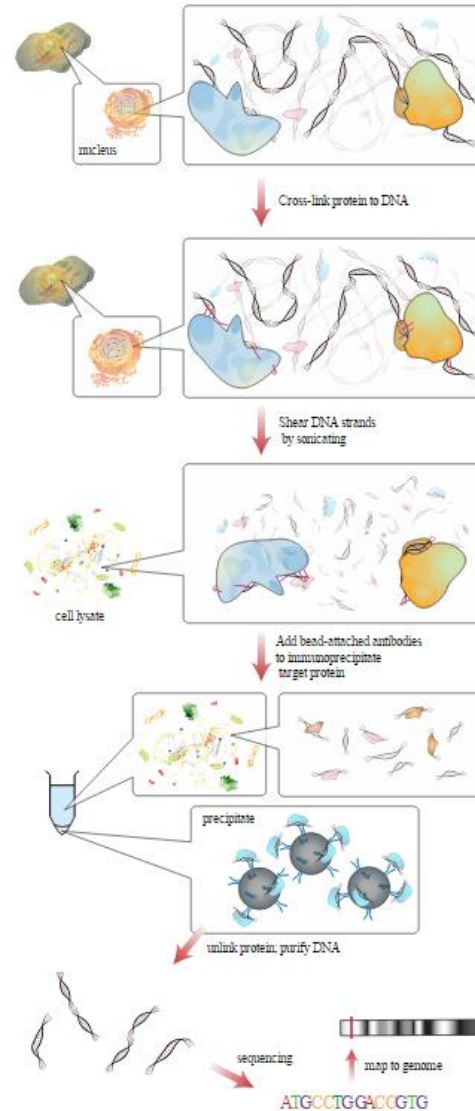


Comparison of ChIP-Seq Peak calling programs

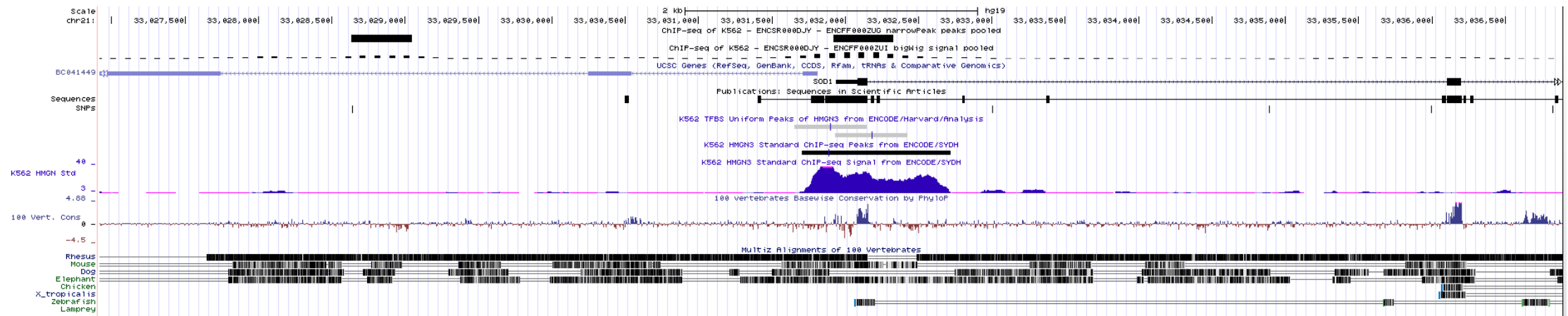
Kanishk Asthana

kasthana@eng.ucsd.edu

Overview of ChIP-seq



A genome wide histogram



<https://www.encodeproject.org/experiments/ENCSR000DOB/>

ChIP-seq calling programs

- SISSRS: only get start position, end position and number of TAGs, no filtering method can be applied because of no score being returned.
- Spp: not user friendly, documentation inadequate.
- Quest: not user friendly, documentation inadequate.
- PeakSeq: documentation inadequate
- MACS: extremely user friendly and easy to use
- HOMER: a little less user friendly, but still very useful. The entire package is what makes it so useful, a lot of associated functions and functionality.
- Sole-search: documentation inadequate, not user friendly
- Hpeak: only works for human genome

Pair wise comparison of shared peaks

Data Set	HOMER (% of Peaks shared with MACS)	MACS (% of Peaks shared with HOMER)
PU.1	93.24609	90.06348
ENCODE	74.55422	89.26311

Data Set	HOMER (% of Peaks shared with SISR5)	SISR5 (% of Peaks shared with HOMER)
PU.1	82.0557	75.01613
ENCODE	80.47605	63.70693

Data Set	SISR5 (% of Peaks shared with MACS)	MACS (% of Peaks shared with SISR5)
PU.1	82.8864	87.56901
ENCODE	63.42505	95.92693

Comparison of HOMER vs MACS: Does more peaks called mean more false positives?

Data Set	Type of Peaks	HOMER (% of Peaks shared with MACS)	MACS (% of Peaks shared with HOMER)
PU.1	Average	93.24609	90.06348
	Top 25 Percentile	90.48544	85.42430
	Top 5000 Peaks	82.57576	80.66000
	Bottom 25 Percentile	57.85524	51.35473
ENCODE	Average	74.55422	89.26311
	Top 25 Percentile	77.50584	91.39446
	Top 5000 Peaks	81.40521	80.64000
	Bottom 25 Percentile	24.38621	25.17400

Motif Analysis: Peaks Called by HOMER and MACS

Figure 1. PU.1 is the most ENRICHED known Motif found in peaks called by HOMER (Enrichement~44%)



Rank	Motif	Name	P-value	log P-pvalue	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File	PDF
1		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-22448	-5.169e+04	0.0000	38652.0	44.72%	5065.1	6.10%	motif file (matrix)	pdf

Figure 2. PU.1 is the most ENRICHED known Motif found in peaks called by MACS (Enrichement~44%)

Rank	Motif	Name	P-value	log P-pvalue	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File	PDF
1		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-23216	-5.346e+04	0.0000	39619.0	44.28%	5071.3	5.94%	motif file (matrix)	pdf

Motif Analysis: Peaks Called HOMER or MACS but not the other

Figure 3. PU.1 is the most ENRICHED known Motif found in peaks found in HOMER and not MACS (Enrichment ~32%)



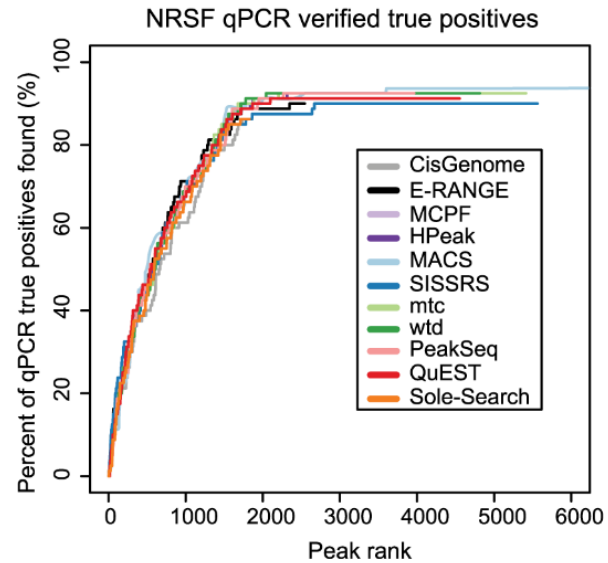
Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File	PDF
1		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-829	-1.910e+03	0.0000	1874.0	32.09%	2492.7	5.81%	motif file (matrix)	pdf

Figure 4. PU.1 is the most ENRICHED known Motif found in peaks found in MACS and not HOMER (Enrichment ~35%)

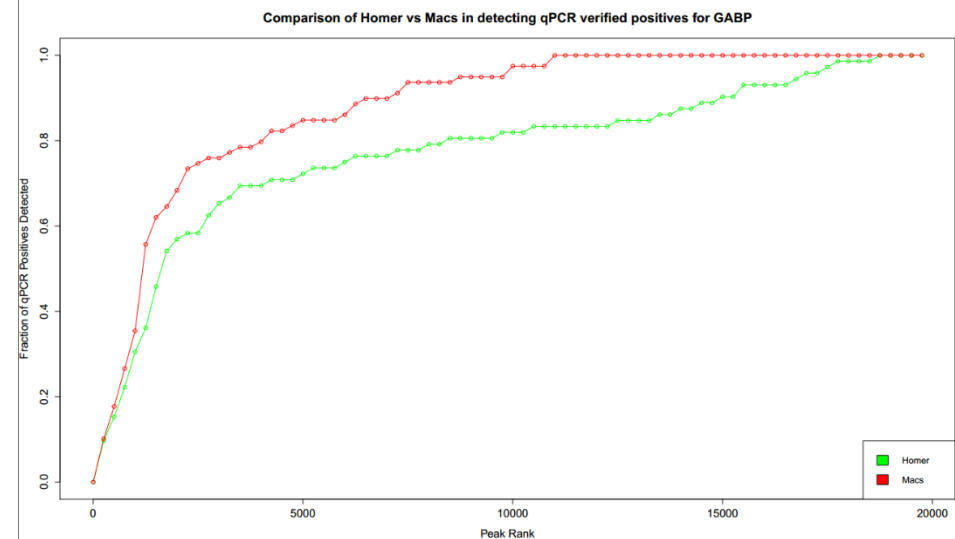
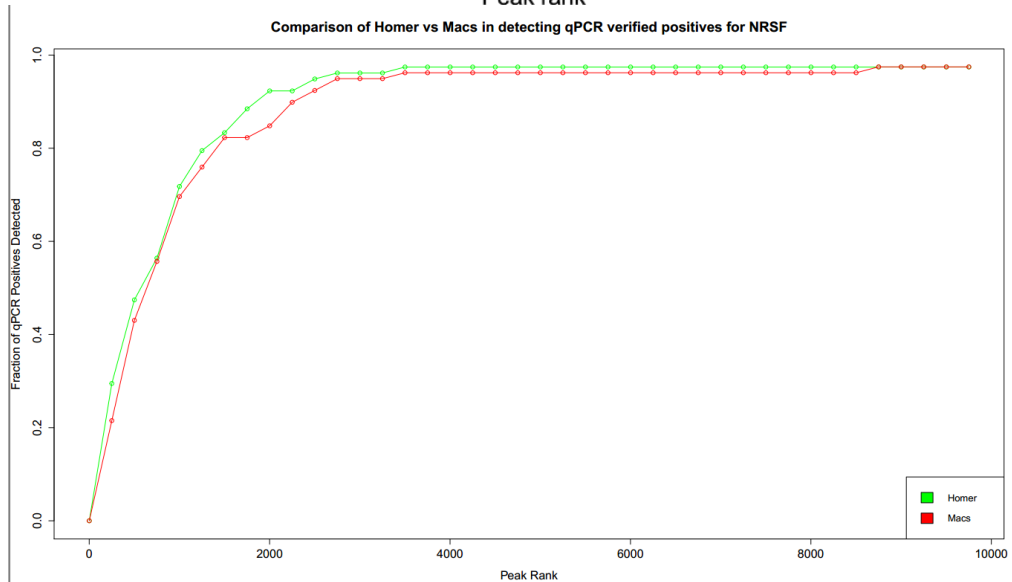
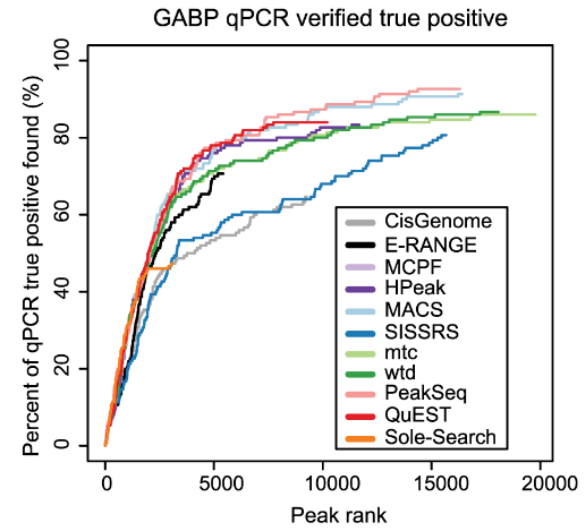
Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File	PDF
1		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-1441	-3.320e+03	0.0000	3166.0	35.45%	2341.5	6.31%	motif file (matrix)	pdf

Detection of qPCR verified True Positives

A



Top Two figures
are from
Wilbanks et. al
(2010) Plos One.



Conclusion: Which is better MACS or HOMER?

Not enough information to decide.

References

- Wilbanks, E. G., & Facciotti, M. T. (2010). Evaluation of algorithm performance in ChIP-seq peak detection. *PloS One*, 5(7), e11471. doi:10.1371/journal.pone.0011471