

---

# **blockify**

***Release 0.1.0***

**Arnav Moudgil**

**Sep 13, 2019**



**CONTENTS:**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Command: segment</b>	<b>3</b>
<b>3</b>	<b>Command: call</b>	<b>5</b>
<b>4</b>	<b>Command: normalize</b>	<b>7</b>
<b>5</b>	<b>Command: downsample</b>	<b>9</b>
<b>6</b>	<b>Indices and tables</b>	<b>11</b>



## INTRODUCTION

blockify is a genomic peak caller for one-dimensional data (e.g. CCF).



## COMMAND: SEGMENT

Segment a .ccf file using Bayesian blocks

Required parameters:

- -i, -input INPUT: Input .ccf file
- -o, -output: Output file (BED format)

Optional parameters:

- -prior PRIOR: Explicit prior on the number of blocks (*not recommended for general use*)
- -p0 P0: Empirical prior based on a specified false-positive rate; must be between 0 and 1 (default: 0.05)
- -method {OP,PELT}: Segment using the optimal partitioning (OP) or pruned exact linear time (PELT) algorithm (default: PELT)





## COMMAND: CALL

Call peaks in a .ccf file

Required parameters:

- -i, -input INPUT: input .ccf file
- output: output bed file
- Either:
  - -p, -pValueCutoff PVALUECUTOFF: p-value cutoff (NOTE: This is a straight cutoff and will not take into account multiple hypothesis correction!), OR
  - -a, -alpha ALPHA: alpha for multiple hypothesis correction (must be between 0 and 1) AND
  - -correction CORRECTION: if alpha provided, need to specify method of multiple hypothesis correction. See [statsmodels.stats.multitest](#) for a complete list of choices (default: bonferroni)
- -bg, -background BACKGROUND: Background .ccf file

Optional parameters:

- -r, -regions REGIONS: Regions over which to normalize event counts; should be supplied as a BED file. If not provided, the input file will be segmented using Bayesian blocks. (all options from blockify segment are available)
- -intermediate INTERMEDIATE: Intermediate file to write verbose output (CSV format)
- -d, -distance DISTANCE: Merge features closer than this distance (bp)
- -min MIN: Report peaks larger than this cutoff (bp)
- -max MAX: Report peaks smaller than this cutoff (bp)
- -t, -tight: Shrink peak boundaries to overlap data points
- -c, -pseudocount PSEUDOCOUNT: Pseudocount for background regions (default: 1)



## COMMAND: NORMALIZE

Calculate normalized rates of events in a .ccf file

Required parameters:

- -i, -input INPUT: input .ccf file
- -o, -output OUTPUT: Output file (bedGraph format)

Optional parameters:

- -r, -regions REGIONS: Regions over which to normalize event counts; should be supplied as a BED file. If not provided, the input file will be segmented using Bayesian blocks. (all options from blockify segment are available)
- -k, -libraryFactor LIBRARYFACTOR: Normalization factor for library size (default: 1000000)
- -l, -lengthFactor LENGTHFACTOR: Normalization factor for the length of regions; used to calculate scaled rates of events per interval (default: None)



## COMMAND: DOWNSAMPLE

Downsample a .ccf file in proportion to the value column

Required parameters:

- -i INPUT, --input INPUT: Input .ccf file
- -o, --output OUTPUT: Output file (CCF format)
- -n, --number NUMBER: Number of entries to downsample to (cannot exceed length of input file)

Optional parameters:

- -s SEED, --seed SEED: Random seed
- --naive: Sample every row with equal likelihood



## INDICES AND TABLES

- `genindex`
- `modindex`
- `search`