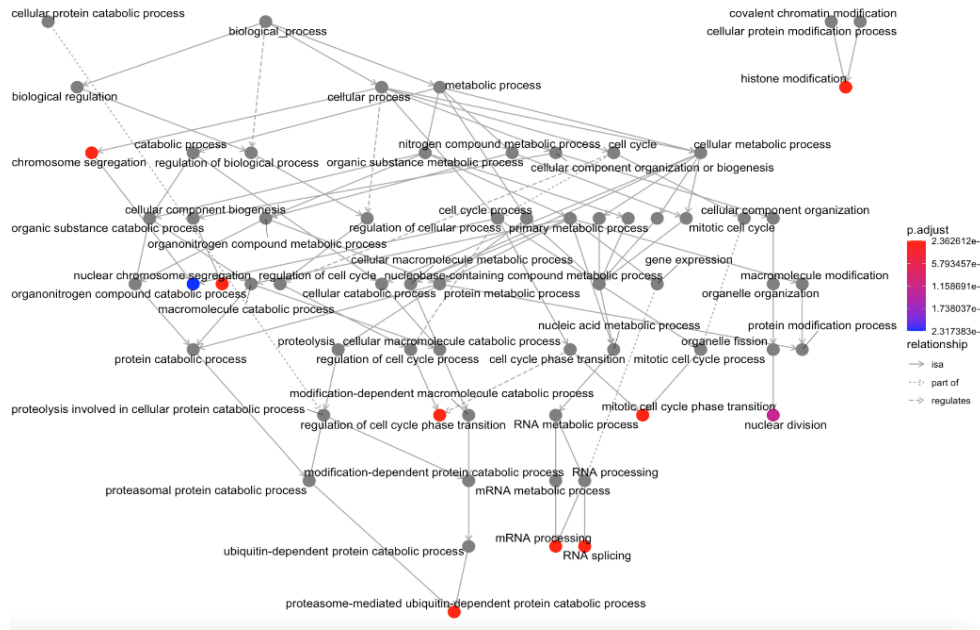


ii. Functions of genes with differential chromatin patterns

From our analysis of gene function in regions of the genome which are more accessible we found there were a few main gene ontologies that were commonly identified. First were genes that are known to be involved in the mitotic cell cycle and chromosome segregation. This is to be expected since we are looking at cell lines which reflect different steps in blood cell differentiation, meaning these cells are going to be actively dividing and differentiating. We also see accessibility changes to genes which are involved in RNA processing, such as splicing. This could reflect the increase in transcript processing that is typically occurring because of gene expression changes.

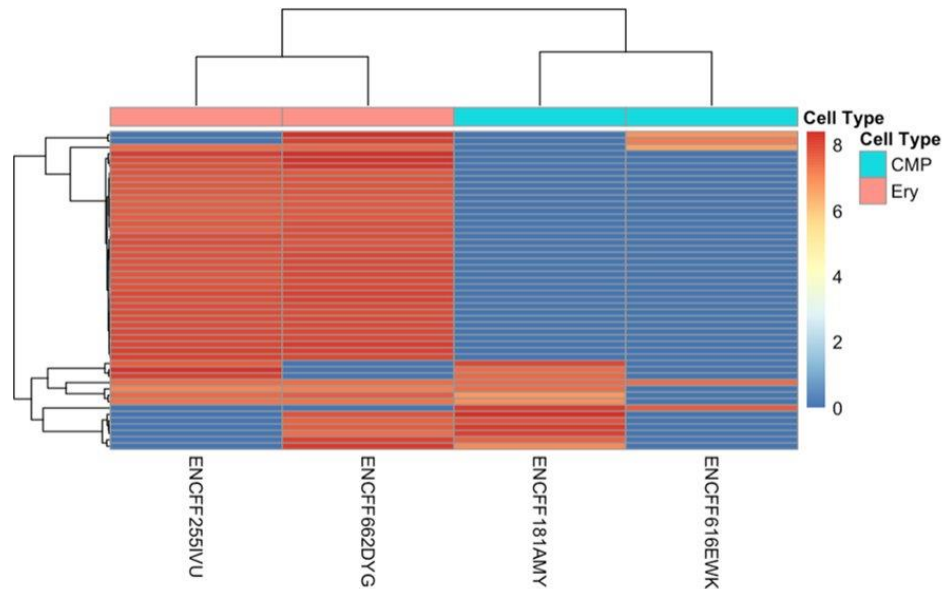


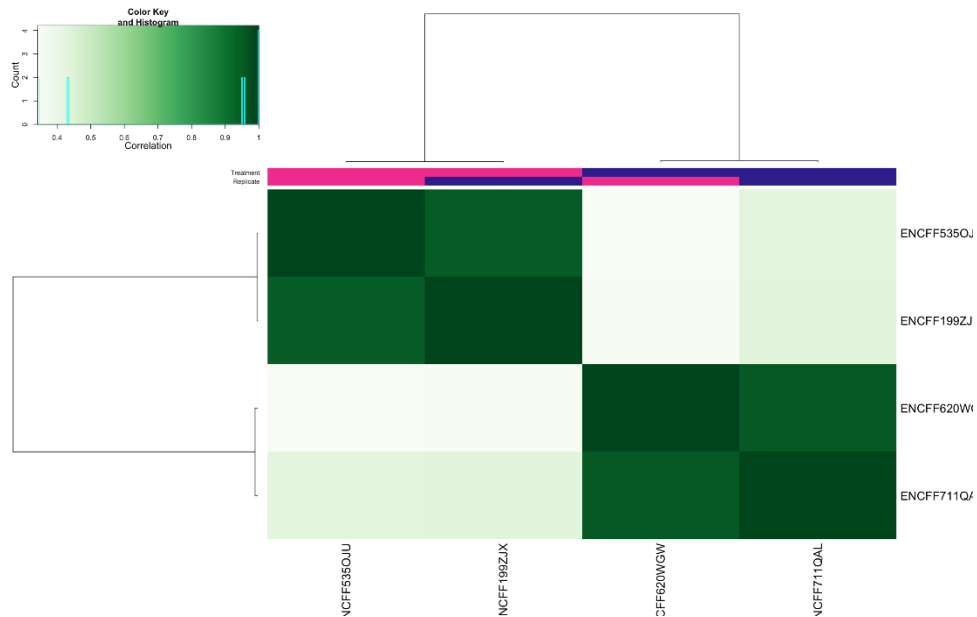




iii. Hierarchical tree using ATAC seq data

Based on our hierarchical tree shown below, we can see that in general the Ery genome is much more accessible than CMP cells. This maps with what we found in our RNA seq data, that in general Ery cells have higher levels of gene expression compared to CMP cells. Together, this data works in complement to suggest that the actively expressed genes are more accessible to transcriptional machinery.



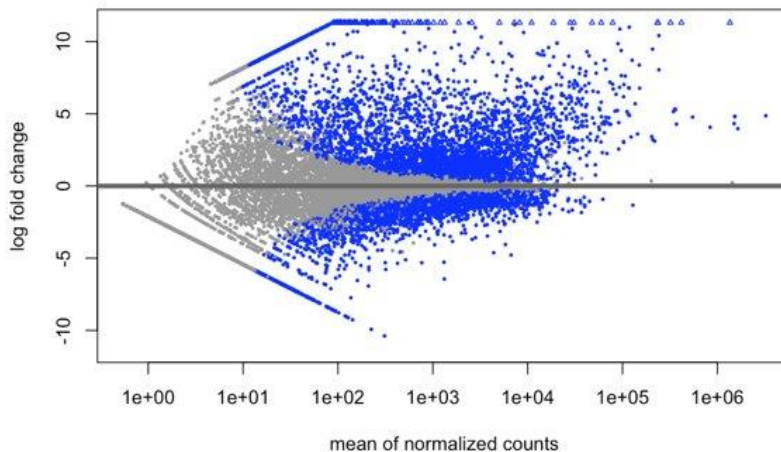


II.) Cell line comparison II – CMP vs HSC

Background: Our second pairwise analysis compares common myeloid progenitor cells (CMPs) and hematopoietic stem cells (HSC).

- a. RNA seq
 - i. DEseq2 differentially expressed genes

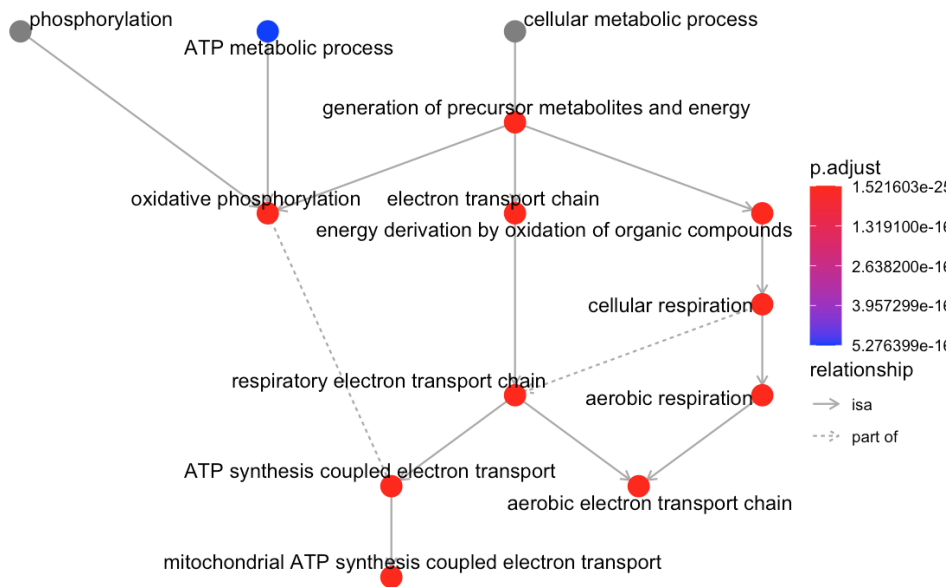
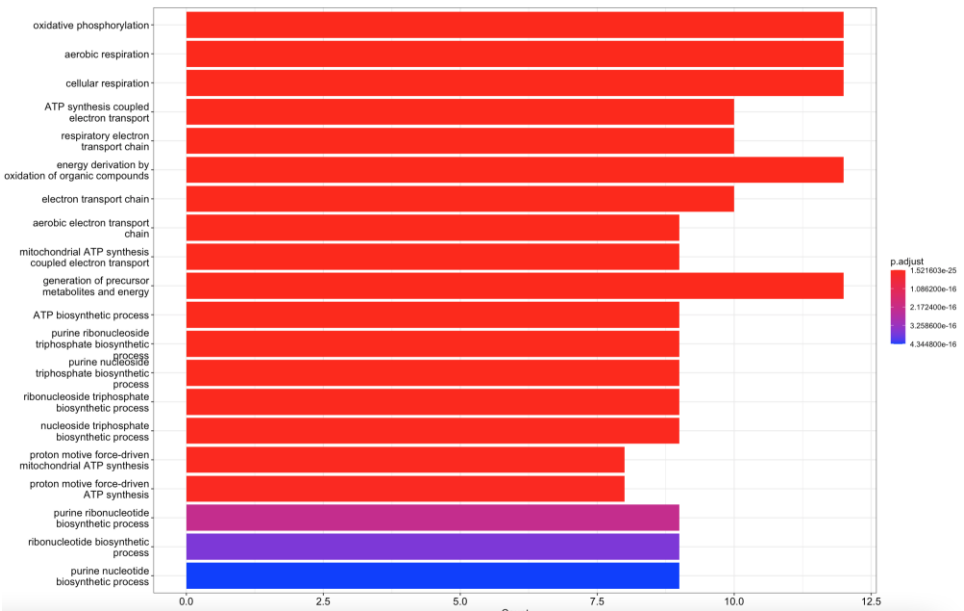
Our analyses found that DEseq2 identified 6287 differentially expressed genes (see MA plot below).

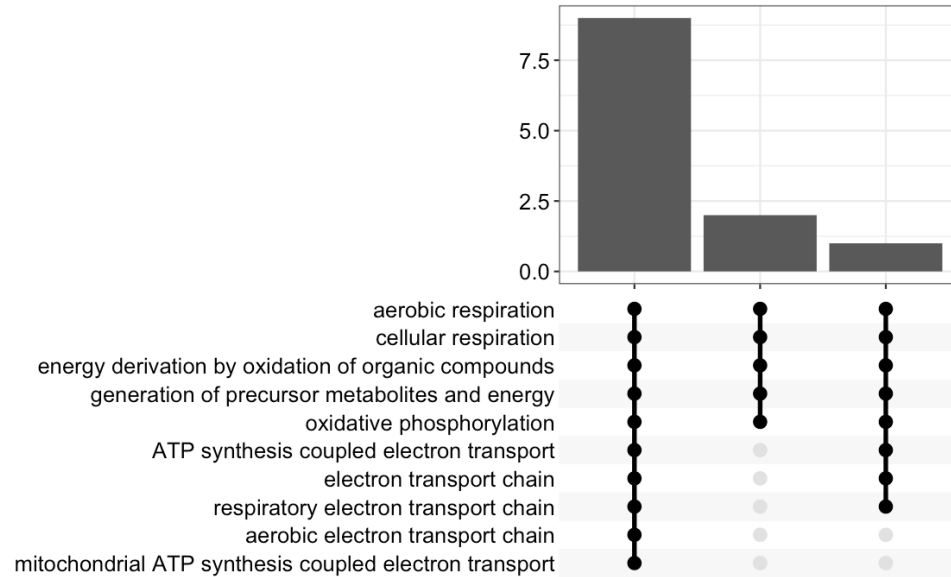


- ii. Functions of differentially expressed genes

Several genes were identified to be differentially expressed when comparing CMPs and HSCs. The major ontologies of these genes fall into the categories listed below. Again, we see that processes important for energy production, mitochondrial function, and

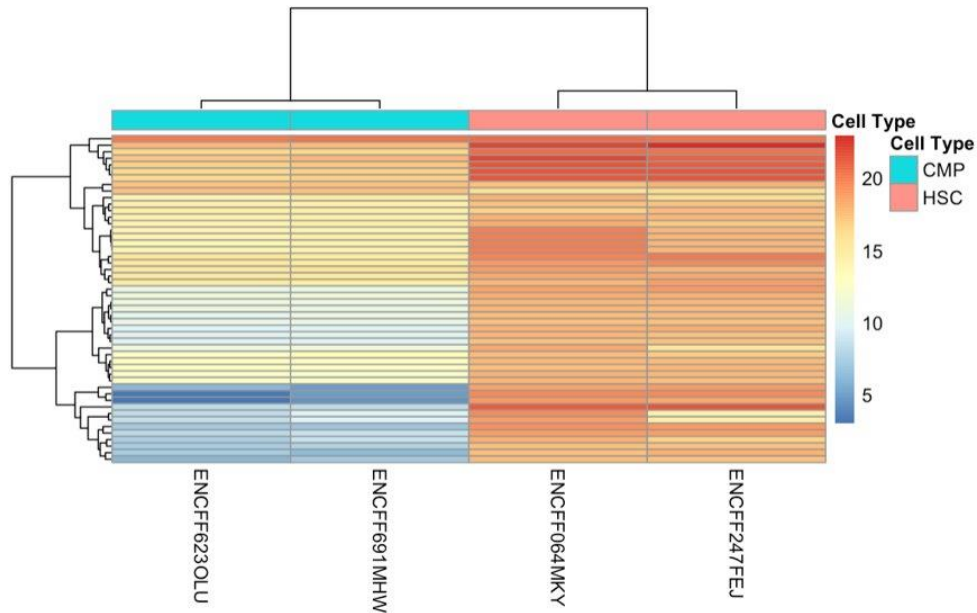
nucleotide synthesis are being identified as having increased expression. Stem cells are known to have increased energetic and metabolic demands compared to differentiated cell types. In fact, activation of energy metabolism pathways has been linked to fate determination (Hu et al., 2016).





iii. Hierarchal tree using RNA seq data

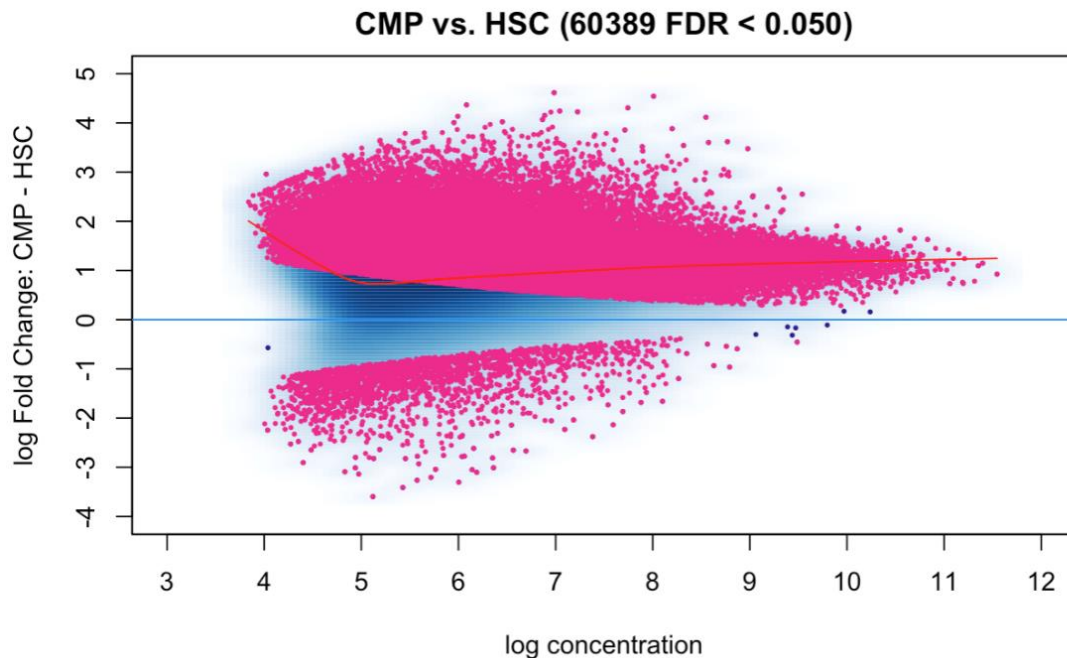
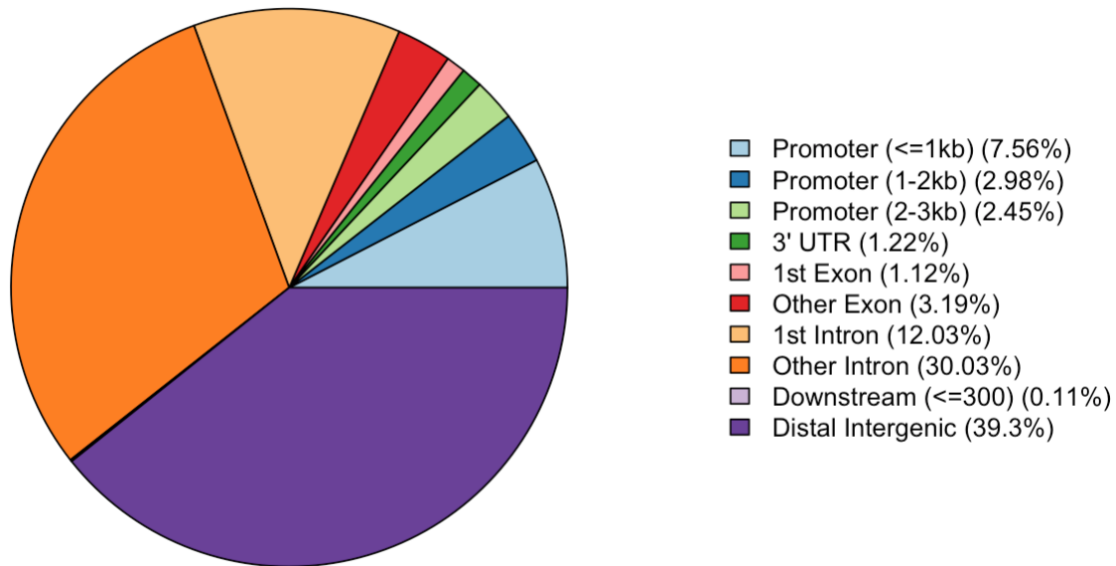
Genes which have higher expression are coded in red, whereas genes that have lower expression are coded in blue. We can see that in general HSCs have higher levels of gene expression compared to CMPs which have lower levels of gene expression (see hierarchal tree below).

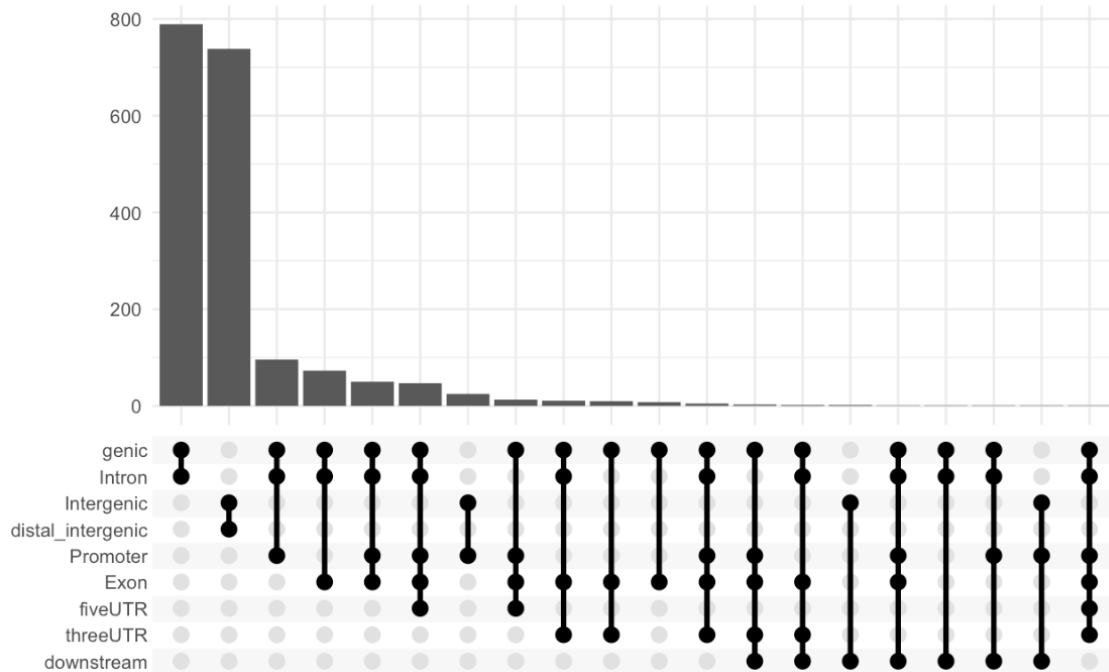


b. ATAC seq

i. Regions of differential chromatin patterns and gene expression implications

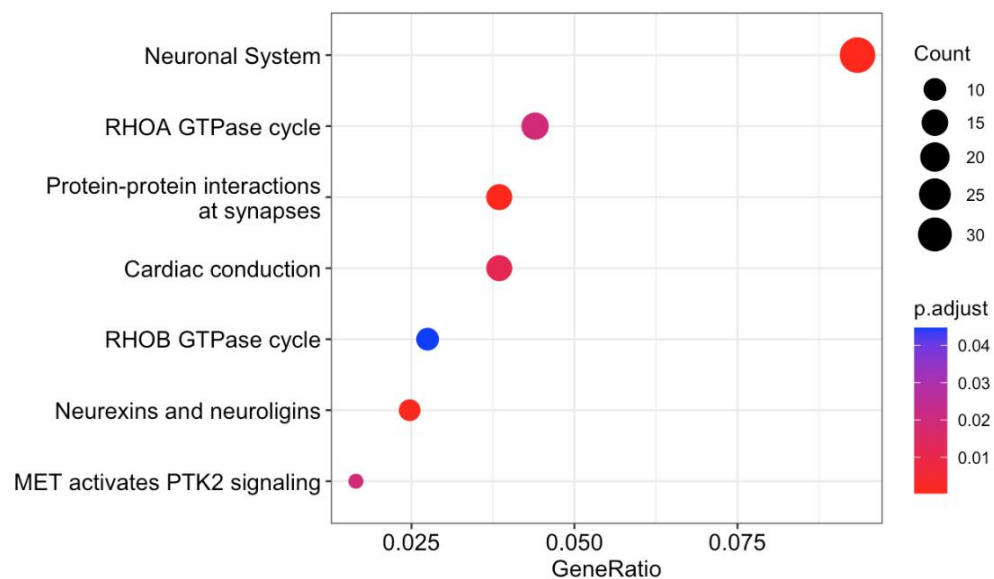
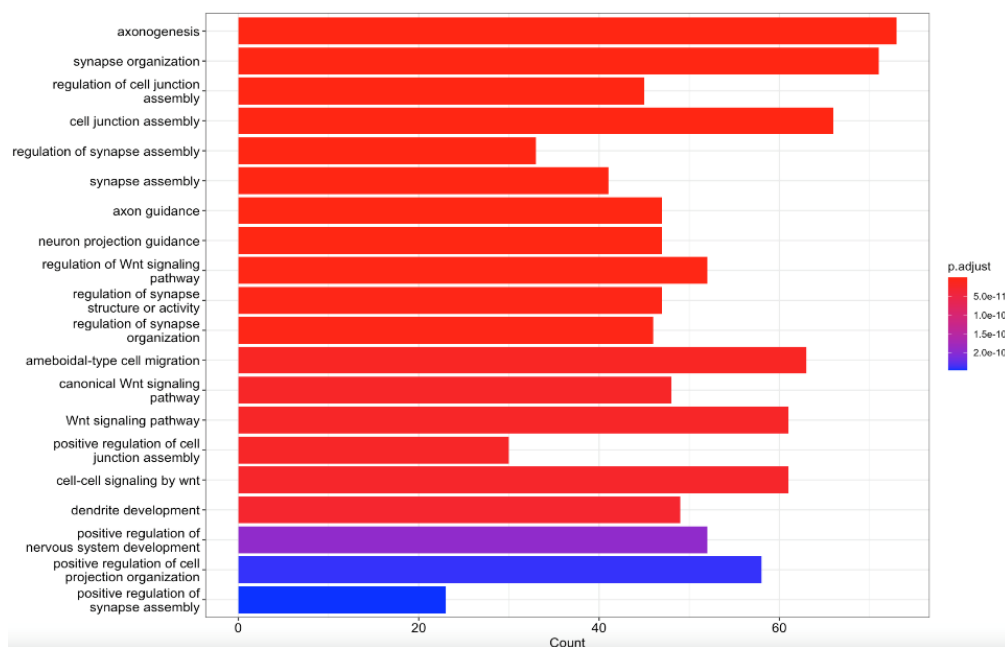
As visualized below in our peak annotation pie chart, around a third of the peaks fall into distal intergenic regions. This suggests that enhancer regions may be regulating promoters of our identified differentially expressed genes. Enhancers are known to be distal sequences which are far upstream of promoters. Due to the flexibility of DNA, enhancers can end up being spatially close to promoters, despite many nucleotides separating the two regions. Gene regulatory machinery can then bind to enhancers and promoters, which can then activate gene transcription. For this specific result, a HiC analysis would be useful to understand more about the chromosome conformation and 3D architecture.

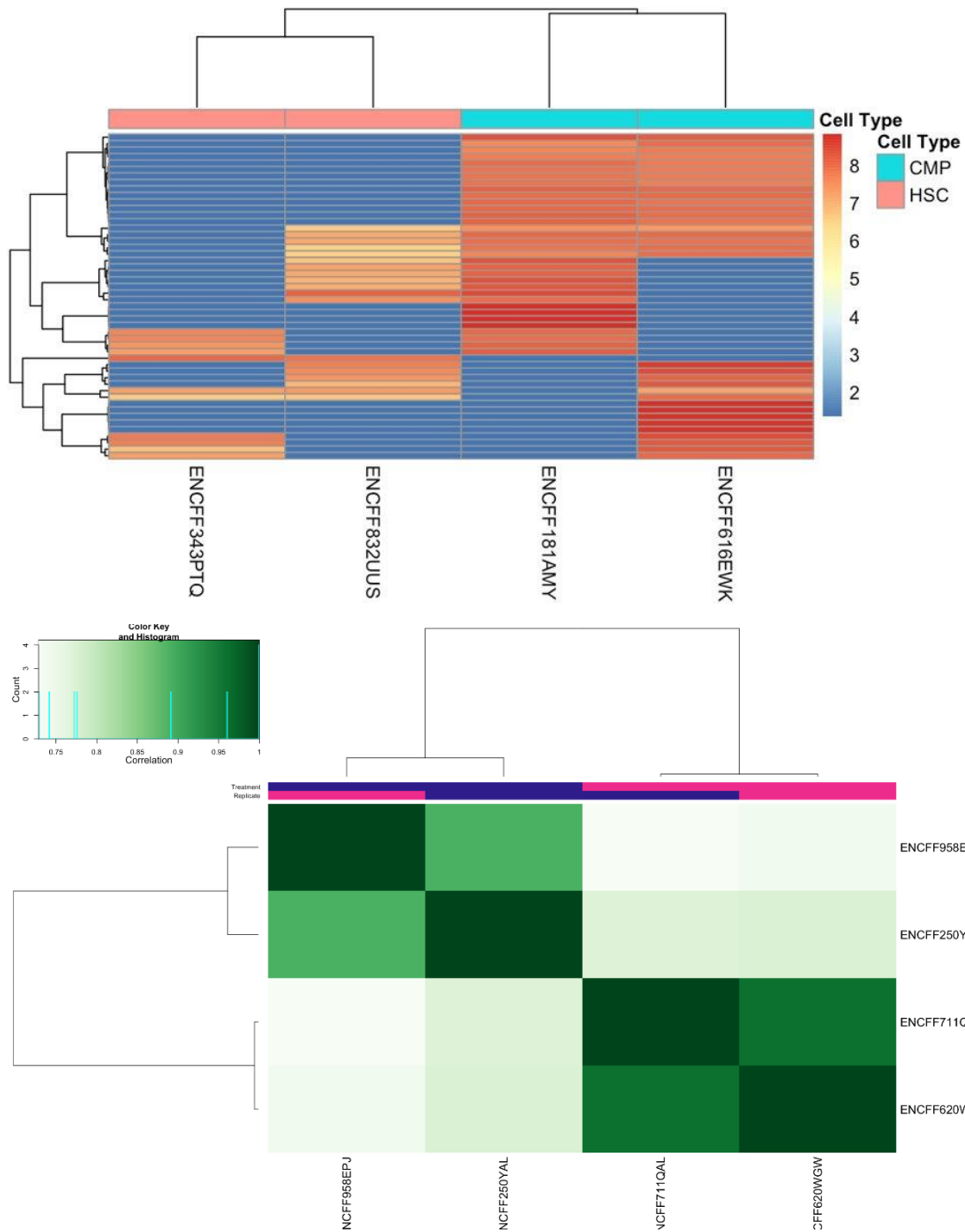




ii. Functions of genes with differential chromatin patterns

Interestingly, gene expression for genes associated with neuronal functions were increased such as axonogenesis and synapse organization. This could reflect that the genes which fall into these categories have yet to be tested for their role in cell differentiation and have identities outside that of nervous system development. Axonogenesis and synapse organization rely on dynamic changes to a cell's cytoskeleton. We predict that the genes which fall into the identified categories are involved in cytoskeletal organization broadly. During differentiation and development, the cytoskeleton is actively changing in response to both cellular morphology changes and cell division. We also saw an increase in genes involved with Wnt signaling, which has been identified for its importance in developmental biology and cellular differentiation. We also see that Rho a and b regions are more accessible. These two GTPases are important in cytoskeletal organization (Ghalloussi, Dhenge, and Bergmeier, 2019) further supporting our hypothesis that cytoskeleton changes play an essential role in blood cell differentiation. Likewise, the Rho pathway has been linked to cellular differentiation, cell motility, adhesion.





Citations:

1.) Ghalloussi, D., Dhenge, A., & Bergmeier, W. (2019). New insights into cytoskeletal remodeling during platelet production. *Journal of thrombosis and haemostasis : JTH*, 17(9), 1430–1439. <https://doi.org/10.1111/jth.14544>

2.) Hu C, Fan L, Cen P, Chen E, Jiang Z, Li L. Energy Metabolism Plays a Critical Role in Stem Cell Maintenance and Differentiation. *Int J Mol Sci*. 2016 Feb 18;17(2):253. doi: 10.3390/ijms17020253. PMID: 26901195; PMCID: PMC4783982.