Images and Plots used in Workflow and Their Brief Description

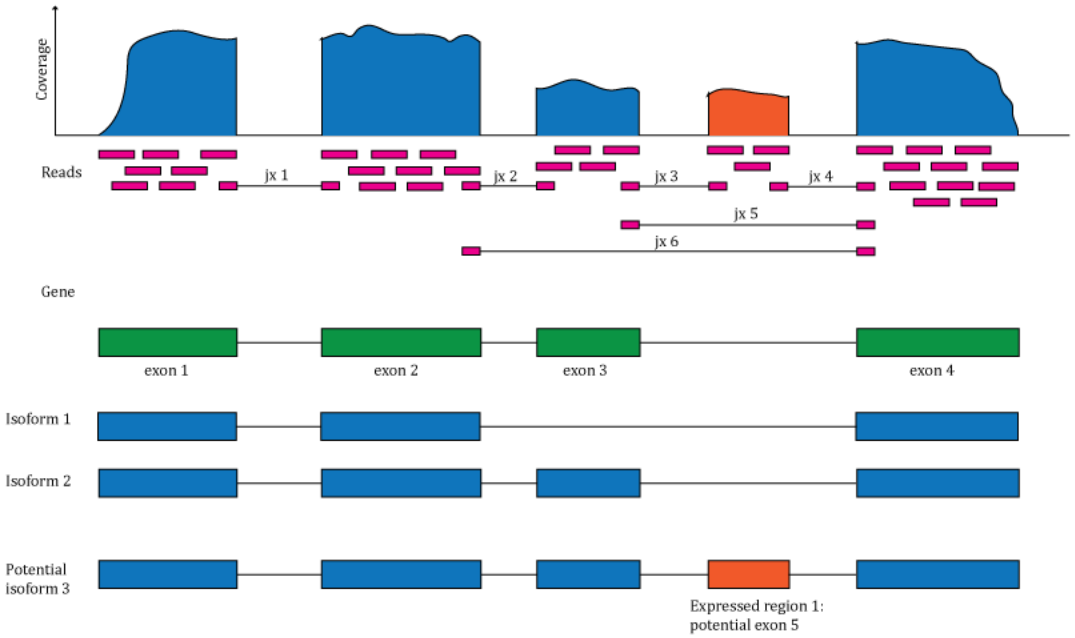

Figure 1: Overview of the data available in recount2. Reads (pink boxes) aligned to the reference genome can be used to compute a base-pair coverage curve and identify exon-exon junctions (split reads). Gene and exon count matrices are generated using annotation information providing the gene (green boxes) and exon (blue boxes) coordinates together with the base-level coverage curve. The reads spanning exon-exon junctions (jx) are used to compute a third count matrix that might include unannotated junctions (jx 3 and 4). Without using annotation information, expressed regions (orange box) can be determined from the base-level coverage curve to then construct data-driven count matrices.

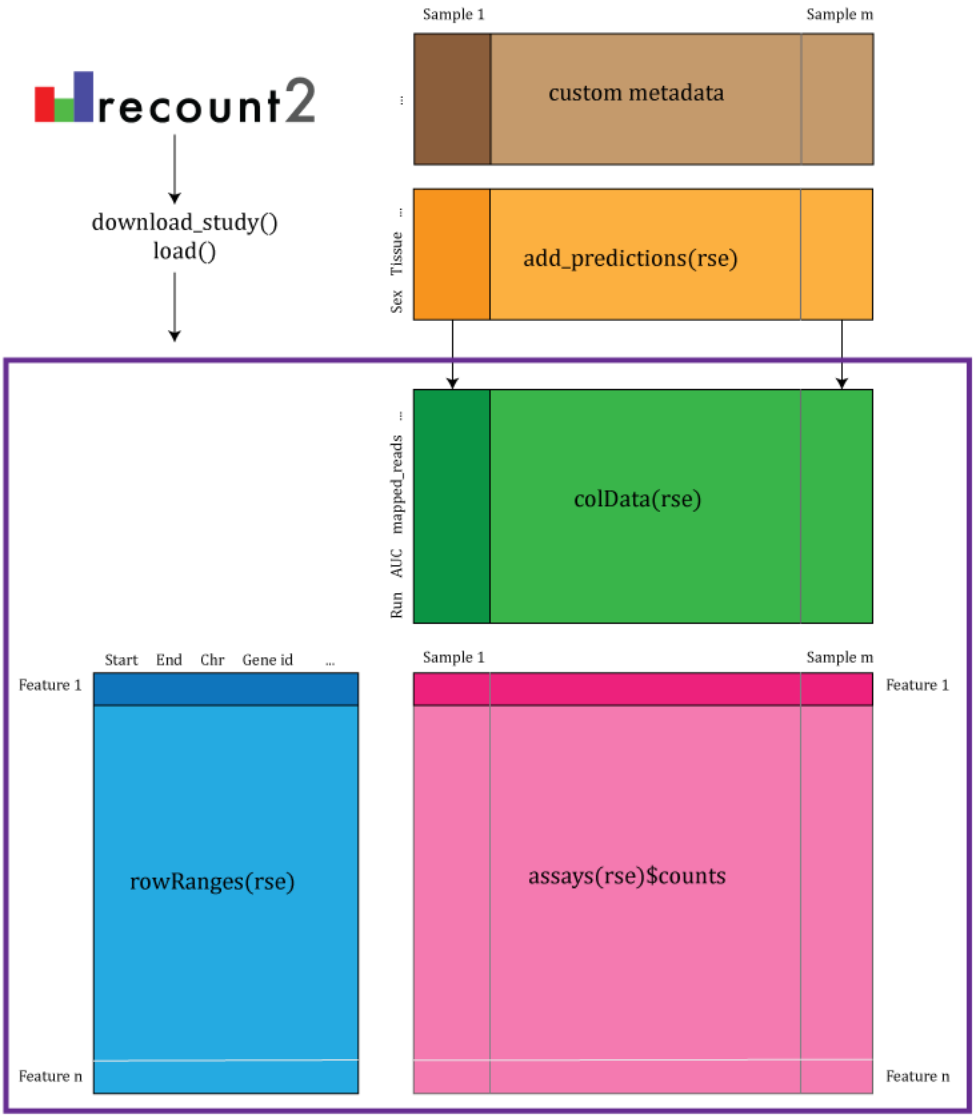

Figure 2: recount2 provides coverage count matrices in RangedSummarizedExperiment (rse) objects. Once the rse object has been downloaded and loaded into R, the feature information is accessed with rowRanges(rse) (blue box), the counts with assays(rse)$counts (pink box) and the sample metadata with colData(rse) (green box). The sample metadata can be expanded using add_predictions(rse) (orange box) or with custom code (brown box) matching by a unique sample identifier such as the SRA Run ID. The rse object is inside the purple box and matching data is highlighted in each box.

Figure 3: RNA-seq starting data. 16 RNA-seq un-aligned RNA-seq reads 3 base-pairs long are shown (pink boxes) alongside a reference genome that is 16 base-pairs long (white box).

Figure 3: RNA-seq starting data. 16 RNA-seq un-aligned RNA-seq reads 3 base-pairs long are shown (pink boxes) alongside a reference genome that is 16 base-pairs long (white box).

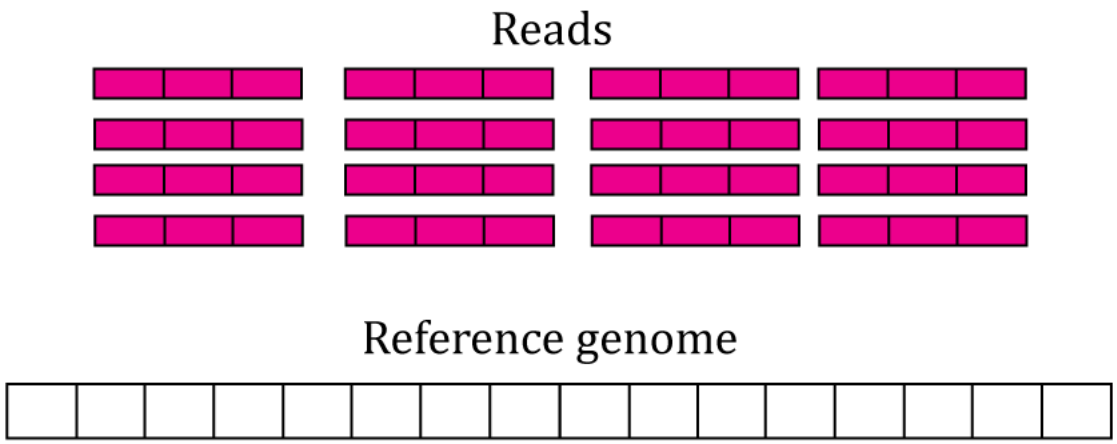

Figure 3: RNA-seq starting data. 16 RNA-seq un-aligned RNA-seq reads 3 base-pairs long are shown (pink boxes) alongside a reference genome that is 16 base-pairs long (white box).

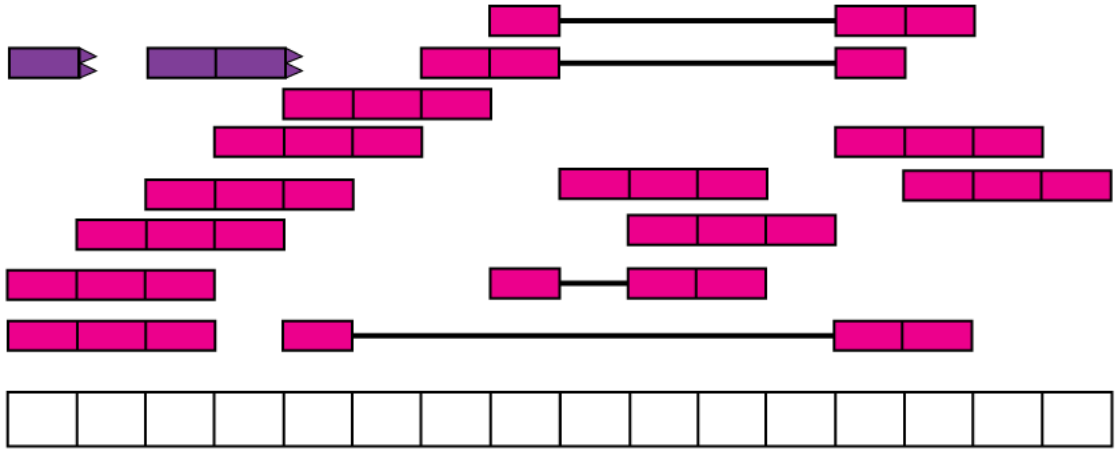

Figure 4: Aligned RNA-seq reads. Spice-aware RNA-seq aligners such as Rail-RNA are able to find the coordinates to which the reads map, even if they span exon-exon junctions (connected boxes). Rail-RNA soft clips some reads (purple boxes with rough edges) such that a portion of these reads align to the reference genome.

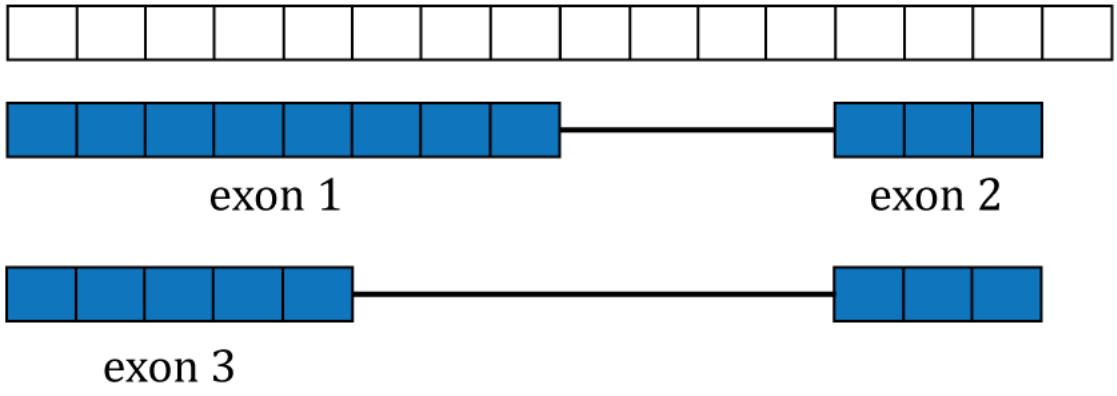

Figure 5: Gene annotation. A single gene with two isoforms composed by three distinct exons (blue boxes) is illustrated. Exons 1 and 3 share the first five base-pairs while exon 2 is common to both isoforms.

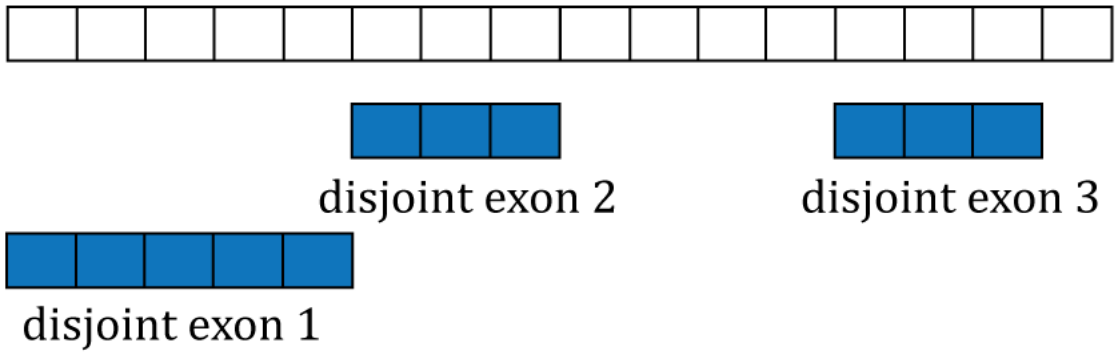

Figure 6: Disjoint exons. Windows of distinct exonic sequence for the example gene. Disjoint exons 1 and 2 form exon 1.

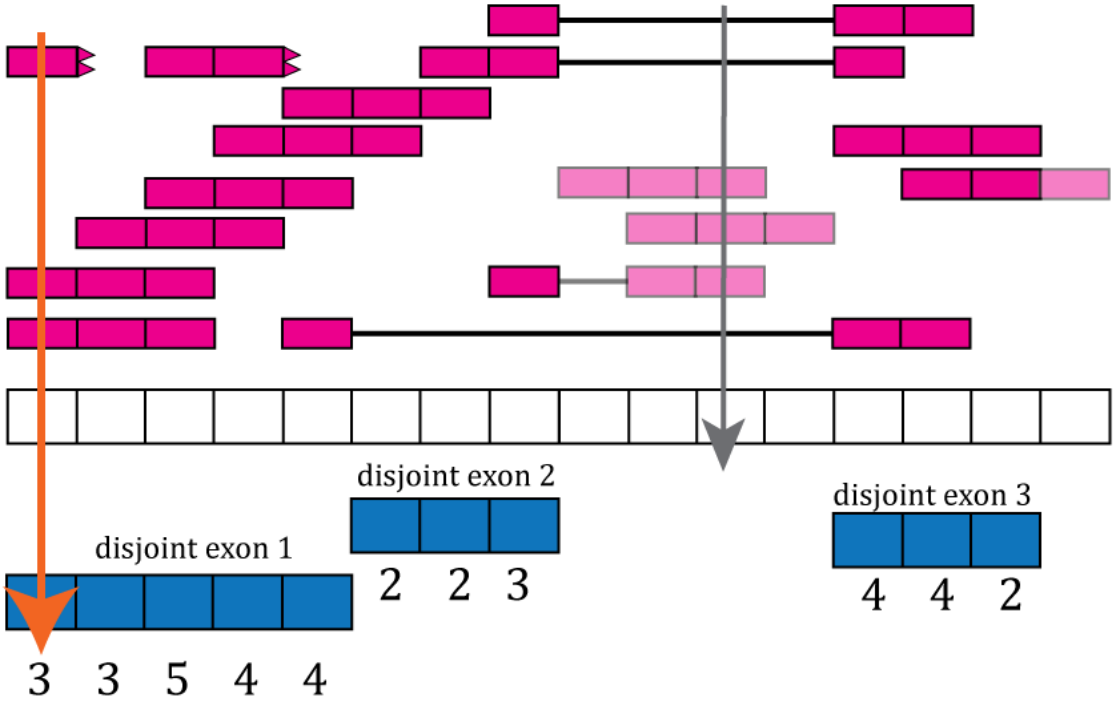

Figure 7: Base-pair coverage counting for exonic base-pairs. At each exonic base-pair we compute the number of reads overlapping that given base-pair. The first base (orange arrow) has 3 reads overlapping that base-pair. Base-pair 11 has a coverage of 3 but does not overlap known exonic sequence, so that information is not used for the gene and exon count matrices (grey arrow). If a read partially overlaps exonic sequence, only the portion that overlaps is used in the computation (see right most read).

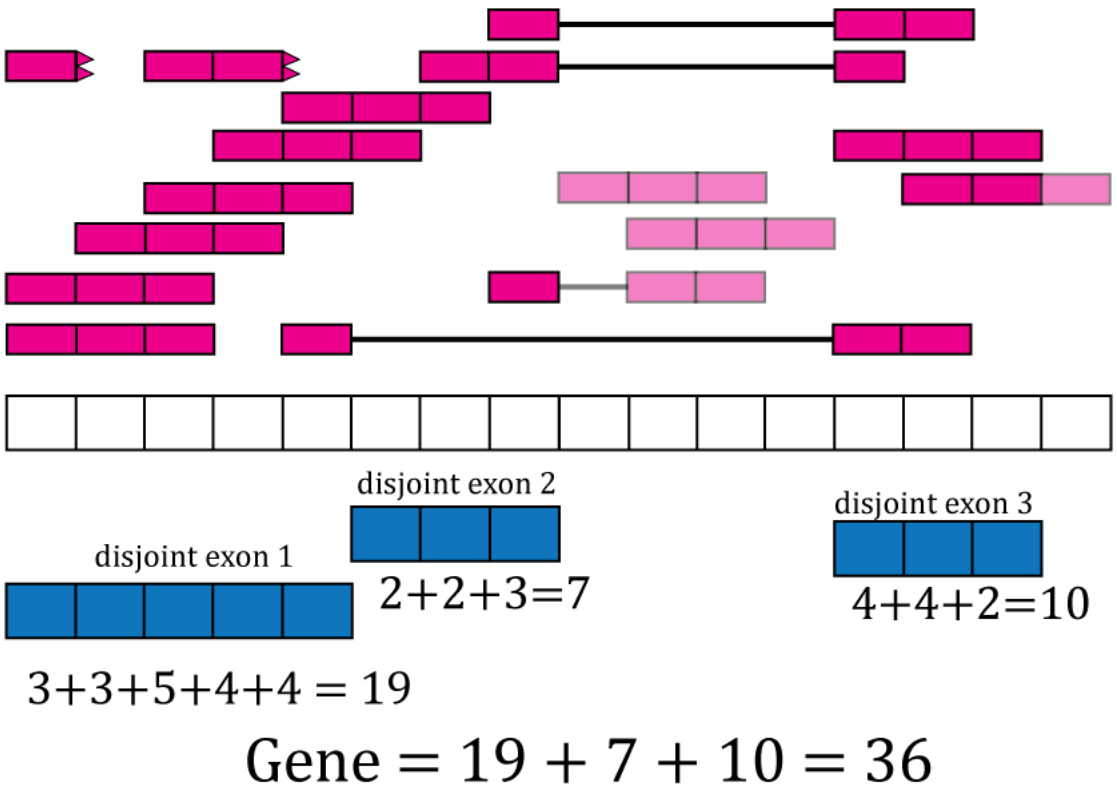

Figure 8: Exon and gene coverage counts. The coverage counts for each disjoint exon are the sum of the base-pair coverage. The gene coverage count is the sum of the disjoint exons coverage counts.

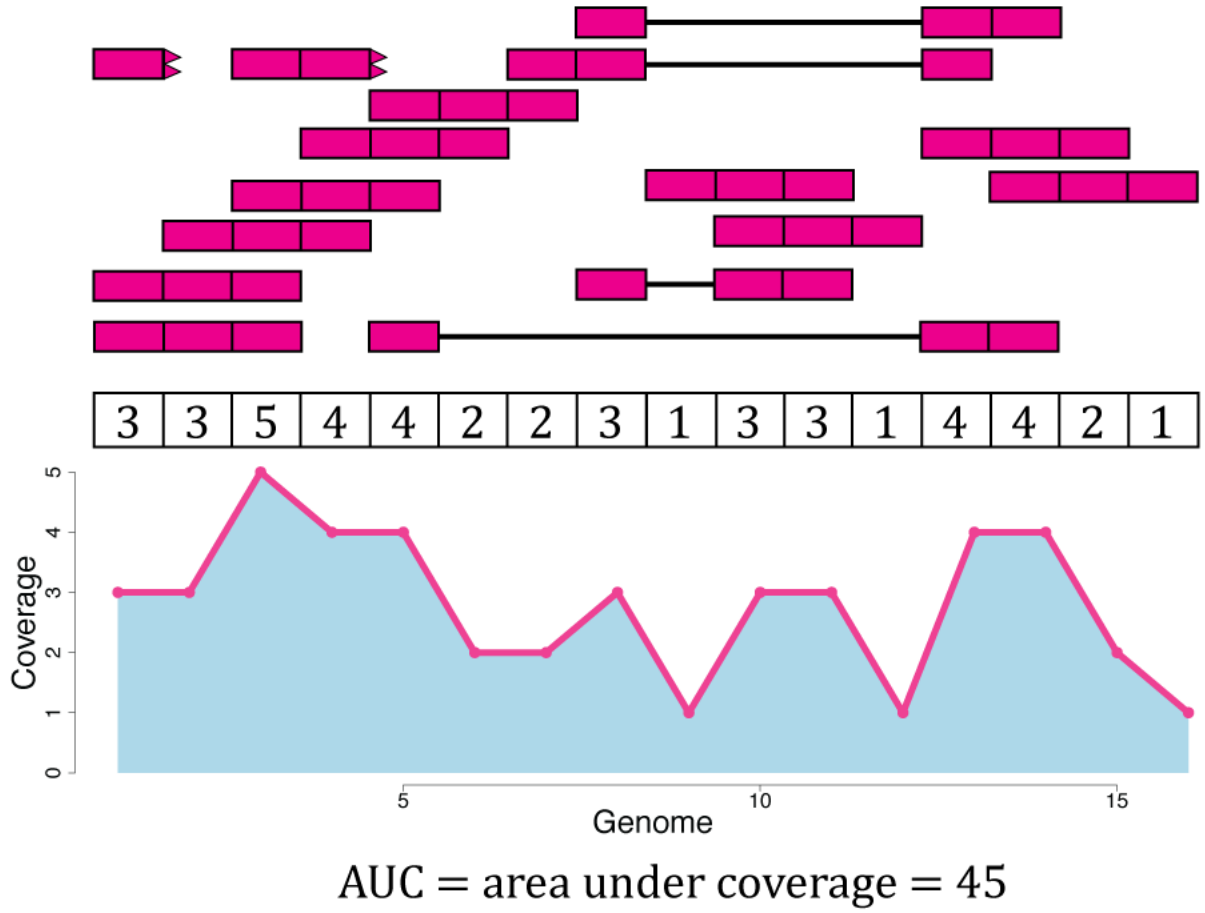

Figure 9: Area under coverage (AUC). The area under coverage is the sum of the base-pair coverage for all positions in the genome regardless of the annotation. It is the area under the base-level coverage curve shown as the light blue area under the pink curve.

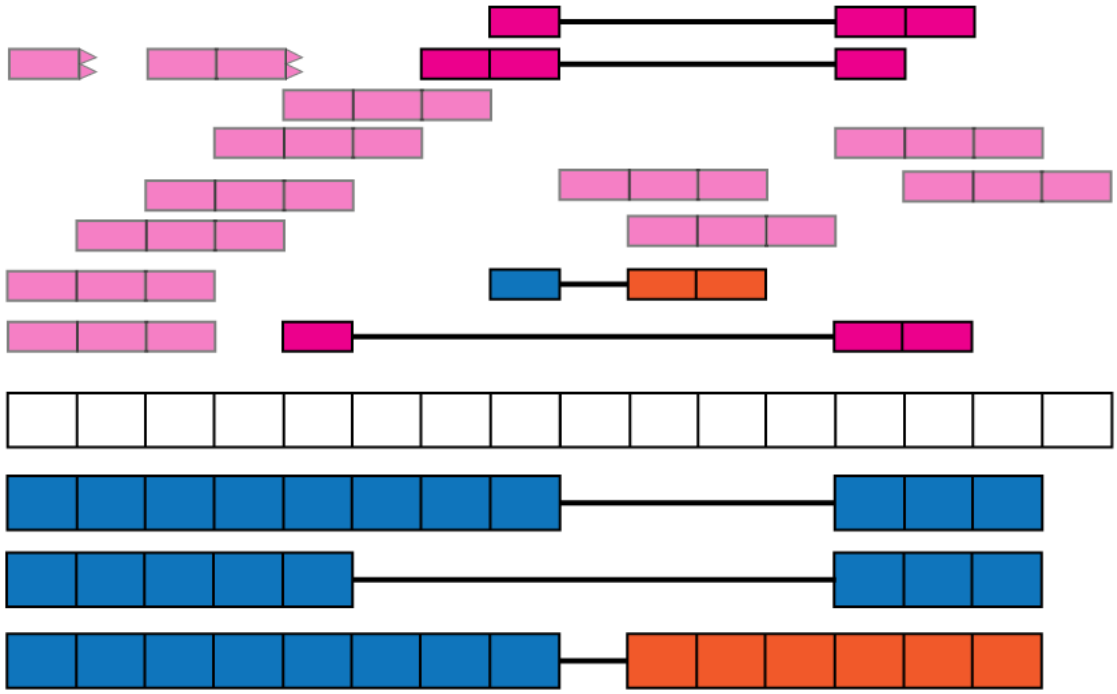

Figure 10: Exon-exon junctions go beyond the annotation. Reads spanning exon-exon junctions are highlighted and compared against the annotation. Three of them match the annotated junctions, but one (blue and orange read) spans an unannotated exon-exon junction with the left end matching the annotation and the right end hinting at a possible new isoform for this gene (blue and orange isoform).

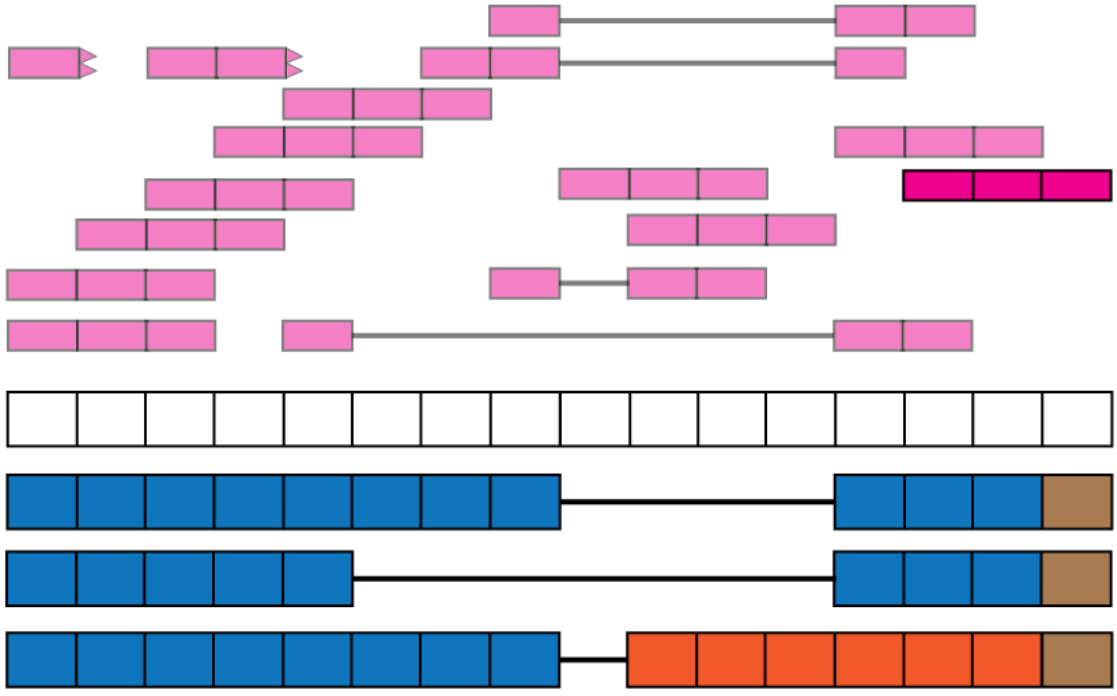

Figure 11: Intron retention events. Some reads might align with known intronic segments of the genome and provide information for exploring intron retention events (pink read). Some might support an intron retention event or a new isoform when coupled with exon-exon junction data (orange read).

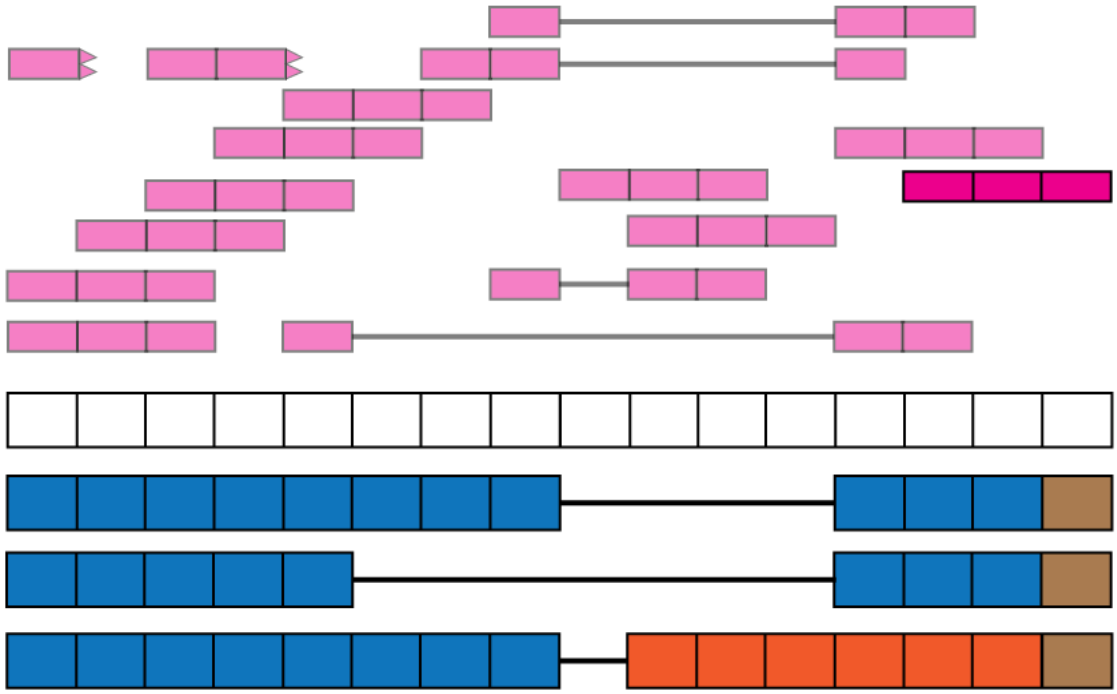

Figure 12: Exon boundaries. Reads that go beyond the known exon boundaries can inform us of whether the annotated boundaries are correct or if there was a run-off transcription event.

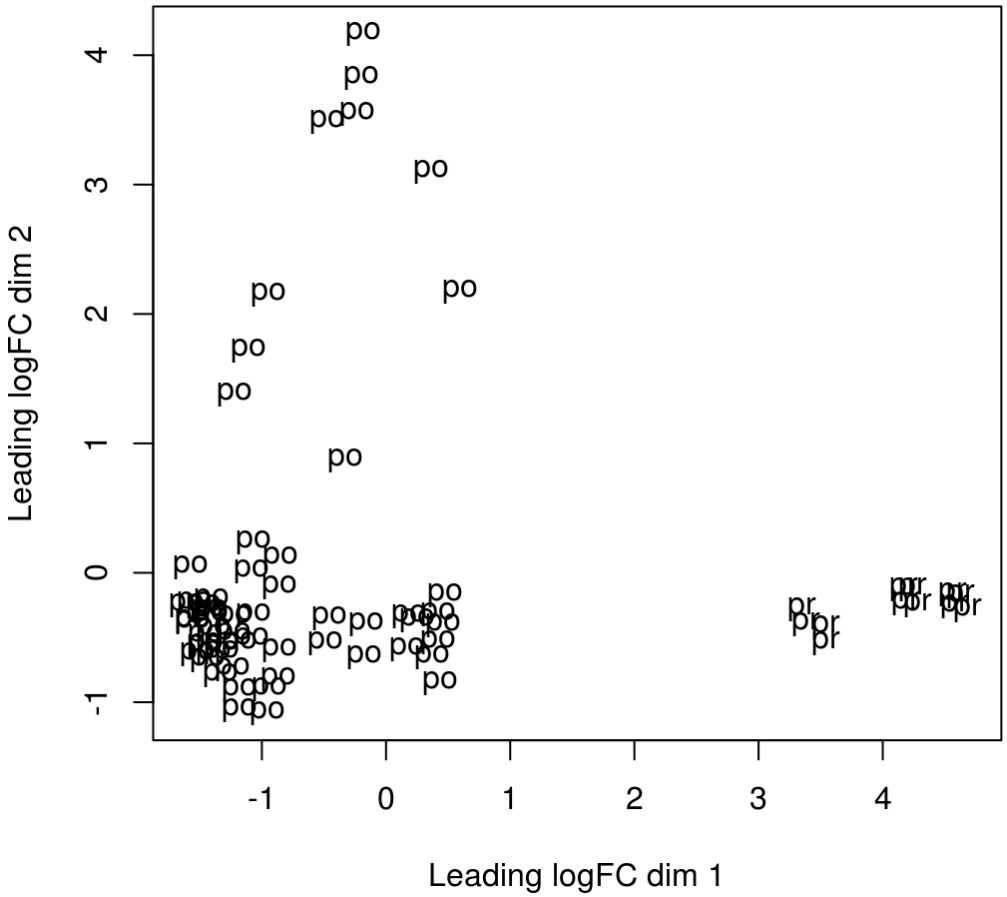

Figure 13: Multi-dimensional scaling plot of the gene-level data by age group.

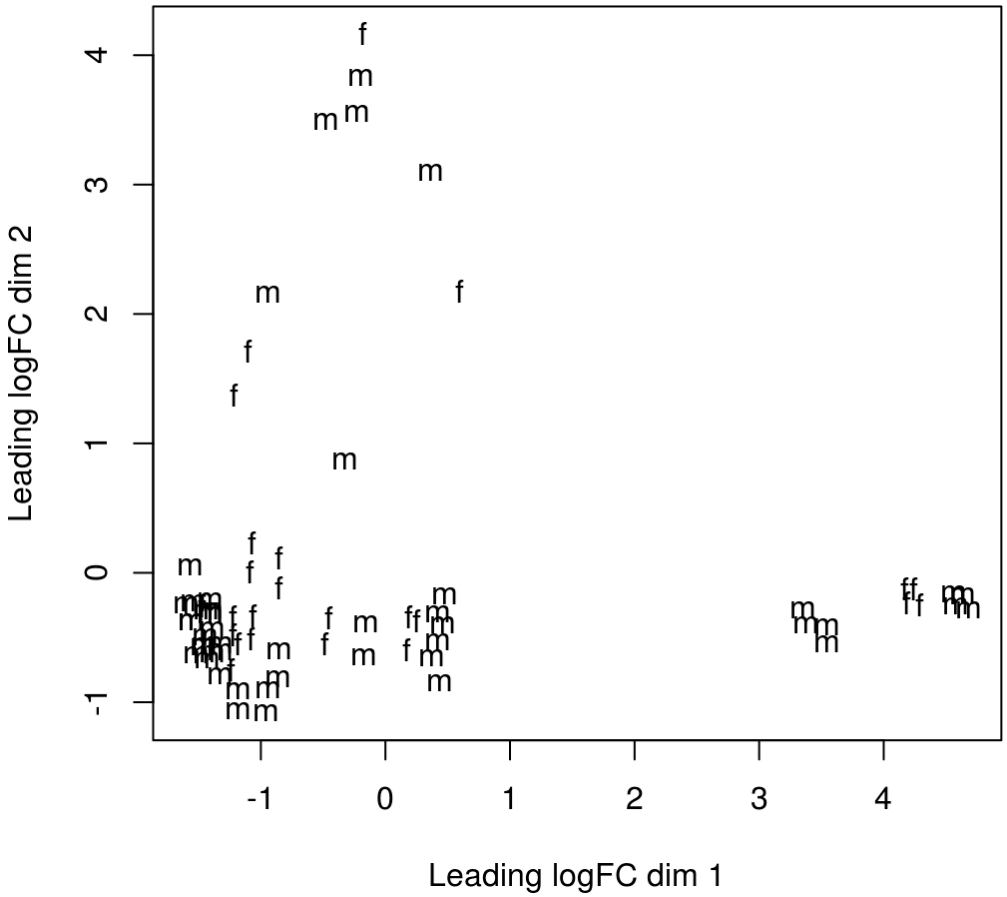

Figure 14: Multi-dimensional scaling plot of the gene-level data by sex.

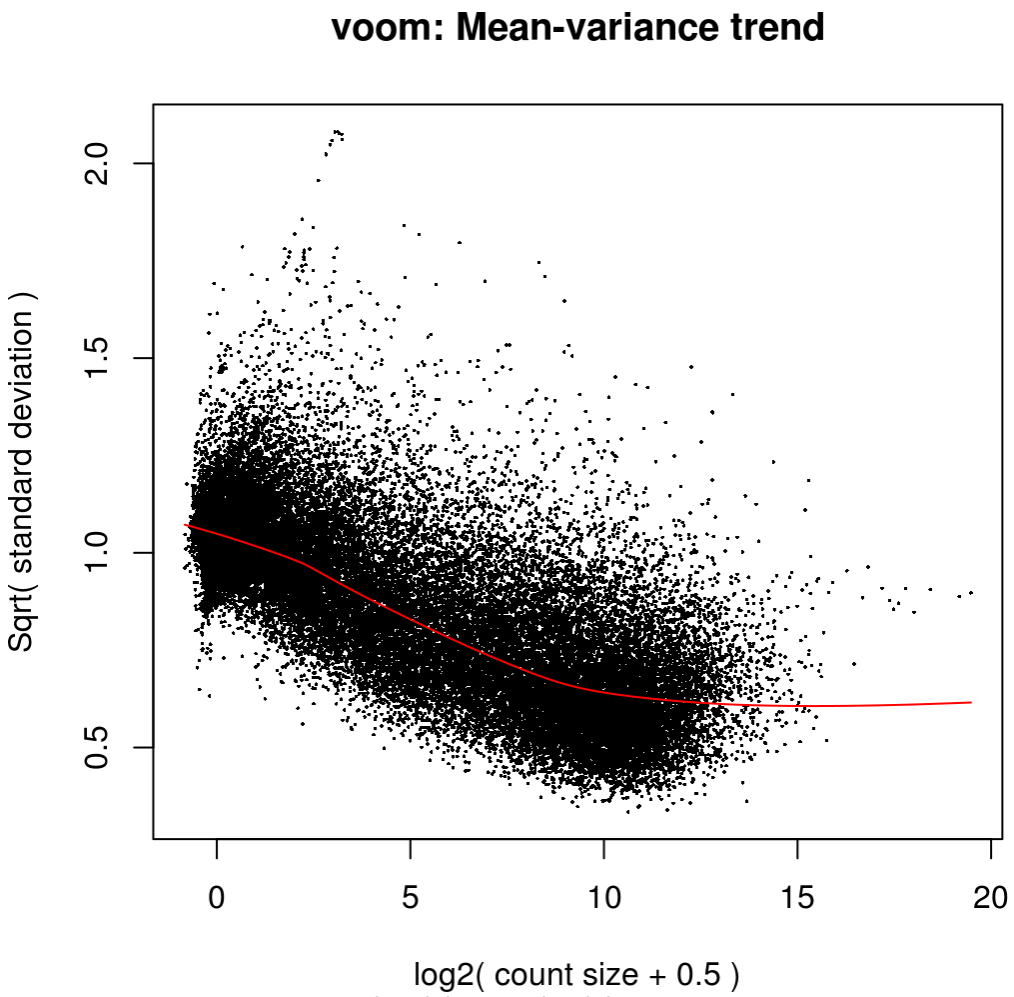

Figure 15: voom mean-variance plot of the gene-level data.

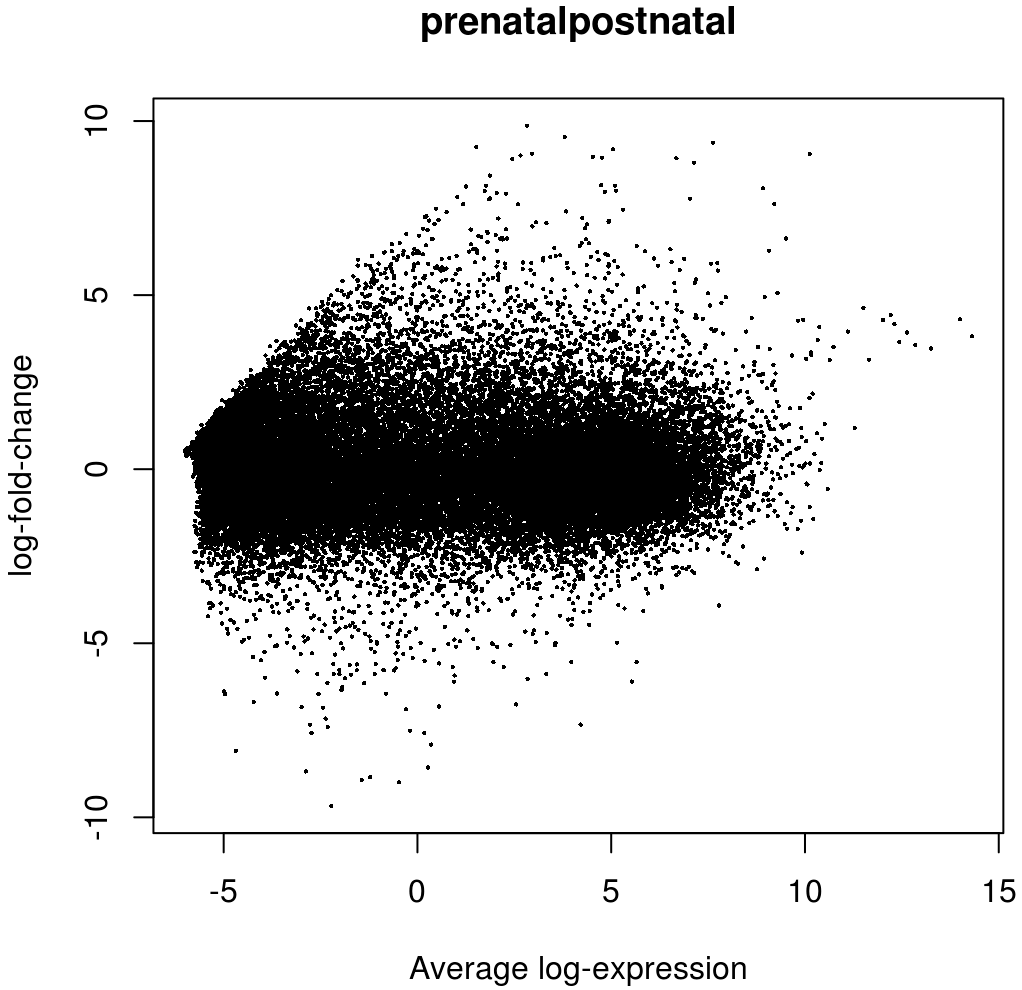

Figure 16: MA plot of the gene-level data. Testing for prenatal and postnatal DE adjusting for sex and RIN.

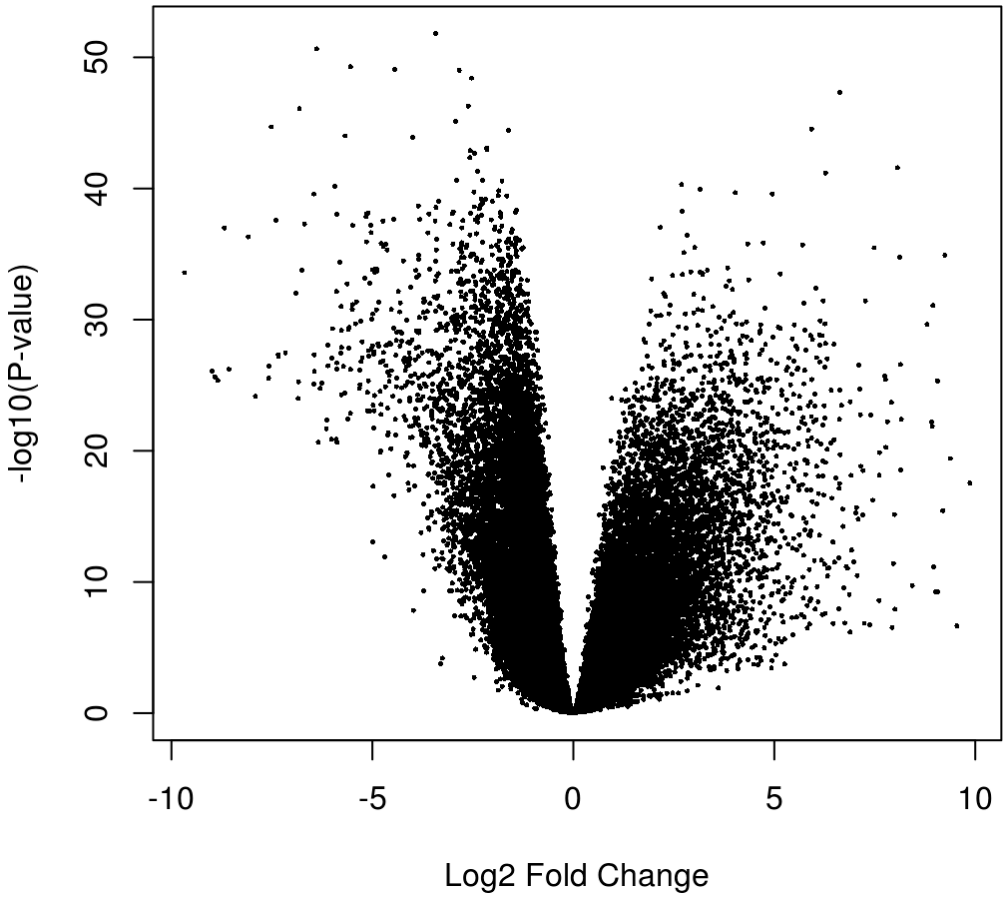

Figure 17: Volcano plot of the gene-level data. Testing for prenatal and postnatal DE adjusting for sex and RIN.

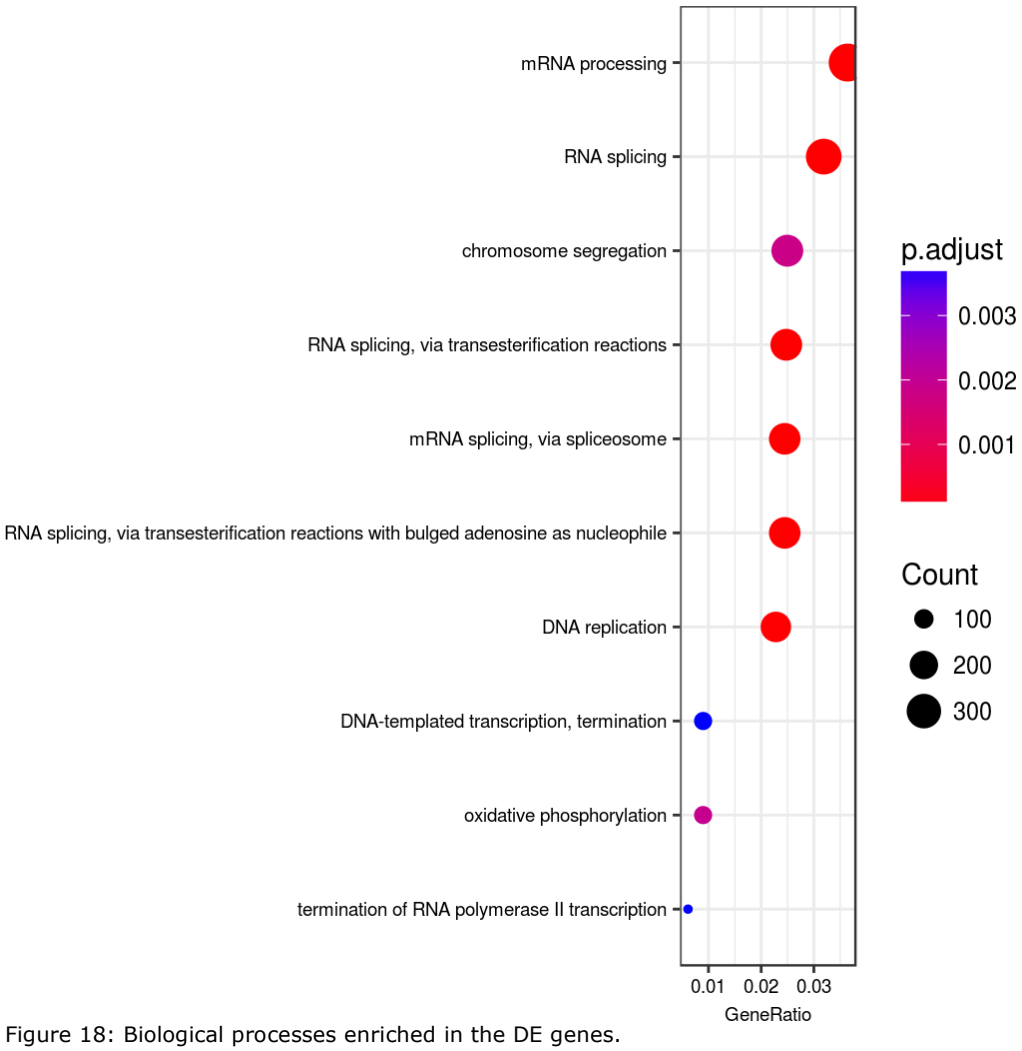

Figure 18: Biological processes enriched in the DE genes.

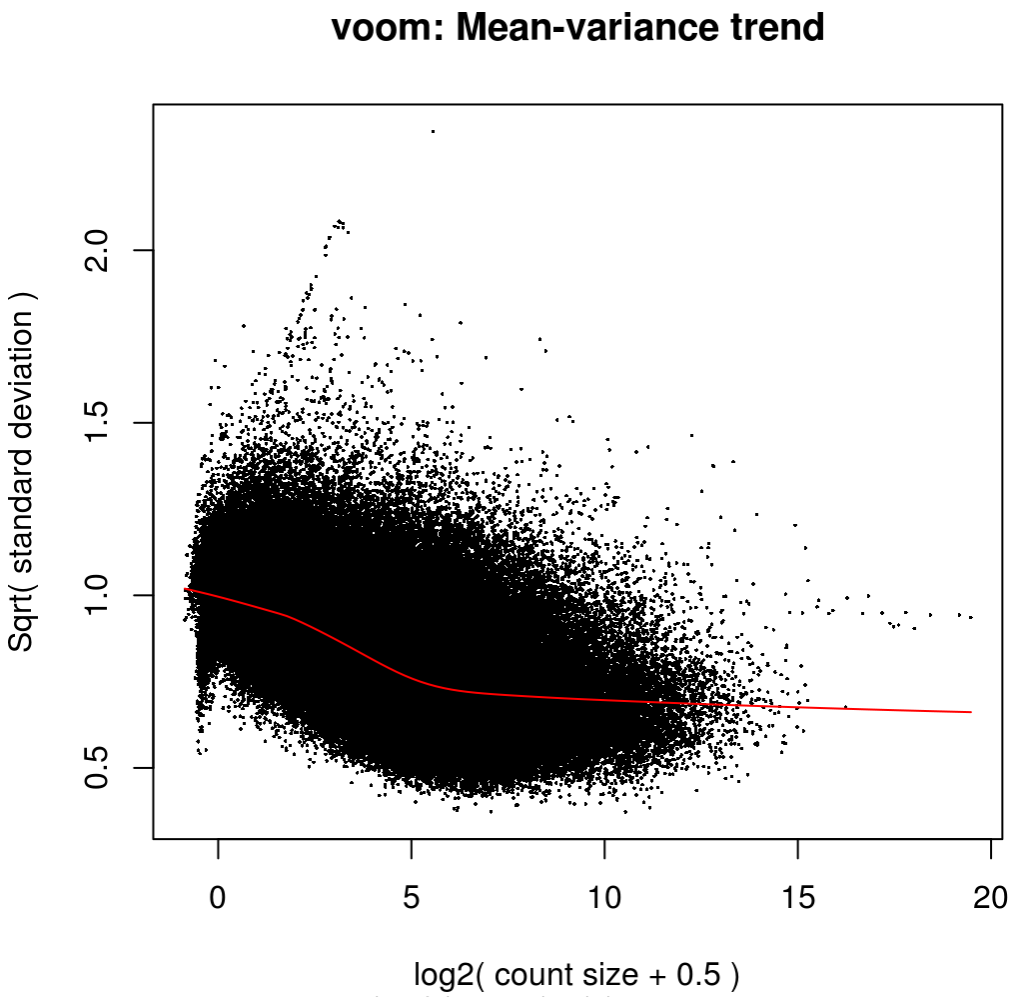

Figure 19: voom mean-variance plot of the exon-level data.

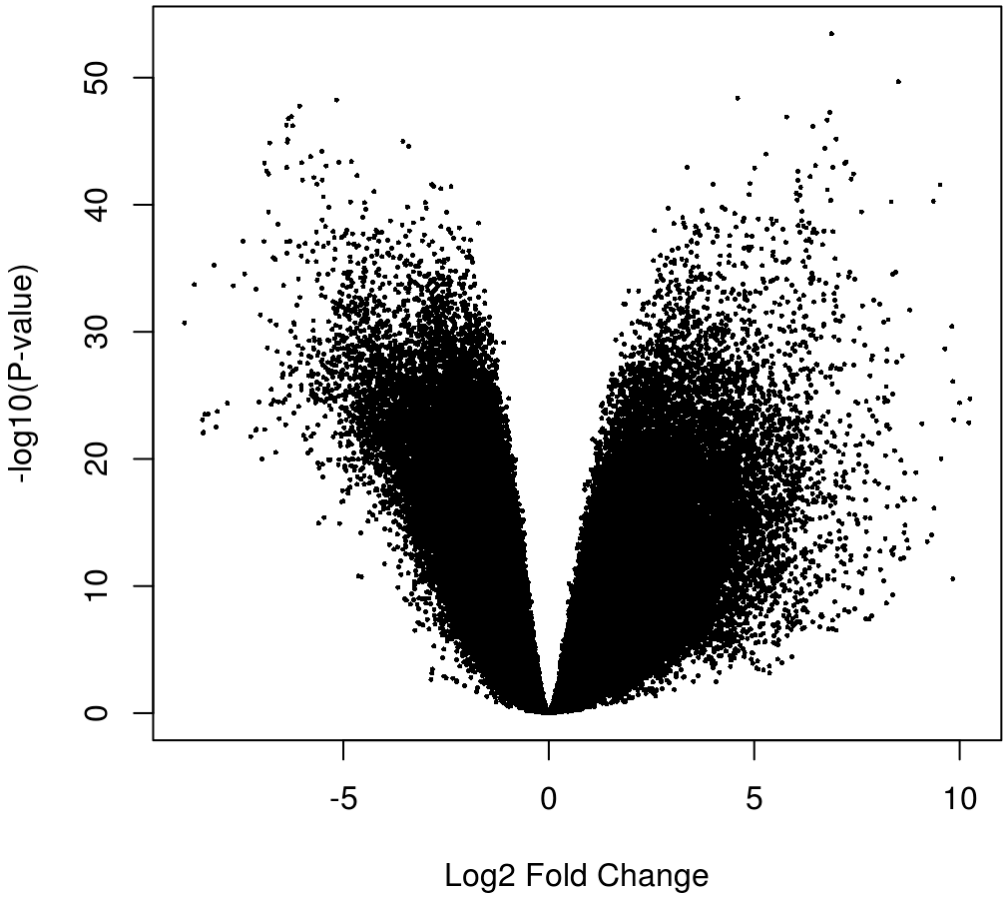

Figure 20: Volcano plot of the exon-level data. Testing for prenatal and postnatal DE adjusting for sex and RIN.

Genes/exons with DE signal

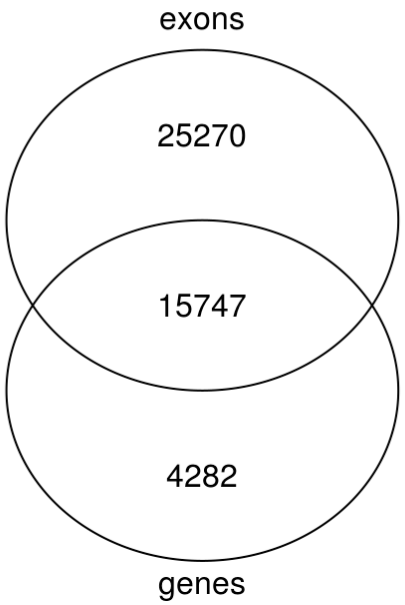

Figure 21: Venn diagram of the overlap between DE genes and genes with at least one exon DE.

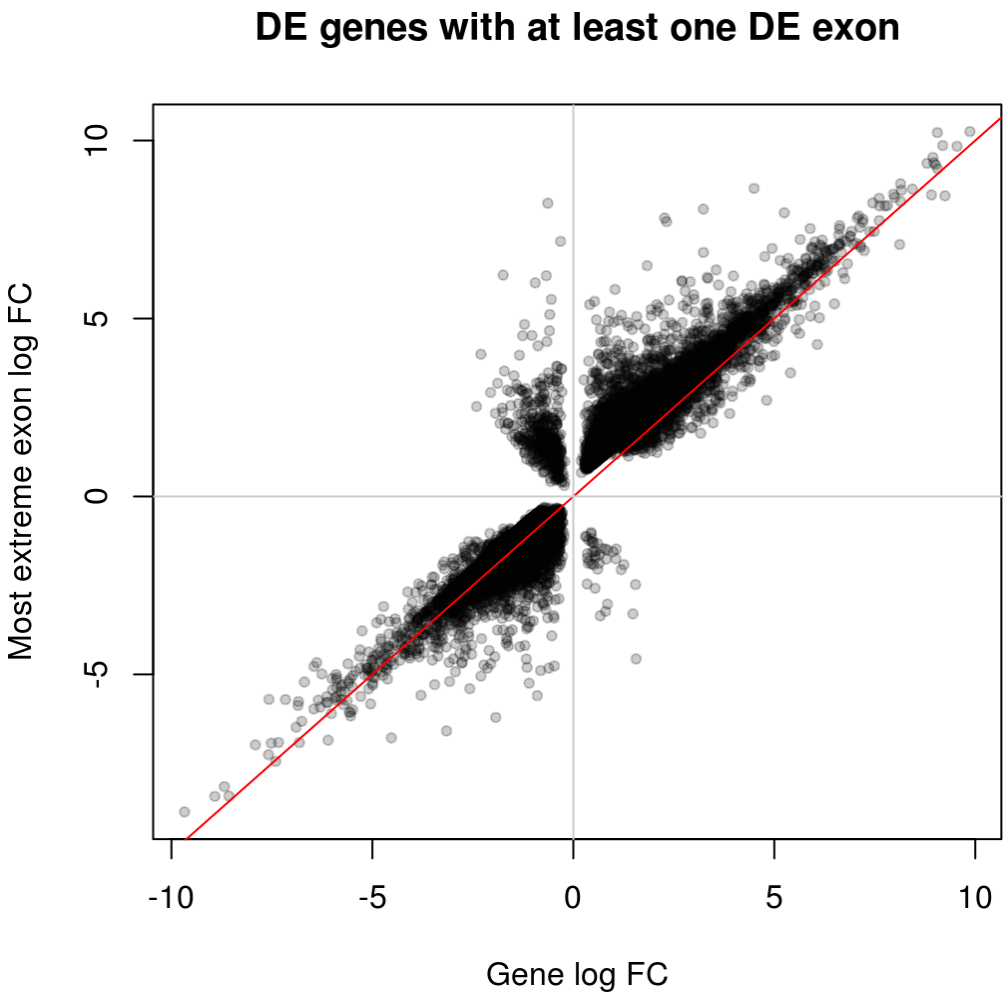

Figure 22: Log fold change (FC) for DE genes compared against the most extreme exon log FC among exons that are DE for the given gene.

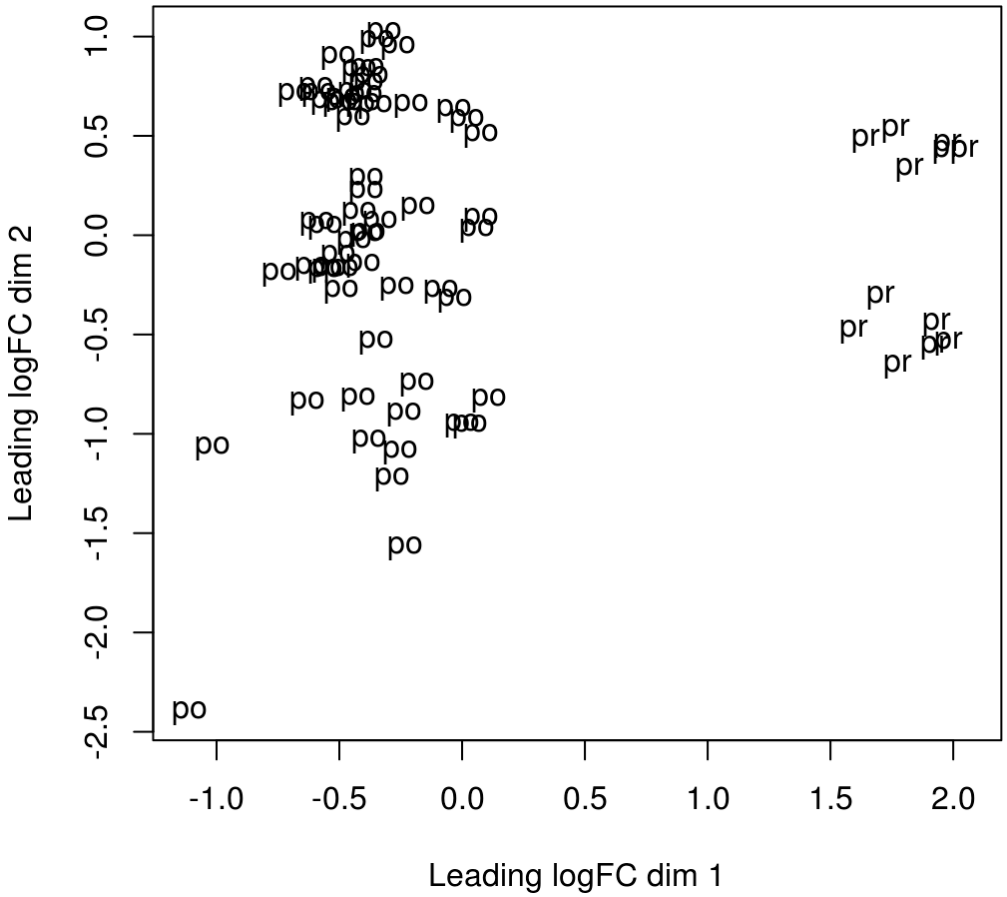

Figure 23: Multi-dimensional scaling plot of the expressed regions level data by age group.

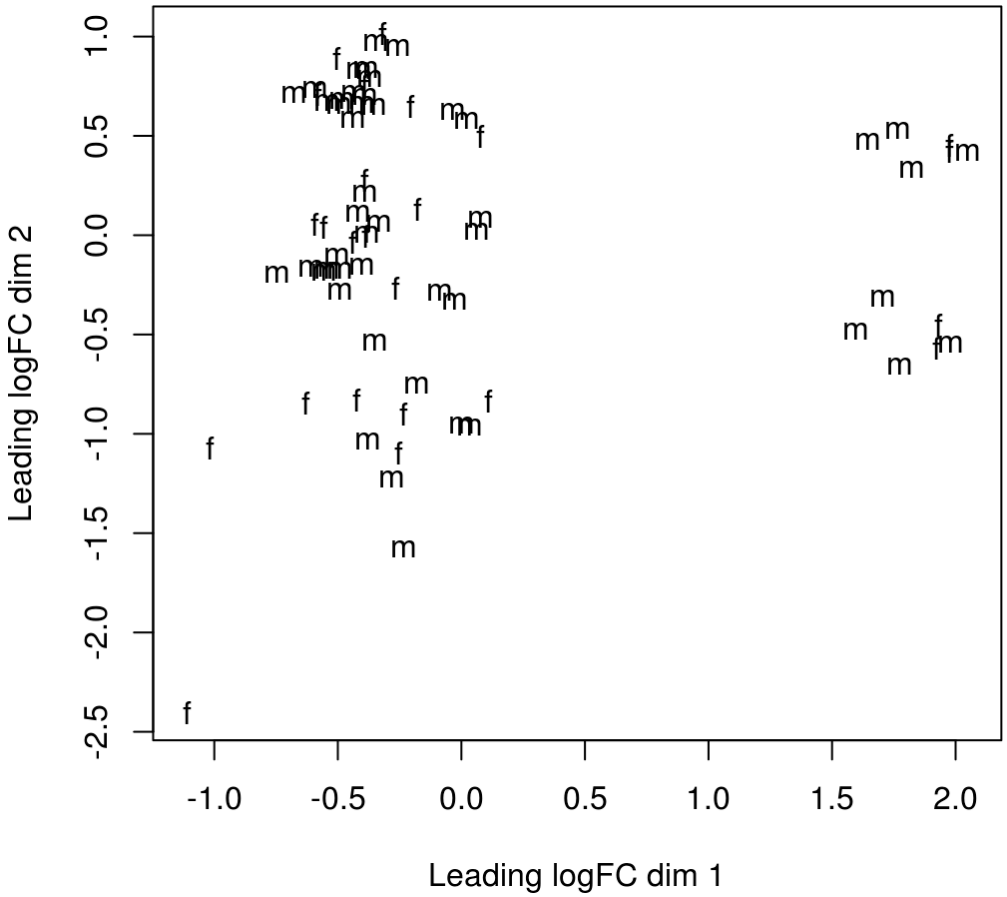

Figure 24: Multi-dimensional scaling plot of the expressed regions level data by sex.

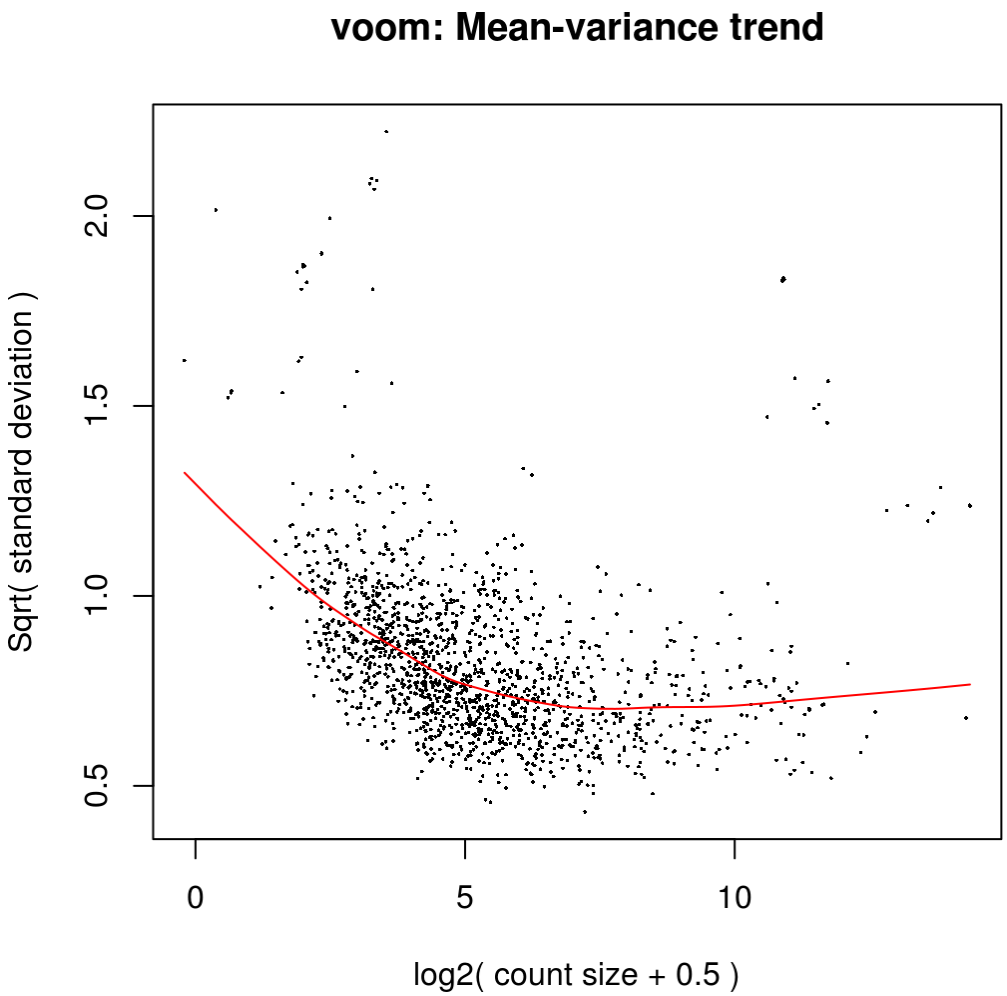

Figure 25: voom mean-variance plot of the expressed regions level data.

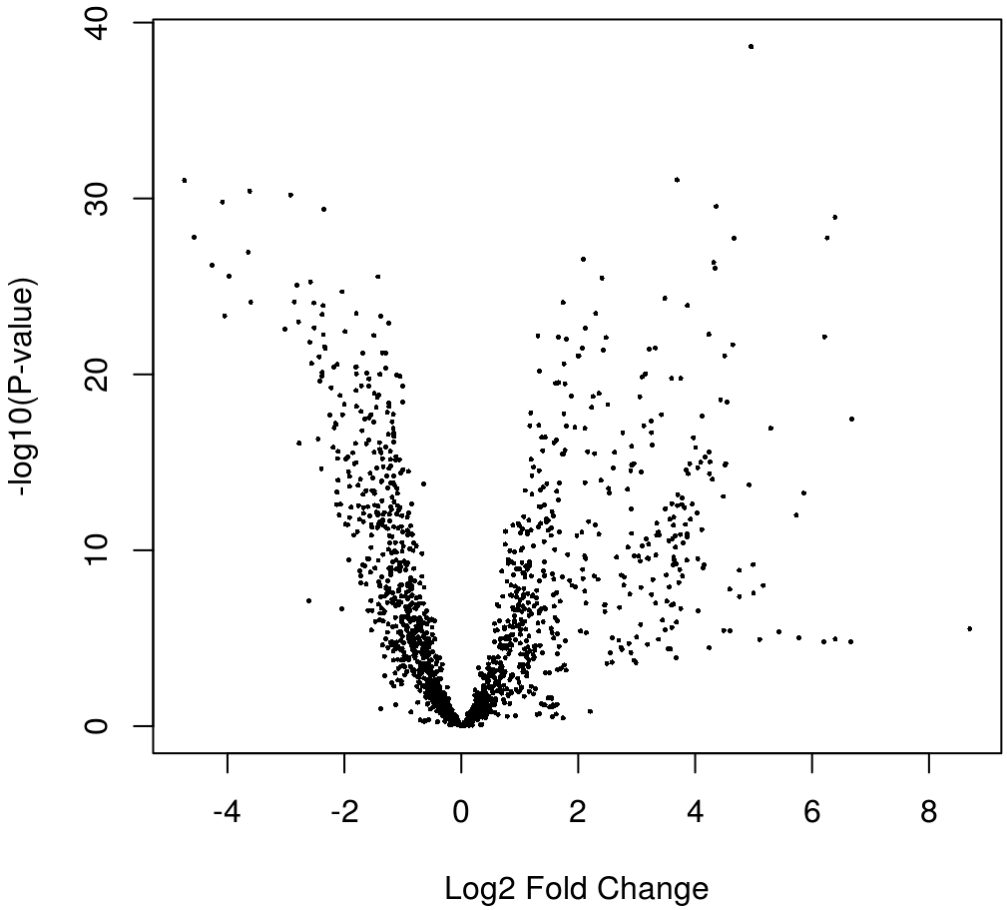

Figure 26: Volcano plot of the expressed regions level data. Testing for prenatal and postnatal DE adjusting for sex and RIN.

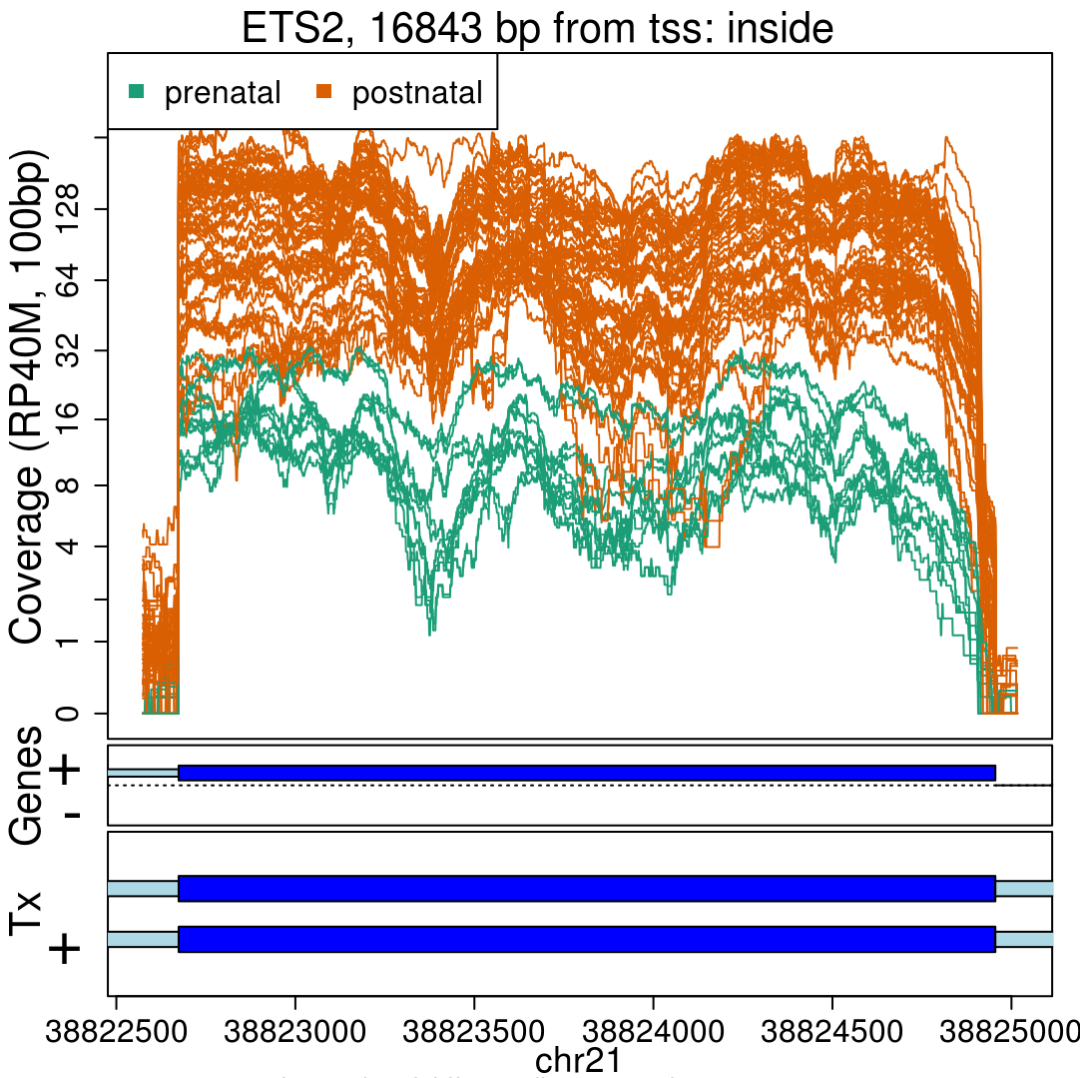

Figure 27: Base-pair resolution plot of differentially expressed region 2.