



Online data supplement for pooled CRISPR screens in *Drosophila* cells, related to Viswanatha et al., “Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in *Drosophila* cells” 2018, eLife

Pooled CRISPR screening is a method for cell-based forward genetic screening first developed in mammalian cells (Joung et al.;doi: 10.1038/nprot.2017.016). First, a large heterogenous pool of mutant cell-lines is generated, each cell barcoded by a genome-encoded CRISPR sgRNAs. Following a perturbation, the genome-encoded CRISPR sgRNAs in the population are harvested using a single PCR step, creating a diverse amplicon. This amplicon is then sequenced using next-generation sequencing, and the number of reads (“readcount”) of each unique sgRNA in the amplicon is determined. Comparing the relative readcount of an sgRNA sequence in amplicons from pre- versus post-selection populations generates a hypothesis about the function of the targeted gene.

The data presented here enables users to perform additional analyses using raw sequence data beyond those reported in our study. In this package, we include raw sequencing of CRISPR sgRNAs from the plasmid pool (“REF”) and from two biological replicates of a genome-wide pooled fitness screen (“rep1”, “rep2”). Each folder contains three FASTQ sequencing files because we conduct pooled screens using three distinct sublibraries (“Group 1” or “g1”, “Group 2” or “g2”, “Group 3” or “g3”), and PCR and sequencing is performed separately for each. Together, the library targets all protein-coding genes in *Drosophila*. To achieve genome-wide coverage, we perform readcount extraction and analysis on each sublibrary and combine the results at the gene level.

The software package we use to analyze CRISPR sequencing data is MAGeCK (Li et al., 2014; <https://doi.org/10.1186/s13059-014-0554-4>). In our analysis, we used two subprograms within `mageck`, `count` and `mle`. The `count` subprogram extracts readcounts for each sgRNA sequence, while `mle` creates gene-level Z-scores using the readcount file generated by `count`. Importantly, a requirement for `count` is a “library file”, which lists the sgRNA ID, sequence, and gene name (in this exact column order) for each CRISPR sublibrary. These have been provided in the folder called “library_files”.

Detailed instructions on the installation and use of MAGeCK can be found on the software’s homepage (<https://sourceforge.net/p/mageck/wiki/Home/>). For comparison, we also include the expected result of the `count` and `mle` commands in the “`mageck_count_results`” and “`mageck_mle_results`” folders.