**How to run BRIE with your own reference genome with 10x genomics data for alternative splicing with short read sequencing**

**Summary**

Github: <https://github.com/huangyh09/briekit>

<https://github.com/huangyh09/brie>

BRIE is a Bayesian hierarchical model that estimates exon inclusion ratios by learning an informative prior distribution from sequence features. It also provides an effective tool for differential isoform quantification between scRNA-seq data sets.

I like BRIE because it is easy to carry out, it does not take a long time to run, and I can find the solutions for my problems all the time in the github. The result is very helpful for me. Here, I want to acknowledge the author for his hard work.

Ease of use: 4.5 out of 5

Runtime: 5 out 5 (does not take hours or days)

Summary of how the algorithm works

* Fine genome reference file (.fa) and annotation files (.gff3. .gtf)
* Set up conda environments
* Generate you alternative splicing event files with annotation files (from my experience, .gff3 work for gencode annotation, whereas .gtf works for Ensemble)
* Filter your event file with skipping exon event file and gtf annotation file and fa genome reference file
* Generate a feature file
* Estimate exon inclusion level with brie
* Calculate differential splicing with brie-diff
* Generate sashimi plot

**Step 1: Find your genome reference and annotation files**

<https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build#grch38_3.0.0>

wget ftp://ftp.ensembl.org/pub/release-93/fasta/homo\_sapiens/dna/Homo\_sapiens.GRCh38.dna.primary\_assembly.fa.gz

gunzip Homo\_sapiens.GRCh38.dna.primary\_assembly.fa.gz

wget ftp://ftp.ensembl.org/pub/release-93/gtf/homo\_sapiens/Homo\_sapiens.GRCh38.93.gtf.gz

gunzip Homo\_sapiens.GRCh38.93.gtf.gz

wget ftp://[ftp.ensembl.org/pub/release-93/gff3/homo\_sapiens/Homo\_sapiens.GRCh38.93.gff3.gz](http://ftp.ensembl.org/pub/release-93/gff3/homo_sapiens/Homo_sapiens.GRCh38.93.gff3.gz)

gunzip Homo\_sapiens.GRCh38.93.gff3.gz

You don’t need to gunzip the file unless you want to look at the file and edit them, since some of the inputs require .gz file

**Step 2: Set up conda environments**

conda create -n briekit python=2.7 numpy=1.13.0

conda activate briekit

Download the briekit github

python setup.py install

Package Version

----------------------------- ----------

backports.functools-lru-cache 1.6.1

briekit 0.2.2

certifi 2019.11.28

cycler 0.10.0

kiwisolver 1.1.0

matplotlib 2.2.5

misopy 0.5.4

numpy 1.16.6

pip 20.0.2

pyparsing 2.4.6

pysam 0.15.4

python-dateutil 2.8.1

pytz 2019.3

scikit-learn 0.16.1

scipy 1.2.3

setuptools 44.1.0

six 1.14.0

subprocess32 3.5.4

wheel 0.34.2

**Step 3:Generate you alternative splicing event files with annotation files**

There are 5 alternative splicing events, but the algorithm only works for SE (skipping exon) for now.

briekit-event -a /home/lingting/brie-master/BRIE/cell\_ranger\_reference/Homo\_sapiens.GRCh38.93.gtf.gz -o /home/lingting/brie-master/BRIE/cell\_ranger\_reference/AS\_events

**Step 4: Filter your event file with skipping exon event file and gtf annotation file and fa genome reference file**

I am using a more strict filtering method for alternative splicing events as the author suggested. You can have a more lenient filtering method. Check author’s documentation.

briekit-event-filter -a /home/lingting/brie-master/BRIE/cell\_ranger\_reference/AS\_events/SE.gff3.gz --anno\_ref=/home/lingting/brie-master/BRIE/cell\_ranger\_reference/Homo\_sapiens.GRCh38.93.gtf.gz -r /home/lingting/brie-master/BRIE/cell\_ranger\_reference/Homo\_sapiens.GRCh38.dna.primary\_assembly.fa --add\_chrom=X -o /home/lingting/brie-master/BRIE/cell\_ranger\_reference/AS\_events/SE.gold.gtf --no\_gene\_version

--add\_chrom=X for genome reference with X for X chromosome instead of chrX

--no\_gene\_version without this line, transcript id was versioned, and the filter provided by author does not have it.

Then you will find an output file, which only contains splicing events passing the following (default) constrains:

* located on autosome and input chromosome
* not overlapped by any other AS-exon
* surrounding introns are no shorter than a fixed length, e.g., 100bp
* length of alternative exon regions, say, between 50 and 450bp
* with a minimum distance, say 500bp, from TSS or TTS
* surrounded by AG-GT, i.e., AG-AS.exon-GT

**Step 5:Generate a feature file**

Now this step is a little tricky, you need to go to the .gff3 file (/home/lingting/brie-master/BRIE/cell\_ranger\_reference/AS\_events/SE.gold.LS.gff3), change chromosome label, 1,2,3…. To chr1,chr2,chr3….Then you need to change the index file of genome reference (Note, I keep a copy of this file for running of brie, this chr chromosome annotation is only need in this step) (/home/lingting/brie-master/BRIE/cell\_ranger\_reference/Homo\_sapiens.GRCh38.dna.primary\_assembly.fa.fai) by changing 1,2,3...to chr1,chr2,chr3……….. (After this step, I change this file back to 1,2,3 for chromosome annotation)

To do so, it also requires a genome sequence file (the same as above), and a [phast](http://compgen.cshl.edu/phast/) conservation file in [bigWig](https://genome.ucsc.edu/goldenpath/help/bigWig.html) format besides the annotation file.

<http://hgdownload.cse.ucsc.edu/goldenPath/hg38/phastCons100way/hg38.100way.phastCons/> is the link for human phast conservation file, and you can find the link for mouse in the documentation. You can find the bigWig file here <http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/bigWigSummary>.

Cd the location of your interest

wget http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86\_64/bigWigSummary

chmod +x bigWigSummary

So you have access to run this file.

Finally, you can extract your feature file now.

briekit-factor -a /home/lingting/brie-master/BRIE/cell\_ranger\_reference/AS\_events/SE.gold.LS.gff3 -r /home/lingting/brie-master/BRIE/cell\_ranger\_reference/Homo\_sapiens.GRCh38.dna.primary\_assembly.fa -c /home/lingting/brie-master/BRIE/hg38.phastCons100way.bw -o human\_features.csv -p 10 --bigWigSummary /home/lingting/brie-master/BRIE/bigWigSummary

Now congrats! you just made all the annotation and feature files that you need for BRIE.

You can run BRIE. briekit only runs in python 2, so I have a separate environment for it, but for BRIE, I just used my base environment, and I attached my packages at the end of articles.

**Step 6:Estimate exon inclusion level with brie**

conda deactivate

brie -a /home/lingting/brie-master/BRIE/cell\_ranger\_reference/AS\_events/SE.gold.LS.copy.gff3 -s /media/lingting/KiRAID/wgetdown/genomecenter.columbia.edu/ngs/release/singleCell/190312\_LANCE\_LINGTING\_2\_HUMAN\_10X/LL001/bam/possorted\_genome\_bam.bam -f /home/lingting/brie-master/BRIE/cell\_ranger\_reference/human\_features.csv.gz -o out\_dir5 -p 15

brie -a /home/lingting/brie-master/BRIE/cell\_ranger\_reference/AS\_events/SE.gold.LS.copy.gff3 -s /media/lingting/KiRAID/wgetdown/genomecenter.columbia.edu/ngs/release/singleCell/190312\_LANCE\_LINGTING\_2\_HUMAN\_10X/LL002/bam/possorted\_genome\_bam.bam -f /home/lingting/brie-master/BRIE/cell\_ranger\_reference/human\_features.csv.gz -o out\_dir6 -p 15

**Step 7: Calculate differential splicing with brie-diff**

I renamed my output file, then ran a differential splicing step.

brie-diff -i /home/lingting/brie-master/BRIE/cell\_ranger\_reference/out\_dir5\_LL001/samples.csv.gz,/home/lingting/brie-master/BRIE/cell\_ranger\_reference/out\_dir6\_LL002/samples.csv.gz -o hard\_soft.diff3.tsv -p 1 --minBF 0

I selected genes with BF greater or equal to 10, then most of them have 0.1 difference in psi between both samples. Those genes will be the alternative splicing genes.

**Step 8: Generate sashimi plot with MISO**

It requires a python 2 environment, so you can use your briekit environment. You also need to install the miso package. You can just update the .sh file and the sashimi setting file in the example section of brie github. Then allow the execution on the file. You can get your sashimi plot afterward.

brie\_sashimi.sh

GFF\_FILE --- your gtf or gff3 annotation file

GFF\_DIR ---- create a folder to put files from miso

PLOT\_DIR ---- assign your plot directory

SASHIMI----- update the folder location of your sashimi\_plot.py file

Update the plot events and name of the gene according

Sashimi\_setting.txt file

Update you bam file and miso file location

Change the label

Change exon scale intron scale and ymax if need

Update color and coverage based on your data

Pip install miso

conda activate briekit

chmod +x /home/lingting/brie-master/example/gastrulation/brie\_sashimi\_LRR1.sh

/home/lingting/brie-master/example/gastrulation/brie\_sashimi\_LRR1.sh

Package Version

---------------------------------- -------------------

alabaster 0.7.12

anaconda-client 1.7.2

anaconda-navigator 1.9.12

anaconda-project 0.8.3

anndata 0.6.22.post1

arboreto 0.1.5

argh 0.26.2

asn1crypto 1.3.0

astroid 2.3.3

astropy 3.2.3

atomicwrites 1.3.0

attrs 19.3.0

autopep8 1.4.4

Babel 2.8.0

backcall 0.1.0

backports.os 0.1.1

backports.shutil-get-terminal-size 1.0.0

beautifulsoup4 4.8.2

biograpy 1.0b0.dev0

biopython 1.76

bitarray 1.2.1

bkcharts 0.2

bleach 3.1.0

bokeh 2.0.0

boltons 20.0.0

boto 2.49.0

Bottleneck 1.3.2

brie 0.2.2

certifi 2019.11.28

cffi 1.14.0

chardet 3.0.4

Click 7.0

cloudpickle 1.3.0

clyent 1.2.2

cmake 3.16.3

colorama 0.4.3

conda 4.8.3

conda-build 3.17.8

conda-package-handling 1.6.0

conda-verify 3.1.1

contextlib2 0.6.0.post1

cryptography 2.8

cycler 0.10.0

Cython 0.29.6

cytoolz 0.10.1

dask 2.12.0

decorator 4.4.2

defusedxml 0.6.0

diff-match-patch 20181111

dill 0.3.1.1

distributed 1.28.1

docutils 0.16

entrypoints 0.3

et-xmlfile 1.0.1

fastcache 1.1.0

filelock 3.0.12

flake8 3.7.9

Flask 1.1.1

frozendict 1.2

fsspec 0.6.3

future 0.18.2

gevent 1.4.0

glob2 0.7

gmpy2 2.0.8

greenlet 0.4.15

h5py 2.9.0

HeapDict 1.0.1

html5lib 1.0.1

hypothesis 5.5.4

idna 2.9

imageio 2.8.0

imagesize 1.2.0

importlib-metadata 1.5.0

interlap 0.2.6

intervaltree 3.0.2

ipykernel 5.1.4

ipython 7.13.0

ipython-genutils 0.2.0

ipywidgets 7.5.1

isort 4.3.21

itsdangerous 1.1.0

jdcal 1.4.1

jedi 0.15.2

jeepney 0.4.3

jgraph 0.2.1

Jinja2 2.11.1

joblib 0.14.0

json5 0.9.3

jsonschema 3.2.0

jupyter 1.0.0

jupyter-client 6.1.0

jupyter-console 6.1.0

jupyter-core 4.6.1

jupyterlab 1.2.6

jupyterlab-server 1.0.7

keyring 21.1.0

kiwisolver 1.1.0

lazy-object-proxy 1.4.3

libarchive-c 2.8

lief 0.9.0

llvmlite 0.28.0

locket 0.2.0

loompy 3.0.6

louvain 0.6.1

lxml 4.5.0

MarkupSafe 1.1.1

matplotlib 3.0.3

mccabe 0.6.1

mkl-fft 1.0.15

mkl-random 1.1.0

mkl-service 2.3.0

mock 4.0.1

more-itertools 8.2.0

mpmath 1.1.0

msgpack 1.0.0

msgpack-python 0.5.6

MulticoreTSNE 0.1

multipledispatch 0.6.0

multiprocessing-on-dill 3.5.0a4

natsort 6.0.0

navigator-updater 0.2.1

nbconvert 5.6.1

nbformat 5.0.4

networkx 2.4

nltk 3.4.5

nose 1.3.7

notebook 6.0.3

numba 0.43.1

numexpr 2.7.1

numpy 1.16.2

numpy-groupies 0+unknown

numpydoc 0.9.2

olefile 0.46

openpyxl 3.0.3

packaging 20.3

pandas 1.0.3

pandocfilters 1.4.2

parso 0.5.2

partd 1.1.0

path 13.1.0

pathlib2 2.3.5

pathtools 0.1.2

patsy 0.5.1

pep8 1.7.1

pexpect 4.8.0

pickleshare 0.7.5

Pillow 7.0.0

pip 20.0.2

pkginfo 1.5.0.1

pluggy 0.13.1

ply 3.11

prometheus-client 0.7.1

prompt-toolkit 3.0.3

psutil 5.7.0

ptyprocess 0.6.0

py 1.8.1

pyarrow 0.16.0

pycodestyle 2.5.0

pycosat 0.6.3

pycparser 2.20

pycrypto 2.6.1

pycurl 7.43.0.5

pydocstyle 4.0.1

pyflakes 2.1.1

Pygments 2.6.1

pylint 2.4.4

pyodbc 4.0.0-unsupported

pyOpenSSL 19.1.0

pyparsing 2.4.6

pyrsistent 0.15.7

pysam 0.15.3

pyscenic 0.9.19+12.gbee2576

PySocks 1.7.1

pytest 5.4.1

pytest-arraydiff 0.3

pytest-astropy 0.8.0

pytest-astropy-header 0.1.2

pytest-doctestplus 0.5.0

pytest-openfiles 0.4.0

pytest-remotedata 0.3.2

python-dateutil 2.8.1

python-igraph 0.7.1.post6

python-jsonrpc-server 0.3.4

python-language-server 0.31.9

pytz 2019.3

PyWavelets 1.1.1

pyxdg 0.26

PyYAML 5.3.1

pyzmq 18.1.1

QDarkStyle 2.8

QtAwesome 0.7.0

qtconsole 4.7.2

QtPy 1.9.0

requests 2.23.0

rope 0.16.0

Rtree 0.9.3

ruamel-yaml 0.15.87

scanpy 1.4.4.post1

scikit-image 0.16.2

scikit-learn 0.20.3

scipy 1.3.1

seaborn 0.10.0

SecretStorage 3.1.2

Send2Trash 1.5.0

setuptools 46.1.1.post20200323

simplegeneric 0.8.1

singledispatch 3.4.0.3

six 1.14.0

snowballstemmer 2.0.0

sortedcollections 1.1.2

sortedcontainers 2.1.0

soupsieve 2.0

Sphinx 2.4.4

sphinxcontrib-applehelp 1.0.2

sphinxcontrib-devhelp 1.0.2

sphinxcontrib-htmlhelp 1.0.3

sphinxcontrib-jsmath 1.0.1

sphinxcontrib-qthelp 1.0.3

sphinxcontrib-serializinghtml 1.1.4

sphinxcontrib-websupport 1.2.1

spyder 4.1.1

spyder-kernels 1.9.0

SQLAlchemy 1.3.15

statsmodels 0.11.0

sympy 1.5.1

tables 3.6.1

tblib 1.6.0

terminado 0.8.3

testpath 0.4.4

toolz 0.10.0

tornado 6.0.4

tqdm 4.43.0

traitlets 4.3.3

typing-extensions 3.7.4.1

ujson 1.35

umap-learn 0.3.9

unicodecsv 0.14.1

urllib3 1.25.8

velocyto 0.17.17

watchdog 0.10.2

wcwidth 0.1.8

webencodings 0.5.1

Werkzeug 1.0.0

wheel 0.34.2

widgetsnbextension 3.5.1

wrapt 1.12.1

wurlitzer 2.0.0

xlrd 1.2.0

XlsxWriter 1.2.8

xlwt 1.3.0

xmltodict 0.12.0

yapf 0.28.0

zict 2.0.0

zipp 2.2.0