

chipseq-CTCF-03242016-cp Quality Metrics

Stephen M. Kelly

Dr. Aristotelis Tsirigos

PI: Aris Tsirigos

Genome Technology Center,

NYU Langone Medical Center, New York, NY 10016

stephen.kelly@nyumc.org

April 1, 2016



Analysis Sample Sheet

Stephen M. Kelly

Sample Sheet

Alignment Summary
Stats

Quality Control

PCA

Session Information

	sample	control	group	genome	fragmentation.size
1	CD34-no.treat-CTCF-Zhao2011-rep1		CD34-no.treat-CTCF	hg19	400
2	CD34-no.treat-CTCF-rep2		CD34-no.treat-CTCF	hg19	400
3	CD34-no.treat-Input-Zhao2011-rep1		CD34-no.treat-Input	hg19	400
4	CD34-no.treat-Input-rep2		CD34-no.treat-Input	hg19	400
5	CD4+no.treat-CTCF-rep1		CD4+no.treat-CTCF	hg19	400
6	CD4.naive.T-no.treat-CTCF-rep1		CD4.naive.T-no.treat-CTCF	hg19	400
7	CD4.naive.T-no.treat-Input-rep1		CD4.naive.T-no.treat-Input	hg19	400
8	DND41-no.treat-CTCF-Bernstein2010-rep1		DND41-no.treat-CTCF	hg19	400
9	DND41-no.treat-CTCF-Bernstein2010-rep2		DND41-no.treat-CTCF	hg19	400
10	DND41-no.treat-Input-Bernstein2010-rep1		DND41-no.treat-Input	hg19	400
11	DND41-no.treat-Input-Bernstein2010-rep2		DND41-no.treat-Input	hg19	400
12	H1-no.treat-CTCF-Bernstein-rep1		H1-no.treat-CTCF	hg19	400
13	H1-no.treat-CTCF-Bernstein-rep2		H1-no.treat-CTCF	hg19	400
14	H1-no.treat-Input-Bernstein-rep1		H1-no.treat-Input	hg19	400
15	H1-no.treat-Input-Bernstein-rep2		H1-no.treat-Input	hg19	400
16	HEK293T-no.treat-CTCF-rep1		HEK293T-no.treat-CTCF	hg19	400
17	HEK293T-no.treat-CTCF-rep2		HEK293T-no.treat-CTCF	hg19	400
18	HEK293T-no.treat-Input-rep1		HEK293T-no.treat-Input	hg19	400
19	IMR90-no.treat-CTCF-Snyder2012-rep1		IMR90-no.treat-CTCF	hg19	400
20	IMR90-no.treat-CTCF-Snyder2012-rep2		IMR90-no.treat-CTCF	hg19	400
21	IMR90-no.treat-Input-Snyder2012-rep1		IMR90-no.treat-Input	hg19	400
22	IMR90-no.treat-Input-Snyder2012-rep2		IMR90-no.treat-Input	hg19	400
23	Jurkat-no.treat-CTCF-rep1		Jurkat-no.treat-CTCF	hg19	400
24	Jurkat-no.treat-Input-rep1		Jurkat-no.treat-Input	hg19	400
25	K562-no.treat-CTCF-Stam-rep1		K562-no.treat-CTCF	hg19	400
26	K562-no.treat-CTCF-Stam-rep2		K562-no.treat-CTCF	hg19	400
27	K562-no.treat-CTCF-rep1		K562-no.treat-CTCF	hg19	400
28	K562-no.treat-Input-Stam-rep1		K562-no.treat-Input	hg19	400
29	K562-no.treat-Input-rep1		K562-no.treat-Input	hg19	400
30	THP1-no.treat-CTCF-rep1		THP1-no.treat-CTCF	hg19	400
31	THP1-no.treat-CTCF-rep2		THP1-no.treat-CTCF	hg19	400
32	THP1-no.treat-Input-rep1		THP1-no.treat-Input	hg19	400
33	THP1-no.treat-Input-rep2		THP1-no.treat-Input	hg19	400

Default Pipeline Parameters

Sample Sheet

Alignment Summary Stats

Quality Control

PCA

Session Information

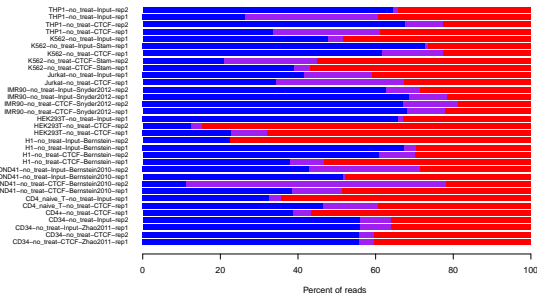
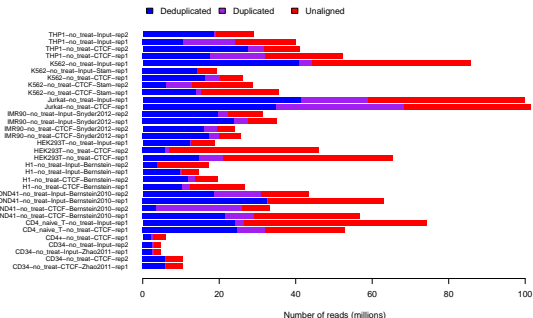
```
#!/bin/tcsh
# load all tools
module unload gcc
module unload samtools
module unload java
module unload r
module load samtools/1.2.1
module load bedtools/2.22.0
module load java/1.7
module load picard-tools
module load bowtie2
#module load macs/2.0.10.20131216
module load r/3.2.0
# set sample sheet
set sheet = inputs/sample-sheet.tsv
```

Alignment Summary Stats

[Sample Sheet](#)[Alignment Summary
Stats](#)[all-samples](#)[Quality Control](#)[PCA](#)[Session Information](#)

Create barplots to show the number of reads that are aligned versus unaligned.

Sequencing Reads



Check the samples for quality control analysis, produce fingerprint plots.

Stephen M. Kelly

Sample Sheet

Alignment Summary
Stats

Quality Control

CD34-no.treat-CTCF

CD34-no.treat-Input

CD4+-no.treat-
CTCFCD4_naive_T-
no.treat-CTCFCD4_naive_T-
no.treat-InputDND41-no.treat-
CTCFDND41-no.treat-
Input

H1-no.treat-CTCF

H1-no.treat-Input

HEK293T-no.treat-
CTCFHEK293T-no.treat-
InputIMR90-no.treat-
CTCFIMR90-no.treat-
InputJurkat-no.treat-
CTCF

Jurkat-no.treat-Input

K562-no.treat-CTCF

K562-no.treat-Input

THP1-no.treat-
CTCF

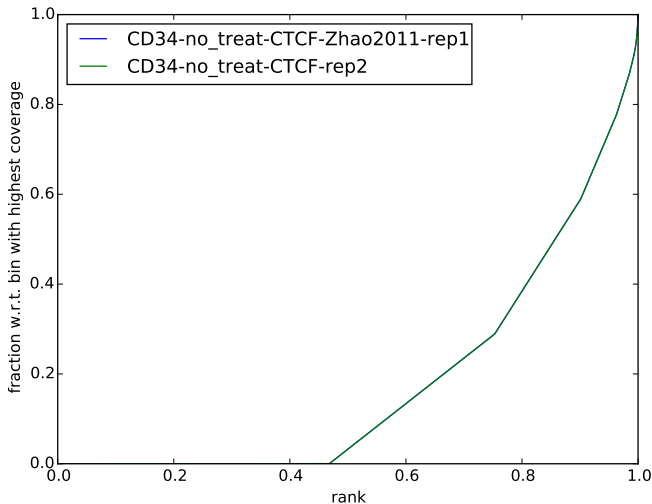
THP1-no.treat-Input

PCA

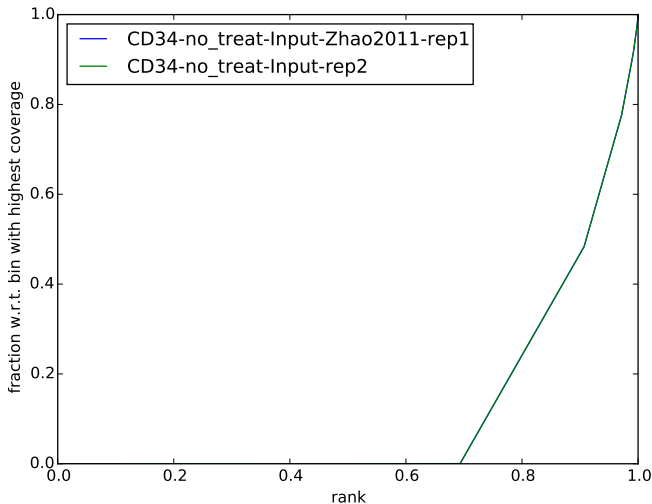
Session Information

```
params.default.tcsh
#!/bin/tcsh
source ./inputs/params/params.tcsh
module unload gcc
module load python/2.7
```

results/fingerprint.by_group.default/align.by_sample.bowtie2



results/fingerprint.by_group.default/align.by_sample.bowtie2



Sample Sheet

Alignment Summary
Stats

Quality Control

CD34-no_treat-CTCF

CD34-no_treat-Input

CD4+-no_treat-
CTCFCD4_naive_T-
no_treat-CTCFCD4_naive_T-
no_treat-InputDND41-no_treat-
CTCFDND41-no_treat-
Input

H1-no_treat-CTCF

H1-no_treat-Input

HEK293T-no_treat-
CTCFHEK293T-no_treat-
InputIMR90-no_treat-
CTCFIMR90-no_treat-
InputJurkat-no_treat-
CTCF

Jurkat-no_treat-Input

K562-no_treat-CTCF

K562-no_treat-Input

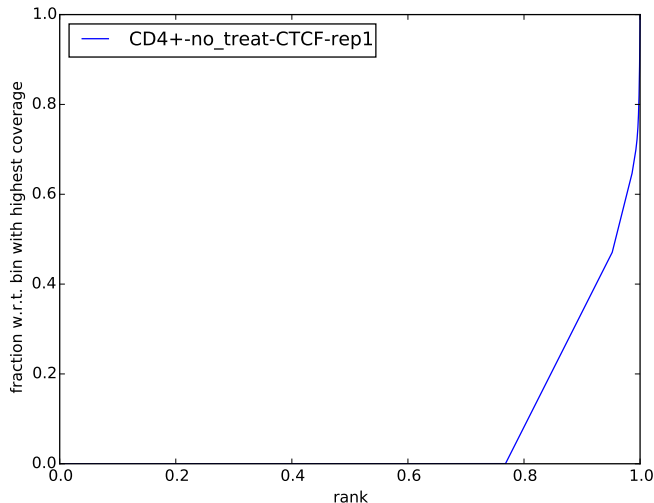
THP1-no_treat-
CTCF

THP1-no_treat-Input

PCA

Session Information

results/fingerprint.by_group.default/align.by_sample.bowtie2



Sample Sheet

Alignment Summary
Stats

Quality Control

CD34-no.treat-CTCF

CD34-no.treat-Input

CD4+-no.treat-CTCF

CD4_naive_T-no.treat-CTCF

CD4_naive_T-no.treat-Input

DND41-no.treat-CTCF

DND41-no.treat-Input

H1-no.treat-CTCF

H1-no.treat-Input

HEK293T-no.treat-CTCF

HEK293T-no.treat-Input

IMR90-no.treat-CTCF

IMR90-no.treat-Input

Jurkat-no.treat-CTCF

Jurkat-no.treat-Input

K562-no.treat-CTCF

K562-no.treat-Input

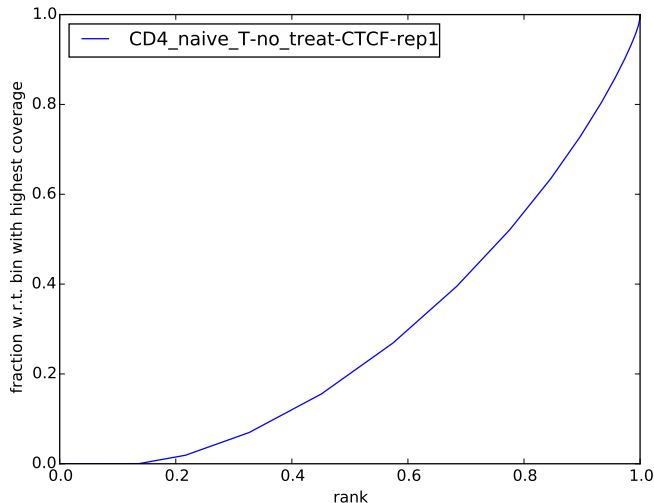
THP1-no.treat-CTCF

THP1-no.treat-Input

PCA

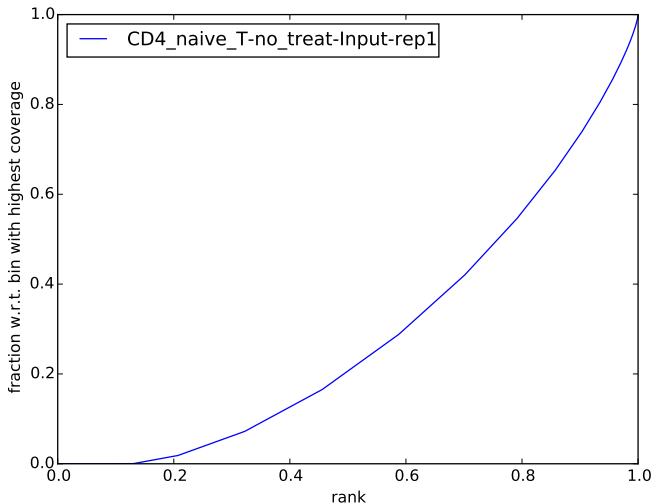
Session Information

results/fingerprint.by_group.default/align.by_sample.bowtie2

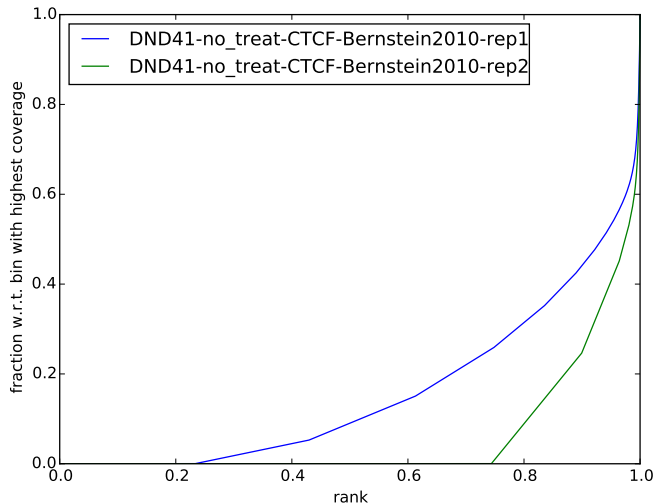


[Sample Sheet](#)[Alignment Summary](#)
[Stats](#)[Quality Control](#)[CD34-no.treat-CTCF](#)[CD34-no.treat-Input](#)[CD4+-no.treat-CTCF](#)[CD4_naive_T-no.treat-CTCF](#)[CD4_naive_T-no.treat-Input](#)[DND41-no.treat-CTCF](#)[DND41-no.treat-Input](#)[H1-no.treat-CTCF](#)[H1-no.treat-Input](#)[HEK293T-no.treat-CTCF](#)[HEK293T-no.treat-Input](#)[IMR90-no.treat-CTCF](#)[IMR90-no.treat-Input](#)[Jurkat-no.treat-CTCF](#)[Jurkat-no.treat-Input](#)[K562-no.treat-CTCF](#)[K562-no.treat-Input](#)[THP1-no.treat-CTCF](#)[THP1-no.treat-Input](#)[PCA](#)[Session Information](#)

results/fingerprint.by_group.default/align.by_sample.bowtie2



results/fingerprint.by_group.default/align.by_sample.bowtie2



Sample Sheet

Alignment Summary
Stats

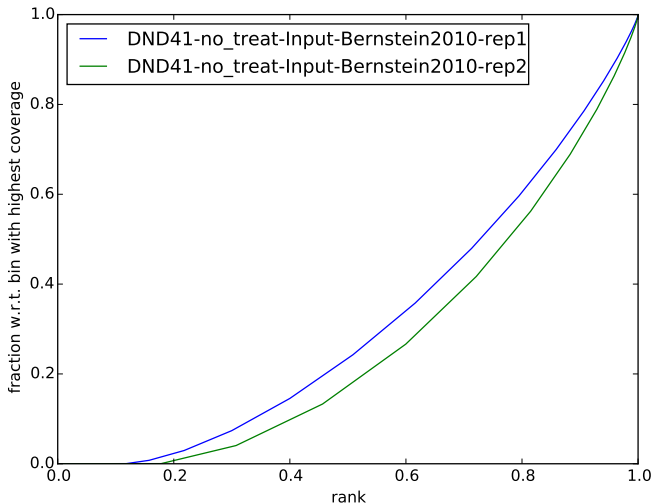
Quality Control

CD34-no_treat-CTCF
CD34-no_treat-Input
CD4+-no_treat-CTCF
CD4_naive_T-no_treat-CTCF
CD4_naive_T-no_treat-Input
DND41-no_treat-CTCF
DND41-no_treat-Input
H1-no_treat-CTCF
H1-no_treat-Input
HEK293T-no_treat-CTCF
HEK293T-no_treat-Input
IMR90-no_treat-CTCF
IMR90-no_treat-Input
Jurkat-no_treat-CTCF
Jurkat-no_treat-Input
K562-no_treat-CTCF
K562-no_treat-Input
THP1-no_treat-CTCF
THP1-no_treat-Input

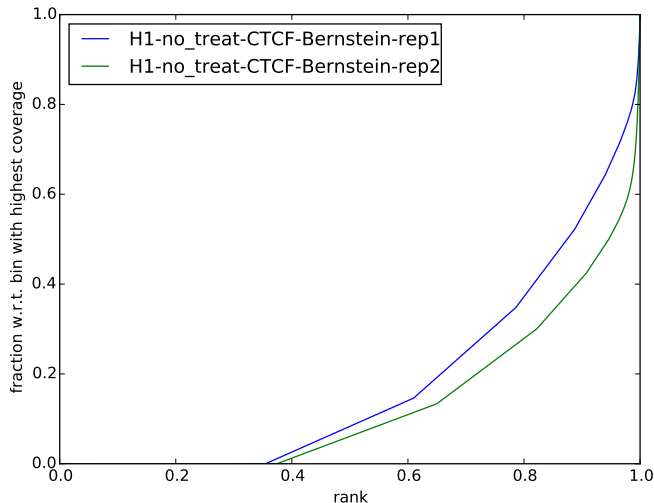
PCA

Session Information

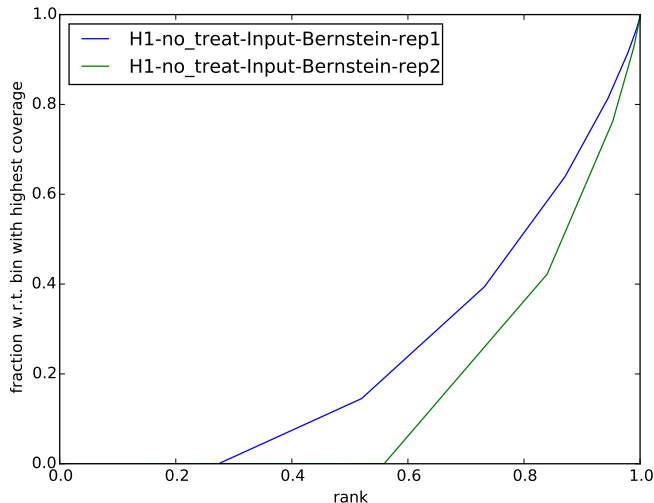
results/fingerprint.by_group.default/align.by_sample.bowtie2



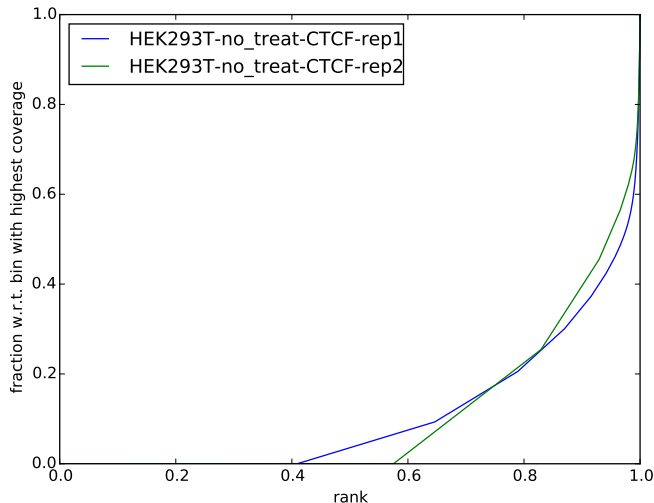
results/fingerprint.by_group.default/align.by_sample.bowtie2



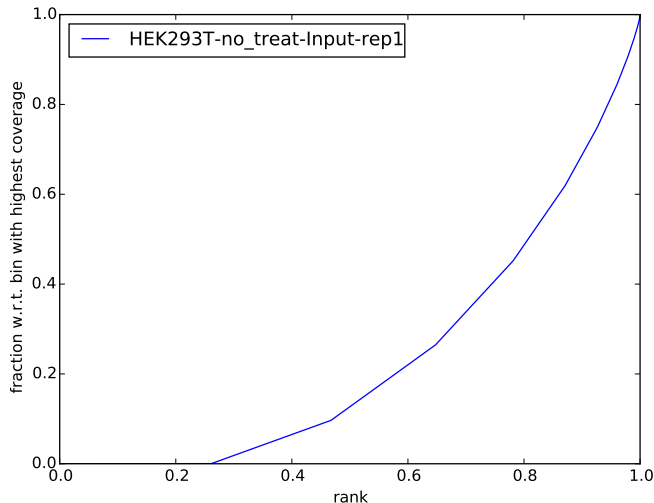
results/fingerprint.by_group.default/align.by_sample.bowtie2



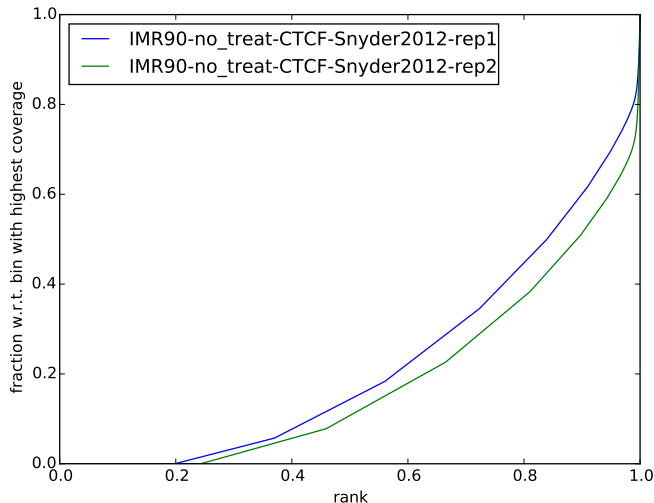
results/fingerprint.by_group.default/align.by_sample.bowtie2



results/fingerprint.by_group.default/align.by_sample.bowtie2



results/fingerprint.by_group.default/align.by_sample.bowtie2



Sample Sheet

Alignment Summary
Stats

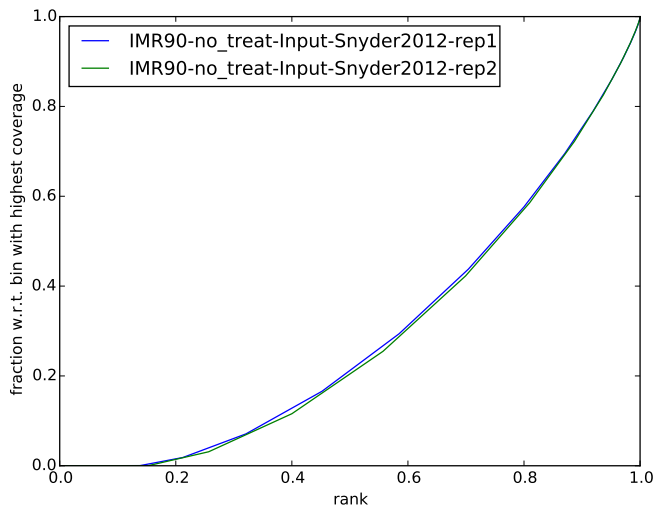
Quality Control

CD34-no_treat-CTCF
CD34-no_treat-Input
CD4+-no_treat-CTCF
CD4_naive_T-no_treat-CTCF
CD4_naive_T-no_treat-Input
DND41-no_treat-CTCF
DND41-no_treat-Input
H1-no_treat-CTCF
H1-no_treat-Input
HEK293T-no_treat-CTCF
HEK293T-no_treat-Input
IMR90-no_treat-CTCF
IMR90-no_treat-Input
Jurkat-no_treat-CTCF
Jurkat-no_treat-Input
K562-no_treat-CTCF
K562-no_treat-Input
THP1-no_treat-CTCF
THP1-no_treat-Input

PCA

Session Information

results/fingerprint.by_group.default/align.by_sample.bowtie2



CD34-no_treat-CTCF

CD34-no_treat-Input

CD4+-no_treat-CTCF

CD4_naive_T-no_treat-CTCF

CD4_naive_T-no_treat-Input

DND41-no_treat-CTCF

DND41-no_treat-Input

H1-no_treat-CTCF

H1-no_treat-Input

HEK293T-no_treat-CTCF

HEK293T-no_treat-Input

IMR90-no_treat-CTCF

IMR90-no_treat-Input

Jurkat-no_treat-CTCF

Jurkat-no_treat-Input

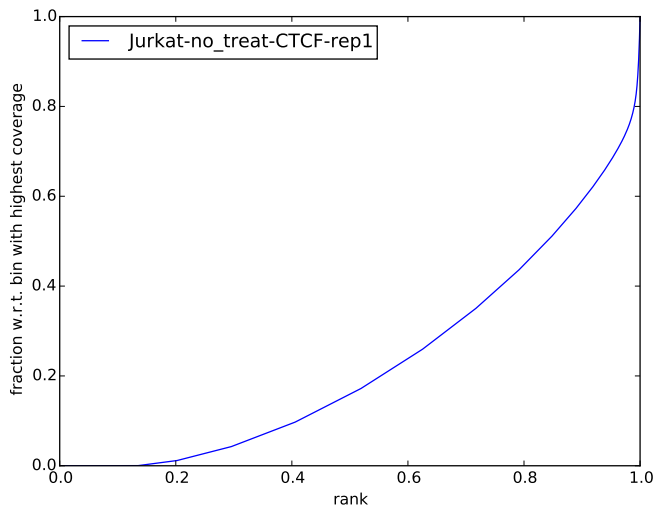
K562-no_treat-CTCF

K562-no_treat-Input

THP1-no_treat-CTCF

THP1-no_treat-Input

results/fingerprint.by_group.default/align.by_sample.bowtie2



Sample Sheet

Alignment Summary
Stats

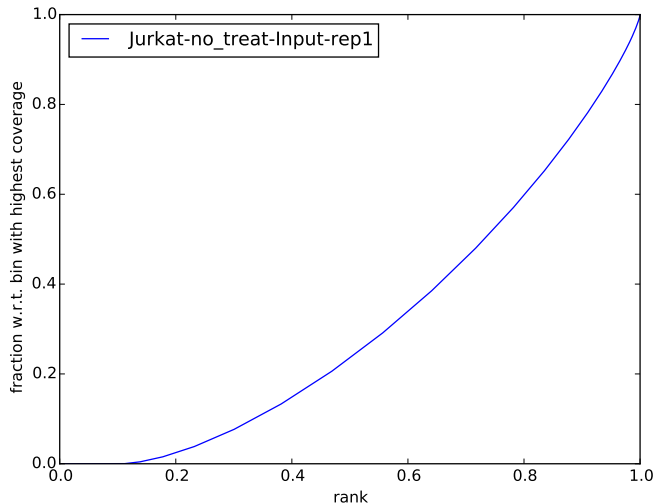
Quality Control

CD34-no_treat-CTCF
CD34-no_treat-Input
CD4+-no_treat-CTCF
CD4_naive_T-no_treat-CTCF
CD4_naive_T-no_treat-Input
DND41-no_treat-CTCF
DND41-no_treat-Input
H1-no_treat-CTCF
H1-no_treat-Input
HEK293T-no_treat-CTCF
HEK293T-no_treat-Input
IMR90-no_treat-CTCF
IMR90-no_treat-Input
Jurkat-no_treat-CTCF
Jurkat-no_treat-Input
K562-no_treat-CTCF
K562-no_treat-Input
THP1-no_treat-CTCF
THP1-no_treat-Input

PCA

Session Information

results/fingerprint.by_group.default/align.by_sample.bowtie2



K562-no_treat-CTCF

Sample Sheet

Alignment Summary
Stats

Quality Control

CD34-no_treat-CTCF

CD34-no_treat-Input

CD4+-no_treat-CTCF

CD4_naive_T-no_treat-CTCF

CD4_naive_T-no_treat-Input

DND41-no_treat-CTCF

DND41-no_treat-Input

H1-no_treat-CTCF

H1-no_treat-Input

HEK293T-no_treat-CTCF

HEK293T-no_treat-Input

IMR90-no_treat-CTCF

IMR90-no_treat-Input

Jurkat-no_treat-CTCF

Jurkat-no_treat-Input

K562-no_treat-CTCF

K562-no_treat-Input

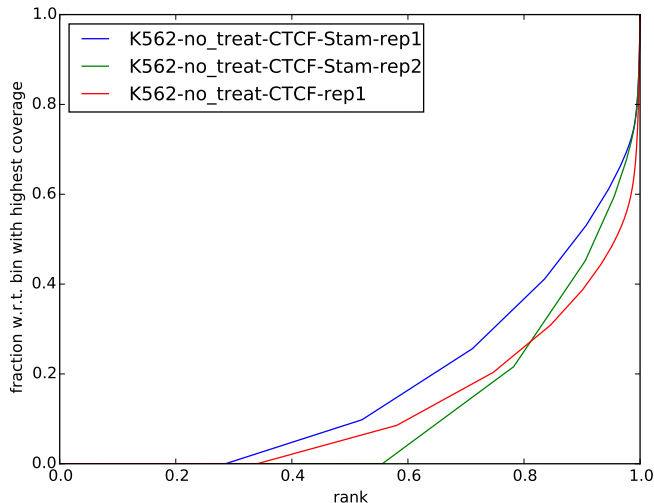
THP1-no_treat-CTCF

THP1-no_treat-Input

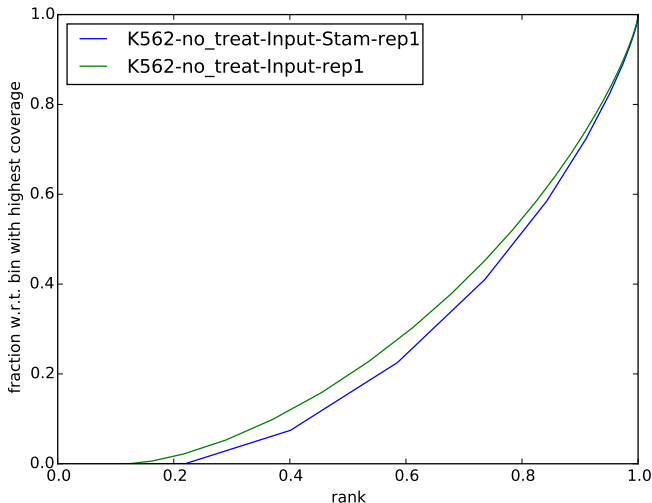
PCA

Session Information

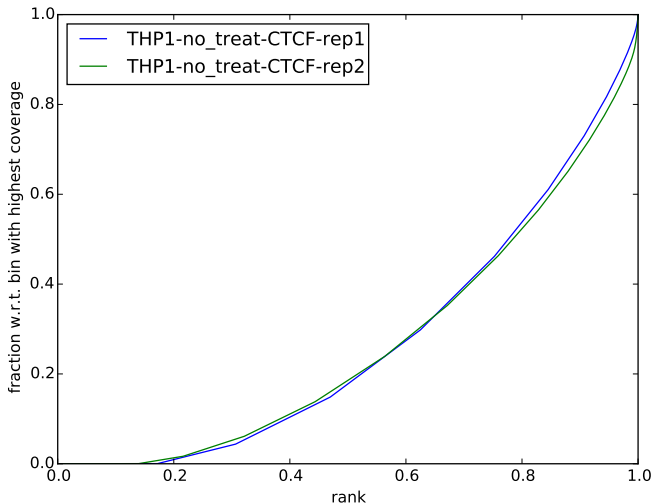
results/fingerprint.by_group.default/align.by_sample.bowtie2



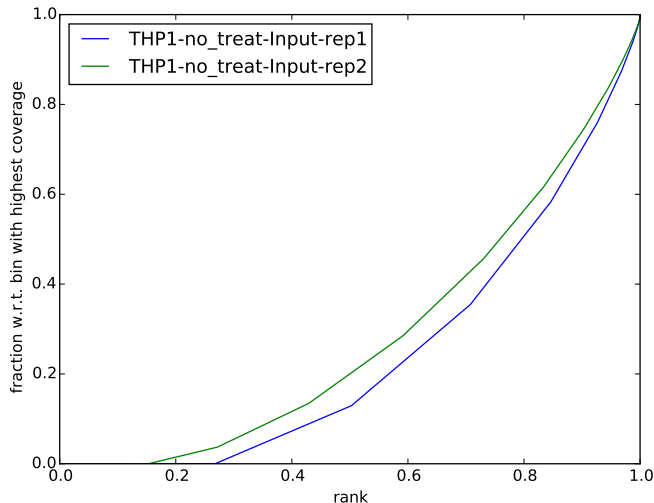
results/fingerprint.by_group.default/align.by_sample.bowtie2



results/fingerprint.by_group.default/align.by_sample.bowtie2



results/fingerprint.by_group.default/align.by_sample.bowtie2



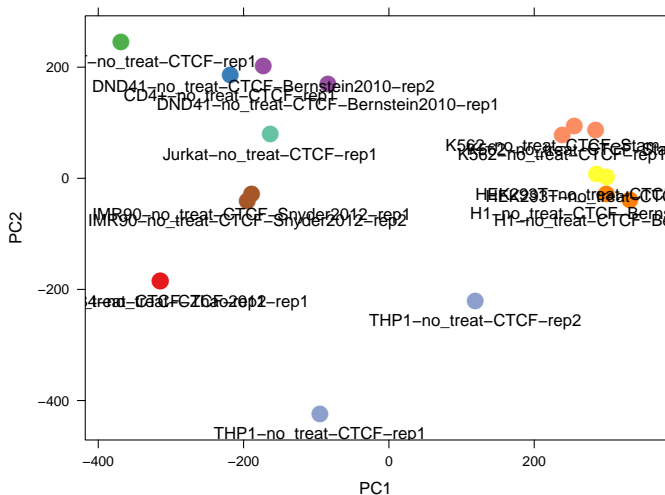
[Sample Sheet](#)[Alignment Summary
Stats](#)[Quality Control](#)[PCA](#)[all-samples](#)[all-samples](#)[all-samples](#)[all-samples](#)[Session Information](#)

This step of the pipeline performs PCA.

PCA Parameters

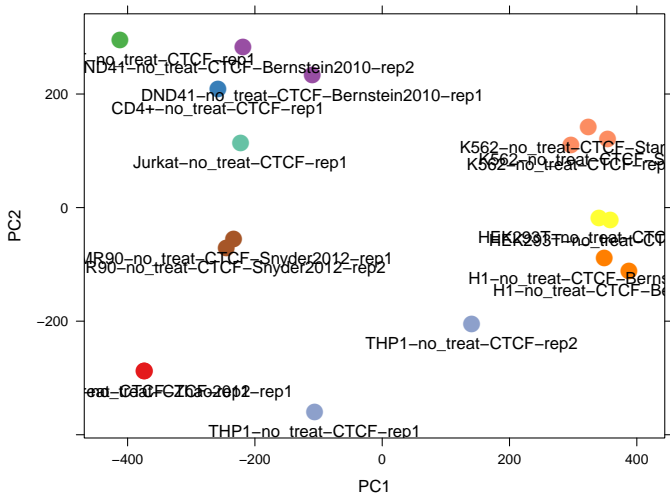
[Sample Sheet](#)[Alignment Summary
Stats](#)[Quality Control](#)[PCA](#)[all-samples](#)[all-samples](#)[all-samples](#)[all-samples](#)[Session Information](#)

```
params.standard.tcsh
#!/bin/tcsh
source ./inputs/params/params.tcsh
set group = 'group'
set include_input = 'false'
set chrom_excluded = 'chr[MXY]' # excluded chromosomes
```



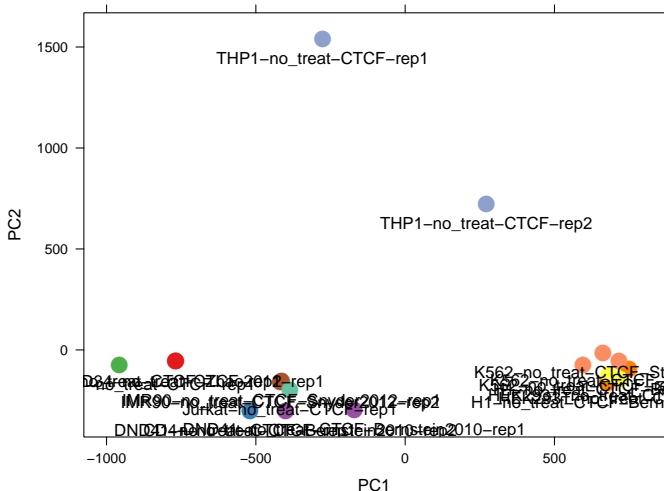
all-samples

results/pca.standard/matrices.by_sample.flank_500nt.nbins.1/peaks.by_sample.macs_narrow/align.by_sample.bowtie2

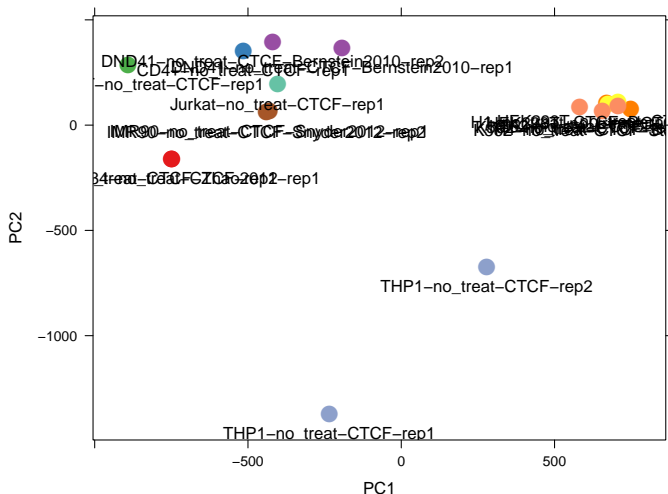


all-samples

results/pca.standard/matrices.by_sample.win_1kb.nbins.1/peaks.by_sample.macs.broad/align.by_sample.bowtie2



results/pca.standard/matrices.by_sample.win_1kb.nbins.1/peaks.by_sample.macs_narrow/align.by_sample.bowtie2




```

system('uname -srv',intern=T)

## [1] "Linux 2.6.32-573.3.1.el6.x86_64 #1 SMP Thu Aug 13 22:55:16 UTC 2015"

sessionInfo()

## R version 3.2.3 (2015-12-10)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: CentOS release 6.3 (Final)
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  base
##
## other attached packages:
## [1] xtable_1.8-2      Hmisc_3.17-2      ggplot2_2.1.0     Formula_1.2-1
## [5] survival_2.38-3   lattice_0.20-33   knitr_1.12.22
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.4        cluster_2.0.3      magrittr_1.5
##  [4] splines_3.2.3      munsell_0.4.3      colorspace_1.2-6
##  [7] stringr_1.0.0      highr_0.5.3        plyr_1.8.3
## [10] tools_3.2.3        nnet_7.3-12        grid_3.2.3
## [13] gtable_0.2.0       latticeExtra_0.6-28 gridExtra_2.2.1
## [16] RColorBrewer_1.1-2 formatR_1.3.2       acepack_1.3-3.3
## [19] rpart_4.1-10       evaluate_0.8.3     stringi_1.0-1
## [22] methods_3.2.3      scales_0.4.0       foreign_0.8-66

save.image()

```