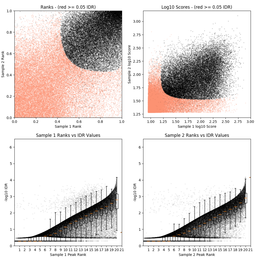
QC Report

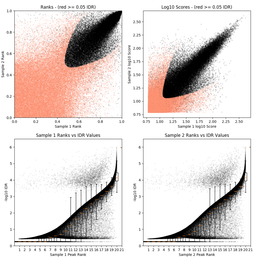
general	
Report generated at	2020-12-23 13:56:41
Title	ENCSR230LKR vs. ENCSR230YZS
Description	Comparison of ENCFF852 and ENCFF409 (PWPE1)
Pipeline version	v1.4.1
Pipeline type	ATAC
Genome	hg38
Organism	human
Sequencing endedness	{"rep1": [paired_end?, True], "rep2": [paired_end?, True], "ctl1": [paired_end?, True], "ctl2": [paired_end?, True]}
Peak caller	macs2

Replication quality metrics

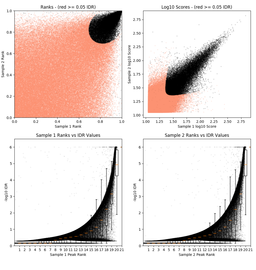
IDR (Irreproducible Discovery Rate) plots



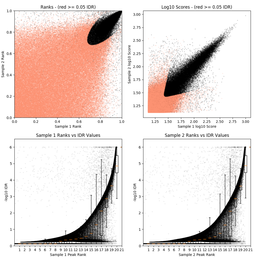
rep1_vs_rep2



rep1 pr1_vs_rep1 pr2



rep2 pr1_vs_rep2 pr2



pooled pr1_vs_pooled pr2

Reproducibility QC and peak detection statistices

	overlap	idr
Nt	89207	27062
Np1	83978	43091
Np2	84848	30278
N conserved	149849	70124
N optimal	149849	70124
N conservative	89207	27062
Optimal Set	pooled-pr1_vs_pooled-pr2	pooled-pr1_vs_pooled-pr2
Conservative Set	rep1_vs_rep2	rep1_vs_rep2
Recall Ratio	0.5726876514877335	2.5861098971622
SelfConsistency Ratio	1.7167972132809011	2.7563424416416175
ReplicateRatio Test	Irreducibe	pass

Reproducibility QC

- Nt: 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)
- Np1: Pooled pseudoreplicate consistent peaks (comparing two pseudoreplicates generated by subsampling Rep1 reads)
- Np: Pooled pseudoreplicate consistent peaks (comparing two pseudoreplicates generated by subsampling pooled reads from Rep1 and Rep2)
- Recall Ratio: overlap.Np / idr.Np (Np=Np1+Np2)
- Recall Ratio: overlap.Nt / idr.Nt
- Reproducibility Test: If Self consistency Ratio (2 AND Recall Ratio < 2, then Pass' else 'Pass'

Number of raw peaks

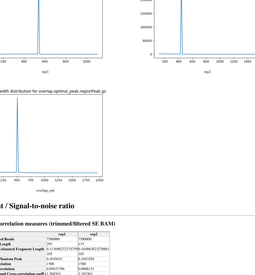
	rep1	rep2
Number of peaks	2903177	2509962

Top 300000 raw peaks from rep1 with FDR 0.01

Peak calling statistics

Peak region size

	rep1	rep2	ctl1_rep1	ctl2_rep1
Min size	146.0	150.0	128.0	128.0
25 percentile	346.0	449.0	347.0	350.0
50 percentile (median)	546.0	648.0	546.0	550.0
75 percentile	746.0	848.0	746.0	750.0
Max size	14918	13911.0	19618.0	19617.0
Mean	571.8597278079746	629.4897164968767	671.4574	675.0221



Enrichment / Signal-to-noise ratio

Strand cross-correlation measures (trimmed/filtered SE BAM)

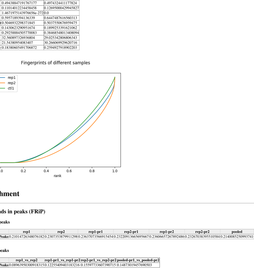
	rep1	rep2
Number of Subsampled Reads	7000000	7000000
Estimated Fragment Length	295	135
Cross-correlation at Estimated Fragment Length	0.11969122324798	0.0463063423238861
Phantom Peak	195	185
Cross-correlation at Phantom Peak	0.1016016	0.0463956
Argmin of Cross-correlation	5.500	5.500
Minimum of Cross-correlation	0.0963759	0.0406151
NSC (Normalized Strand Cross-correlation coeff.)	1.204765	1.101963
RSC (Relative Strand Cross-correlation coeff.)	1.178017	2.314872

Performed on subsampled (7000000) 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 and R1 (if each read pair is selected and trimmed to 50) the reads are then randomly subsampled to 15000000.

- Normalized strand cross-correlation coefficient (NSC) = ccff at ccmax / ccff at fragLen
- Relative strand cross-correlation coefficient (RSC) = ccff at ccmax / ccff at phaseLen
- Estimated fragment length = ccff at ccmax / ccff at fragLen, take the top value

Phantom Peak



Jensen-Shannon distance (filtered/deduped BAM)

	rep1	rep2
AUC	0.277366857497197	0.222217897410345
Synthetic AUC	0.494386471917877	0.497432441217928
X intercept	5.11414012154456e-06	0.0499866299658
Synthetic X intercept	0.00147514387869e-06	0.0
Elbow Point	0.000000000000000	0.000708563968217
Synthetic Elbow Point	0.666512788171041	0.327170367059435
JS Distance	0.143062129691678	0.109525139423482
Synthetic JS Distance	0.666512788171041	0.327170367059435
N. Genome Enriched	1.76000712493606	20.2152352800001
DHS_Enrichment	21.5430954814837	20.3660495296237
CHANCE Divergence	43.5596245917669	22.2594027910922

Peak enrichment

Fraction of reads in peaks (FRiP)

FRiP for opt_peak

	rep1	rep2	rep1-pr1	rep2-pr1	rep1-pr2	rep2-pr2	pooled	pooled-pr1	pooled-pr2
Fraction of Reads in Peaks	0.2301 (0.2369/0.2234)	0.2417 (0.2478/0.2356)	0.2462 (0.2517/0.2407)	0.2423 (0.2481/0.2365)	0.2386 (0.2437/0.2335)	0.2426 (0.2479/0.2373)	0.2404 (0.2455/0.2353)	0.2445 (0.2493/0.2397)	0.2412 (0.2461/0.2363)

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	0.4804 (0.5034/0.4583)	0.5132 (0.5349/0.4925)	0.5197 (0.5422/0.4982)	0.5187 (0.5412/0.4968)

FRiP for IDR peaks

	rep1_vs_rep2	rep1-pr1_vs_rep1-pr2	rep2-pr1_vs_rep2-pr2	pooled-pr1_vs_pooled-pr2
Fraction of Reads in Peaks	0.7780296197726411	0.8097971968172416	0.8017717374024011	0.8129981387054

For opt set peaks:

- rep1: Peak from true replicate 1
- rep2: Peak from true replicate 2
- 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, ...)

For overlap/IDR peaks:

- rep1_vs_rep2: Comparing two peaks from true replicates 1 and 2
- rep1-pr1_vs_rep1-pr2: Comparing two peaks from both pseudoreplicates from replicate A
- pooled-pr1_vs_pooled-pr2: Comparing two peaks from 1st and 2nd pooled pseudo replicates