|  |  |  |
| --- | --- | --- |
| **What**? | **pipeline** | **For**? |
| **Alignment**  Control | vast-tools align /storage2/stav/jona/control/all\_R1\_control.fastq /storage2/stav/jona/control/all\_R3\_control.fastq -sp Hsa --expr | Input: FASTQ R1 and R3 (PE)  **-sp Hsa**: to generate the list of gene IDs, humen genome reference  **-expr** :to enable gene expression analysis |
| Alignment  KO | vast-tools align /storage2/stav/jona/sample/all\_R1\_sample.fastq /storage2/stav/jona/sample/all\_R3\_sample.fastq -o /storage2/stav/jona/vast-tools/output -sp Hsa --expr | Input: FASTQ R1 and R3 (PE)  **-sp Hsa**: to generate the list of gene IDs, humen genome reference  **-expr** :to enable gene expression analysis |
| **Combine** all control alignment output with all sample aligment output | vast-tools combine /storage2/stav/jona/vast-tools/output/vast\_out/\* -o /storage2/stav/jona/vast-tools/output -sp Hsa ir\_version 2 | Input: inclusion\_table :all control and KO alignment output  **sp Hsa:** to generate the list of gene IDs, humen genome reference  **ir\_version 2**: incorporates alternative exon-exon junction for skipping reads to obtain a more representative Percent Intron Retention at the transcript level. |
| **Comparing**  **Option A:**  PSI score between control and KO | vast-tools compare INCLUSION\_LEVELS\_FULL-Hsa2-hg19.tab -a all\_R1\_control -b all\_R1\_sample --paired --GO -sp Hsa | Input: inclusion\_table from combine output.  **--paired:** because its PE  **--GO**: produce list of gene IDs for the selected events to run Gene Ontology (GO) analyses  **-sp Hsa** : to generate the list of gene IDs, humen genome reference |
| **Comparing**  **Option B:**  PSI score between control and KO | vast-tools compare INCLUSION\_LEVELS\_FULL-Hsa2-hg19.tab -a all\_R1\_control -b all\_R1\_sample --paired --GO -sp Hsa --min\_dPSI 25 --min\_range 5 | All above+  **--min\_dPSI 25** : For valid AS events, vast-tools compare then requires that the absolute value of ΔPSI  **--min\_range 5** :PSI distribution of the two groups do not overlap. To provide higher or lower stringency |
| **Diff**  PSI score between control and KO | vast-tools diff -a all\_R1\_control -b all\_R1\_sample -o /storage2/stav/jona/vast-tools/output > /storage2/stav/jona/vast-tools/output/diff\_output.tab | Input: inclusion\_table from combine output.  **-r:** Probability threshold for P( (psi1 - psi2) > x ) > threshold [default 0.95] . minimal probability of acceptance that is required to consider a comparison to be 'believable' .  **-m:** Threshold for min diff where P( (psi1 - psi2) > threshold ) > --prob [default 0.1] .minimum value of difference between psi1 and psi2 that you will accept, does not currently alter the output sent to STDOUT, but does filter what is plotted to PDF and printed to file.  **Paired**: Samples are paired, -a pairOneA,pairTwoA,.. -b pairOneB,pairTwoB,.. [default FALSE] In the case that you have paired samples, where NormalA is dependent on PerturbationA, it is appropriate to use the --paired=TRUE flag. For example when considering NormalA and NormalB, to compare to PerturbationA and PerturbationB, the probability that P( joint\_psi1 - joint\_psi2 > -m ) is calculated such that NormalA is only compared to PerturbationA, and then NormalB is compared to PerturbationB. No MLE fitting is used in this case.  In all multireplicate cases where --paired=FALSE, the posterior distributions of the individual replicates are used to estimate a 'best fit joint posterior' distribution over psi for each sample. |
| **Plot**  PSI/cRPKM plot per page **(**PDF file) | #PSI data from combine output  vast-tools plot INCLUSION\_LEVELS\_FULL-Hsa2-hg19.tab | Input:  1**) PSI data** - one AS event per row - using the standard PSI format  e.g. GENE EVENT COORD LENGTH FullCO COMPLEX Tissue1 Tissue1\_Q ...  \*Recommended to use only a subset of AS events instead of the full table otherwise the resulting PDF file will be very large. Use option -m/--max to limit the maximum number of plots to generate.  \*PSI values that are "NA" or have "NA" quality scores will not be plotted (not point will be drawn).  2) **cRPKM data** - one gene per row - using the standard cRPKM format Use option "--expr TRUE" to enable this mode. If no input file is provided, standard input will be used. |