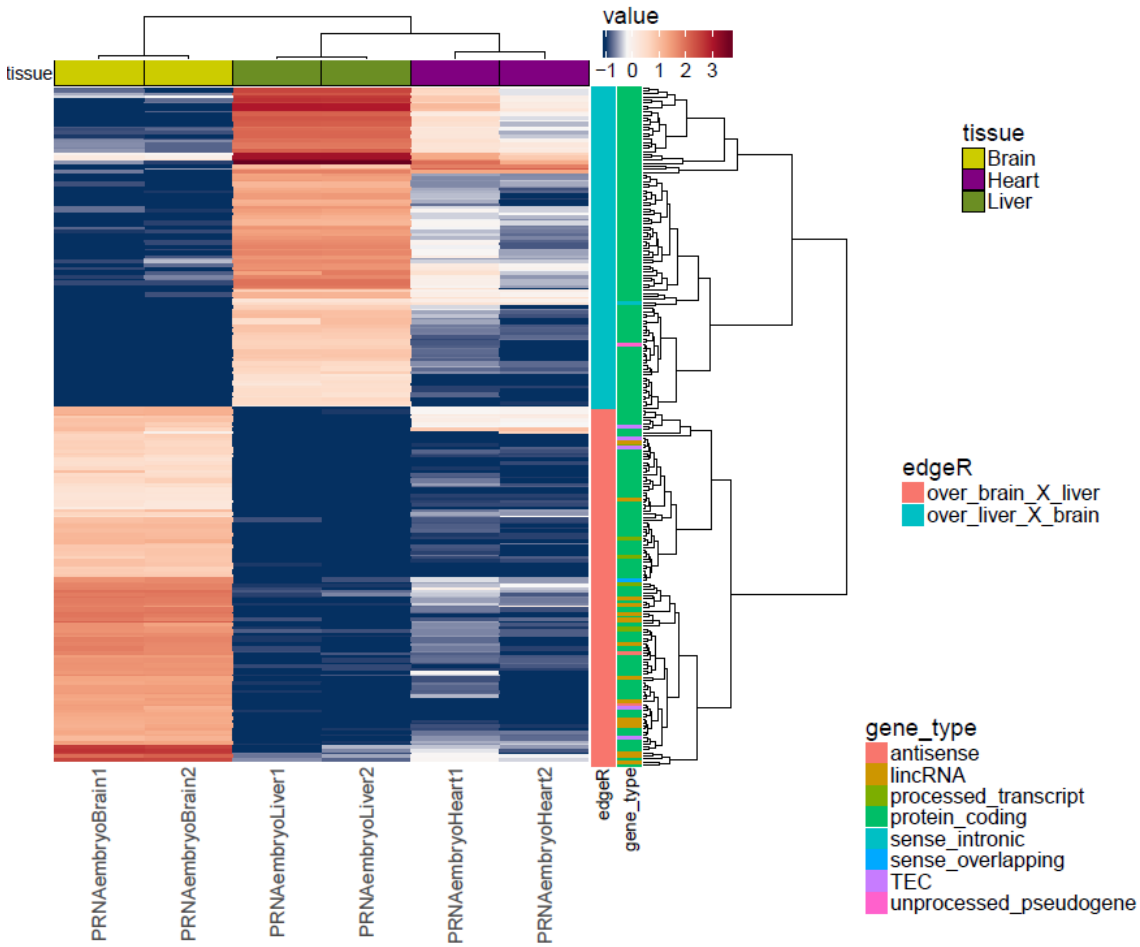
1. Perform differential expression analysis between brain and liver using the EdgeR. Present results using a heatmap with hierarchical clustering in rows and columns and colored classification of differentially expressed genes (DEGs), i.e. overexpressed in brain versus liver and the other way around.

There is a total of 225 genes differentially expressed between liver and brain, 118 are overexpressed in brain and 107 in liver.

On the one hand, we can see, in the clustering by columns, how the heart and liver are closer to each other with respect to the brain, even using differentially expressed genes between the liver and the brain, probably because brain presents a different gene expression pattern.

In the clustering by genes we can see, in pink, the 118 genes over-expressed in brain compared to liver's expression and, in blue, the 107 over-expressed in the liver, in addition to the type of gene to which they correspond. These results agree with what we see in the heatmap where the expression levels of these genes correctly correspond to each of the two groups.

Finally comment that the separation in the heart samples is not correct because these samples have not been used to make this graph.

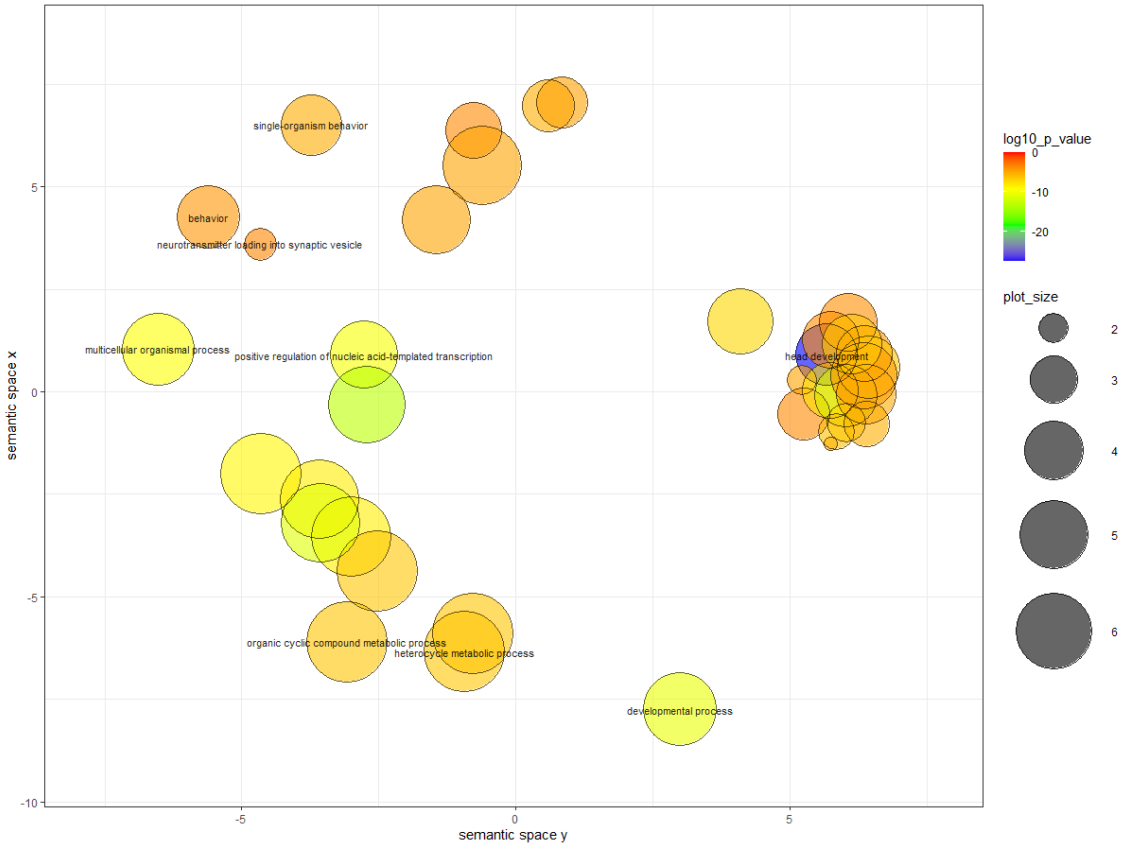


2. Perform gene ontology enrichment analysis of the two sets of DEGs using the command line wrapper of GOstats R package for biological processes. Plot results using any graphical representation and discuss results.

In over-expressed genes in the brain, there is, in a clear cluster, many genes related to embryonic development, such as sex differentiation, endoderm development or retina development. The most significant ontological term in this group is head development, which makes a lot of sense since we are using brain samples from mouse embryos. There are also other terms related to development a little further from this cluster, such as the developmental process or multicellular organismal process.

Other terms that present significant p-values are related to behavior and transport of neurotransmitters in synaptic vesicles, which clearly agrees with the type of tissue we are working with.

Finally, it should be noted that there are also many genes related to the regulation of transcription, both positively and negatively.



A significant number of ontological terms also appear in the genes overexpressed in the liver.

We can see, at the bottom, a very clear cluster with various elements related to the immune response and defense systems such as acute-phase response, inflammatory response or response to stress, among others. The liver, in humans, is a very important barrier against the entry of pathogens into the body, detecting those that can be found in the guts, what we could be observing here in mice (Immune Responses in the Liver. Kubes P. and Jenne C. 2018. Annu Rev Immunol). Also, the liver produces some of these elements, for example the acute-phase proteins.

There are also many genes related to the regulation of the metabolism and transport of different biomolecules such as lipids (sterol transport, triglyceride homeostasis or acylglycerol transport) or proteins (protein metabolism or peptide secretion). It is known the great number of functions that the liver performs in controlling the levels of certain substances in the blood to maintain homeostasis (Source).

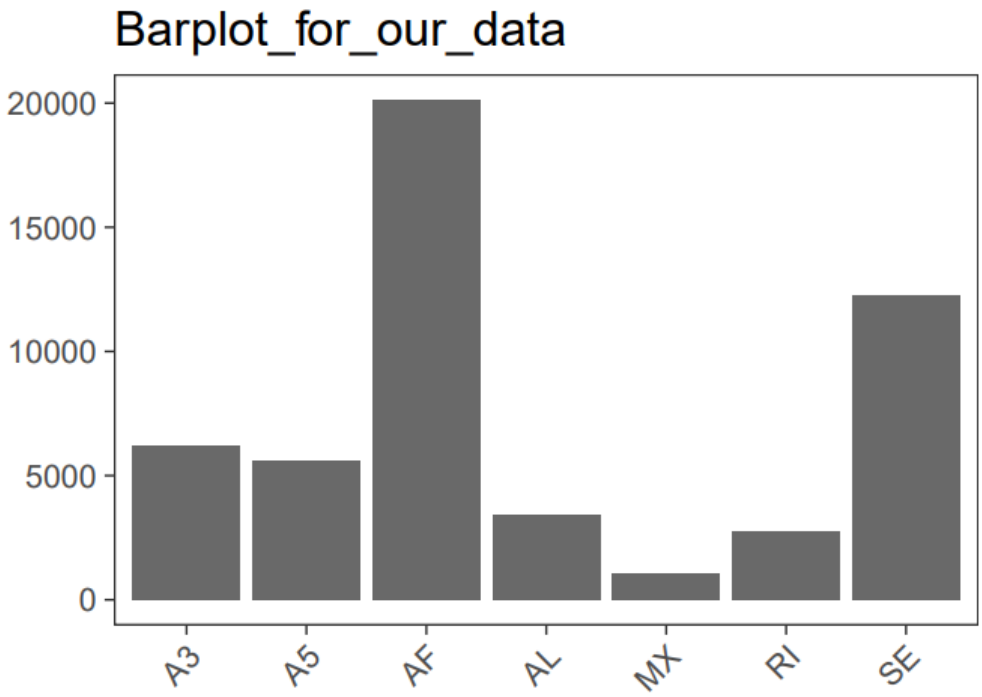
We also find different receptors as members of MAP kinase (ERK1 and ERK2) or the JAK2 protein, linked to cytokine receptors.

Finally, mention three of the terms with smaller p-values which are related to hemostasis (negative regulation of hemostasis, negative regulation of wound healing, negative regulation of coagulation) which could be preventing blood clotting when it is not necessary.

3. Analyze differential splicing using SUPPA between brain and liver for skipping exon, intron retention, mutually exclusive exon and alternative first exon. Plot top results using heatmaps. Different thresholds may be chosen for each event type.

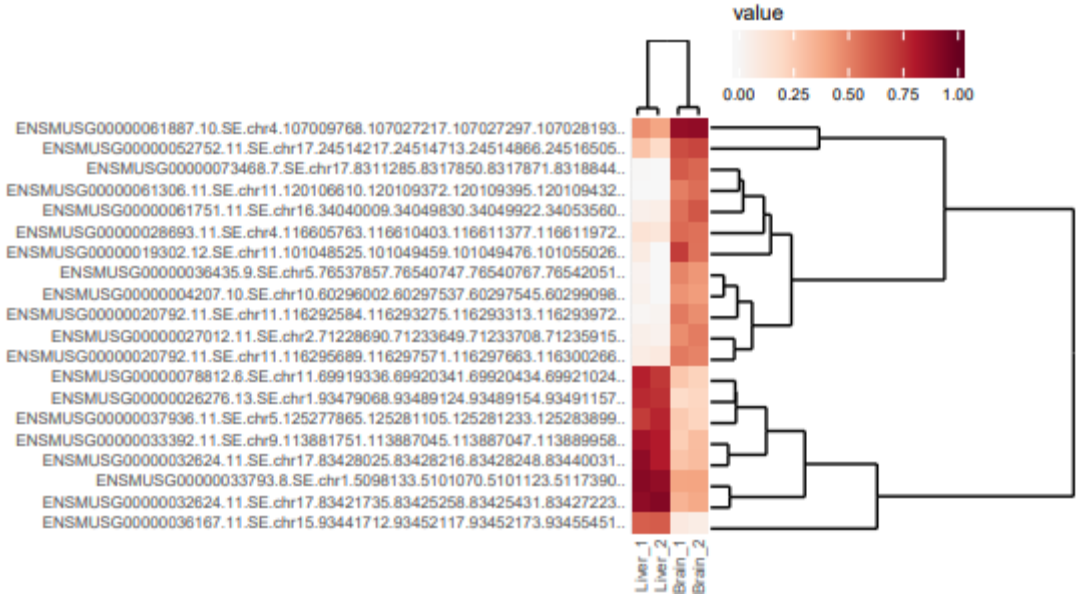
We obtained the following: 12,254 of Skipping exon, 2,763 of retained intron, 1,054 of mutually exclusive exons and 20,128 of alternative first exon counting lines of the files.

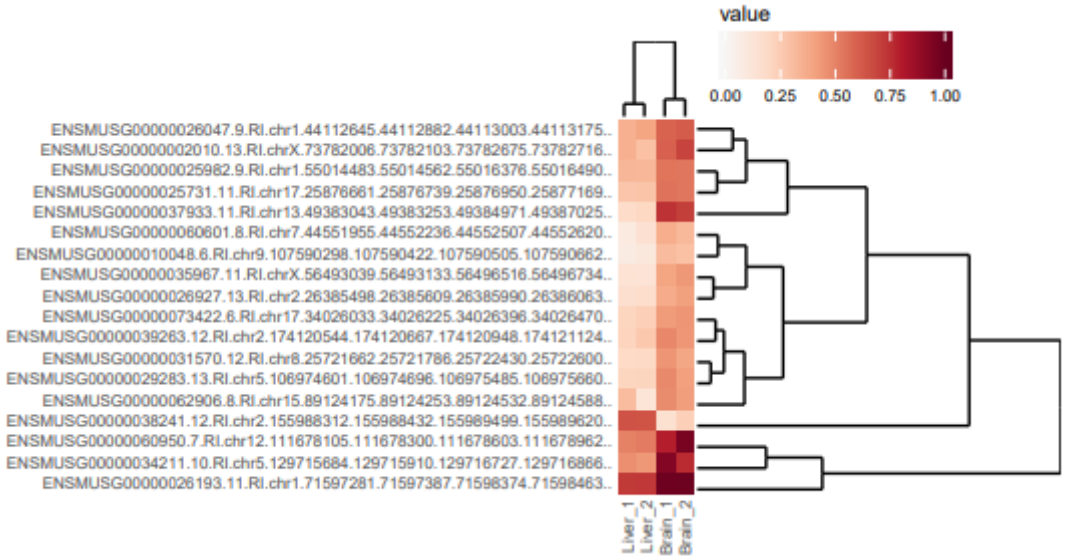
We create a bar plot with the different alternative splicing types used in this exercise and including also A3 and A5 ones.

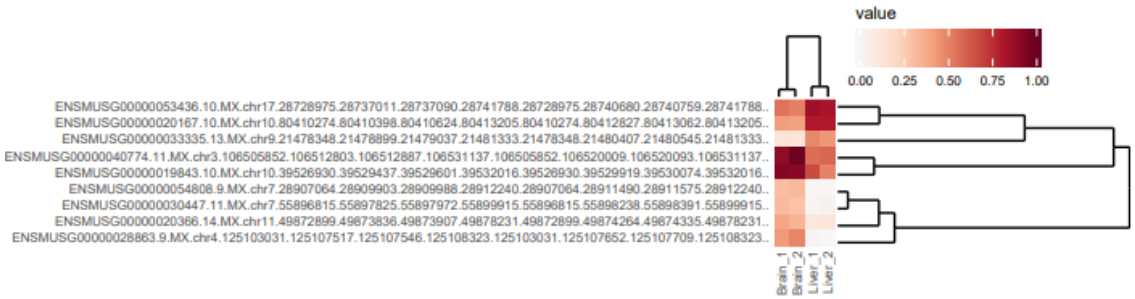


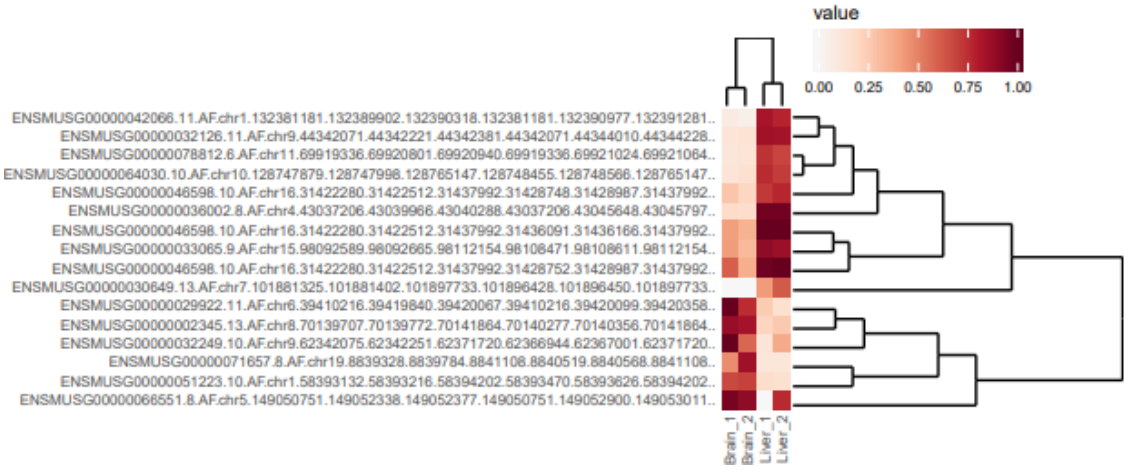
In the creation of the .tsv files, the thresholds had to be adjusted in the PSI and p-value values to achieve similar results in number of transcripts. For example, in the MX event, where not many transcripts were available, the levels had to be lowered to 0.2 PSI (column 2) and 0.1 p-value (column 3). On the other hand, for other events, such as AF, other thresholds were used since a greater number of elements were available.

In all event types we can see a correct separation between the two replicas of each tissue, which suggests a different alternative splicing pattern between brain and liver. On the other hand, and as we have seen previously, the number of times that each of the four events takes places is not the same, with SE and AF being the majority and RI and MX occurring less frequently.


value
0.00 0.25 0.50 0.75 1.00
ENSMUSG00000061887.10.SE.chr4.107009768.107027217.107027297.107028193..
ENSMUSG00000052752.11.SE.chr17.24514217.24514713.24514866.24516505..
ENSMUSG00000073468.7.SE.chr17.8311285.8317850.8317871.8318844..
ENSMUSG00000061306.11.SE.chr11.120106610.120109372.120109395.120109432..
ENSMUSG00000061751.11.SE.chr16.34040009.34049830.34049922.34053560..
ENSMUSG00000028693.11.SE.chr4.116605763.116610403.116611377.116611972..
ENSMUSG00000019302.12.SE.chr11.101048525.101049459.101049476.101055026..
ENSMUSG00000036435.9.SE.chr5.76537857.76540747.76540767.76542051..
ENSMUSG00000004207.10.SE.chr10.60296002.60297537.60297545.60299098..
ENSMUSG00000020792.11.SE.chr11.116292584.116293275.116293313.116293972..
ENSMUSG00000027012.11.SE.chr2.71228690.71233649.71233708.71235915..
ENSMUSG00000020792.11.SE.chr11.116295689.116297571.116297663.116300266..
ENSMUSG00000078812.6.SE.chr11.69919336.69920341.69920434.69921024..
ENSMUSG00000026276.13.SE.chr1.93479068.93489124.93489154.93491157..
ENSMUSG00000037936.11.SE.chr5.125277865.125281105.125281233.125283899..
ENSMUSG00000033392.11.SE.chr9.113881751.113887045.113887047.113889958..
ENSMUSG00000032624.11.SE.chr17.83428025.83428216.83428248.83440031..
ENSMUSG00000033793.8.SE.chr1.5098133.5101070.5101123.5117390..
ENSMUSG00000032624.11.SE.chr17.83421735.83425258.83425431.83427223..
ENSMUSG00000036167.11.SE.chr15.93441712.93452117.93452173.93455451..
Liver_1
Liver_2
Brain_1
Brain_2


value
0.00 0.25 0.50 0.75 1.00
ENSMUSG00000026047.9.RI.chr1.44112645.44112882.44113003.44113175..
ENSMUSG00000002010.13.RI.chrX.73782006.73782103.73782675.73782716..
ENSMUSG00000025982.9.RI.chr1.55014483.55014562.55016376.55016490..
ENSMUSG00000025731.11.RI.chr17.25876661.25876739.25876950.25877169..
ENSMUSG00000037933.11.RI.chr13.49383043.49383253.49384971.49387025..
ENSMUSG00000060601.8.RI.chr7.44551955.44552236.44552507.44552620..
ENSMUSG00000010048.6.RI.chr9.107590298.107590422.107590505.107590662..
ENSMUSG00000035967.11.RI.chrX.56493039.56493133.56496516.56496734..
ENSMUSG00000026927.13.RI.chr2.26385498.26385609.26385990.26386063..
ENSMUSG00000073422.6.RI.chr17.34026033.34026225.34026396.34026470..
ENSMUSG00000039263.12.RI.chr2.174120544.174120667.174120948.174121124..
ENSMUSG00000031570.12.RI.chr8.25721662.25721786.25722430.25722600..
ENSMUSG00000029283.13.RI.chr5.106974601.106974696.106975485.106975660..
ENSMUSG00000062906.8.RI.chr15.89124175.89124253.89124532.89124588..
ENSMUSG00000038241.12.RI.chr2.155988312.155988432.155989499.155989620..
ENSMUSG00000060950.7.RI.chr12.111678105.111678300.111678603.111678962..
ENSMUSG00000034211.10.RI.chr5.129715684.129715910.129716727.129716866..
ENSMUSG00000026193.11.RI.chr1.71597281.71597387.71598374.71598463..
Liver_1
Liver_2
Brain_1
Brain_2

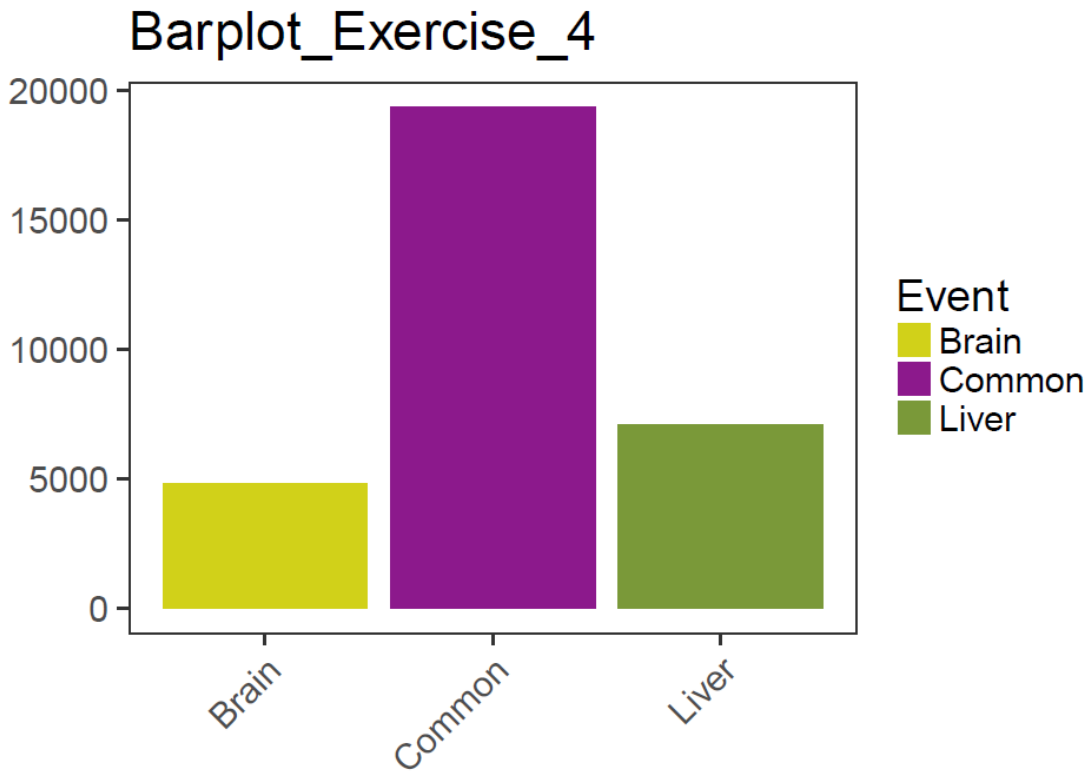

value
0.00 0.25 0.50 0.75 1.00
ENSMUSG00000053436.10.MX.chr17.28728975.28737011.28737090.28741788.28728975.28740680.28740759.28741788..
ENSMUSG00000020167.10.MX.chr10.80410274.80410398.80410624.80413205.80410274.80412827.80413062.80413205..
ENSMUSG00000033335.13.MX.chr9.21478348.21478899.21479037.21481333.21478348.21480407.21480545.21481333..
ENSMUSG00000040774.11.MX.chr3.106505852.106512803.106512887.106531137.106505852.106520009.106520093.106531137..
ENSMUSG00000019843.10.MX.chr10.39526930.39529437.39529601.39532016.39526930.39529919.39530074.39532016..
ENSMUSG00000054808.9.MX.chr7.28907064.28909903.28909988.28912240.28907064.28911490.28911575.28912240..
ENSMUSG00000030447.11.MX.chr7.55896815.55897825.55897972.55899915.55896815.55898238.55898391.55899915..
ENSMUSG00000020366.14.MX.chr11.49872899.49873836.49873907.49878231.49872899.49874264.49874335.49878231..
ENSMUSG00000028863.9.MX.chr4.125103031.125107517.125107546.125108323.125103031.125107652.125107709.125108323..
Brain_1
Brain_2
Liver_1
Liver_2



4. Find H3K4me3 peaks shared by brain and liver and the ones exclusively found in each tissue using the narrow peaks found in /tutorial/results using bedtools intersect. Show results using a bar plot colored by the color code used during the hands-on. Palette is available at /tutorial/palettes/palTissue.txt. Any color may be chosen for shared peaks.

Methylation at the H3K4me3 position is a triple methylation at the fourth lysine of histone H3 and is normally associated with increased transcription.

It is surprising that most methylation peaks are shared between the two tissues. On the other hand, if we focus on the specific peaks, we can see that the liver has a greater number of these ones. These results could be indicating that the methylation patterns in H3K4me3 are quite similar between the two tissues since they share a significant number of peaks.



5. Create a BED file of 200bp up/downstream TSS of genes and overlap DEGs (step 1) with the 3 sets of H3K4me3 peaks classified in the previous step (4). Show three examples in the UCSC genome browser, including RNA-seq, ChIP-seq and ATAC-seq tracks. Ideally, one example of each peak set (i.e. shared peak, peak exclusively called in brain and peak exclusively called in liver). Discuss the integration of the three datasets in the TSS of the selected cases.

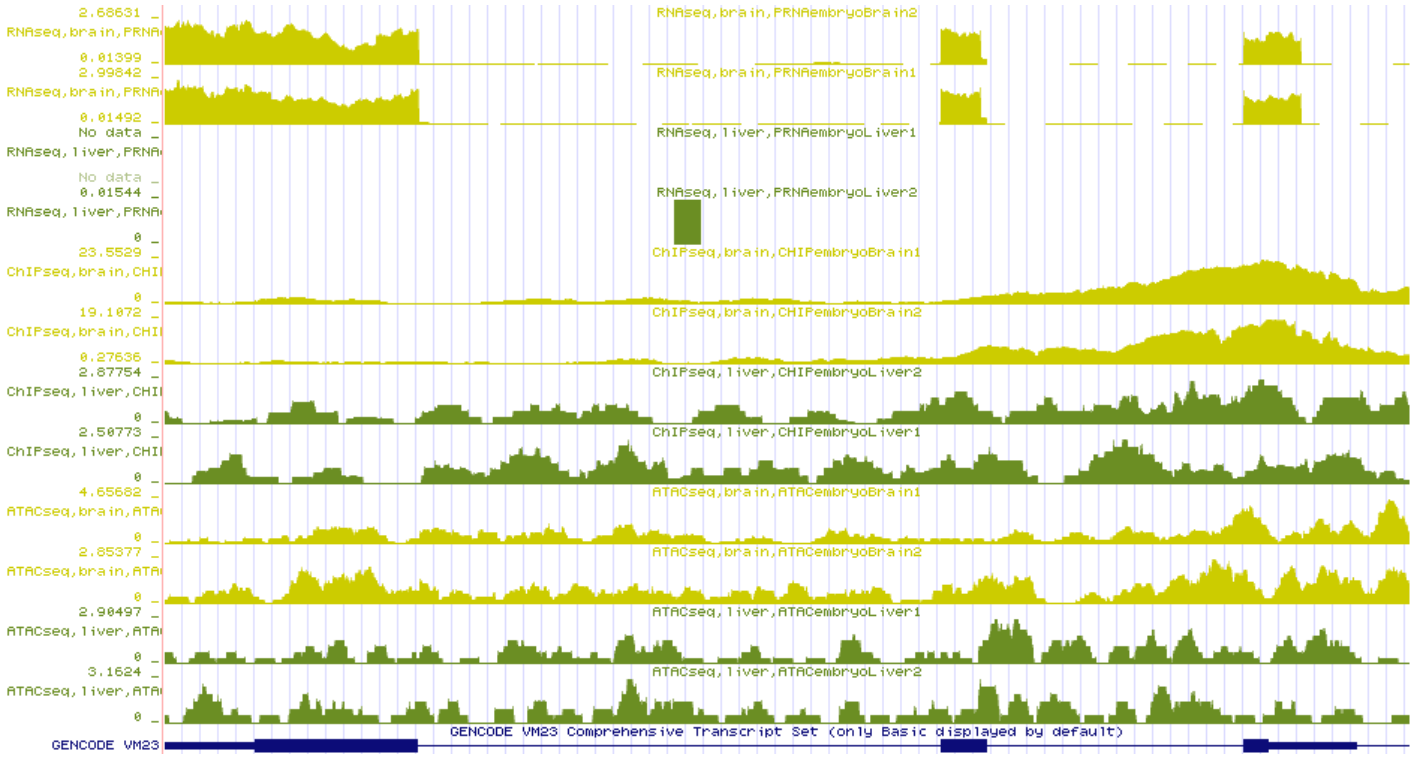
Counting the rows of the files created, we can see that there are no genes that are over-expressed in one tissue and only marked in the other. The rest of the results have been the following: 15 over expressed in brain and methylated in brain, 58 marked in both tissues and over expressed in brain, 19 expressed in liver and marked in it and 5 marked in both tissues and over expressed in liver .

1. OVEREXPRESSED BRAIN:

The protein chosen for genes expressed in the brain and marked with H3K4me3 in this same tissue has been the Homeobox protein (OTX2). The gene that codes for this protein is found in the - strand.

The RNA-seq results show clearly how it is not expressed in liver but it is expressed in brain. In the ChIP-seq results we can see how in brain the TSS region is abundantly methylated, whereas, in liver, methylation is distributed throughout the gene with some peaks and not abundant in the TSS region. Finally, in the Atac-seq results we can see that there are no major differences in the accessibility to chromatin in both tissues, although it is slightly higher in the brain and, above all, it also increases in the TSS region, which usually implies a higher transcription.

All these results coincide with what we had determined, the OTX2 protein is over-expressed in the brain and is not present in liver.

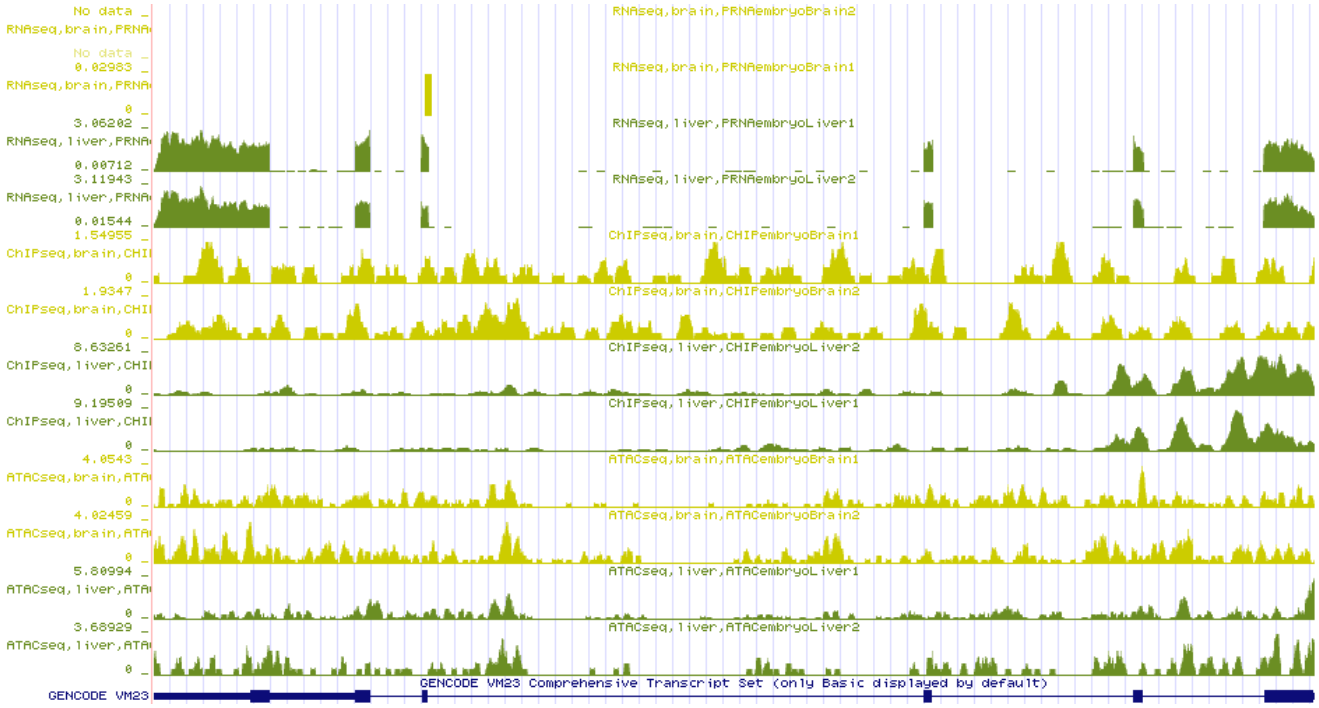


2. OVEREXPRESSED LIVER:

The protein over-expressed in liver with methylation peaks has been UDP-glucuronosyltransferase (Ugt2b34), a protein with a glycosyltransferase function. The gene of it is found in the - strand.

The results are quite similar to those found with the brain-specific protein but in liver. In RNA-Seq we can see how it is expressed only in the liver with abundant reads in exonic regions. In ChIP-seq, we can see how methylation in the brain is generalized throughout the gene and, in the liver, it is much more abundant in TSS and practically null in the body of the gene. Finally, in Atac-seq, the accessibility to chromatin does not seem to show great differences between the two tissues.

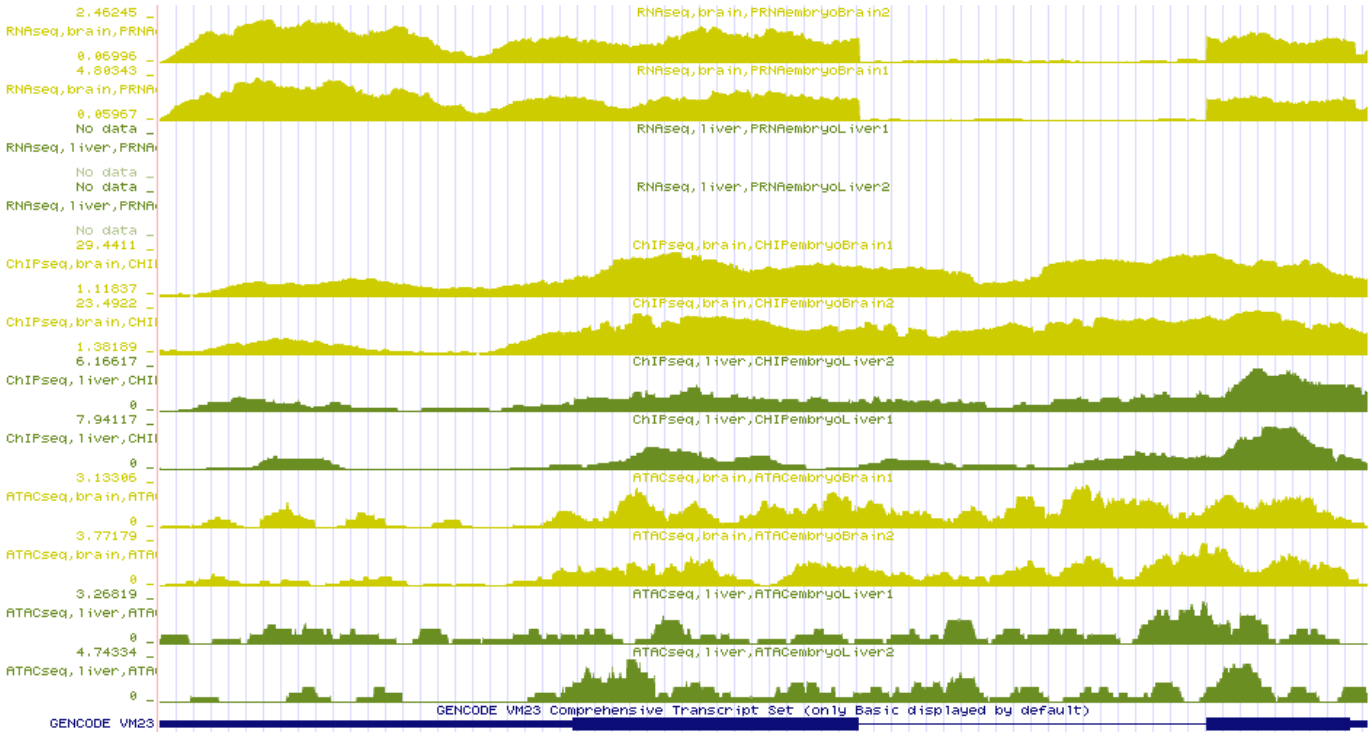
Although the results of Atac-seq are not entirely clear, the other two experiments are enough to corroborate the results obtained.



3. OVEREXPRESSED IN BRAIN TISSUES, PEAKS IN BOTH TISSUES:

Finally, a protein that has been over-expressed in the brain but that has peaks in both tissues has been chosen. The protein is NK2 homeobox 1 (NKX2-1) the gene of which is found in the - strand and acts as a regulator of transcription in various tissues.

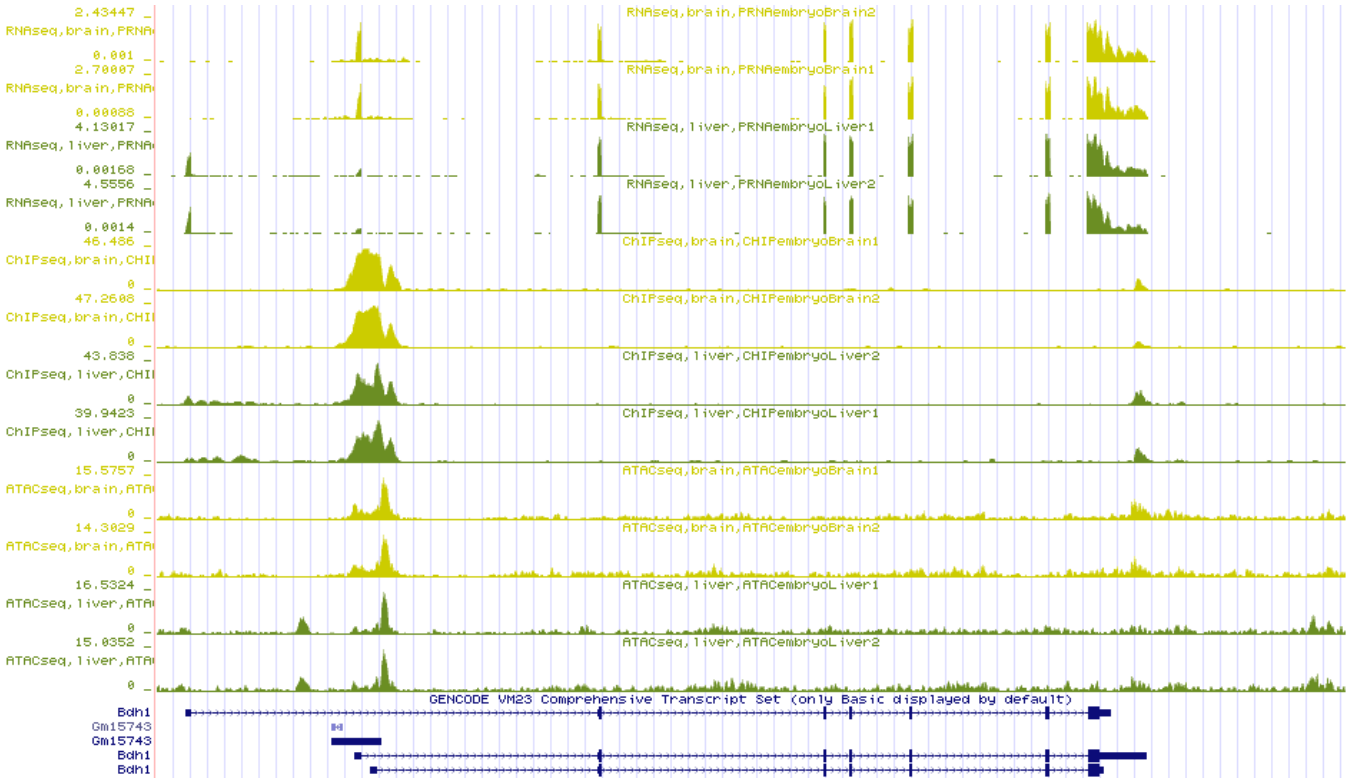
In RNA-seq, we can see how it is only detected in the brain but not in the liver. In contrast, methylation is present in both tissues, although more abundantly in brain samples. This could be related to the fact that it is only expressed in the brain. Finally, in Atac-seq we can see that accessibility seems to be greater in the brain, although it is visible in both tissues. This fact could also be linked to the lack of expression in the liver. The results, therefore, coincide with those obtained in the BED file.


RNAseq,brain,PRNAembryoBrain2
RNAseq,brain,PRNAembryoBrain1
RNAseq,liver,PRNAembryoLiver1
RNAseq,liver,PRNAembryoLiver2
ChIPseq,brain,CHIPembryoBrain1
ChIPseq,brain,CHIPembryoBrain2
ChIPseq,liver,CHIPembryoLiver2
ChIPseq,liver,CHIPembryoLiver1
ATACseq,brain,ATACembryoBrain1
ATACseq,brain,ATACembryoBrain2
ATACseq,liver,ATACembryoLiver1
ATACseq,liver,ATACembryoLiver2
GENCODE VM23 Comprehensive Transcript Set (only Basic displayed by default)
GENCODE VM23

6. Show two examples of alternative first exons in the UCSC genome browser, including RNA-seq, ChIP-seq and ATAC-seq tracks. Discuss the integration of the three datasets in the TSS of the selected cases.

One of the genes that appeared as an example of alternative first exon was Bdh1. When using the UCSC genome browser, we can see, in RNA-seq, how an exon appears in the liver that is not present in brain. The ChIP-seq results show a similar pattern of methylation in various regions, but, in the liver, there is a peak located at the exclusive exon of that tissue which could be the TSS. The Atac-seq results do not show great differences between the two tissues.

Finally, comment, that there is an lnRNA (Gm15743) that is located just in the region of the first brain exon. The second methylation peak in liver could be caused by this sequence in the same way that the second peak in RNA-seq would be. However, the fact that there are several isoforms of the gene with different exons and that the RNA-seq results show a liver exon that is not present in the brain seems enough information to corroborate our results.



A second example has been the Tmcc2 gene found in the - strand. In the same way as before we can see that in the brain RNA-seq an exon appears that is not present in the liver. In the ChIP-seq of the brain a peak, that would correspond to its TSS region, can be seen, while in the liver, we can see two of these peaks. The second of these would correspond to the TSS of the isoform expressed in this tissue and it is using another exon. The Atac-seq results do not show great differences, although greater peaks appear in the liver.

The existence of several isoforms and the absence of the first exon in the liver corroborates our results.