Each R script is extensively annotated, below explains the overall goal of each script. The scripts are in the order in which they would be used within the pipeline.

VCFmerge.R
The goal of the script is to merge together multiple VCF files
Outputting a single "super" SNP list used for filtering in RNAidentfie.R below.

RNAmappe.R
The overall goal of this script is to identify wildtype SNPs and use this info to plot the frequency of the SNP in mutants
This allows for the identification of regions of linkage to the mutation
The general steps are:
 # STEP1 - ID SNPs in .vcf files - I've found that variant callers such as bcftools -v throw away many RNA-Seq SNPs
  # .vcf files MUST be v4.0 or v4.1 and only for indivdual chromsomes labeled wt_chr[1-25] and mut_chr[1-25]
  # a lynix for loop is used to input each chromosome sequentially into this script
  # --> output ALL SNPs and INDELs identified in widltype (wt) and mutant (mut) for user access
 # STEP2 - count allele frequency - this data is buried in the INFO column of the .vcf file and must be extracted
 # STEP3 - find high quality SNPs (markers) in wt useful for mapping
 # STEP4 - count the allele frequency of SNPs in mut at the marker positions
 # STEP5 - graph the data to identify regions of linkage in the mutant
  # --> output mutant SNP frequency at markers across ENTIRE genome
  # --> output mutant SNP frequency at markers on individual chromsomes

RNAidentifie.R
The overall goal of this script is to extract ALL SNP/INDEL information from mutants within a defined region of the genome
The region will most likely be that defined by RNAmappe.R as being linked
Th mutant SNPs are then filtered against known widltype SNPs to identify potentially deleterious changes
The general steps are:
 # STEP1 - Extract ALL the SNP information from the region of linkage/interest
  # this will be a "dirty" list that needs to be filtered, but also want to retain all possible info for user to evaluate
  # vcf formatted file corresponding to the chromsome or interest is imported, must have DP4=... or I16=... in INFO column
  # importing SNPs from chrX named mut_chrX_allALT.vcf and wt_chrX_allALT.vcf from RNAmappe.R
  # and extracting information between positions Y and Z
  # X, Y, and Z are defined either by user or will be automated within pipeline
 # STEP2 - Identify INDELs within the region and filter against known INDELs
  # --> output mutant INDEL list
 # STEP3 - Filter mutant SNPs against known wildtype SNPs

single vcf formatted file of SNPs to filter against is imported, must have DP4=... or I16=... in INFO column
ideally this is a vcf created from VCFmerge.R - which concatenates many vcfs together and outputs a more confident single list
--> output filtered mutant SNP list in vcf format allowing for piping into Variant Effect Predictor
--> output UNfiltered mutant SNP with wildtype SNP appended on - allows user to evaluate directly

VEPsorte.R
The goal of this script is to sort the VEP output file
The sorting is prioritzed based on effect of the SNP to the gene

RNAeffecto.R
The goal of this script is to extract and sort info on genes within linked regions from the cufflinks output files
must be a subdirectory _settings as described in RNAidentifier above
must be in the subdirectory _cuffdiff containing files from cuffdiff program
files must be called cds_exp.diff, cds.diff, gene_exp.diff, isoform_exp.diff, promoters.diff, splicing.diff, tss_group_exp.diff