

Differential peak calling in Chip-Seq experiments. Developing a comprehensive test set



Advisor: Oleg Shpynov
JetBrains Research

Student: Chernov Sergey
Academic University



A comprehensive comparison of tools for differential ChIP-seq analysis

Sebastian Steinhauser, Nils Kurzawa, Roland Eils and Carl Herrmann

Corresponding author. Carl Herrmann, IPMB Universität Heidelberg and Department of Theoretical Bioinformatics, DKFZ, Im Neuenheimer Feld 364, D-69120 Heidelberg, Tel.: (+49) 6221 423612. E-mail: carl.herrmann@uni-heidelberg.de

Many different tools have been developed and published in recent years

diffReps
HOMER DiffBind
MultiGPS RSEG
ODIN
DBChip SICER
MACS2
MAnorm MMDiff
PePr QChIPat
ChIPComp

Which is the best one?

Some require detection of enriched
regions
(by external peakcallers)

While others implement their own
method

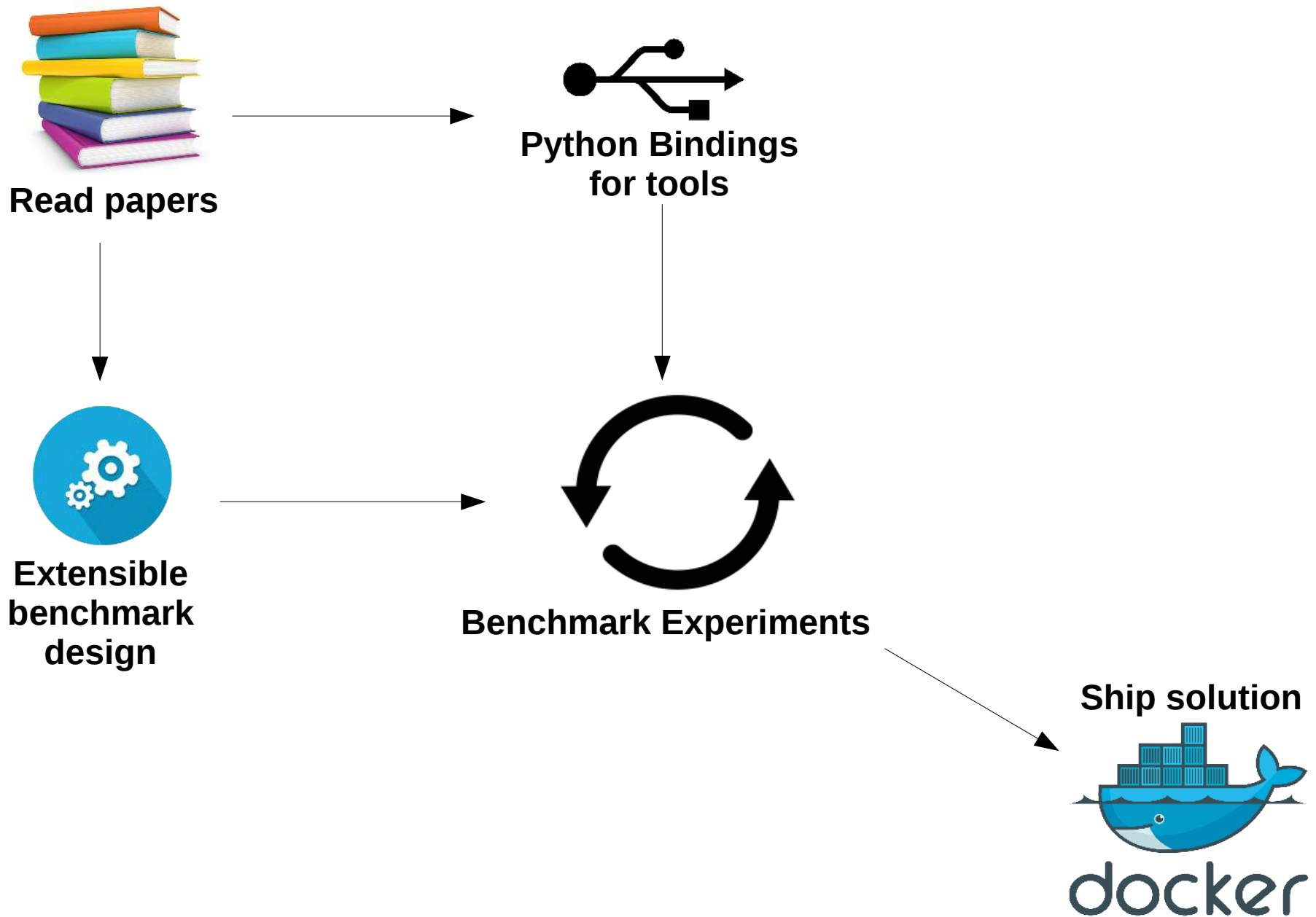
Differ in the underlying
statistical modeling of the signal
distribution

(Poisson vs. Neg. Binomial)

Some work in absence of **replicates**

While others require
ones to provide analysis

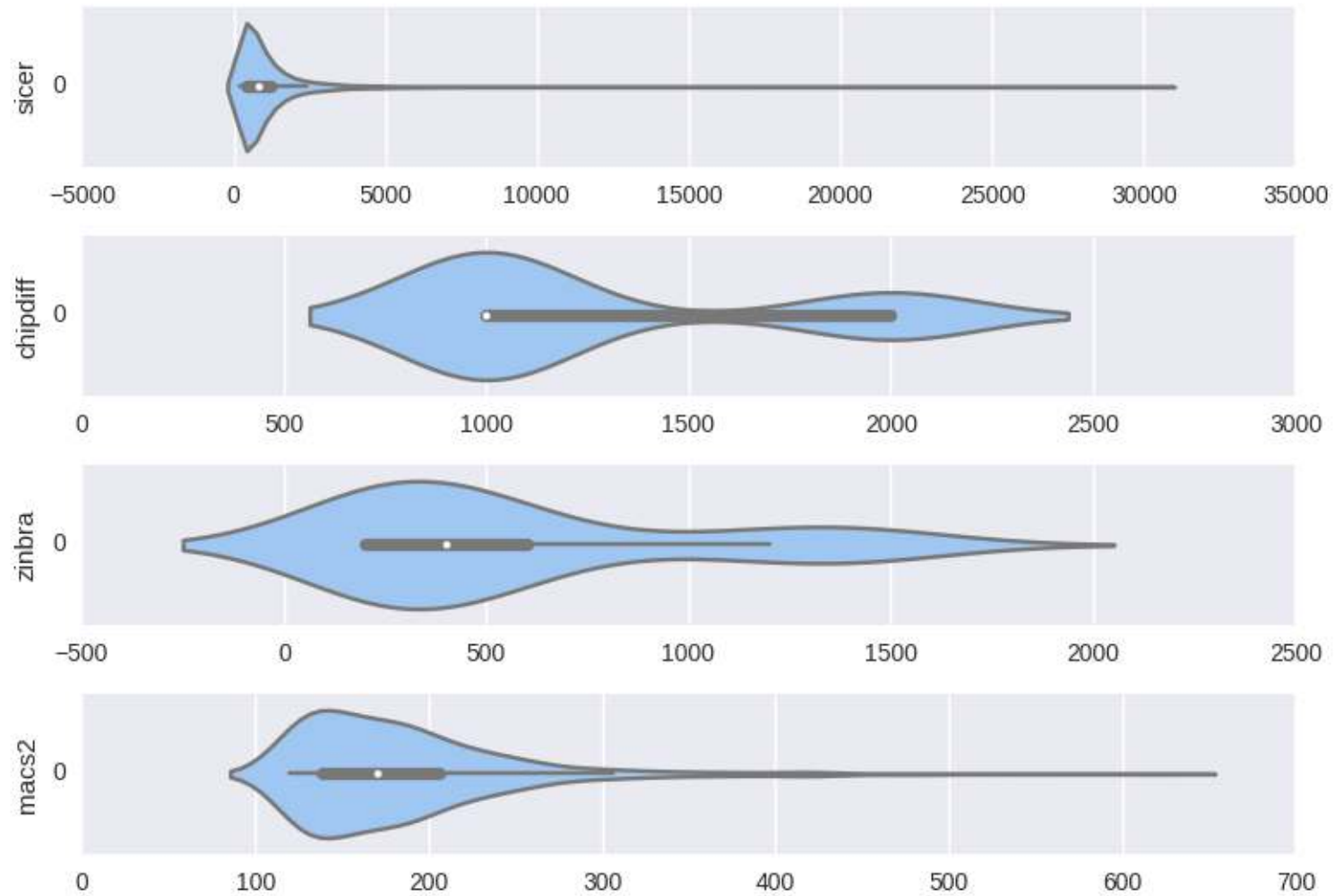
Project Roadmap



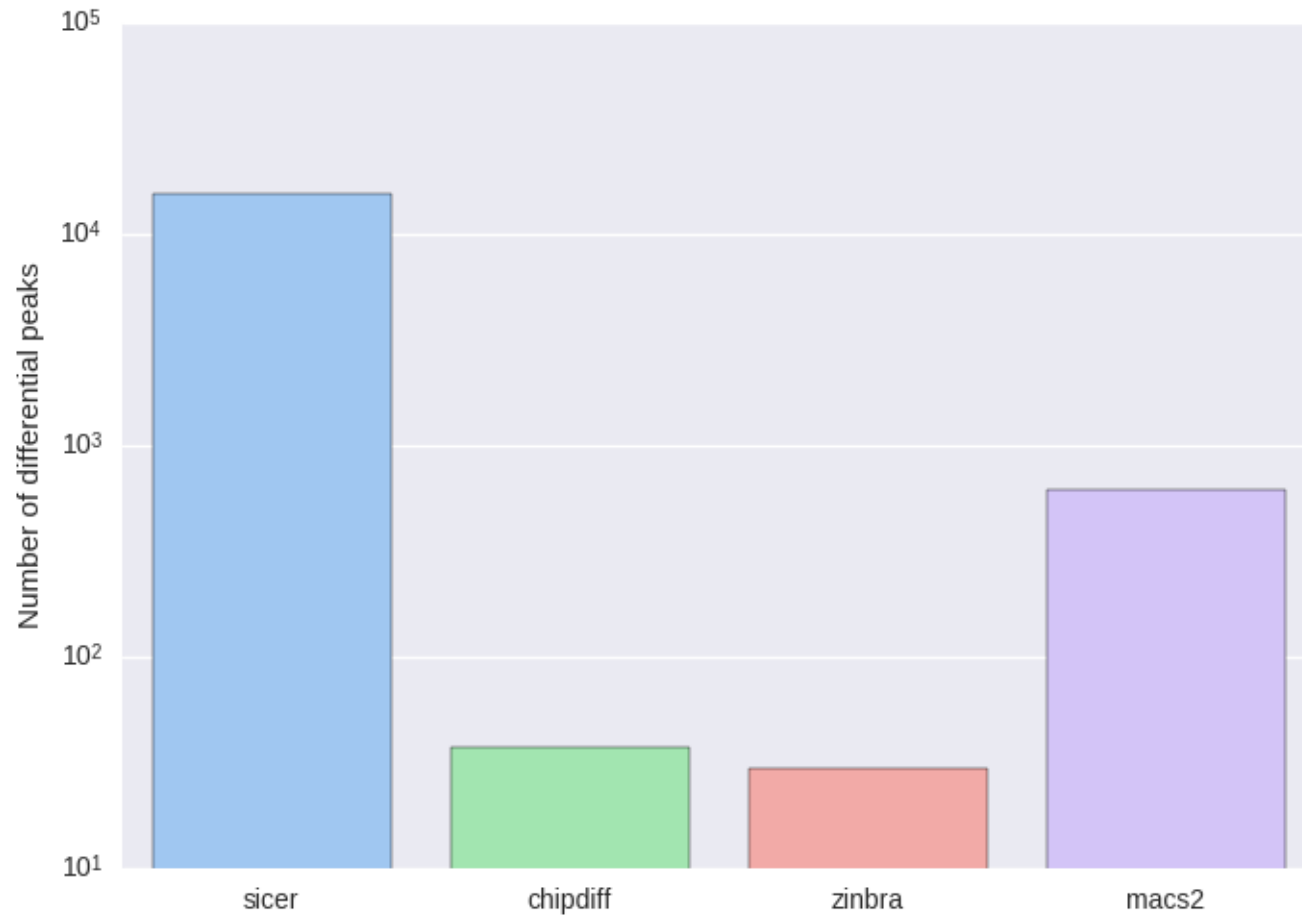
Design expectations

- Easy configurable
- Obvious how to add N+1-th tool for comparison
- Simply extensible benchmark set
- User-friendly «interface» and logging
- Provide HowTo and examples of usage

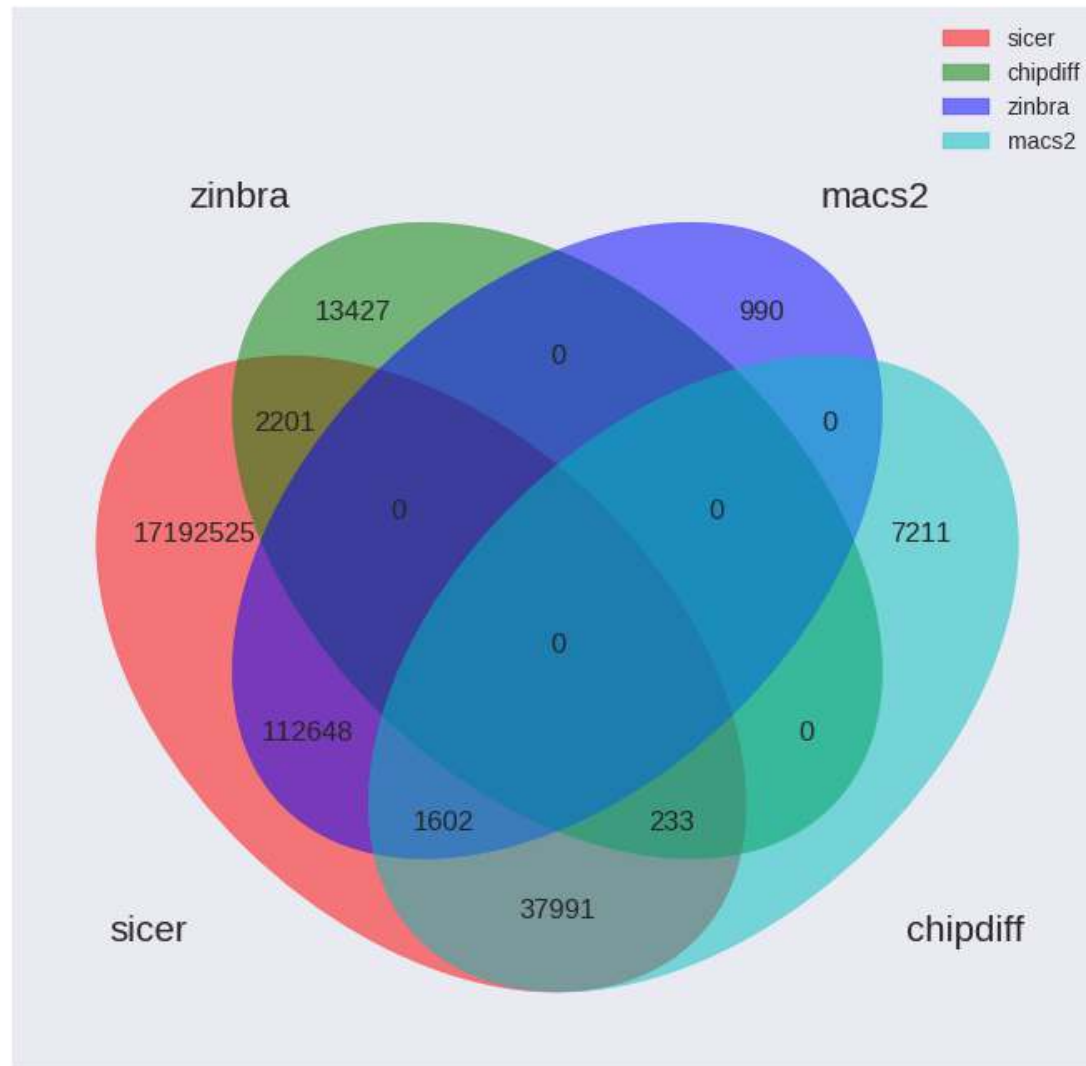
Compare number of peaks & their length distribution



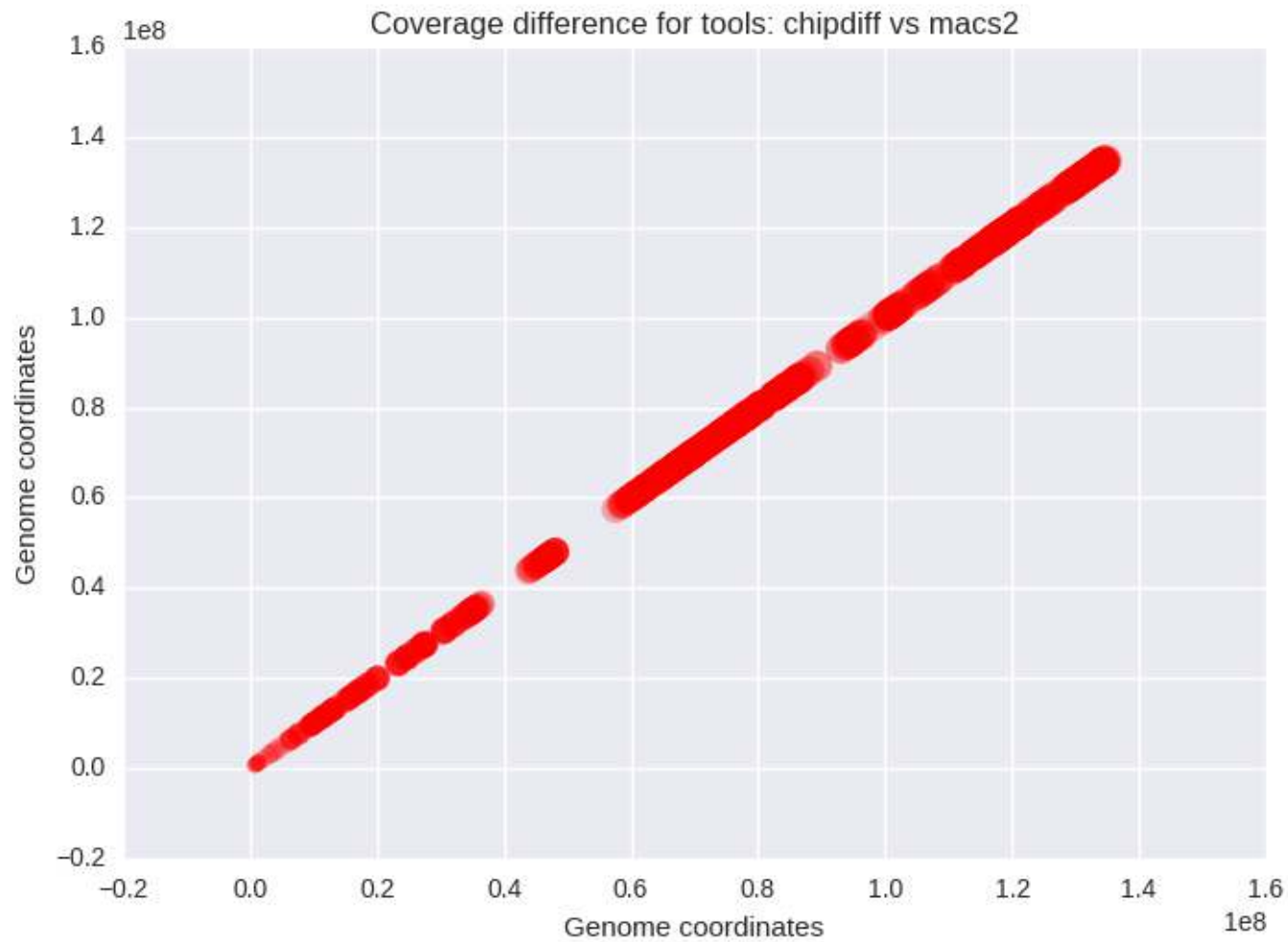
Compare number of peaks & their length distribution



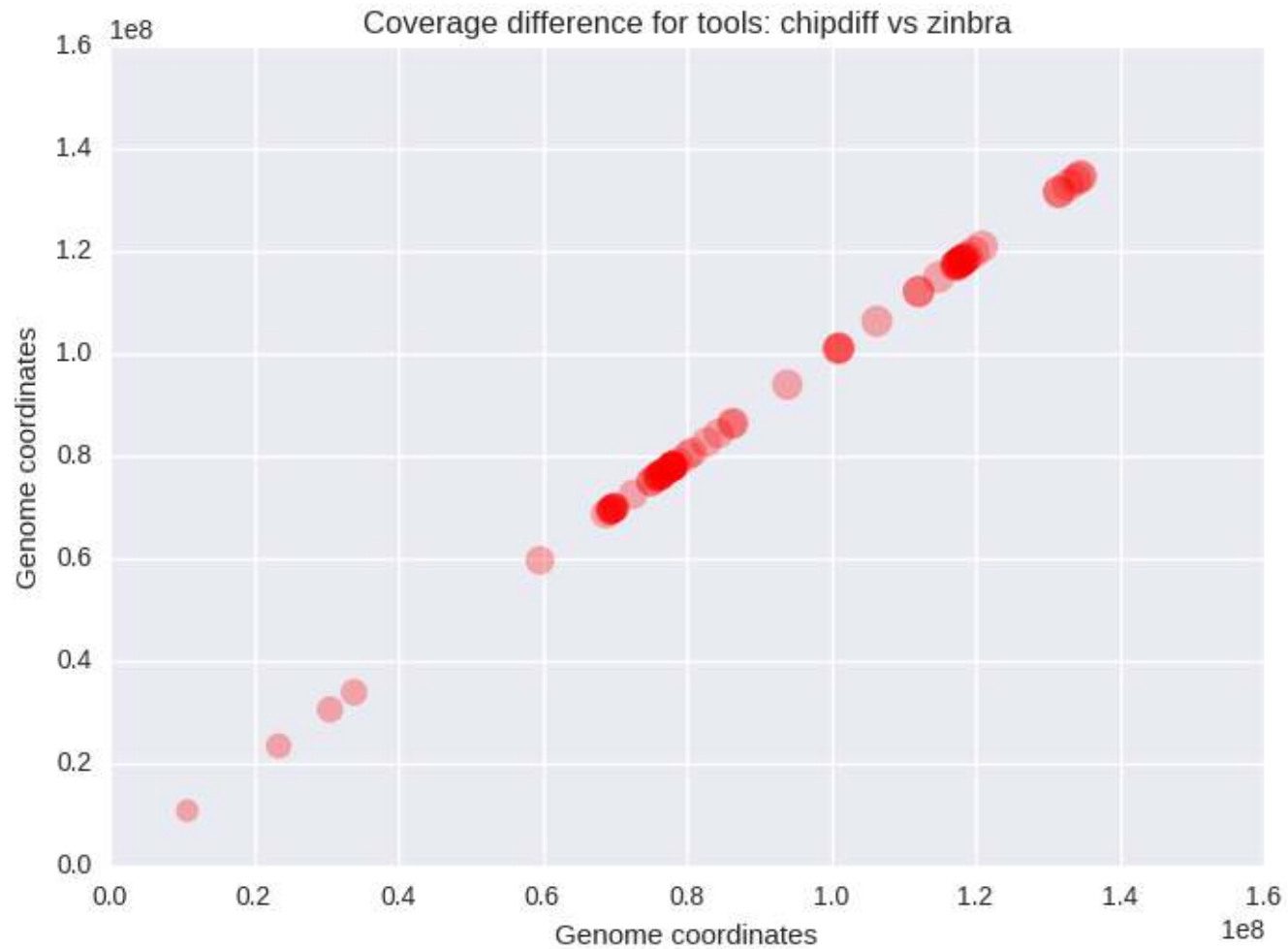
Consistency Test



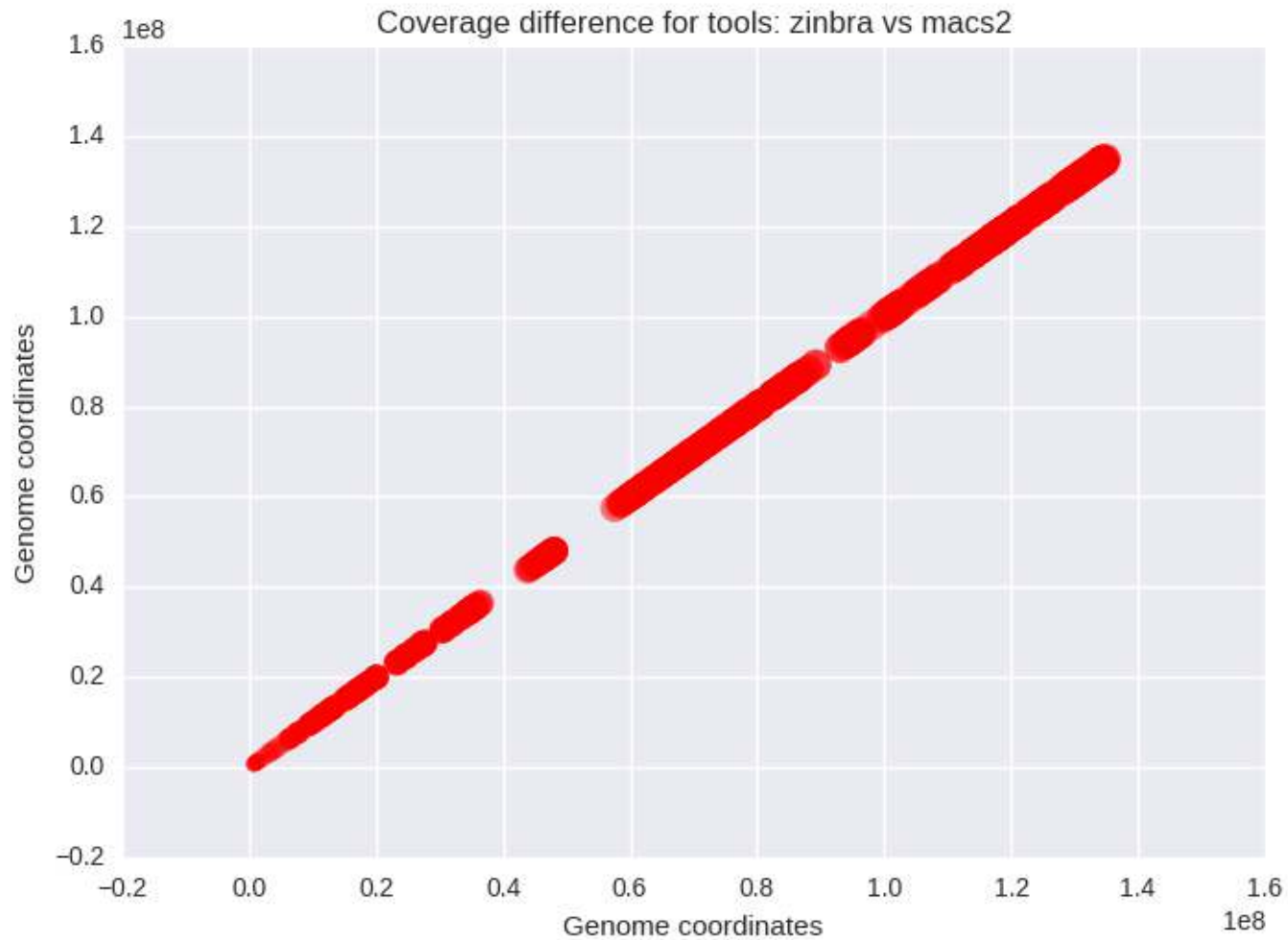
Yet Another Consistency Test



Yet Another Consistency Test



Yet Another Consistency Test



Data configuration(YAML)

15 lines (12 sloc) | 545 Bytes

```
1 # YAML
2
3 humangenome: /home/dirname/hg19.2bit
4
5 condition_first:
6     replicate_first: /home/dirname/condition_one/rep1.bed
7     input_replicate_first: /home/dirname/input1_rep1.bed
8     replicate_second: /home/dirname/condition_one/rep2.bed
9     input_replicate_second: /home/dirname/input1_rep1.bed
10
11 condition_second:
12     replicate_first: /home/dirname/condition_two/rep1.bed
13     input_replicate_first: /home/dirname/input2_rep1.bed
14     replicate_second: /home/dirname/condition_two/rep2.bed
15     input_replicate_second: /home/dirname/input2_rep1.bed
```

Tools configuration(YAML)

30 lines (24 sloc) | 679 Bytes

```
1 # YAML
2
3 # Here just an example of how your tool config file may look like
4
5 sicer:
6   outdir: /home/tools_output/sicer
7   path: /home/sicer/bin
8   projname: diff_cond1_cond2
9
10 manorm:
11   outdir: /home/tools_output/manorm
12   path: /home/manorm/bin
13   projname: diff_cond1_cond2
14
15 macs2:
16   outdir: /home/tools_output/macs2
17   path: /home/macs2/bin
18   projname: diff_cond1_cond2
19
20 chipdiff:
21   outdir: /home/tools_output/chipdiff
22   path: /home/chipdiff/bin
23   chromosomes: /home/denovo/AU/Research/Data/bed/chrom_descr.txt
24   projname: diff_cond1_cond2
25
26 zinbra:
27   outdir: /home/tools_output/zinbra
28   path: /home/zinbra/bin
29   projname: diff_e2_vh
30   only_chr: chr11
```


For Everything Else, there's....

GitHub

[chernovsergey / term2-research](#)

