ACER: Analysis of CRISPR Essentiality in R

Elizabeth Hutton[1]

[1]Cold Spring Harbor Laboratory; Syosset, NY, USA; orcid.org/0000-0003-3422-5278

This version was compiled on December 23, 2020

gene essentiality | computational biology | CRISPR | differential essentiality | gene addiction

Tool Applications. ACE (Analysis of CRISPR Essentiality) was designed to test for differential essentiality between sets of samples from a CRISPR knockout screen. Please see the ACE paper for a detailed discussion of methods. (Hutton *et al.*, 2020)

Input Data. A minimum of two data files must be provided for ACE:

- `countFile`: a single text file with unique sgRNA identifiers in the first column and gene names in the second column. Subsequent columns must contain either depleted read counts (if a master library sequencing is provided), or alternating initial and depleted read counts.
- `negCtrlFile`: a text file of the gene names to use as negative controls.

Other files necessary for analysis may include:

- `masterFiles`: a vector of names of files containing sequenced master libraries.
- `sampleMasterInfoFile`: Text file containing a column of master library file names, and a column with the corresponding sample name derived from that master library (same as the `countFile` headers).
- `sampleInfoFile`: Required for differential essentiality prediction, this file contains one column listing sample names and a second column containing sample annotations to use in test partitioning (for instance, 'KRAS_WT' vs. 'KRAS_Mut').
- `guideCovarFile`: File used to estimate guide efficiency (optional).

```
library(ACER)
```

```
#> Loading required package: data.table
```

```
#> Loading required package: R6
```

Tool Applications. ACE (Analysis of CRISPR Essentiality) was designed to test for differential essentiality between sets of samples from a CRISPR knockout screen. Please see the ACE paper for a detailed discussion of methods. (Hutton *et al.*, 2020)

Input Data. A minimum of two data files must be provided for ACE: `countFile`: a single text file with unique sgRNA identifiers in the first column and gene names in the second column. Subsequent columns must contain either depleted read counts (if a master library sequencing is provided), or alternating initial and depleted read counts. `negCtrlFile`: a text file of the gene names to use as negative controls.

Other files necessary for analysis may include: `masterFiles`: a vector of names of files containing sequenced master libraries. `sampleMasterInfoFile`: Text file containing a column of master library file names, and a column with the corresponding sample name derived from that master library (the same sample name used as a header in the `countFile`). `sampleInfoFile`: Required for differential essentiality prediction, this file contains one column listing sample names and a second column containing sample annotations to use in test partitioning (for instance, 'KRAS_WT' vs. 'KRAS_Mut'). `guideCovarFile`: File used to estimate guide efficiency (optional).

```
head(fread(system.file('extdata','countData.csv', package='ACER')))
```

```
#>            sgrna       gene init_Sample_1 dep_Sample_1 init_Sample_2
#> 1: gene_0_1_0_g1 gene_0_1_0           178          115           108
#> 2: gene_0_1_0_g2 gene_0_1_0           196          156           207
#> 3: gene_0_1_0_g3 gene_0_1_0             0            4             0
#> 4: gene_0_1_0_g4 gene_0_1_0            97           47            93
#> 5: gene_0_2_0_g1 gene_0_2_0            92           77            37
#> 6: gene_0_2_0_g2 gene_0_2_0           423          438           353
#>    dep_Sample_2 init_Sample_3 dep_Sample_3 init_Sample_4 dep_Sample_4
#> 1:           89           198          136            58          162
#> 2:          196           166          232           101          178
```

```
#> 3:           10             3           0            96          156
#> 4:          118            52          92           130           84
#> 5:          172           109         107            37          103
#> 6:          493           435         553           382          442
#>    init_Sample_5 dep_Sample_5 init_Sample_6 dep_Sample_6
#> 1:            55           56           115          147
#> 2:           226          246           152          132
#> 3:             0            0             0            0
#> 4:           228           93            85          163
#> 5:            51           68            13           19
#> 6:           414          488           474          544
```

Basic Analysis. The basic workflow for using ACE is three commands. First, all input data files are loaded into an R object (`DataObj`) and checked for problems. Next, this raw data is preprocessed according to the desired analysis to create a collection of parameters, forming a `ModelObj`. Both the raw data and the estimated parameters are then provided to the main function, `optimizeModelParameters`, which uses expectation maximization to infer gene essentiality, guide efficiency, and sample effects.

```
newDataObj <- DataObj$new(masterFiles = system.file('extdata','masterLibraryCounts.csv', package='ACER'),
                          countFile = system.file('extdata','countData.csv', package='ACER'),
                          negCtrlFile = system.file('extdata','negCtrlGenes.txt', package='ACER'),
                          sampleInfoFile=system.file('extdata','sampleAnnotations.txt', package='ACER'),
                          hasInitSeq = T)
```

```
#> Using 6 samples.
```

```
#> Not calculating guide efficiency. No covariate file.
```

```
#> Using these columns as sgrna and gene labels:
#> sgrna    gene
```

```
#> Warning in removeNullData(self, private$write_log): Some guides have no counts
#> in any initial samples, removing.
```

```
#> Warning in removeNullData(self, private$write_log): Some guides have no counts
#> in any depleted samples or masterlib, removing.
```

```
#> DataObj initialized.
```

```
newModelObj <- ModelObj$new(user_DataObj = newDataObj,
                            use_neg_ctrl=T,
                            test_samples='test',
                            use_master_library = T)
```

```
#> Summation performed over vector of length 1800
```

```
#> extracting sample subtype:test
```

```
#> Using the following number of samples as our test set:
```

```
#> 3
```

```
#> Using initial counts in calculation of LFC.
```

```
#> Scaling samples relative to matched masterlibrary
```

```
#> Excess of guides seem depleted; recommend using negative controls.
```

```
#> Using controls?: TRUE
```

```
newResultsObj <- optimizeModelParameters(user_DataObj = newDataObj,
                                         user_ModelObj = newModelObj)
```

```
#> ==========================================================================
```

```
#> Initial by-gene parameter optimization complete
```

```
#> Optimization complete, iteration: 0
```

```
writeResObj(newResultsObj)
```

```
#> writing results from
```

```
#> newResultsObj
```

```
#> ##------ Wed Dec 23 13:35:46 2020 ------##
```

```
#> newResultsObj written
```

```
#> $gene_results
#> [1] "ACE_output_data/newResultsObj_2020-12-23_13_35_46_gene_results.txt"
#>
#> $diff_genes
#> [1] "ACE_output_data/newResultsObj_2020-12-23_13_35_46_diff_genes.txt"
#>
#> $readMe
#> [1] "ACE_output_data/newResultsObj_2020-12-23_13_35_46_readMe.txt"
```

Differential Essentiality. To obtain estimates of differential essentiality estimated by sample subtype, as opposed to global essentiality across all samples, the file `sampleInfoFile` must be provided with sample annotation information to the original `DataObj`. Essentiality will be compared between samples with and without the annotation provided in the `ModelObj`'s `test_samples` parameter.

Acknowledgments. This tool was designed in collaboration with Prof. Adam Siepel and Prof. Chris Vakoc at Cold Spring Harbor Laboratory.

References

Hutton ER, Vakoc CR, Siepel A (2020). "ACE: A Probabilistic Model for Characterizing Gene-Level Essentiality in CRISPR Screens." *bioRxiv*. doi:10.1101/868919. URL https://www.biorxiv.org/content/10.1101/868919v2.