* Download 5’, 3’, exon skipping, intron retention, and mutually exclusive exon splice events data from <https://gdc.cancer.gov/about-data/publications/PanCanAtlas-Splicing-2018>.
* Organize the data input and format it for SNP-splicing event association analysis. Use script 1.PrepareData.sh. This script keeps the genes that mapped to the most significant SNPs (i.e. Suggestive and GW). Example: grep –f ListUniqueGenes.txt splice3prime > Data\_3prime. It creates two more files with TCGA subject IDs and splicing event IDs and types as well. It finally runs an R script Analyze.R that performs association analysis between SNP genotypes and splicing events.
* Outcome: An output file containing association results as follows: Chromosome, Position, Ensemble ID, Gene Name, Splicing event ID, Pan-cancer sample size, Pan-cancer effect size, pan-cancer P-value, and per-cancer sample size, effect size, and p-value for all cancer types.