Venn diagram of analyses results

The python script RRBS_HOME/Venn/get_venn.py allows user to compare differents set of results.

We will illustrate usage of this script with the comparison of differentially methylated cytosine (DMC) obtained either by methylSig or methylKit analyses (samples {A1;A2} of condition A versus samples {B1;B2} of condition B).

Output file of these analyses have the following format:

DMC position

Samples: A1 A2 B1 B2 — Statistical results of A vs B comparison

Chromosome	Start	End	CovA1	FreqCA1	CovA2	FreqCA2	CovB1	FreqCB1	CovB2	FreqCB2	pvalue	Methyl diff	Methylation state in Condition_A
1	178480	178481	13	4	18	6	19	0	13	0	5.86E-05	32.26	hypermeth
1	178484	178485	13	4	18	6	19	0	13	0	5.86E-05	32.26	hypermeth
1	178489	178490	13	4	18	7	19	0	13	0	2.18E-05	35.48	hypermeth
1	178492	178493	13	4	18	6	18	1	13	0	0.001447853	29.03	hypermeth
1	178525	178526	12	5	17	8	19	0	13	0	1.38E-06	44.83	hypermeth
1	178528	178529	13	4	17	4	19	0	13	0	0.000330623	26.67	hypermeth
1	178531	178532	13	6	17	2	19	0	13	0	0.000330623	26.67	hypermeth
1	178540	178541	13	4	18	5	18	1	12	0	0.003945532	25.70	hypermeth
1	178545	178546	13	5	18	8	18	0	12	0	4.50E-06	41.94	hypermeth
1	178549	178550	13	1	18	10	18	0	12	0	3.28E-05	35.48	hypermeth
1	178553	178554	13	7	18	10	18	1	12	0	2.04E-06	51.51	hypermeth
1	178595	178596	12	3	18	9	19	0	13	0	5.83E-06	40.00	hypermeth
1	214593	214594	20	18	30	27	22	20	30	12	0.000591235	28.46	hypermeth
1	252480	252481	23	7	23	13	16	16	16	16	1.68E-09	-56.52	hypometh
1	535487	535488	17	8	36	25	14	14	22	22	7.10E-07	-37.74	hypometh
1	558926	558927	12	10	23	22	12	3	21	3	1.15E-10	73.25	hypermeth

(Cf. RRBS_HOME/Differential_analysis/analysis_examples/out_methylKit/MethylKit - Condition A vs B - pvalue0.01 - with obvious DMCs.txt)

We want to produce:
- a Venn diagram showing the number of DMCs detected by methylKit only, methylSig only or by the two methods
- a list of DMCs belonging to each one of these three sets

get_venn.py accepts several arguments to specify such goal:

Option	Constaint	Comment
--set-file	mandatory (2 or 3 files)	list of input files pathname (one for each entry set)
--set-name	optional (same number as input files, if specified) by default, entry sets are named with the name of the corresponding input file	list of names used to identify entry sets
--key-columns	optional (default: 1,2,3)	Column number used to build the key on which comparison between sets will be performed
--txt-output-file	optional (if not specified no text output file is produced)	pathname to the text file containing results of Venn comparison
--img-output-file	optional (if not specified no image output file is produced)	pathname to the image file containing the Venn diagram. Format of the image is specified by the suffix of this file name. (Supported format : eps, jpeg, jpg, pdf, pgf, png, ps, raw, rgba, svg, svgz, tif, tiff)
--venn-title	optional	Title displayed at the top of Venn diagram image
--keep-columns	optional	List of columns kept from input files and added to text output file

For our example, we have two input files (file paths are relative to RRBS_HOME/Venn) :
1. ../Differential_analysis/analysis_examples/out_methylKit/MethylKit - Condition A vs B - pvalue0.01 - with obvious DMCs.txt
2. ../Differential_analysis/analysis_examples/out_methylSig/MethylSig - Condition A vs B - pvalue0.01 - with obvious DMCs.txt

Corresponding to these two input files, we will associate two set names :
1. methylKit
2. methylSig

Therefore, the beginning of our command should be (we suppose we are in RRBS_HOME/Venn directory) :

```
python get_venn.py \
        --set-file "../Differential_analysis/analysis_examples/out_methylKit/MethylKit - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-file "../Differential_analysis/analysis_examples/out_methylSig/MethylSig - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-name methylKit --set-name methylSig
```

We want to produce a text and an image output file (we also want to have a title for the Venn diagram) :

```
python get_venn.py \
        --set-file "../Differential_analysis/analysis_examples/out_methylKit/MethylKit - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-file "../Differential_analysis/analysis_examples/out_methylSig/MethylSig - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-name methylKit --set-name methylSig \
        --txt-output-file example_output/Venn_methylKit_methylSig.txt \
        --img-output-file example_output/Venn_methylKit_methylSig.jpg \
        --venn-title "methylKit vs methylSig results"
```

We want to compare list of DMCs in the two entry sets. A DMC can be specified by its genomic position using the columns « Chromosome » and « Start ». They are the first two columns of the entry set files :

```
python get_venn.py \
        --set-file "../Differential_analysis/analysis_examples/out_methylKit/MethylKit - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-file "../Differential_analysis/analysis_examples/out_methylSig/MethylSig - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-name methylKit --set-name methylSig \
        --txt-output-file example_output/Venn_methylKit_methylSig.txt \
        --img-output-file example_output/Venn_methylKit_methylSig.jpg \
        --venn-title "methylKit vs methylSig results" \
        --key-columns 1,2
```

Finally, we want to keep the columns « Methyl diff » and « Methylation state …» of each entry set file, therefore we add argument :

```
        --keep-columns "Methyl diff" "Methylation state in Condition_A"
```

At the end, our command looks like this :

```
python get_venn.py \
        --set-file "../Differential_analysis/analysis_examples/out_methylKit/MethylKit - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-file "../Differential_analysis/analysis_examples/out_methylSig/MethylSig - Condition A vs B - pvalue0.01 - with obvious DMCs.txt" \
        --set-name methylKit --set-name methylSig  \
        --txt-output-file example_output/Venn_methylKit_methylSig.txt \
        --img-output-file example_output/Venn_methylKit_methylSig.jpg \
        --venn-title "methylKit vs methylSig results"  \
        --key-columns 1,2 \
        --keep-columns "Methyl diff" "Methylation state in Condition_A"
```

See next slide, an example of output of this command.

Output image :

methylKit vs methylSig results
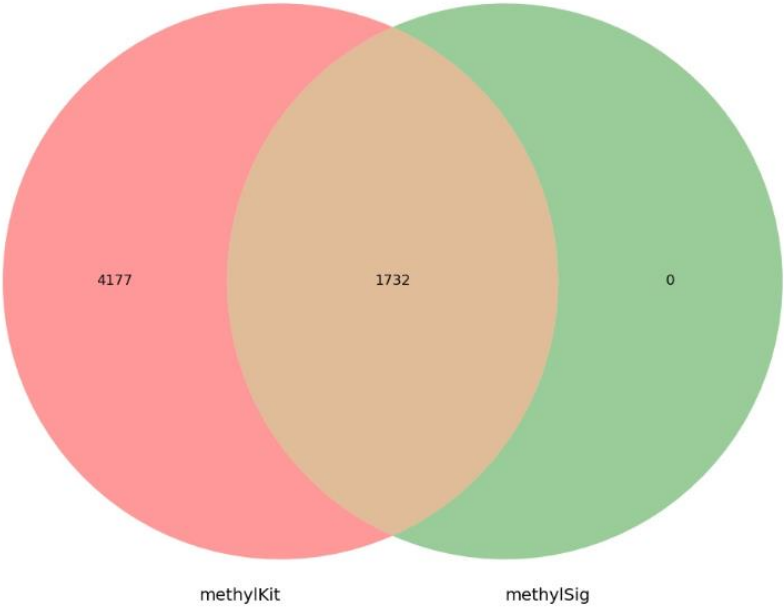
Output text file:

Chromosome	Start	Only in methylKit	Only in methylSig	Common methylKit methylSig	methylKit		methylSig	
					Methyl diff	Methylation state in Condition_A	Methyl diff	Methylation state in Condition_A
1	104146767	*			88.59259259	hypermeth		
1	104146784	*			84.88888889	hypermeth		
1	104146791	*			55.55555556	hypermeth		
1	104146800	*			88.2962963	hypermeth		
1	104146803			*	100	hypermeth	100	hypermeth
1	104271966	*			35.48717949	hypermeth		
1	104576531	*			26.99564586	hypermeth		
1	105020522	*			66.42857143	hypermeth		
1	105020559	*			49.04761905	hypermeth		
1	10520299	*			-95.45454545	hypometh		
1	108085926	*			28.26086957	hypermeth		
1	108237337	*			61.9047619	hypermeth		
1	108377225	*			-50	hypometh		
1	108377239			*	-100	hypometh	-100	hypometh
1	108377253	*			-80	hypometh		
1	109543375	*			-43.76303713	hypometh		