

cutpointcondenser.pl

```
#!/usr/bin/env perl -w
use Math::Complex;

#feed the large set of derived cutpoints into the condenser to assess the
distribution of derived cutpoints from the leave-5-out assessment. Take the
frequency mean as the derived upper and lower cutpoints.

for $gene (qw/ CEBPG E2F1 CAT ERCC4 ERCC5 GPX1 GPX3 GSTM3 GSTP1 GSTT1 GSTZ1 MGST1
SOD1 XRCC1 /) {

my %upper = ();
my %lower = ();

open GENEDATAERROR, "<Validation-$gene.txt";
my @genedataset = ();
while (my $line = <GENEDATAERROR>) { #push the genedataset into an array
  chomp ($line);
  my @linearray = split(",", $line);

  if (exists($upper{@linearray[0]})) {
    $upper{@linearray[0]}[1] = $upper{@linearray[0]}[1] + 1;
  } else {
    $upper{@linearray[0]} = [ "@linearray[0]", "1" ];
  }

  if (exists($lower{@linearray[1]})) {
    $lower{@linearray[1]}[1] = $lower{@linearray[1]}[1] + 1;
  } else {
    $lower{@linearray[1]} = [ "@linearray[1]", "1" ];
  }

# $numKeys = keys(%upper);
  if (exists($upper{@linearray[0]})) {
    print "$upper{@linearray[0]}[1]\n";
  }
# print "$numKeys\n";
}
close GENEDATAERROR;

print "\n ----- \n";

open HISTOGRAMOUT, ">Validation-$gene-histogram.csv";

print HISTOGRAMOUT "Upper-Cutoffs\n";
for $logTA ( keys %upper) {
  print HISTOGRAMOUT "$upper{$logTA}[0], $upper{$logTA}[1]\n";
}

print HISTOGRAMOUT "Lower-Cutoffs\n";
for $logTA ( keys %lower) {
  print HISTOGRAMOUT "$lower{$logTA}[0], $lower{$logTA}[1]\n";
}

close HISTOGRAMOUT;

}
```

cutpointcondenser.pl

The output will be readable in excel comma-delimited format.