|  |  |  |
| --- | --- | --- |
| Folder | Script | Description |
| StringTie | Stringtie\_looper\_v2 2.pl | Create GTF for each bam file |
| 2018\_10\_31\_stringtie\_merge.pl | Merge GTF of samples with the same SF mutation |
| 2018\_11\_01\_stringtie\_merge\_all.pl | Merge the merged GTF files |
| rMATS | 2018\_11\_02\_rMATSlooper\_statoff.pl | rMATS –statoff for all samples |
| Data Filtering  Scripts | 2018\_11\_08\_deletions\_filtered.Rmd | Filter for samples with WGS and RNA-Seq datasets |
| 2018\_11\_26\_rmats\_join\_filterlowcov\_V4.R  2018\_11\_26\_rmats\_join\_filterlowcov\_V4\_SE.R  2018\_11\_26\_rmats\_join\_filterlowcov\_V4\_RI.R  2018\_11\_26\_rmats\_join\_filterlowcov\_V4\_MXE.R  2018\_11\_26\_rmats\_join\_filterlowcov\_V4\_A5SS.R  2018\_11\_26\_rmats\_join\_filterlowcov\_V4\_A3SS.R  2018\_12\_03\_rmats\_join\_filterlowcov\_V4\_SE\_chr1-3.R | Remove reads that have greater than 10 percent of samples with read counts 10 or lower for SJC or IJC |
| 2019\_01\_09\_filteringforbiologicalrelevance.Rmd | remove any splicing changes that are not >10% or <90% in a single sample: canonical splicing events |
| 2019\_01\_17\_filteringmisidentifiedspliceevents.Rmd | Junction coverage filtering: IJC + SJC ≥ 10 reads |
| 2019\_01\_24\_filteringforjunctioncoverage | 1) Skipped exons that contain an intron will be removed. 2) I need to remove A3SS and A5SS that are doubly labeled as SE |
| 2019\_02\_20\_MLLdatafiltering\_2.Rmd | Vera’s filtering scheme for somatic variants |
| 2019\_03\_01\_finalizingdatasets.Rmd | Data tidying   1. Remove samples that do not have both RNA Seq and GWAS data 2. Confirm CNV calls with karyotype 3. Combine somatic mutation data with deletion data, sex and age. |
| Clustering | 2019\_04\_11\_paralellclustering.R | Unsupervised k-means hierarchical clustering of AS signatures |
| rMATS STAT | 2019\_04\_09\_rMATStorMATSSTATS\_hotspot\_all\_vs\_healthycontrols.Rmd  2019\_04\_09\_rMATStorMATSSTATS\_hotspot\_vs\_healthycontrols.Rmd  2019\_04\_09\_rMATStorMATSSTATS\_hotspot\_vs\_sfWT.Rmd  2019\_04\_09\_rMATStorMATSSTATS\_lowexp\_vs\_healthybm.Rmd  2019\_04\_09\_rMATStorMATSSTATS\_lowexp\_vs\_highexp.Rmd | Creates in the input files for rMATS\_STAT, the bashscripts joining multiple jobs for the HPC, and the batchscripts needed to submit them to the HPC |
| batchscript\_U2AF1.sh | Example rMATS STAT batchscript (U2AF1) to be submitted on HPC |
| Bigwigs | 2018\_11\_06\_bigwigloop.pl | Make bigwigs from bam files |
| rMATS SURVIV | 2019\_04\_17\_rMATS\_SURVIV.Rmd | Creating input files for rMATS SURVIV |
| surviv\_bash.sh | rMATS SURVIV bash script |
| Neojunctions | 2019\_05\_06\_identifyingneojuncitons.Rmd | Filtering process for identifying neojunctions |
| GSEA | 2019\_06\_18\_GenerateGSEAfiles.Rmd | Create input files for GSEA and loop through GSEA input files |
| rMATS\_17186\_MDS\_gseainput.txt | Example input GSEA file |
| rMATS\_17186\_MDS\_gseainput.cls | Example cls GSEA file |
| ExonOntology | exont.obo.txt | Exon ontology trees |
| exont\_rMATS\_intersect.txt | rMATS IDs with intersected exon ontologies |
| Exont\_annotations\_v1.5.0.tsv | Exon ontology IDs and coordinates |
| 2019\_06\_21\_exont.Rmd | Joins rMATS data with exon ontology |