Intervene Documentation

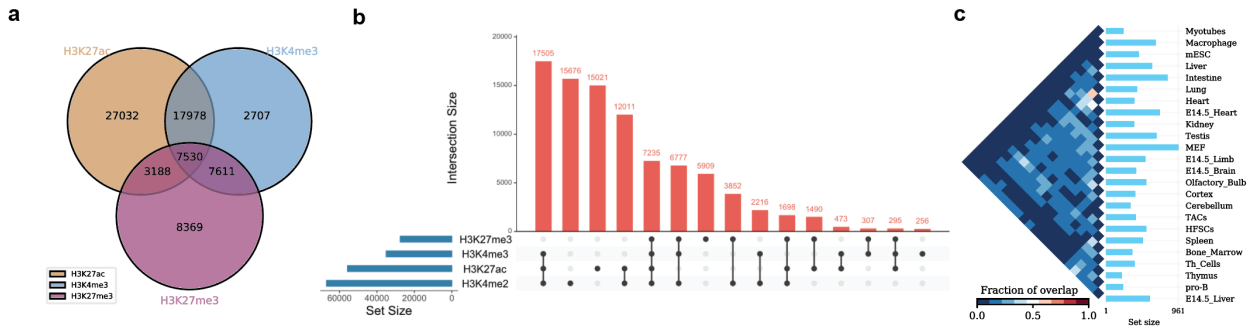
Release v0.41.0

Aziz Khan

Mar 08, 2017

Table of contents

Welcome to Intervene - a tool for intersection and visualization of multiple genomic region sets



CHAPTER 1

Introduction

Intervene is a tool for intersection and visualization of multiple genomic region and gene sets.

Intervene, provides an easy and automated interface for effective intersection and visualization of genomic region sets, thus facilitating their analysis and interpretations. Intervene contains three modules: venn to compute Venn diagrams of up-to 6 sets, upset to compute UpSet plots of more than 3 sets, and pairwise to compute and visualize intersections of genomic sets as clustered heatmap. Intervene gives user flexibility to choose figure colors, labels, size, quality, and type to make them as publication standard.

CHAPTER 2

Installation

Prerequisites

Intervene requires the following Python modules and R packages:

- Python (=> 2.7): https://www.python.org/
- BedTools (Latest version): https://github.com/arq5x/bedtools2
- pybedtools (>= 0.7.9): https://daler.github.io/pybedtools/
- Pandas (>= 0.16.0): http://pandas.pydata.org/
- R (>= 3.0): https://www.r-project.org/
- R packages including UpSetR, corrplot

Install BEDTools

Intervene is using pybedtools, which is Python wrapper for BEDTools. So, BEDTools should be installed before using Intervene It's recomended to have a latest version, but if you have an older version already install, it should be fine. Please read the instructions at https://github.com/arq5x/bedtools2 to install BEDTools, and make sure it is on your path and you are able to call bedtools from any directory.

Install required Python modules

Intervene takes care of the installation of all the required Python modules. If you already have a working installation of Python, the easiest way to install required Python modules is by installing Intervene using `pip`. If you're setting up Python for the first time, we recommend to install it using Anaconda Python distribution http://continuum.io/downloads. These come with several helpful scientific and data processing libraries. These are available for platforms including Windows, Mac OSX and Linux.

If you want to install requires Python modules individually, you can use the following commands, else you can install Intervene directly.

Install pybedtools

Install it from PyPi

```
pip install pybedtools
```

or using conda

```
conda install -c bioconda pybedtools
```

Read more details about ''pybedtools" installation: https://daler.github.io/pybedtools/main.html

Install Pandas

Install it from PyPi

```
pip install pandas
```

Or install with conda

```
conda install pandas
```

Install required R packages

Intervene requires two R packages, `UpSetR` https://cran.r-project.org/package=UpSetR and `corrplot` https://cran.r-project.org/package=corrplot for visualization. To install these open R/RStudio and use the following command.

```
install.packages(c("UpSetR", "corrplot"))
```

Install Intervene

You can install a stable version of Intervene by using `pip` from PyPi or a development version by using `git` from GitHub.

Install using *pip*

You can install InterVene either from PyPi using pip or install it from the source. Please make sure you have already installed the above mentioned python libraries required to run InterVene.

Install from PyPi:

```
pip install intervene
```

Install development version from *GitHub*

If you have *git* installed, use this:

```
git clone https://github.com/asntech/intervene.git
cd intervene
python setup.py install
```

CHAPTER 3

How to use Intervene

Once you have installed Intervene, you can type:

```
intervene --help
```

This will show the main help, which list three subcommands/modules, including `venn`, `upset`, `pairwise`.

To view the help for the individual subcommands, please type:

To view `venn` module help, type this;

```
intervene venn --help
```

To view `upset` module help, type this;

```
intervene upset --help
```

To view `pairwise` module help, type this;

```
intervene pairwise --help
```

Run Intervene on test data

To run Intervene's each module using example data use the following commands.

To run `venn` module with test data, type this;

```
intervene venn --test
```

To run `upset` module with test data, type this;

```
intervene upset --test
```

To run `pairwise` module with test data, type this;

```
intervene pairwise --test
```

These commands will save the results in the current working directory with a folder named `Intervene_results`. If you wish to save the results in a specific folder, you can type:

```
intervene <module_name> --test --output ~/path/to/your/results/folder
```

CHAPTER 4

Intervene modules

Intervene provides three types of plots to visualize intersections of genomic regions and list sets. These are pairwise heatmap of N genomic region sets, classic Venn diagrams of genomic regions and list sets of up to 6-way and UpSet plots.

Venn diagram module

Once you have installed Intervene, you can type:

Usage:

```
intervene venn [options]
```

Note: Please scroll down to see a detailed summary of available **options**.

Help:

```
intervene venn --help
```

Example:

```
intervene venn -i path/to/BED/files/*.bed --type jaccard --htype tribar
```

This will save the results in the current working directory with a folder named `Intervene_results`. If you wish to save the results in a specific folder, you can type:

```
intervene venn -i path/to/BED/files/*.bed --type jaccard --htype tribar --output ~/
↪results/path
```

Summary of options

Option	Description
-h, —help	To show the help message and exit
-i	Input genomic regions in (BED/GTF/GFF) format or lists of genes/SNPs IDs. For files in a directory use *.<extension>. e.g. *.bed
–type	{genomic,list}. Type of input data sets. Genomic regions or lists of genes/SNPs. Default is `genomic`
–names	Comma-separated list of names as labels for input files. Default is: –names=A,B,C,D,E,F
–filenames	Use file names as labels instead. Default is `False`
–colors	Comma-separated list of matplotlib-valid colors. E.g., –colors=r,b,k
-o, –output	Output folder path where results will be stored. Default is current working directory.
–figtype	{pdf,svg,ps,tiff,png} Figure type for the plot. e.g. –figtype svg. Default is `pdf`
–figsize	Figure size as width and height.e.g. –figsize 12 12.
–dpi	Dots-per-inch (DPI) for the output. Default is: `300`
–fill	{number,percentage} Report number or percentage of overlaps (Only if –type=list). Default is `number`
–test	This will run the program on test data.

UpSet plot module

Once you have installed Intervene, you can type:

Usage:

```
intervene upset [options]
```

Note: Please scroll down to see a detailed summary of available **options**.

Help: You can also see list of options by typing this on the terminal.

```
intervene upset --help
```

Example:

```
intervene upset -i path/to/BED/files/*.bed --type jaccard --htype tribar
```

This will save the results in the current working directory with a folder named `Intervene_results`. If you wish to save the results in a specific folder, you can type:

```
intervene upset -i path/to/BED/files/*.bed --type jaccard --htype tribar --output ~/
↪results/path
```

Summary of options

Option	Description
-h, –help	show this help message and exit
-i, –input	Input genomic regions in <BED/GTF/GFF/VCF> format or list files. For files in a directory use *.<ext>. e.g. *.bed
–type	Type of input sets. Genomic regions or lists of genes sets {genomic,list}. Default is `genomic`
–names	Comma-separated list of names for input files. Default is‘‘–names=A,B,C,D,E,F‘‘
–filenames	Use file names as labels instead. Default is `False`
-o, –output	Output folder path where plots will store. Default is current working directory.
–order	The order of intersections of sets {freq,degree}. e.g. –order degree. Default is `freq`
–ninter	Number of top intersections to plot. Default is `40`
–showzero	Show empty overlap combinations. Default is `False`
–showsize	Show intersection sizes above bars. Default is `False`
–mbcolor	Color of the main bar plot. Default is `gray23`
–sbcolor	Color of set size bar plot. Default is `#56B4E9`
–mblabel	The y-axis label of the intersection size bars. Default is `No of Intersections`
–sxlabel	The x-axis label of the set size bars. Default is `Set size`
–figtype	Figure type for the plot. e.g. –figtype svg {pdf,svg,ps,tiff,png} Default is `pdf`
–figsize	Figure size for the output plot (width,height)
–dpi	Dots-per-inch (DPI) for the output. Default is `300`
–run	Run Rscript if R and UpSetR package is installed. Default is `True`

Pairwise intersection module

Once you have installed Intervene, you can type:

Usage:

```
intervene pairwise [options]
```

Note: Please scroll down to see a detailed summary of available **options**.

Help:

```
intervene pairwise --help
```

Example:

```
intervene pairwise -i path/to/BED/files/*.bed --type jaccard --htype tribar
```

This will save the results in the current working directory with a folder named `Intervene_results`. If you wish to save the results in a specific folder, you can type:

```
intervene pairwise -i path/to/BED/files/*.bed --type jaccard --htype tribar --output ~
↪/results/path
```

Summary of options

Option	Description
-h, –help	show this help message and exit
-i	Input genomic regions in (BED/GTF/GFF) format. For files in a directory use *.<extension>. e.g. *.bed
–type	Report count/fraction of overlaps or statistical relationships. {`count frac jaccard fisher reldist`}
	–type=count - calculates the number of overlaps.
	–type=frac - calculates the fraction of overlap.
	–type=jaccard - calculate the Jaccard statistic.
	–type=reldist - calculate the distribution of relative distances.
	–type=fisher - calculate Fisher's statistic.
	Default is `frac`
–htype	{tribar,color,pie,circle,square,ellipse,number,shade}. Heatmap plot type. Default is `pie`.
–names	Comma-separated list of names for input files. Default is base name of input files.
–filenames	Use file names as labels instead. Default is `False`.
–sort	Set this only if your files are not sorted. Default is `False`.
–genome	Required argument if –type=fisher. Needs to be a string assembly name such as `mm10` or `hg38`
-o, –output	Output folder path where results will be stored. Default is current working directory.
–barlabel	x-axis label of boxplot if –htype=tribar. Default is `Set size`
–barcolor	Boxplot color (hex vlaue or name, e.g. blue). Default is `#53cfff`.
–fontsize	Label font size. Default is `8`.
–title	Heatmap main title. Default is `Pairwise intersection`
–space	White space between barplt and heatmap, if –htype=tribar. Default is `1.3`.
–figtype	{pdf,svg,ps,tiff,png} Figure type for the plot. e.g. –figtype svg. Default is `pdf`
–figsize	Figure size for the output plot (width,height). e.g. –figsize 8 8
–dpi	Dots-per-inch (DPI) for the output. Default is: `300`.
–test	This will run the program on test data.

CHAPTER 5

Example gallery

Here we listed some examples to demonstrate how Intervene can be used to generated different types of set intersection plots.

Venn module examples

In this example, a 3-way Venn diagram of ChIP-seq peaks of histone modifications (H3K27ac, H3Kme3 and H3K27me3) in hESC from ENCODE data (Dunham et al., 2012).

```
intervene venn -i ~/ENCODE/data/H3K27ac.bed ~/ENCODE/data/H3Kme3.bed ~/ENCODE/data/
↪H3K27me3.bed --filenames
```

By adding one more BED file to `-i` argument, Intervene will generate a 4-way Venn diagram of overlap of ChIP-seq peaks.

```
intervene venn -i ~/ENCODE/data/H3K27ac.bed ~/ENCODE/data/H3Kme3.bed ~/ENCODE/data/
↪H3K27me3.bed  ~/ENCODE/data/H3Kme2.bed --filenames
```

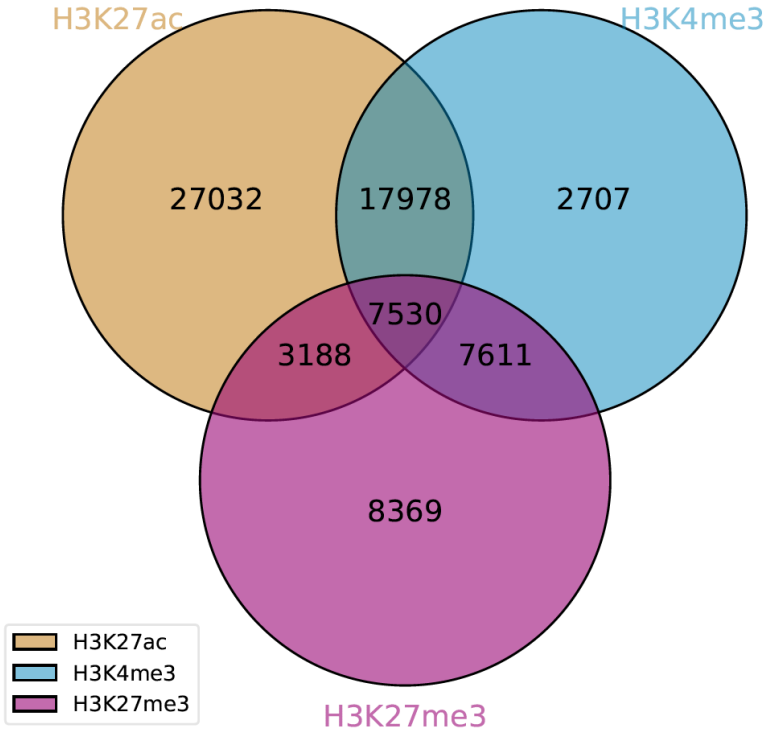
Read more about the `venn` diagrams module here:

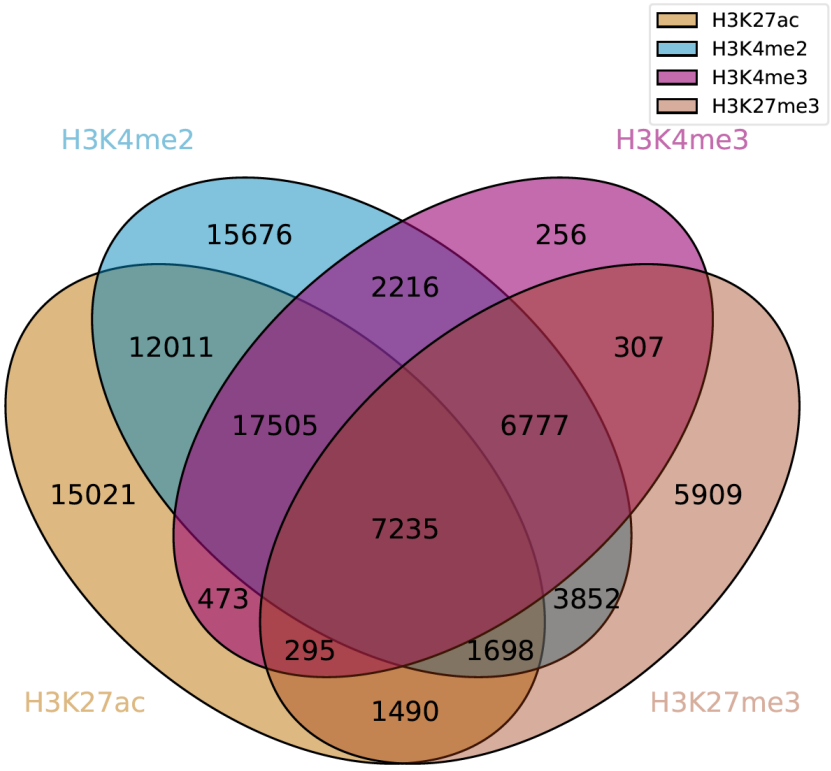
```
intervene venn --help
```

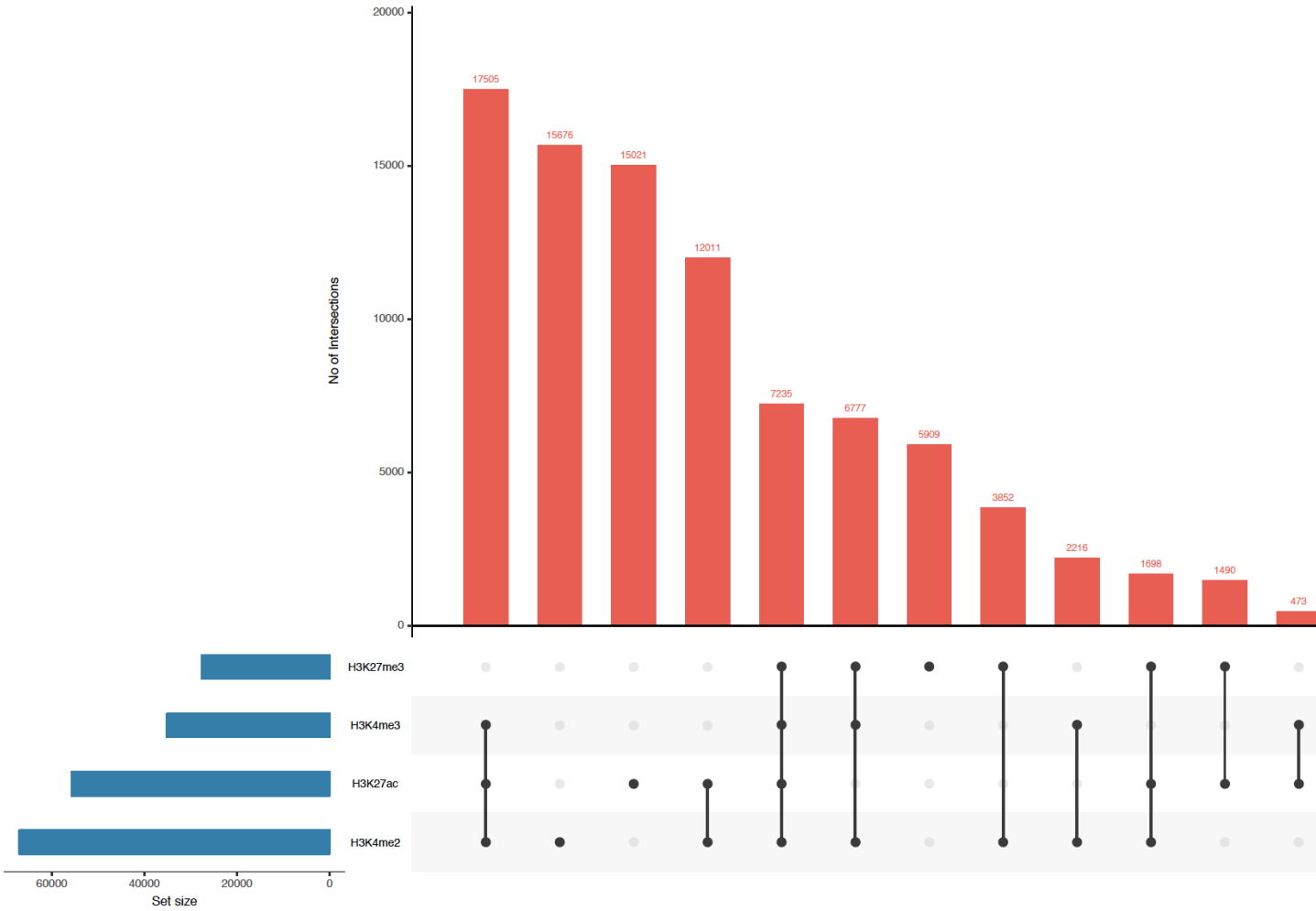
UpSet module examples

In this example, a UpSet plot of ChIP-seq peaks of four histone modifications (H3K27ac, H3Kme3 H3Kme2, and H3K27me3) in hESC from ENCODE data (Dunham et al., 2012).

```
intervene upset -i ~/ENCODE/data/H3K27ac.bed ~/ENCODE/data/H3Kme3.bed ~/ENCODE/data/
↪H3K27me3.bed ~/ENCODE/data/H3Kme2.bed --filenames
```


H3K27ac
H3K4me3
27032
17978
2707
7530
3188
7611
8369
H3K27ac
H3K4me3
H3K27me3
H3K27me3


H3K27ac
H3K4me2
H3K4me3
H3K27me3
H3K4me2
H3K4me3
15676
256
2216
12011
307
17505
6777
15021
5909
7235
473
3852
295
1698
1490
H3K27ac
H3K27me3


20000
15000
10000
5000
0
No of Intersections
17505
15676
15021
12011
7235
6777
5909
3852
2216
1698
1490
473
H3K27me3
H3K4me3
H3K27ac
H3K4me2
60000
40000
20000
0
Set size

Read more about the `upset` module:

In this example ...

```
intervene upset --help
```

Pairwise module examples

In this example, we performed a pairwise intersections of super-enhancers in 24 mouse cell and tissue types from dbSUPER(Khan and Zhang, 2016) and showed the fraction of overlap in heatmap.

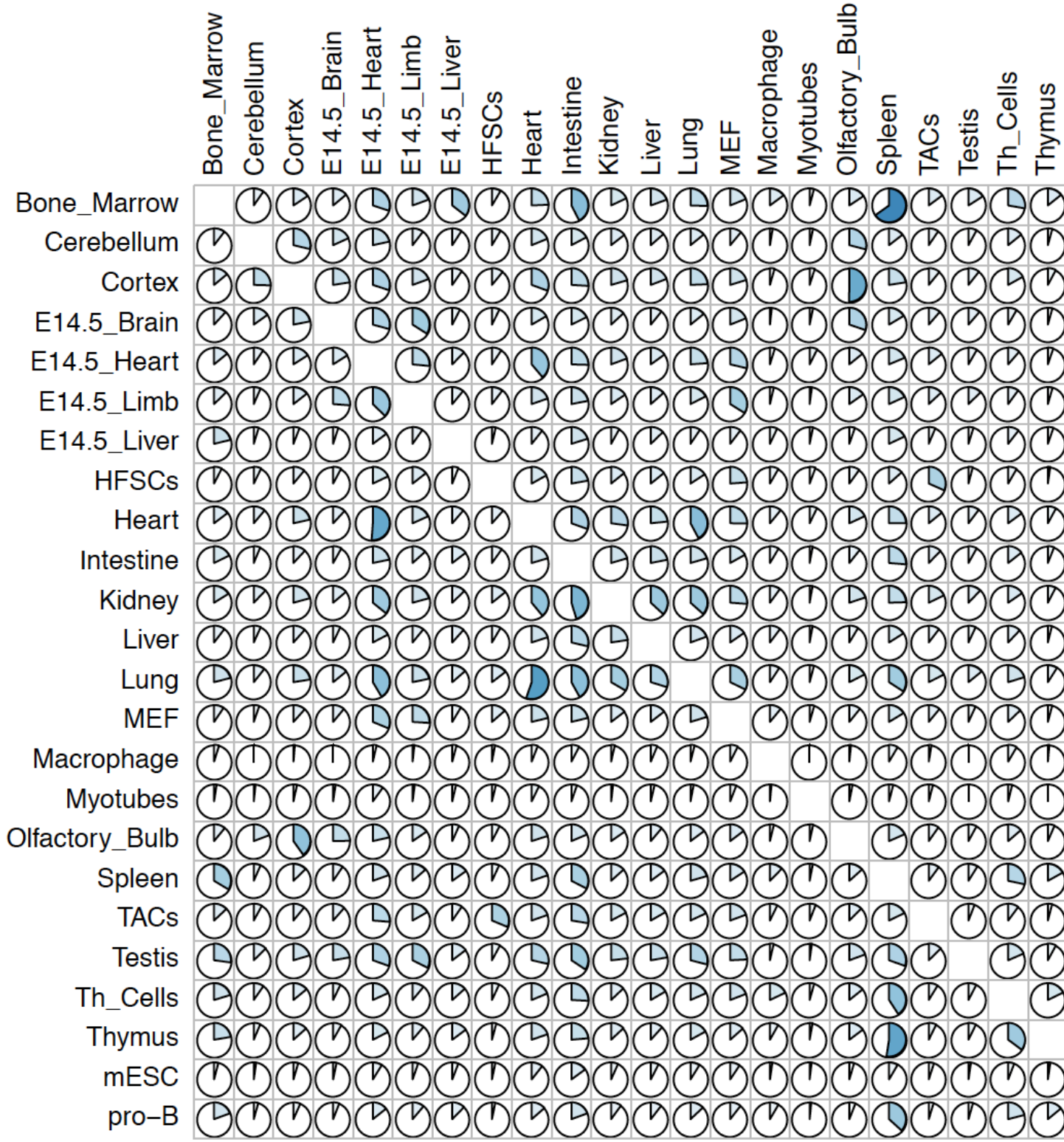
```
intervene upset -i ~/dbSUPER/mm9/*.bed --filenames --type frac --htype pie
```

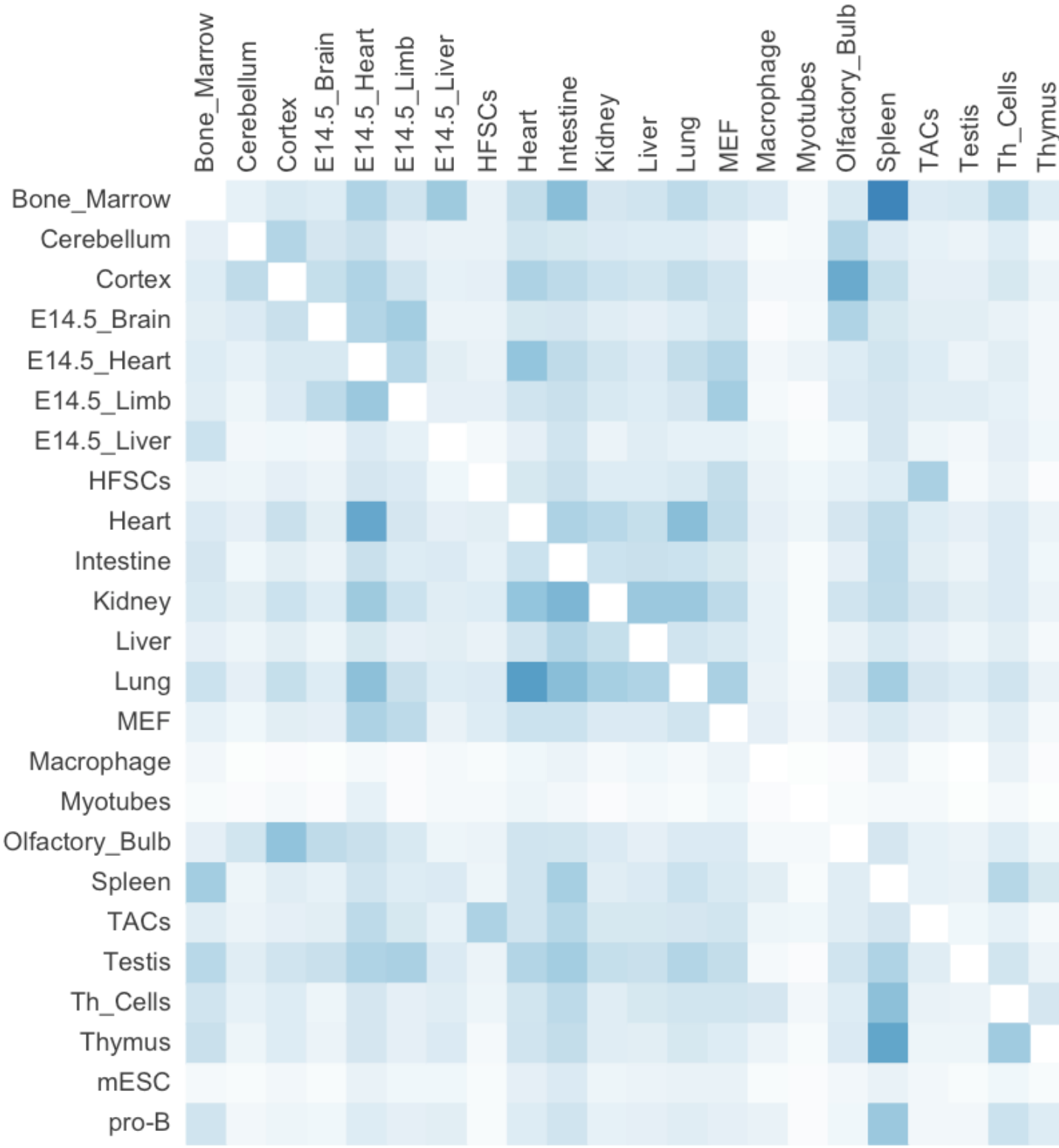
By setting the `--htype` to `color` will produce this plot.

```
intervene upset -i ~/dbSUPER/mm9/*.bed --filenames --type frac --htype color
```

Read more about the `pairwise` module here:

```
intervene pairwise --help
```


Bone_Marrow
Cerebellum
Cortex
E14.5_Brain
E14.5_Heart
E14.5_Limb
E14.5_Liver
HFSCs
Heart
Intestine
Kidney
Liver
Lung
MEF
Macrophage
Myotubes
Olfactory_Bulb
Spleen
TACs
Testis
Th_Cells
Thymus
mESC
pro−B


Bone_Marrow
Cerebellum
Cortex
E14.5_Brain
E14.5_Heart
E14.5_Limb
E14.5_Liver
HFSCs
Heart
Intestine
Kidney
Liver
Lung
MEF
Macrophage
Myotubes
Olfactory_Bulb
Spleen
TACs
Testis
Th_Cells
Thymus
mESC
pro-B

CHAPTER 6

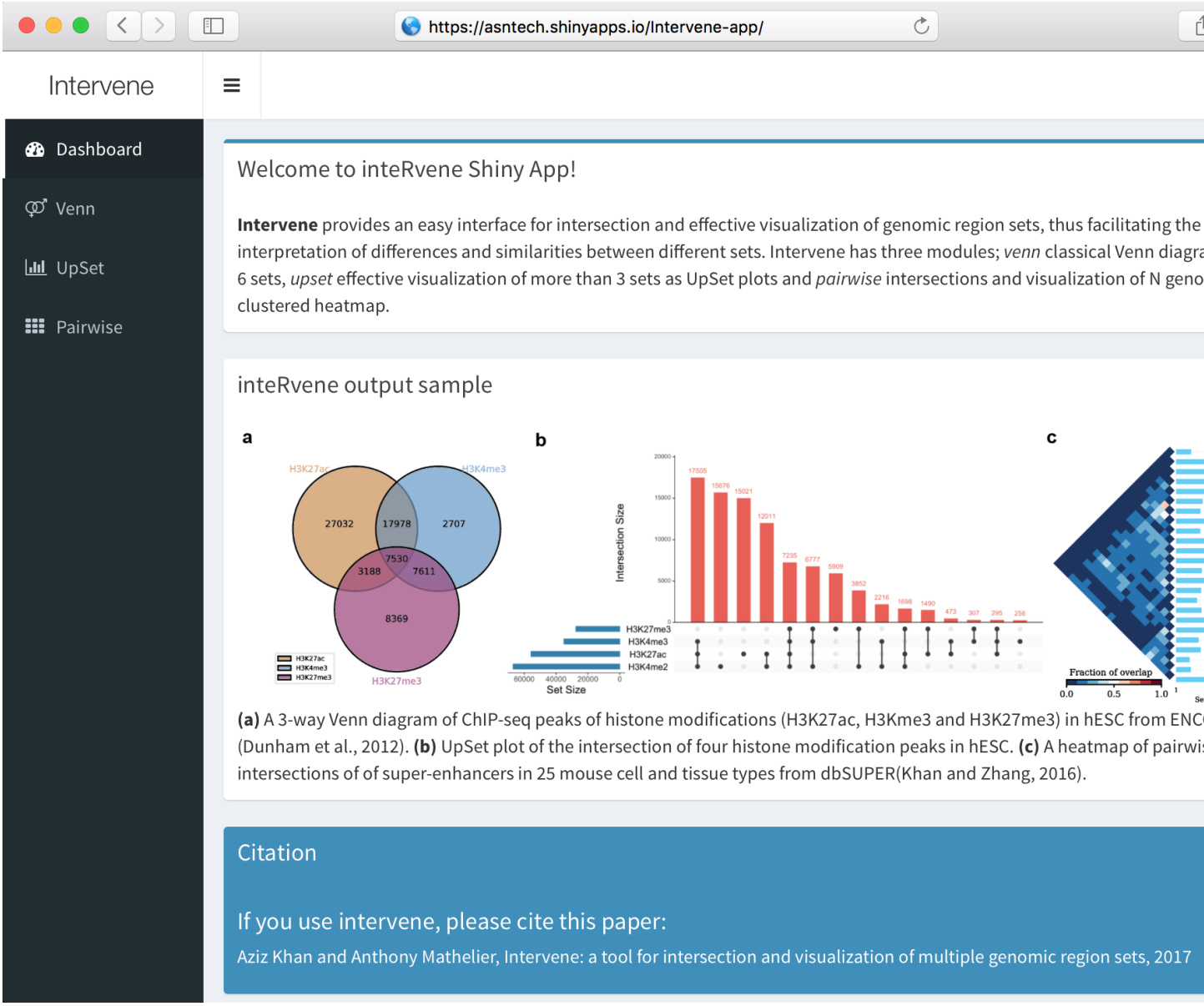
Interactive Shiny App

Introduction

Intervene also comes with an interactive Shiny App to further explore and filter the results in a more interactive way. Intervene command line interface also gives option to produce results as text files, which can be easily import to the Shiny App for interactive visualization and customization of plots.

Availability

The Intervene Shiny App is freely available at https://asntech.shinyapps.io/Intervene-app



(a) A 3-way Venn diagram of ChIP-seq peaks of histone modifications (H3K27ac, H3Kme3 and H3K27me3) in hESC from ENC (Dunham et al., 2012). **(b)** UpSet plot of the intersection of four histone modification peaks in hESC. **(c)** A heatmap of pairwi intersections of of super-enhancers in 25 mouse cell and tissue types from dbSUPER(Khan and Zhang, 2016).

CHAPTER 7

Support

If you have questions, or found any bug in the program, please write to us at `aziz.khan[at]ncmm.uio.no`

CHAPTER 8

Citation

If you use Intervene in a paper, please cite:

- *Aziz Khan and Anthony Mathelier,* **Intervene: a tool for intersection and visualization of multiple genomic region sets**, 2017