Order of processing steps for polyA data (after alignment via blat)

1) processGenomeSeqs.sh, takes the \*.psl alignment file and creates files that break down the alignments into

(presumption is that all seqs initially aligned were named as “seq\_n\_count”)

\*\_alnCount.txt

\*\_tgtCount.txt

\*\_tgt\_aln.txt

\*\_siteCount.txt

2) convertCountsToBed.sh \*\_siteCount.txt

processes the site count file to create the bed file \*\_siteCount.bed, which is in a format that can be used to pull sequences from the genome

3) getSequencesAndReformat.sh \*\_siteCount.bed

processes the siteCount.bed file to create a new siteCount file that also has upstream and downstream sequences

4) assignAndSummarize.sh \*\_siteCount\_withSeqs.txt

processes the \*\_siteCount\_withSeqs.txt file to use genomic data, along with annotations to make best assignments of each site to known genes and features

creates the following files

\*.feature.txt

\*.gene.txt

\*\_assigned.txt

\*.feature.txt and \*.gene.txt files can be used for expression analysis, and includes a Q3 value for normalization purposes as the final line in the file; multiple columns are created (to be described later) that give counts based on treatment of genomic Arich regions as well as multihit tags.

5) collectSiteCounts \*\_assigned.txt prefix

processes \*\_assigned.txt files from above (usually multiple files concatenated into one, identified by sampleID in the first column) to characterize the sites across multiple samples.

Creates following files

prefix\_gene.txt

prefix\_site.txt

6) processAllPolyASites \*\_site.txt [outputPrefix] [multiSwitch(0:1:2)=0] [AGcount=8] [UScutoff=80] [DScutoff=1000] [PseudoWeight=0.05] [paProbCutoff=30]

processes the prefix\_site.txt file generated in last step and for each sample calculates cumulative probability, as well as directional polyA processing probability (based on assignments from step 4).

For sequence analysis- FDR < 0.35, |log2ratio|>0.25, genomically unique sites => gives 82 increased and 705 suppressed sites