GLIPH1 Scoring Overview

Taken from GLIPH1 source code found at https://github.com/immunoengineer/gliph, in the Perl file 'gliph-group-scoring.pl'

GLIPH1 is used because the open source code is readily available on Github, whereas GLIPH2 would require correspondence with the author.

Differences between GLIPH1 and GLIPH2 can be found on GLIPH2 website: GLIPH 2

The major difference between the scoring is the addition of Fisher Score in GLIPH2

The GLIPH1 score report contains, <u>for each convergence group</u>:

- CDR3s
- Subjects
- Clones
- CRG_score (Final Score)
- Vb_score
- CDR3_p (Length Score)
- HLA_p
- Expansion_p (Expansion Score)
- Motif_p
- Size_p (Cluster_size)
- HLA
- Motifs

Detailed description of each, this is for each convergence group:

CDR3s: Number of unique CDR3s

Subjects: Number of patients

Clones: Number of unique clones

CRG_score (Final Score):

Vb_score * CDR3_p * HLA_p * expansion_p * motif_p * size_p * 64

(This is why it's such a tiny number, the final score is many p-values multiplied together)

Vb_score: p-value as probability of surpassing the observed simpson index of diversity for Vb-genes in this convergence group.

CDR3_p (length score): p-value as probability of surpassing the observed simpson index of diversity for the CDR3 length in the convergence group.

HLA_p: This value I am less certain of, but seems to be the probability of getting this specific HLA depending on population size, cluster size, observed HLAs in cluster, HLAs in population, and the number of simulations.

Expansion_p (Expansion score): p-value for the bias in clonal expansion in the indicated convergence group. Calculated by taking 1000 random permutations via Fisher-Yates shuffle of different clones and counts from the annotation reference file. The p-value is the percentage of these random samplings that achieve a better score than the clone we are testing. Because this permutation test is done 1000 times, the best possible (most expanded) expansion score is 1/1000 (0.001 or $1*10^{-3}$).

Motif_p: This value always seems to be 0.001 in GLIPH1.

```
139    # get motif enrichment - which motifs were found here, and what are their values?
140    my $motif_p=0.001;
```

HLA: The specific HLA of this convergence group

Motifs: Motifs present in this convergence group

Size_p (Cluster size): p-value of convergence group size. Determined from a defined set of key-value pairs (shown below) of unique convergence group CDR3 counts and p-values. If the number of unique CDR3s exceeds 50,then the lowest p-value of $1.0*10^{-5}$ is used.

```perl
my %counts2scores = (
                    1,0.954980692,
                    2,0.029106402,
                    3,0.006190808,
                    4,0.002347221,
                    5,0.001456437,
                    6,0.001212588,
                    7,0.000862264,
                    8,0.000829941,
                    9,0.000476289,
                    10,0.000313723,
                    11,0.000240521,
                    12,0.00015401,
                    13,7.70048E-05,
                    14,7.41528E-05,
                    15,9.50677E-05,
                    16,8.55609E-05,
                    17,9.50677E-05,
                    18,8.55609E-05,
                    19,0.000111229,
                    20,8.08075E-05,
                    21,9.60184E-05,
                    22,7.51035E-05,
                    23,8.46103E-05,
                    24,8.36596E-05,
                    25,5.22872E-05,
                    26,3.80271E-05,
                    27,2.56683E-05,
                    28,3.99284E-05,
                    29,3.99284E-05,
                    30,3.04217E-05,
                    31,2.37669E-05,
                    32,2.85203E-05,
                    33,3.70764E-05,
                    34,3.42244E-05,
                    35,3.04217E-05,
                    36,3.32737E-05,
                    37,3.04217E-05,
                    38,3.32737E-05,
                    39,3.99284E-05,
                    40,2.9471E-05,
                    41,3.04217E-05,
                    42,3.2323E-05,
                    43,2.6619E-05,
                    44,1.14081E-05,
                    45,1.42602E-05,
                    46,1.52108E-05,
                    47,1.52108E-05,
                    48,1.71122E-05,
                    49,2.18656E-05,
                    50,1.42602E-05);
```