- Visualization of Funseq2 output
  - both BED and VCF formats
  - 32 features in total
  - Among which
    - GERP scores, noncoding features (Promoter, TFP, TMP, Intron, Enhancer, ncRNA, UTR, Pseudogene), mutation type (transition vs transversion), reference and variant genotype, strand +/- information
    - Top N (default N=50) target genes, MOTIFG, MOTIBR, hot regions
    - Variant distribution across and within chromosomes
  - More features can be added by the user
- Three steps
    - Statistics
    - Plot generation
      - Flexible *w.r.t.* types of plots
    - HTML Generation
- Accepts additional parameters to qualify specific variants
  - Up to two, unordered ones

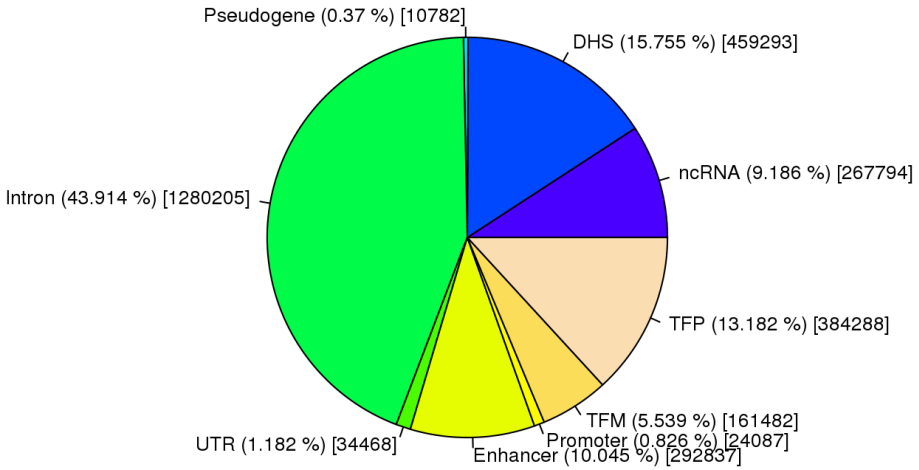
```
[hm444@farnam2 ~]$ awk -v regex_args="Promoter" -f cancer-variant-correlation/scripts/funseq_stats.awk funseq_results/Output.vcf > Funseq2_stats_example.txt
```

Example

file:///Users/husseinmohsen/Desktop/Visualization%20Results/somatic_visualization_results/

Funseq Statistics
[Liver-HCC.out_rare_variants_bed_funseq_stats]

Noncoding Features



- Execution time
  - ~ 10 seconds for 100,000 variants
  - Grows linearly

- Tested on 200+ samples
- Part 1 to be followed by part 2 very soon

- What's next
  - Already generated Part 1 statistics for all PCAWG somatic and germline samples
    - 37 x 2 x 3 samples/files
  - To be presented with across-samples statistics as part of the two-hit hypothesis project