QC Report

3.25

3.00

2.75

1.75

1.50

1.25

1.0

1.5

2.0

Sample 1 log10 Score

Sample 2 Ranks vs IDR Values

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Sample 2 Peak Rank

Log10 Scores - (red \geq 0.05 IDR)

1.5 2.0 Sample 1 log10 Score

Sample 2 Ranks vs IDR Values

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Sample 2 Peak Rank

Log10 Scores - (red \geq 0.05 IDR)

1.00 1.25 1.50 1.75 2.00 2.25 2.50 2.75 Sample 1 log10 Score

Sample 2 Ranks vs IDR Values

3.0

2.5

1.5

1.0

2.5

3.0

1.0

rep1_vs_rep2

Sample 2 log10 Score

1.0 -

-log10 IDR w

rep1-pr1_vs_rep1-pr2

2.75

2.50

1.50

1.25

1.00

1.0

0.8

0.6

Sample 1 Rank

Sample 1 Ranks vs IDR Values

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Sample 1 Peak Rank

Ranks - (red \geq 0.05 IDR)

0.6

Sample 1 Rank

Sample 1 Ranks vs IDR Values

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Sample 1 Peak Rank

Ranks - (red \geq 0.05 IDR)

0.8

1.0

Log10 Scores - (red \geq 0.05 IDR)

general

Sequencing endedness {'rep1': {'paired_end': True}, 'rep2': {'paired_end': True}, 'ctl1': {'paired_end': True}, 'ctl2': {'paired_end': True}}

Ranks - (red \geq 0.05 IDR)

Report generated at 2020-10-23 07:33:08

v1.6.0

hg38

bowtie2

Replication quality metrics

0.2 -

0.0

Sample 2 Rank 0.0 4.0

0.2 -

0.0

-log10 IDR w

0.8

Sample 2 Rank 0.0 4.0

0.2 -

0.0

Sample 2 Rank .0

N1

N2

Np

Peak region size

50 percentile (median) 390.0

Min size

25 percentile

0.2

0.4

Sample 1 Rank

Sample 1 Ranks vs IDR Values

Sample 1 Peak Rank

Ranks - (red \geq 0.05 IDR)

0.6

0.8

0.2

0.0

0.2

IDR (Irreproducible Discovery Rate) plots

Description

Pipeline type

Genome

Aligner

Peak caller

Pipeline version

ENCSR514HBO vs. ENCSR541AMF

Comparision of ChIA-PET and ChIP-seq (SK-N_SH)

Fingerprints of different samples rep1 — rep2 — ctl1

> rep2 rep1-pr1 rep2-pr1 rep1

0.8 1.0 0.6 0.2 0.0 0.4 rank **Peak enrichment** Fraction of reads in peaks (FRiP) FRiP for spp raw peaks rep1-pr2 Fraction of Reads in Peaks 0.35610139898542964 0.3258323498862521 0.367412361885151 0.3208059224478757 0.36748551772167426 0.32073789588745094 0.326701276085476 0.320625988308118 0.3206654619661591 FRiP for overlap peaks

rep1_vs_rep2 rep1-pr1_vs_rep1-pr2 rep2-pr1_vs_rep2-pr2 pooled-pr1_vs_pooled-pr2 Fraction of Reads in Peaks 0.19733516227888412 0.29055993516925804 0.25812210307749456 0.27418757568170016 rep1_vs_rep2 rep1-pr1_vs_rep1-pr2 rep2-pr1_vs_rep2-pr2 pooled-pr1_vs_pooled-pr2

rep2-pr2

FRiP for IDR peaks Fraction of Reads in Peaks 0.1697293641073604 0.2670750084039998 0.1781486958042025 0.2120392510844536 For spp raw peaks: • repX: Peak from true replicate X • repX-prY: Peak from Yth pseudoreplicates from replicate X • pooled: Peak from pooled true replicates (pool of rep1, rep2, ...) pooled-pr1: Peak from 1st pooled pseudo replicate (pool of rep1-pr1, rep2-pr1, ...)
pooled-pr2: Peak from 2nd pooled pseudo replicate (pool of rep1-pr2, rep2-pr2, ...)

200 600 800 1000 400 rep1 Peak width distribution for overlap.optimal_peak.regionPeak.gz 100000 80000 60000 40000 20000 500 1000 1500 2000 2500 overlap_opt Enrichment / Signal-to-noise ratio Strand cross-correlation measures (trimmed/filtered SE BAM) rep1 rep2 **Number of Subsampled Reads** 7500000 7500000 100 **Estimated Fragment Length** 185 Cross-correlation at Estimated Fragment Length 0.20071237168354 0.12659373118077 **Phantom Peak** 0.1495871 **Cross-correlation at Phantom Peak** 0.1261669 1500 **Argmin of Cross-correlation** 1500 **Minimum of Cross-correlation** 0.08329578 0.0822723 NSC (Normalized Strand Cross-correlation coeff.) 2.409634 1.538716 RSC (Relative Strand Cross-correlation coeff.) 1.009725

> correlation 0.12 cross-0.09 90.0 -500 500 1000 1500 strand-shift (185)

0.10 0.08 -500 NSC=2.40963,RSC=1.77122,Qtag=2 rep1 Jensen-Shannon distance (filtered/deduped BAM) rep2 rep1 0.2082230665497578 0.18957259589871933 0.49742086689349213 **Synthetic AUC** 0.49428296172294206 X-intercept 0.12993611469362046 0.13307539265968826 **Synthetic X-intercept** 1.595157160677432e-262 0.0 **Elbow Point** 0.7323775531627725 0.7036801535646802 **Synthetic Elbow Point** 0.5071520152270073 0.4982307418525903 0.29022349776376927 JS Distance 0.31810015902461664 **Synthetic JS Distance** 0.4245693690096325 0.44390456869780576 **% Genome Enriched** 14.080761424872279 24.94386291153035 31.382657871395885 38.123693431544126 Diff. Enrichment

CHANCE Divergence 0.28010600289615206 0.3279409446700306

AUC

Performed on subsampled (15000000) reads mapped from FASTQs that are trimmed to 50. Such FASTQ trimming and subsampling reads are for cross-correlation analysis only. Untrimmed FASTQs are used for all the other analyses. NOTE1: For SE datasets, reads from replicates are randomly subsampled to 15000000. NOTE2: For PE datasets, the first end (R1) of each read-pair is selected and trimmed to 50 the reads are then randomly subsampled to 15000000. • Normalized strand cross-correlation coefficient (NSC) = col9 in outFile • Relative strand cross-correlation coefficient (RSC) = col10 in outFile • Estimated fragment length = col3 in outFile, take the top value ENCFF729SHE.no_chrM.R1.15M.tagAlign.gz s-correlation 0.14 0.16 cross-

1.50 0.2 -1.25 1.00 0.0 0.0 0.2 0.4 0.6 0.8 1.0 Sample 1 Rank Sample 1 log10 Score Sample 1 Ranks vs IDR Values Sample 2 Ranks vs IDR Values -log10 IDI w 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 Sample 1 Peak Rank Sample 2 Peak Rank pooled-pr1_vs_pooled-pr2 Reproducibility QC and peak detection statistics overlap idr 28634 60341 94116 53115 157983 46381 160084 52723 52723 160084 60341 28634 pooled-pr1_vs_pooled-pr2 pooled-pr1_vs_pooled-pr2 rep1_vs_rep2 rep1_vs_rep2 2.6529888467211347 1.8412726129775792

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 Sample 2 Peak Rank rep2-pr1_vs_rep2-pr2 Log10 Scores - (red \geq 0.05 IDR) 2.75 2 1.75 1.00 1.25 1.50 1.75 2.00 2.25 2.50 2.75 3.00

N optimal N conservative **Optimal Set Conservative Set Rescue Ratio Self Consistency Ratio** 1.6785987504781335 1.145188762639874 Reproducibility Test | borderline Reproducibility QC

N1: Replicate 1 self-consistent peaks (comparing two pseudoreplicates generated by subsampling Rep1 reads)
N2: Replicate 2 self-consistent peaks (comparing two pseudoreplicates generated by subsampling Rep2 reads)
Ni: Replicate i self-consistent peaks (comparing two pseudoreplicates generated by subsampling RepX reads)
Nt: True Replicate consistent peaks (comparing true replicates Rep1 vs Rep2)
Np: Pooled-pseudoreplicate consistent peaks (comparing two pseudoreplicates generated by subsampling pooled reads from Rep1 and Rep2)
Self-consistency Ratio: max(N1,N2) / min (N1,N2) • Rescue Ratio: max(Np,Nt) / min (Np,Nt) • Reproducibility Test: If Self-consistency Ratio > 2 AND Rescue Ratio > 2, then 'Fail' else 'Pass'

Number of raw peaks rep1 rep2 **Number of peaks** 299355 299920 Top 300000 raw peaks from spp with FDR 0.01 **Peak calling statistics**

116.0

328.0

464.0

rep2

114.0

456.0

456.0

rep1

98.0

390.0

idr_opt

overlap_opt

250000

116.0

464.0

464.0

390.0 456.0 589.0 464.0 75 percentile Max size 1125.0 2687.0 2694.0 2694.0 383.72704314275694 437.3194318484929 480.2404832805417 452.8888521026461 Mean Peak width distribution for ENCFF729SHE_x_ctl.pooled.300K.bfilt.regionPeak. 250000 200000 150000 100000

Peak width distribution for ENCFF871QIA_x_ctl.pooled.300K.bfilt.regionPeak.c

Peak width distribution for overlap.optimal_peak.regionPeak.gz

100000

80000

500

strand-shift (100,120,110)

NSC=1.53872,RSC=1.00972,Qtag=1

rep2

pooled-pr1

pooled

pooled-pr2

1000

1500

ENCFF871QIA.no_chrM.R1.15M.tagAlign.gz

fraction w.r.t. bin with highest coverage

For overlap/IDR peaks: repX_vs_repY: Comparing two peaks from true replicates X and Y
repX-pr1_vs_repX-pr2: Comparing two peaks from both pseudoreplicates from replicate X
pooled-pr1_vs_pooled-pr2: Comparing two peaks from 1st and 2nd pooled pseudo replicates