Stephen M. Kelly

Sample Sheet

Alignment Summary Stats

Quality Control

\_ \_ .

Session Information

# chipseq-CTCF-03242016-cp Quality Metrics

Stephen M. Kelly

Dr. Aristotelis Tsirigos Pl: Aris Tsirigos Genome Technology Center, NYU Langone Medical Center, New York, NY 10016

> stephen.kelly@nyumc.org April 1, 2016



# **Analysis Sample Sheet**

### Sample Sheet

	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 Parameters

Sample Sheet

Alignment Summar

Quality Control

Session Informati

#!/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

Stephen M. Kelly

ample Shee

Alignment Summary Stats

all-samples

wality Contro

DCA

Session Informatio



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

Stephen M. Kelly

# all-samples

Sample Sheet

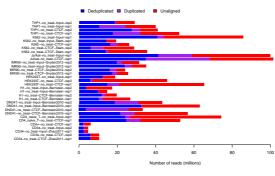
Alignment Summary Stats

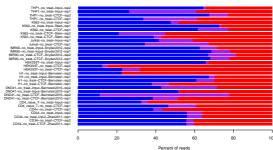
all-samples

Session Information

### results/align-stats.standard/align.by\_sample.bowtie2

### Sequencing Reads





Stephen M. Kelly

----I- CL--

Alignment Summary

Quality Control

CD34-no\_treat-CTCI
CD34-no\_treat-Input
CD4--no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-cput

CTCF

DND41-no\_tre

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat

IMR90-no treat-

IMPOO -- +---

Input

CTCF

VE62 no troot CTCI

K562-no\_treat-Input

THP1-no treat-Inp

PCA

ession Informatio

# **Quality Control**

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

# **Quality Control Parameters**

Sample Sheet

Alignment Summar

### Quality Control

CD34-no\_treat-CTCF CD34-no\_treat-Input CD4+-no\_treat-CTCF CD4.naive\_Tno\_treat-CTCF CD4.naive\_Tno\_treat-Input

CTCF DND41-no\_trea

Input

H1-no\_treat-Input HEK293T-no\_trea

HEK293T-no\_treat-

IMR90-no\_treat-

IMR90-no\_trea

Input

CTCF

K562-no\_treat-CTC

THP1-no\_treat-

THP1-no\_treat-I

**PCA** 

ession Informatio

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

#### .i CF\_Cnir-Seq

Stephen M. Kelly

Sample Sheet

Alignment Summar

#### Quality Control

CD34-no\_treat-CTCF
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treat-

DND41-no..tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat

IMR90-no\_treat-

IMR90-no trea

Input

CTCF
Jurkat-no\_treat-li

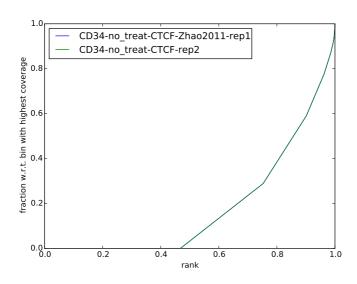
K562-no\_treat-Inpo

HP1-no\_treat-inp

#### PCA

ession informatio

## CD34-no\_treat-CTCF



# CD34-no\_treat-Input

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTCF
CD34-no\_treat-Input
CD4+-no\_treat-CTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treat-CTCF

DND41-no\_tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat-

IMR90-no\_treat-

IMR90-no\_tre

Input

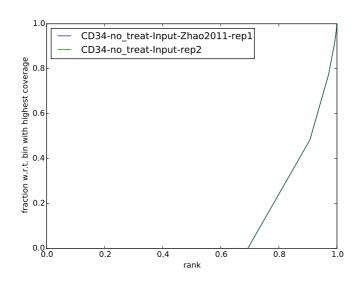
CTCF
Jurkat-no\_treat-li
K562-no\_treat-C

K562-no\_treat-Inp

i HP1-no\_treat-inp

#### PCA

ession Informatio



# CD4+-no\_treat-CTCF

Stephen M. Kelly

Sample Sheet

Alignment Summary Stats

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4-naive\_T-

no\_treat-CTCI
CD4\_naive\_Tno\_treat-Input
DND41-no\_tre
CTCF

DND41-no..tre

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat-

IMR90-no\_treat

IMR90-no\_tre

Input

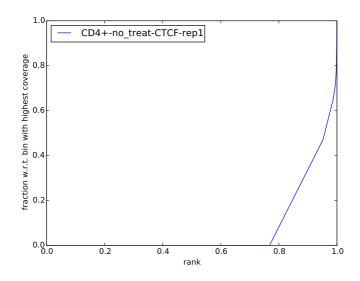
Jurkat-no\_treat-CTCF Jurkat-no\_treat-li

K562-no\_treat-Inpi THP1-no\_treat-

THP1-no\_treat-Inp

#### PCA

ession Informatio



### Stephen M. Kelly

Sample Sheet

Alignment Summary Stats

#### Quality Control

CD34-no\_treat-CTCl CD34-no\_treat-Input CD4+-no\_treat-CTCF CD4\_naive\_T-

no\_treat-CTCF
CD4\_naive\_Tno\_treat-Input

CTCF DND41-no tre

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_trea

HEK293T-no\_treat-

IMR90-no\_treat

IMR90-no\_tre

Jurkat-no\_treat-

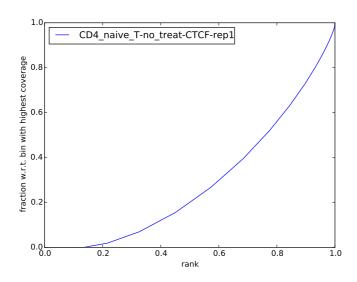
Jurkat-no\_treat-Inpu K562-no\_treat-CTCl K562-no\_treat-Input THP1-no\_treat-

THP1-no\_treat-Inp

#### PCA

ession Informatior

# CD4\_naive\_T-no\_treat-CTCF



# CD4\_naive\_T-no\_treat-Input

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTC CD34-no\_treat-Input CD4+-no\_treat-CTCF CD4\_naive\_Tno\_treat-CTCF CD4\_naive\_Tno\_treat-Input DND41-no\_treat-CTCF

DND41-no\_tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_trea

HEK293T-no\_treat-

IMR90-no\_trea

IMR90-no\_tre

Input Jurkat-no\_tre

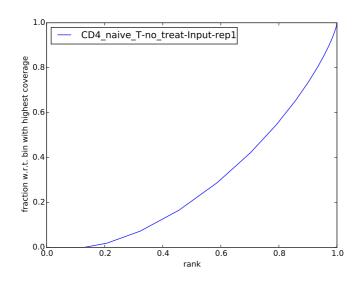
Urkat-no\_treat-lo K562-no\_treat-lo

CTCF

.... 1 ..........

#### PCA

ession Informatio



## DND41-no\_treat-CTCF

Sample Sheet

Alignment Summary

#### Quality Control

CD34-no.treat-CTCI
CD34-no.treat-Input
CD4+-no.treat-CTCI
CTCF
CD4-naive\_Tno.treat-CTCF
CD4-naive\_Tno.treat-CTCF
CD4-no.treat-Input
DND41-no.treatCTCF
DND41-no.treatInput

HEK293 I -no\_treat CTCF

Input

CTCF

IMR90-no\_tr

Jurkat-no\_tr

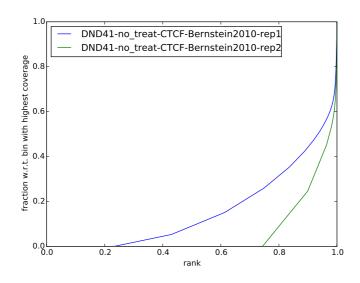
K562-no\_treat-CT

K562-no\_treat-Inp

| HP1-no\_treat-Inp

#### PCA

ession Informatio



# DND41-no\_treat-Input

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTCl
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treatCTCF

### DND41-no\_treat-

Input H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat CTCF

Input

CTCF

IMR90-no\_treation

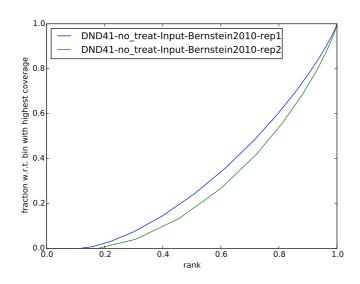
Jurkat-no\_treat-CTCF Jurkat-no\_treat-

K562-no\_treat-Inp

i HP1-no\_treat-inp

#### PCA

ession Informatio



# H1-no\_treat-CTCF

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treatCTCF

DND41-no\_tre

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat-

IMR90-no\_treat-

IMR90-no\_trea

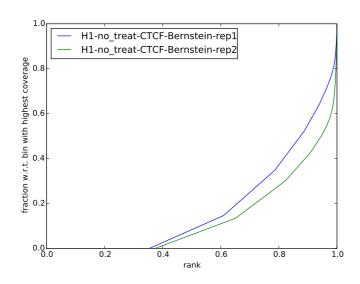
Input

Jurkat-no\_treat-Inpu K562-no\_treat-CTC K562-no\_treat-Input THP1-no\_treat-

| HP1-no\_treat-Inp

#### **PCA**

ession Informatio



# H1-no\_treat-Input

Sample Sheet

Alignment Summar Stats

### Quality Control

CD34-no\_treat-CTCl
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treatCTCF

DND41-no\_tr

H1-no\_treat-CTCI H1-no\_treat-Input HEK293T-no\_trea

HEK293T-no\_treat-Input

IMR90-no\_treat

IMR90-no\_tre

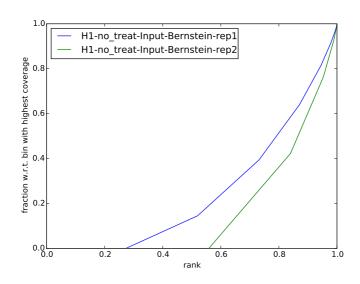
Jurkat-no\_treat-CTCF

K562-no\_treat-CT K562-no\_treat-Inp THP1-no\_treat-

ΓHP1-no\_treat-Inc

#### PCA

ession Informatio



# HEK293T-no\_treat-CTCF

Sample Sheet

Alignment Summar

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treatCTCF

DND41-no\_tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat-

HEK293T-no\_treat-

IMR90-no\_treat

IMR90-no\_tre

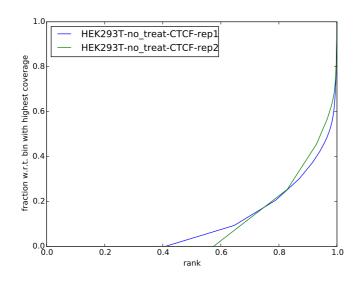
Jurkat-no\_treat-CTCF Jurkat-no\_treat-l

K562-no\_treat-Inp

HP1-no\_treat-Inp

#### PCA

ession Informatior



Stephen M. Kelly

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Inpu
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treatCTCF

DND41-no\_tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat-

IMR90-no\_trea

IMR90-no\_tre

Input Jurkat-no\_trea

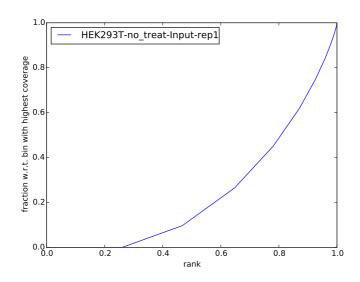
Jurkat-no\_treat-li K562-no\_treat-ln

THP1-no treat-In

#### PCA

Session Information

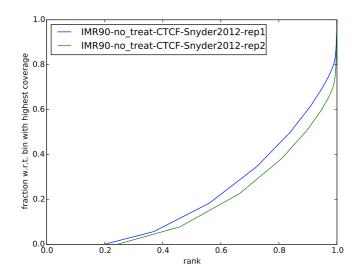
# HEK293T-no\_treat-Input



## Sample Sheet

#### Quality Control

#### IMR90-no\_treat-CTCF



# IMR90-no\_treat-Input

Sample Sheet

Alignment Summar Stats

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input

DND41-no\_tre

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_trea

HEK293T-no\_treat-Input

IMR90-no\_treat

### IMR90-no\_treat-

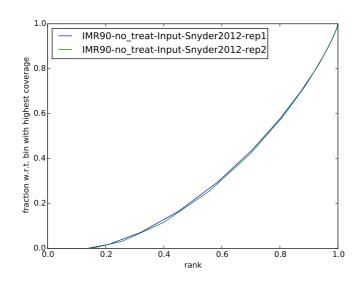
Input
Jurkat-no\_treatCTCF

Jurkat-no\_treat-Inpu K562-no\_treat-CTCI K562-no\_treat-Input THP1-no\_treat-

HP1-no\_treat-inp

#### PCA

ession Informatio



# Jurkat-no\_treat-CTCF

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treat-

DND41-no..tre

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_trea

HEK293T-no\_treat-

IMR90-no\_treat-

IMR90-no\_tre

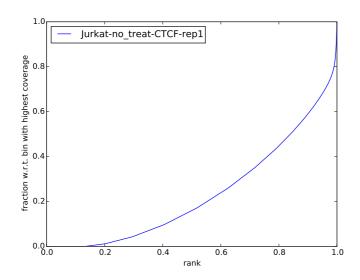
#### Jurkat-no\_treat-CTCF

Jurkat-no\_treat-Inpu K562-no\_treat-CTCF K562-no\_treat-Input THP1-no\_treat-CTCF

THP1-no\_treat-Inp

#### PCA

ession Informatio



# Jurkat-no\_treat-Input

Alignment Summar

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treat-

DND41-no..tre

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat

IMR90-no\_treat

IMR90-no\_tre

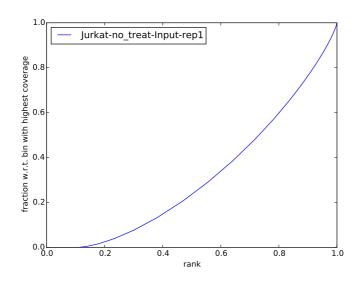
Jurkat-no\_treat-CTCF

Jurkat-no\_treat-Input K562-no\_treat-CTCF K562-no\_treat-Input THP1-no\_treat-

I HP1-no\_treat-Inp

#### PCA

ession Informatio



## Stephen M. Kelly

# Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTCl
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treat-

DND41-no..tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat

IMPO0-no treat-

CTCF IMR90-no trea

Input

CTCF

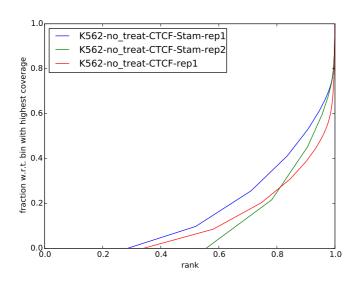
K562-no\_treat-CTCF K562-no\_treat-Input

HP1-no\_treat-Inp

#### PCA

ession Informatio

# K562-no\_treat-CTCF



# K562-no\_treat-Input

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTCl
CD34-no\_treat-Input
CD4+-no\_treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treatCTCF

DND41-no..tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_trea

HEK293T-no\_treat-

IMR90-no\_treat

IMR90-no\_tre

Input

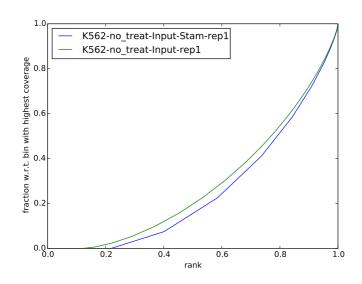
CTCF Jurkat-no\_treat-I

K562-no\_treat-Input

THP1-no\_treat-Inj

#### PCA

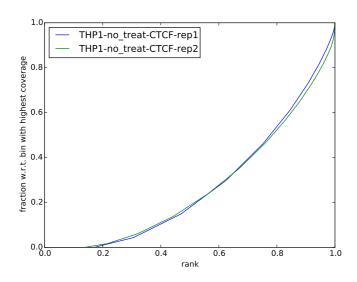
ession Informatio



Sample Sheet

### Quality Control

THP1-no\_treat-



# THP1-no\_treat-Input

Sample Sheet

Alignment Summary

### Quality Control

CD34-no\_treat-CTC
CD34-no\_treat-Input
CD4+-no.treatCTCF
CD4\_naive\_Tno\_treat-CTCF
CD4\_naive\_Tno\_treat-Input
DND41-no\_treat-

DND41-no\_tr

H1-no\_treat-CTCF H1-no\_treat-Input HEK293T-no\_treat

HEK293T-no\_treat

IMR90-no\_treat

IMR90-no\_tre

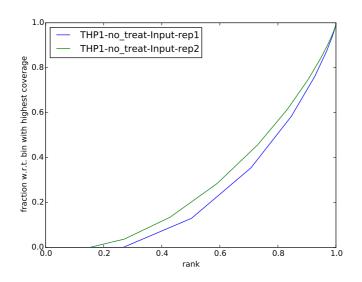
Input lurkat-no tre

Jurkat-no\_treat-Inpu K562-no\_treat-CTC K562-no\_treat-Input

THP1-no\_treat-Input

#### PCA

ession Informatio



PCA

ample Sheet

Alignment Summary Stats

Quality Control

#### PCA

all-samples all-samples all-samples

Session Information

This step of the pipeline performs PCA.

Alignment Summary

#### Quality Contro

#### quanty com

PCA

all-samples all-samples all-samples

Session Informatio

# PCA Parameters

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

Stephen M. Kelly

### Sample Sheet

Alignment Summary

### Quality Control

#### PCA all-samples

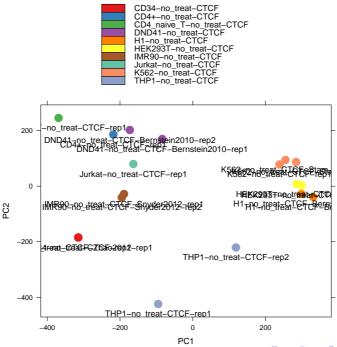
all-sample

all-sample

Session Informatio

# all-samples

results/pca.standard/matrices.by\_sample.flank\_500nt.nbins\_1/peaks.by\_sample.macs\_broad/align.by\_sample.bowtie2



Stephen M. Kelly

### Sample Sheet

Alignment Summary Stats

### Quality Control

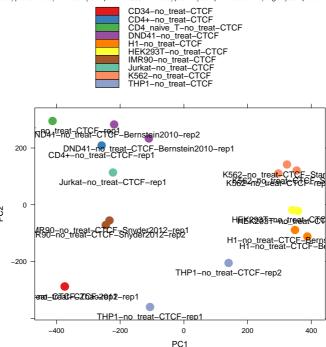
all-samples all-samples all-samples

**PCA** 

Session Information

# all-samples

results/pca.standard/matrices.by\_sample.flank.500nt.nbins\_1/peaks.by\_sample.macs\_narrow/align.by\_sample.bowtie2



Stephen M. Kelly

### Sample Sheet

Alignment Summary Stats

### Quality Control

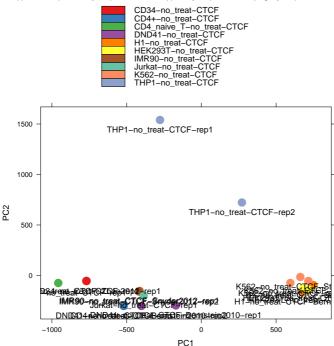
all-samples all-samples all-samples

PCA

Session Information

# all-samples

results/pca.standard/matrices.by\_sample.win\_1kb.nbins\_1/peaks.by\_sample.macs\_broad/align.by\_sample.bowtie2



Stephen M. Kelly

### Sample Sheet

Alignment Summary

### Quality Control

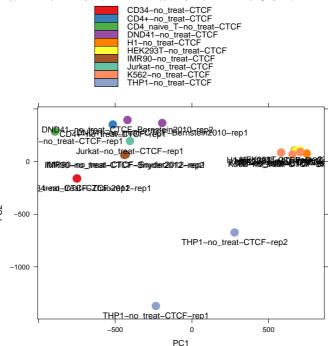
PCA

all-samples all-samples all-samples all-samples

Session Information

# all-samples

results/pca.standard/matrices.by\_sample.win\_1kb.nbins\_1/peaks.by\_sample.macs\_narrow/align.by\_sample.bowtie2



```
system('uname -srv',intern=T)
  CTCF_ChIP-Sea
 Stephen M. Kelly
                     ## [1] "Linux 2.6.32-573.3.1.el6.x86_64 #1 SMP Thu Aug 13 22:55:16 UTC 2015"
Sample Sheet
                     sessionInfo()
                     ## R version 3.2.3 (2015-12-10)
                     ## Platform: x86_64-pc-linux-gnu (64-bit)
Quality Control
                     ## Running under: CentOS release 6.3 (Final)
                     ##
                     ## locale:
Session Information
                     ## [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:
                                      graphics grDevices utils datasets base
                     ## [1] stats
                     ##
                     ## 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
                                                                    plvr 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
```

evaluate 0.8.3

scales\_0.4.0

acepack\_1.3-3.3

stringi 1.0-1

foreign\_0.8-66

save.image()

## [19] rpart 4.1-10

## [22] methods\_3.2.3

LATEX version: LATEX  $2\varepsilon$  2005/12/01

## [16] RColorBrewer 1.1-2 formatR 1.3.2