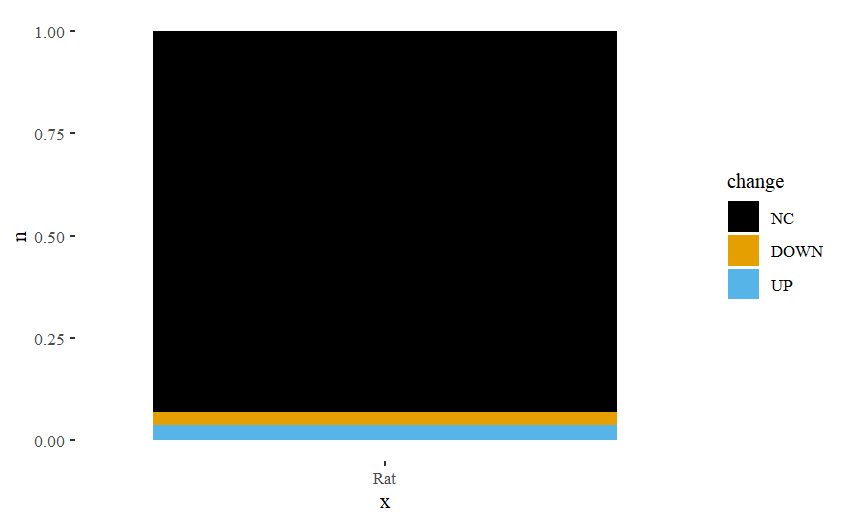
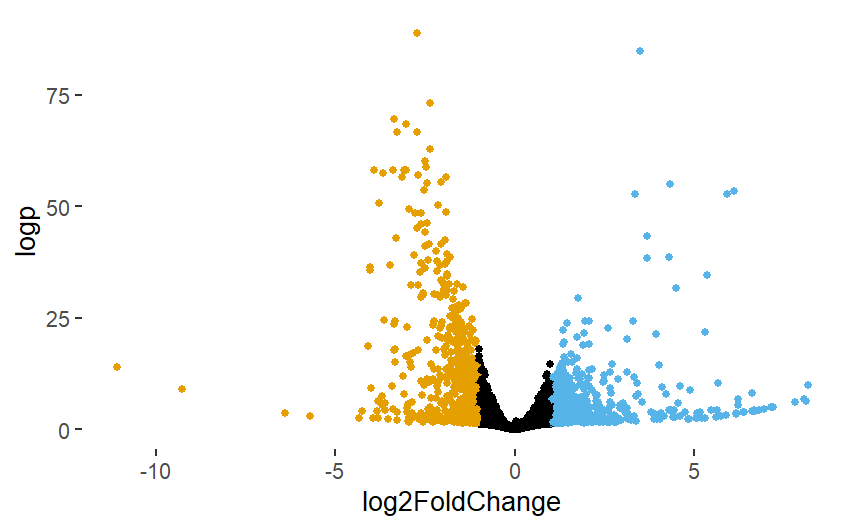
**Introduction**

**Comparing Transcriptome – RNA-SEQ**

The RNA sequence data was aligned to genome via Salmon by the researchers, which output count data and used for our analysis. Comparing the wild-type and EED knockout, one would expect the lack of PRC2 complex lead to a reduction in H3K27me3 which results in greater heterochromatin and a general upregulation of genes. However, upon using DESeq2 to normalize the count data within the sample, and resultant log-fold change, there was a population of both up and downregulated genes. Cut-off being p-value of 0.05 and logarithm base 2 threshold of 1.



**No Change**

93%

**Down**

3.1%

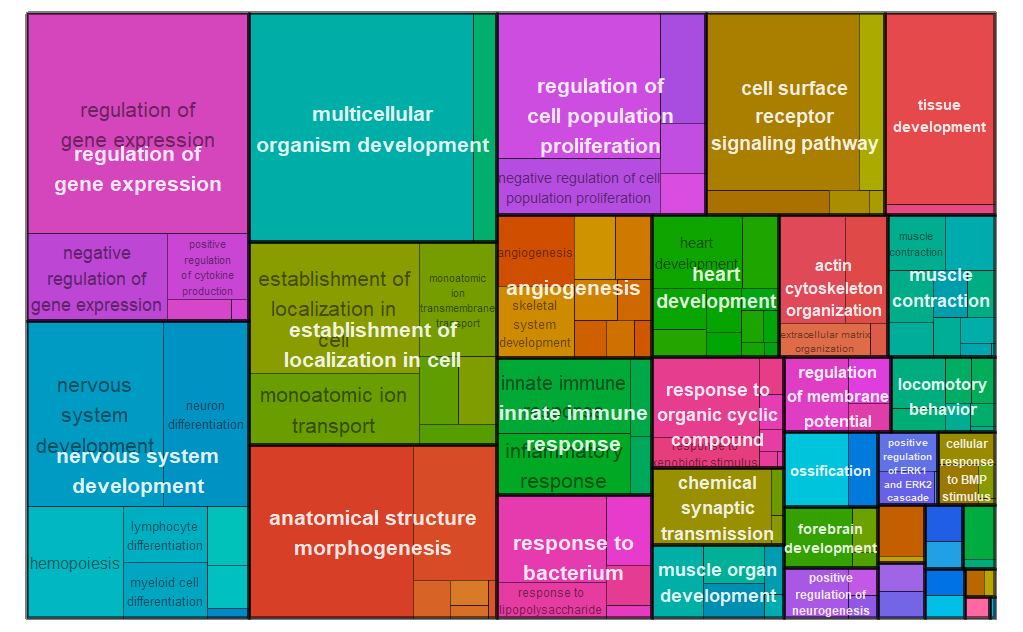
**Up**

3.9%

The upregulated are expected, and hence further investigated to assess whether these genes were affected due to a common biological pathway stemming from EED or instead due to chromatin editing i.e. locality in the genome.

**Upregulated Genes’ Ontology**

Ensembl’s database, *biomaRt*, includes the respective gene ontology entries. With biological process being the broadest, it was analysed first. Mapping of the genes was conducted and enriched using *ClusterProfile* which revealed no clear ontological groupings. The same can be seen for the ontological groups, cellular components and molecular function, as seen in the annex.

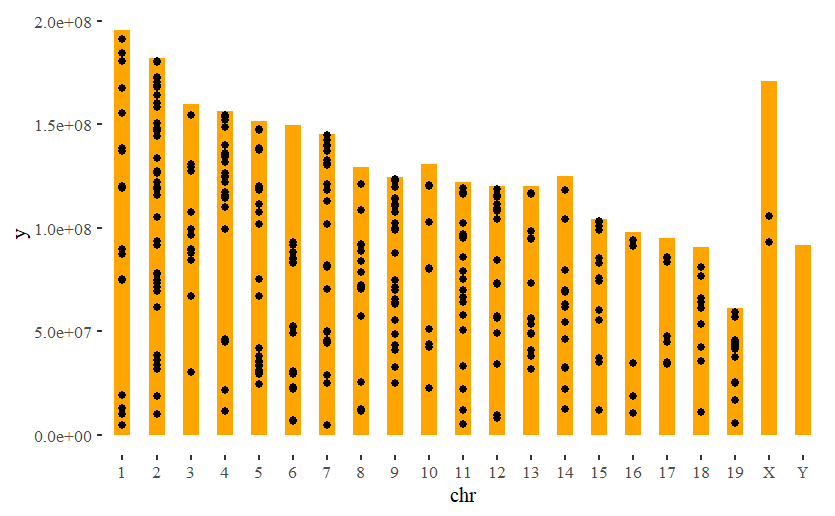


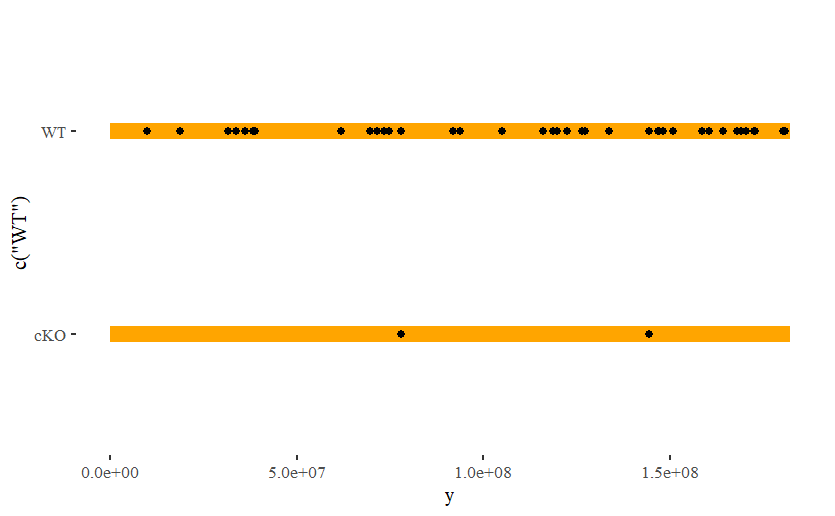
While extremely broad, visually, the ontology can be categorized into mainly the following biological relevance.

|  |  |
| --- | --- |
| Index | System |
| 1 | Neurological |
| 2 | Immunological |
| 3 | Muscular |
| 4 | Embryonic |

Such a widespread impact is to be expected by the knockout of EED due to the importance of PRC2 for the formation of H3K27me3, an established important gene repressor. This generality hints towards the downstream impact of EED-/- on gene upregulation to be based on chromatin state instead of direct biological processes and pathways. Hence to look further into chromatin state via CHIP-Seq. **CHIP-Seq**

The loss of H3K27me3 is observed in the CHIP-seq data provided and correlates with the gene upregulation observed. In fact, almost all the H3K27me3 is lost, with the exception of a few in chromosome 2. The drastic change emphasizes the importance of PRC2 with regards to histone methyltransferase activity in the nucleus.





A line with black dots and numbers

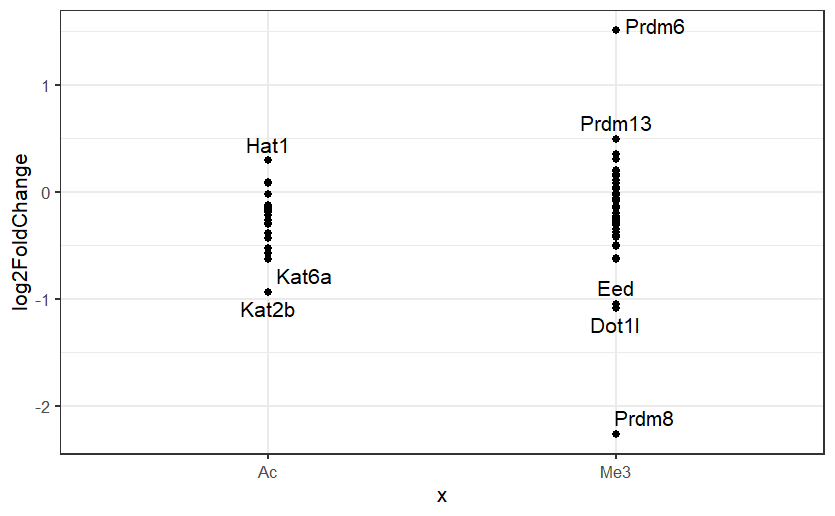
AI-generated content may be incorrect.

This significant loss of H3K27me3 would then lead to an increase in H3K27ac, which was also observed. In fact, the increase in acetylation was much more than the reduction in methylation. This could be due to the positive feedback loop where having acetylation would result in greater unravelling of the chromatin resulting in greater access of the DNA to acetyltransferase.

A black and blue lines with numbers

AI-generated content may be incorrect.

In fact, this is more likely as based on the count data, it was shown there was no significant change in the genes related to histone acetylation activity. Hence the acetylation had to be from the lack of competition with methyltransferase on the substrate and affinity stemming from unpacked heterochromatin. There was also not much change in the histone methylating genes except for EED, Prdm6 and Prdm8.



**Correlation**

However, there are still a large population of genes that were downregulated. While it made sense that having the acetylation results in more accessible DNA, as mentioned earlier, it may not seem direct the reason for gene downregulation as observed by RNA-SEQ count data. Mapping all genes to their histone modification revealed a few counter-intuitive facts.

|  |  |  |  |
| --- | --- | --- | --- |
|  | No Change | Down | Up |
| H3K27ac | 5269 (2.5) | 187 (0.9) | 165 (-4.2) |
| Non-H3K27ac | 13907 (-2.5) | 460 (-0.9) | 629 (4.2) |

p-value = 7.63×10-5

Fisher’s exact test was conducted and weirdly, there is significant evidence to show H3K27ac correlates with a reduced upregulation. There is also more evidence that non-acetylated are more “stable”. Investigating further, the H3K27ac presented above refers to only the KO. However, information regarding its origin is not actually captured. There are 3 main occurrences, the methylation transitioning to acetylation, null to acetylation or acetylation to acetylation. Hence, splitting by the transition / origin reveals the below table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | No Change | Down | Up |
| Ac 🡪 Ac | 793 (0.43) | 40 (2.49) | 10 (-3.26) |
| Me3 🡪 Ac | 15 (-0.94) | 0 (-0.77) | 2 (2.16) |
| Null 🡪 Ac | 4461 (-0.28) | 147 (-2.35) | 153 (2.91) |

p-value=2.29×10-4

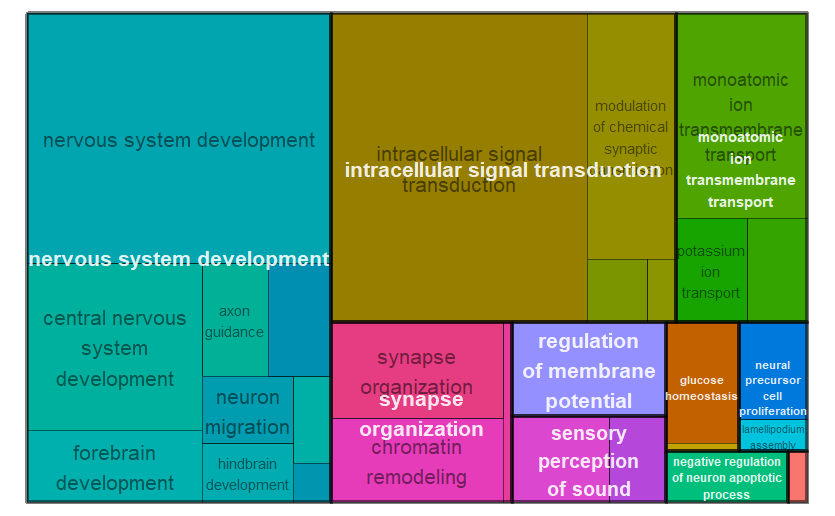
As seen, one of the main types that led to the low p-value is acetylation 🡪 acetylation. However, that makes sense as if there were a general upregulation of genes, the previously upregulated genes would appear to be downregulated.

While this may stem due to competition of resources, it is also possible that it is the effect of normalization. RNA-sequence data is very difficult to normalize cross samples to determine absolute quantity, hence a lot of the data is in relation to other genes. Hence, when a lot of genes get upregulated, some of these previously upregulated genes appear to be downregulated. Case in point Me3/null 🡪 Acetylation, there are a significant number of upregulated genes compared to what was expected. In addition, for the Ac 🡪 Ac, there were more downregulated genes than expected.

However, it is also important to note that there is still a large population of genes that are considered to not have any significant change. So, one would expect majority of the downregulated genes to be related to the wild type’s purpose.

**Downregulated gene ontology**

As expected, when mapping the gene ontology of the downregulated genes, they were mostly relating to neuron activity. With the sample taken from the cerebellum, with the exceptions of glial cells, most of these would be usually upregulated in the Wild type.



Hence, there will be an impact on the neurons of the cerebellum. Such impact can be further scoped via looking into the molecular function as well as the cellular processes. Being neurons of the cerebellum, the largest population of activity characterised by the cell type, neuron, to be excitatory (glutamatergic) and inhibitory (GABAergic) synapse.

Hence, using the KEGG pathway, all the related genes and their products were filtered and found their correspond changes. Below is the labelled excitatory pathway.

A diagram of a machine

AI-generated content may be incorrect.

With the 2 most critical components of a synapse AMPAR (for transmitter acceptance and conversion to electrical) and NMDAR (coincidence detector – major learning mechanism – essential for cerebellum for motor memory), implied cells have poor ability to transmit action potential and perform in the cerebral-motor circuits – poor motor function of the organism.

A diagram of a machine

AI-generated content may be incorrect.

Looking at the GABAergic synapse, the impact is not as obvious, with many components’ transcript variants being simultaneously up and downregulated. Hence not much can be inferred about it.

However, while the data does show that there was some upregulation of genes upon EED knockout, it may not always be due to the chromatin structure. Hence there is not much statistical evidence of any relationship, in fact based on this data, it almost shows a negative correlation. A cell has other ways to repress genes, hence majority of the cell’s genes are still not expressed.

The other upregulated genes can be traced back to the upregulation of genes that controls other genes’ transcription. Using the gene ontology, can filter significantly upregulated genes relevant to controlling other gene expression to be:

|  |  |
| --- | --- |
| Gene | Function |
| Pou4f1 |  |
| Gata4 |  |
| Tlr2 |  |
| Nkx2-5 |  |
| Acta1 |  |
| Mmp12 |  |
| Hand2 |  |
| Ccl3 |  |
| Itgax |  |
| Gata3 |  |
| Actc1 |  |

**References**

**Annex**