BALM GUI

Version 1.0

The BALM GUI (Fig. 1) is a wrapper around the BALM binary, designed to eliminate interaction with the command line. We hope that this will be of aid to biologists who seek to use BALM but lack DOS/UNIX/PowerShell experience.

Running the GUI:

On Mac OS X and Windows, double-click BALM_GUI.jar. On Linux, run "java -jar PATH" (without quotes, substitute PATH with the path to BALM_GUI.jar).

Instructions:

Specify the parameters for your analysis and click Submit. "Choose control file" and "Perform robust linear normalization" will

Figure 1: BALM GUI main window

become available if "Use control file" is selected, "Choose CpG file" if "MBD-seq data" is selected, and "Choose pericentromeric region file" if "Use pericentromeric region filter" is selected. It is also necessary to specify the location of the BALM binary for your system and threading preference that you wish to use (click "Select BALM binary"), and the directory in which BALM will create its output files (click "Select output directory"). Percentiles can be specified as a column-delimited or space-delimited list. All data files must be in the chosen format.

To process data in a custom tabular format, choose "Custom..." from the Format menu. This will bring up a window (Fig. 2) in which you can specify the columns of the chromosome, the start coordinate, the end coordinate and the strand. The leftmost column is considered 1 (unlike the binary's convention).
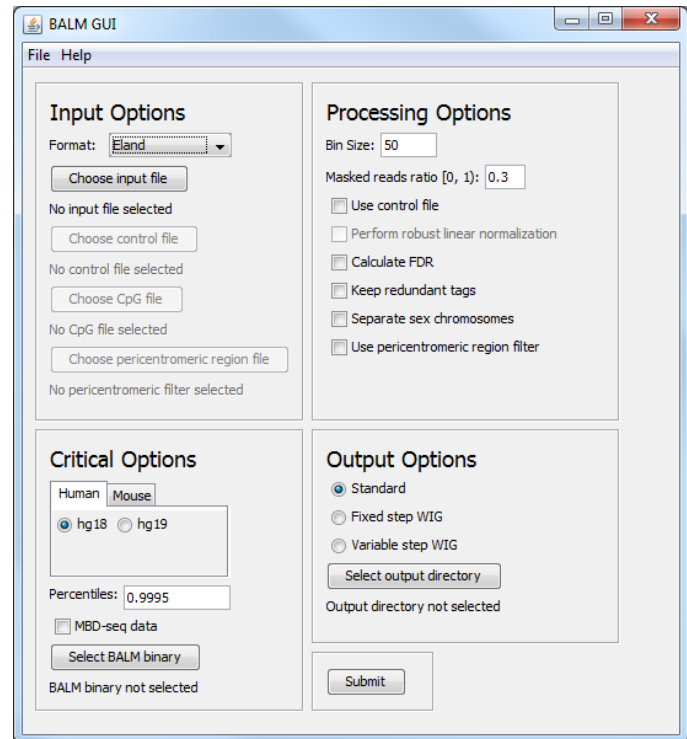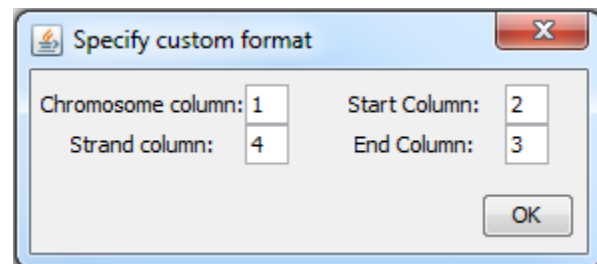

Figure 2: Custom format window

Note that the bin size can be no less than 50 and no greater than 10,000,000. The masked reads ratio can be no less than 0 and no greater than or equal to 1. Percentiles can be no less than 0.8 and no greater than or equal to 1.

After Submit is clicked, a window (Fig. 3) will open allowing you to monitor BALM's progress. You can terminate the run of BALM by closing the window. After the BALM run is complete, the Save button will be enabled, which saves a copy of BALM's progress and summary output.
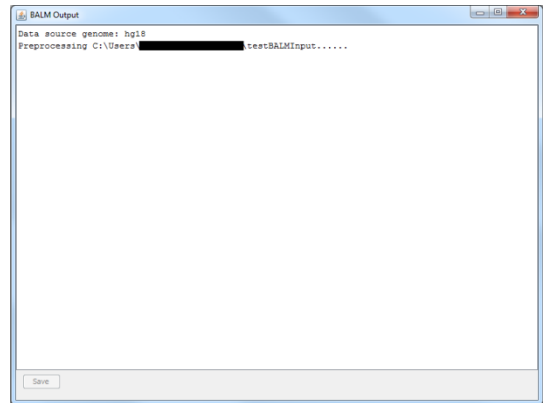

Figure 3: BALM output window

Acceptable Formats:

ELAND: at least five columns.

	1	2	3	4	5
Fields:	read sequence	ignored	ignored	chromosome number:start location (two fields, colon-separated)	strand (F or R)

Example: TGAGTGAGGTGTGGGCTCCACACCC 12500 1 chr19:4124128 F

Extended ELAND: at least four columns.

	1	2	3	4
Fields:	ignored	read sequence	ignored	chromosome file:start location, strand (**chromosome file as chr#.fa**, colon, start location, followed immediately by strand (F or R)

Example: >HWI-EAS435:4:1:2:1706#0/1 GGATGGAGTGCAGTGCTGCAATCATGGTTCACTGAA 0:1:86 chr12.fa:121208577R15G20

BED: at least six columns.

	1	2	3	4	5	6
Fields:	chromosome (chr#)	read start	read end	ignored	ignored	strand (+ or -)

Example: chr2 33925682 33925706 U0 0 -

SAM: at least ten columns.

	1	2	3	4	5	6	7	8	9	10
Fields:	ignored	SAM flag (column must be present)	chromosome (chr#)	read start	ignored	ignored	ignored	ignored	read sequence	ignored

Example: >HWI-EAS435_0007:2:1:0:1314#0 16 chr1 30740220 0 36M * 0 0
GGTATCAGTGATGGAAGACGAAATGAATGGAATGANcccccccccccccccccccccccccccccccccccc XT:A:R NM:i:2
X0:i:7 X1:i:0 XM:i:2 XO:i:0 XG:i:0 MD:Z:29A5A0

GFF: at least six columns, <u>tab separated.</u>

	1	2	3	4	5	6
Fields:	ignored	ignored	chromosome (single number)	strand (+ or -)	read start	read end

Example: Seq_4383668 ChIPSeq 1 + 559765 559799

bowtie: at least five columns.

	1	2	3	4	5
Fields:	ignored	strand (+ or -)	chromosome (chr#)	read start	sequence

Example: 853_8_11_F3 - chr15 71397686 GCTGTCTGCCCCTGTGCTGGGAGCATTCTATCACTGAC 79-
%%*>1)656/:H7)3?BHID<G;7AGECIH25KB 0

Custom: at least four columns
Chromosome column: in chr# format (e.g. chr2)
Start column: numerical format (e.g. 33925682)
End column: numerical format (e.g. 33925706)
Strand column: + or -

<u>Output files:</u>

Enriched region file (named [input file name]_P[percentage]_enrichedRegion]):

	1	2	3	4	5	6	7	8
Fields:	Chromosome number	Region start coordinate	Region end coordinate	Forward strand tags	Reverse strand tags	Region score	Number of tags in control	Region p-value

Example: chr1 559501 560100 53 55 4.99292 50 0.0002534
Note that columns 7 and 8 will only be present if BALM is run using a control file.

Target site file:

	1	2	3	4
Fields:	Chromosome number	Target site coordinate	Target site coordinate + 1	Target site score

Example: chr2 135000322 135000323 5.234432

CpG score file (only output when processing MBD-seq data):

	1	2	3
Fields:	Chromosome number	Coordinate	Methylation Score

Example: chr1 469 0.320838

preProcessInfo.txt: Contains summary statistical data obtained by BALM during preprocessing.

BALM_summary.txt: Contains summary statistical data (such as the FDR if selected) obtained by BALM after processing.

Known Issues:

- Closing the About box may throw an InvocationTargetException and/or StackOverflowError on UNIX systems with older JVM versions. This appears not to affect the functioning of the program in any way though, and we believe that this is caused by an API bug.
- Using custom genome .fa files is currently not supported by this GUI, although it is supported by BALM with the -g option.

We hope that BALM and the BALM GUI may be useful in your analysis and research.