ASKA\_lookup\_map.mat is an input for fastq2barcodeCounts\_v3\_1.m and contains all knockout strain barcodes and strain indentifier information.

backgroundSignal.mat holds background/autofluorescence of colonies on certain media at certain time points. These values are to be input to plot\_colony\_selfRecA.m to subtract from experimental values. Generated with plot\_colonyBackground.m

barcodeAnalysis.R is a function (used in BarcodeScreen.R) that runs DESeq and pathway enrichment anaylsis on barcoded screen fastq files. DESeq is used to determine p-values and fold-change of knockout strains in control and treatment samples. Outputs include hits, volcano plots, and pathway enrichment. It relies on outputs from fastq2barcodeCounts\_v3\_1.m

BarcodeScreen.R is a wrapper script using the barcodeAnalysis.R function to analyze barcoded knockout strain screens. Different comparisons of treatment and control groups can be assigned in this script.

Breseq\_mutation\_list.xlsx is an example output from breseq that is an input to writeMasterMutationList.m

calcMutationTable\_v2.m is a script that filters mutations by certain criteria and removes mutations shared by both treatment and control groups. The output is a .txt file with condition, mutation position, and mutation sequence context. Other outputs are fasta format files of mutation sequence context for both treatment and control, and a randomly generated control.

ctrl\_context.txt is the randomly generated control fasta file from calcMutationTable\_v2.m to be used in downstream analysis with fimo.

COG function Entrez GMT.csv is a file input to BarcodeScreen.R that contains COG pathways and the entrez ID for genes in those pathways.

comparison\_pksP2pksN.csv is an example comparison file that can be input to BarcodeScreen.R

conds.mat is an output of plot\_recASignal.m (by using the segment\_recASignal.m function), but can be used as an input for plot\_recASignal.m if the function has previously been run.

countsOddPlusEven.csv is an example barcoded knockout strain counts file generated by fastq2barcodeCounts\_v3\_1.m and is an input to BarcodeScreen.R

countTriNucs.m is a script that iterates through all possible trinucleotide sequences and identifies the frequency of each sequence in downloaded genomes. Information for each genome is saved in dataTable\_Ecoli.mat

dataTable\_Ecoli.mat is an input to countTriNucs.m and contains information on over 9,000 E. coli genomes, including accession numbers.

EcoCyc Pathways Entrez GMT.csv is a file input to BarcodeScreen.R that contains EcoCyc pathways and the entrez ID for genes in those pathways.

example\_flow\_files is a folder containing example .fcs files to run flowGate.m and flowPlot.m on. Metadata for these files is contained in example\_metadata.xlsx

example\_metadata.xlsx is an example metadata file to use as input for both flowGate.m and flowPlot.m to sort and keep track of .fcs files in the analysis.

example\_recA\_files is a folder containing example .mat files that are examples of outputs from the segment\_recASignal.m function used by plot\_recAsignal.m and can be used in the plotting portion of plot\_recAsignal.m

fastq2barcodeCounts\_v3\_1.m generates knockout barcode counts from fastq files. The script iterates over fastq files, removes reads with low quality scores, and searches each read for known barcodes (provided in ASKA\_lookup\_map.mat). Identified barcodes are summed and normalized to reads per million for each sample. Sample fastq files can be found at SRA project XYZ

fimo.xlsx is the output from a fimo analysis run on mutation sequence context files (i.e. ctrl\_context.txt and pks+\_context\_5.txt). it is an input to findSeqwithMotif.m

findSeqwithMotif.m is a script that finds mutated sequences that match the enriched motif identified with fimo. fimo.xlsx and pks+\_context\_5.txt are inputs.

flowGate.m reads in .fcs files to gate on the cell population, single cell population, YFP positive population, CFP positive population, and mCherry positive population. It requires a metadata spreadsheet to sort files and extract representative samples to perform the gating on and saves the coordinates of each gate to use in downstream analyses (xyCoordinate.mat).

flowPlot.m uses xyCoordinate.mat to extract fluorescent signal for cells within each gate defined in flowGate.m to analyze positive reporter cells.

genomeFeatures.xlsx is a file containing positions of key genome features in E. coli such as the origin and terminus regions. It is an input to plotMutPositions.m

macrodomain\_positions.xlsx is a file containing the end point positions of previously defined macrodomains. It is an input to plotMutPositions.m

master\_mutation\_final.xlsx is a file containing cleaned up breseq results and extra information such as sequence context and mutation type.

metadata\_pksP2pksN.csv is an example metadata file to be input to BarcodeScreen.R to assign samples to control or treatment groups.

mutationContext\_5.txt is a file containing conditions, mutation positions, and mutation sequence context. It serves as an input to plotMutPositions.m

mutationType.m is a script that extracts types of mutations (i.e. sbs, small/medium/large indels) and the frequency of each mutation.

pks+\_context\_5.txt is a fasta format file to be used in fimo and for downstream analysis identifying sequences that match the enriched motifs.

plot\_colony\_selfRecA.m is a script that uses the quant\_selfRecASignal.m function to quantify YFP expression in colonies inflicting self-damage. Background fluorescence of untagged colonies is subtracted from both YFP and CFP channels and then YFP is divided by CFP to normalize to overall protein levels in each colony

plot\_colonyBackground.m is a script that determines the background/autofluorescence of untagged colonies on different media agar plates over time for YFP and CFP. These values are saved in an output to be used in plot\_colony\_selfRecA.m

plot\_recASignal.m is a wrapper script employing the segment\_recASignal.m function that plots signal intensity, calculates distances of signal and between colonies, and separates contacting and non-contacting colonies for downstream comparison.

plotMutPositions.m is a script that plots the genomic location of mutation positions on a circle plot and relative to macrodomains from macrodomain\_positions.xlsx

quant\_backgroundSignal.m is a function used by plot\_colonyBackground.m that segments colonies and measures fluorescent signal intensity

quant\_selfRecASignal.m is a function used by plot\_colony\_selfRecA.m to segment colonies and measure YFP fluorescence.

segment\_recASignal.m is a function that opens a popup image of brightfield and CFP channels overlaid for the user to draw a line between two colony centers and mark the edges of each colony. These coordinates are saved for each image and the fluorescent signal of all channels is determined along the line between colonies, while the intersect of the line and edges of colonies is calculated.

writeMasterMutationList.m is a script that takes a breseq output, cleans up some of the columns, and adds information like mutation type or sequence context.

yfp.mat is an example output file from the quant\_selfRecASignal.m function used in plot\_colony\_selfRecA.m. The file can also be used as an input for the plotting portion of plot\_colony\_selfRecA

xyCoordinate.mat is an example output of flowGate.m to be used as an input for flowPlot.m