FAQ for CAM

Questions for downloading data

Q1: The links you provided for output result and the program are not downlaod link, I can't download directly.

A: Yes the links we provided are not direct link, you need to click to the webpage and click download yourself.

Q2: Why there's two copy of same links for output results?

A: Not totally same. For each of downloading file, we provide Google Drive link for users and Baidu Pan link as a backup, because Google Drive is not accessiable in some place.

Questions for CAM program

Q1: In your pipeline you use bowtie only as alignment tool, have you tried other alignment tool or at least bowtie2?

A: We tried many alignment tools and they show similar performance on most MNase-seq data, though some slight difference exist. The reason we use bowtie1 is because we prefer the unique aligned function of bowtie1 (-m 1, which is no longer exist in bowtie2). We mentioned in the manual section that users can align their MNase-seq data with their prefered software and input the result (SAM format recommended) to CAM and take the following function.

Q2: Followed Q1, bowtie1 don't output mapping quality (MAPQ) when -m 1 is turned on, but CAM still have Q30 filter function. Is this a bug?

A: We understand that bowtie1 only report 0/255 for uniquely map (-m 1) mode. However, q30 is also effective in separating reads with quality=0 or 255 which helps remove non-uniquely mapped reads. What's more, q30 also works for aligned reads (SAM/BED) input which is aligned by other aligner user prefered.

Q3: What should I do if I do want to generate your example output myself?

A: If you want to generate our example output yourself, you can download the corresponded MNase-seq data from GEO (accession ID: GSM907784), annotation file (hg19) and custom region file in the Download section. Then you run CAM with all default parameters to get the example results (you can use both simple mode and standard mode with hg19 as genome version). NOTE that GSM907784 has 4 parts of .sra file, user should download all 4 parts, transform to paired end fastq, and combine them for CAM input. User can also input single part only, in which case the output result will be slightly different.

Q4: The published MNase-seq data I download is not in FASTQ format, but .sra format ?

A: If you download the MNase-seq from public domain (say, GEO), you will get a .sra file. Then you can transform the .sra file to FASTQ format with a free software called fastq-dump (from package SRA toolkit http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software). After you successfully installed SRA toolkit, Type command line:

```
$ fastq-dump.2.2.2 --split-files SRRXXXXX.sra
```

Then the .sra file will be split to 1 FASTQ files (for paired end there are 2 FASTQ files), named as SRRXXXXX_1.fastq and SRRXXXXX_2.fastq (paired end) or SRRXXXXX.fastq (single end).

Q5: I don't have pdflatex, and I don't have root permission to install it. How can I get the CAM QC report?

A: CAM will still process and generate all results without pdflatex, you can check summary/ folder for everything you want. Also, if you do need the summary report, you can copy the summary/plots/ folder to another machine with pdflatex installed and type:

```
$ pdflatex outname_summary.tex
```

In the summary/plots/ folder to generate the summary report.

Q6: The pipeline take too much time, its too slow !

A: Actually CAM take very little time to generate QC report and analysis results. The most time consuming step (which take more than 80% of time) is the mapping step, because MNase-seq data

usually have deep sequencing coverage. Users can modify -p parameter to speed up the mapping step. User can also specify mapping index to skip bowtie-build step (which take time to generate bowtie mapping index).