GRIBCG Version 1.0 Usage Tips

Brian Merritt

Introduction

Here, GRIBCG is run against a test file to show the necessary steps needed in order to generate a balancer chromosome using a single sgRNA for CRISPR/Cas9 breakpoints.

Platform: Linux

Languages:

Perl Version 5

R Version 3.4.4

Dependencies:

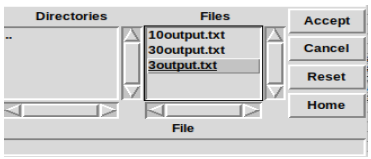
BioPerl (Source: https://bioperl.org/INSTALL.html)

predictSGRNA (Source: http://www.ams.sunysb.edu/~pfkuan/softwares.html#predictsgrna)

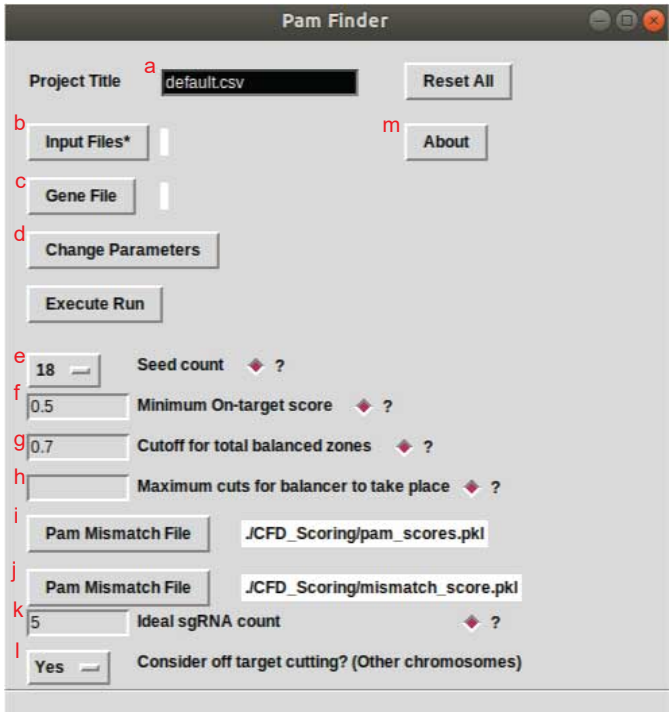
Tk GUI (Source: https://metacpan.org/pod/distribution/Tk/pod/UserGuide.pod)

Tutorial

1. Traverse to directory: GRIBCG from terminal.
2. Input: perl GUI.pl
3. Select Input Files: Traverse to example directory
4. Select 3output.txt
5. Select Gene File and choose exampleGeneFile.fasta



6. Select Change parameters to view current parameters of the system in the design process.
   a) Output .csv filename
   b) FASTA File containing genome
   c) OPTIONAL: FASTA file containing chromosome accessions and all known genes and their respective locations.
   d) Expand parameter options for user-defined custom settings.
   e) Length of Seed Sequence.
   f) Cutoff threshold for average on-target score for binned sgRNAs



g) Percentage of total coverage that is necessary to consider it an ideal balancer.
h) Default: considers the total chromosome size in relation to ideal cut count where each inversion is equal to 4Mbp of coverage. Users may instead choose to set a maximum threshold of cuts.
i) User may upload .pkl file containing mismatch penalties for PAM sequences
j) User may upload .pkl file containing mismatch penalties for spacer sequence
k) Tell GUI to output only top X amount of sgRNAs per chromosome
l) User can choose whether to take into account the minimization of off-target effects in overall process.
m) Learn more about the tool.

7. Allow the script to run. Two output files are generated:
    a. User defined project tile (default: default.csv) which contains top ideal sgRNAs for each chromosome at a given cut count.
    b. File (.csv) containing all sgRNAs that surpass the on-target, coverage, and cut limit thresholds (output/all.csv).

Results of default.csv:

Chr	Sequence	SSV	Avg On Target	CFD	Cuts	Coverage(%)	# Genes Affected	Genes	Off target sites (high probability)
NC_0	ACTGTGTCGGTGGCCACTAGGGG	0.312	0.933	3.2	2	100	1	60133:>ATXY0.1 399414:none	NC_1:ACTGTGTCGGTGGCCACTAGGGG
NC_0	ACTAATTGAGGCCCCTTGACCGG	0.312	0.566	3.2	2	100	2	40454:>ATXY0.0 766769:>ATXY0.2	NC_1:ACTAATTGAGGCCCCTTGACCGG
NC_0	GAGTATGCCACACCGAACGGCGG	0.303	0.983	3.3	2	100	0	914353:none 253824:none	NC_1:GAGTATGCCACACCGAACGGCGG
NC_0	GTATAAGACTGGGGATGGTACGG	0.294	0.824	3.4	2	100	0	817508:none 725879:none	NC_1:GTATAAGACTGGGGATGGTACGG
NC_0	GTGAGGATCGTAACCACTTACGG	0.244	0.54	4.1	2	100	0	672918:none 562649:none	NC_1:GTGAGGATCGTAACCACTTACGG
NC_1	ACTAATTGAGGCCCCTTGACCGG	0.312	0.566	3.2	2	100	0	40454:none 766769:none	NC_2:ACTAATTGAGGCCCCTTGACCGG NC_0:ACTAATTGAGGCCCCTTGACCGG
NC_1	ACTGTGTCGGTGGCCACTAGGGG	0.312	0.933	3.2	2	100	1	60133:>ATXY1.1 399414:none	NC_0:ACTGTGTCGGTGGCCACTAGGGG NC_2:ACTGTGTCGGTGGCCACTAGGGG
NC_1	GAGTATGCCACACCGAACGGCGG	0.303	0.983	3.3	2	100	1	914353:>ATXY1.2 253824:none	NC_2:GAGTATGCCACACCGAACGGCGG NC_0:GAGTATGCCACACCGAACGGCGG
NC_1	GTATAAGACTGGGGATGGTACGG	0.294	0.824	3.4	2	100	0	817508:none 725879:none	NC_2:GTATAAGACTGGGGATGGTACGG NC_0:GTATAAGACTGGGGATGGTACGG
NC_1	GTGAGGATCGTAACCACTTACGG	0.244	0.54	4.1	2	100	0	672918:none 562649:none	NC_0:GTGAGGATCGTAACCACTTACGG NC_2:GTGAGGATCGTAACCACTTACGG
NC_2	ACTAATTGAGGCCCCTTGACCGG	0.312	0.566	3.2	2	100	0	40454:none 766769:none	
NC_2	ACTGTGTCGGTGGCCACTAGGGG	0.312	0.933	3.2	2	100	1	60133:>ATXY2.0 399414:none	
NC_2	GAGTATGCCACACCGAACGGCGG	0.303	0.983	3.3	2	100	1	914353:>ATXY2.2 253824:none	
NC_2	GTATAAGACTGGGGATGGTACGG	0.294	0.824	3.4	2	100	0	817508:none 725879:none	
NC_2	GTGAGGATCGTAACCACTTACGG	0.244	0.54	4.1	2	100	0	672918:none 562649:none	

Figure 1: Default.csv run on example data file. These are top 5 sgRNAs for each chromosome with considerations of both off-target activity and coverage percentage. All parameters used are default.