Structure of Capgenie

Atul Phadke

**Data acquisition Functions (preprocessing)**

1. Trimming low quality reads based on a threshold
2. Count reads based on **known** variants
   1. Giving 21mer sequence
3. Count reads based on **unknown** variants
4. Calculating enrichment scores based on a preinsert library
5. Pruning duplicate reads
6. Removing Common Sequences
7. Saving into .pkl files

**Data visualization functions (non-barcode)**

1. Saving into excel files
2. Calculating Bubble charts
   1. Coloring specific peptides
3. Calculating frequency distribution charts

**Data visualization functions (barcode)**

1. Saving into excel files
2. Normalization to a specific capsid
3. Adding excel heatmaps
4. Adding graphs and saving into pictures

**Technical structure and file accessing (preprocessing)**

1. Has an instructions.pkl metadata file to tell the program where each file is and what to do
   1. Has information on if the reads are pruned, denoised, etc
   2. Gives 21mer sequence of capsid file
   3. Has location of capsid file