Enhancer Predictions - from flies to mice

Anurag Sethi

TECH

March 2015

Wrap up of what we learnt from the pattern recognition method to predict enhancers in fly

Matched filter with different histone marks

Histone Mark (double peak)	**AUC (ROC) S2 cell-line**	**AUC (ROC) BG3 cell line**
H3K27ac	0.88	0.97
H3K4me1	0.85	0.87
H3K4me2	0.85	0.86
H3K4me3	0.71	0.76
H3K9ac	0.88	0.75
H3K27me3	0.24	0.32
H3K27ac (scrambled)	0.49	0.48

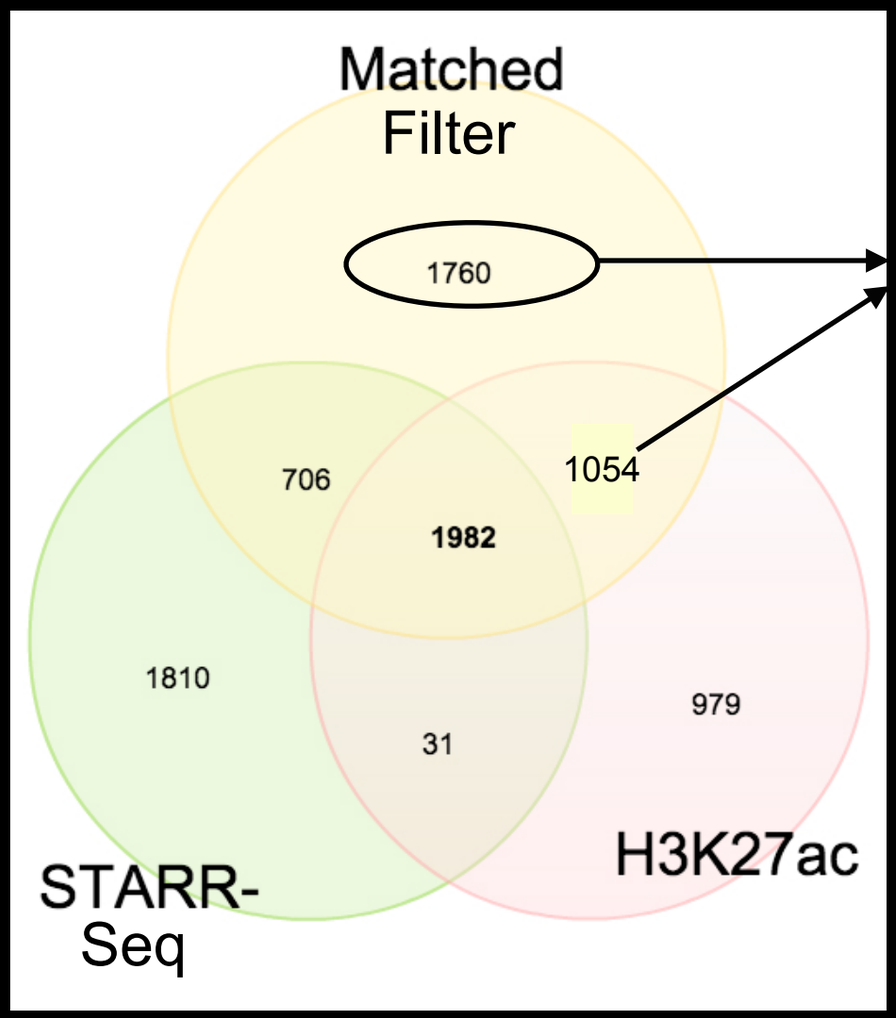
Stability of marks across cell lines

Histone Mark (double peak)	**AUC (ROC) BG3 (S2) profile S2 cell-line**	**AUC (ROC) S2 (BG3) profile BG3 cell line**
H3K27ac	0.89 (0.88)	0.96 (0.97)
H3K4me1	0.86 (0.85)	0.87 (0.87)
H3K4me2	0.85 (0.85)	0.87 (0.86)
H3K4me3	0.71 (0.71)	0.75 (0.76)
H3K9ac	0.88 (0.88)	0.77 (0.75)

>95% of genome positives are same when using either cell-line's metaprofile.

Genome positives comparison across marks/cell-lines

H3K27ac Matched Filter



1964 on promoters

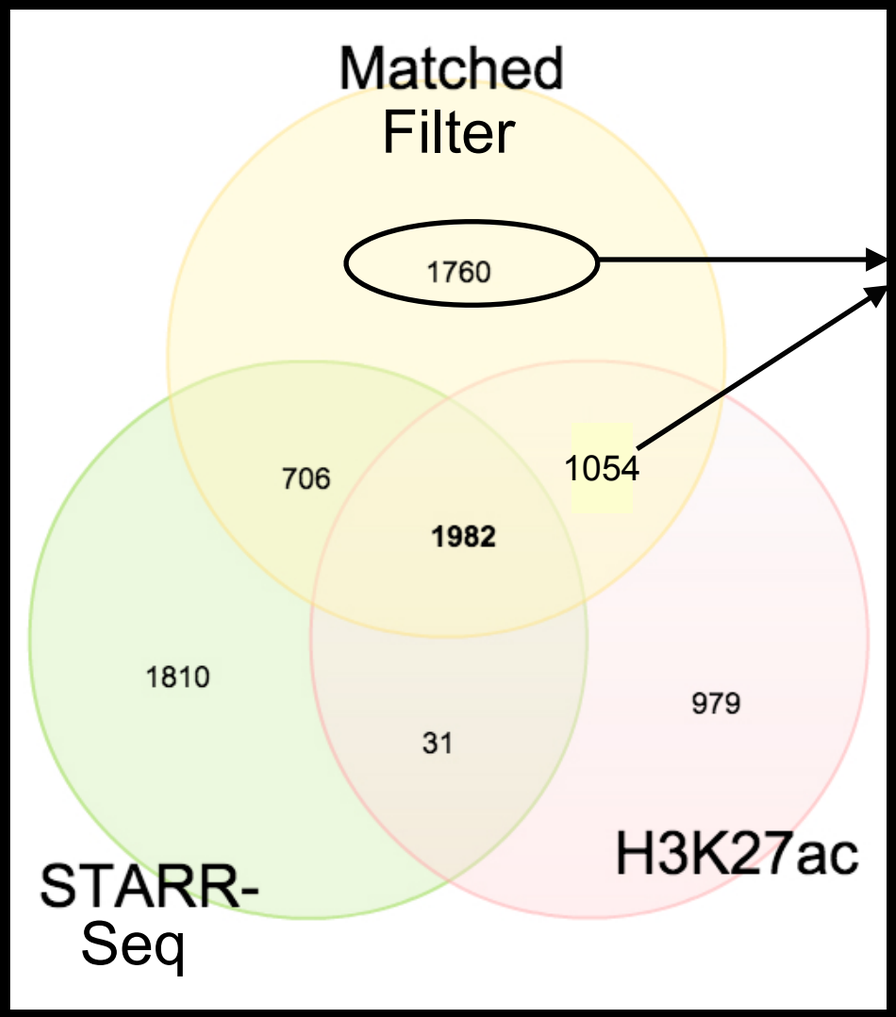
Need to show these promoters are active

1885 (1490) of STARR-seq peaks are positive in at least two (all three) filters

Most of the matched filter positives are positive on multiple histone marks. H3K9ac and H3K4me3 - most different (promoters).

Genome positives comparison across marks/cell-lines

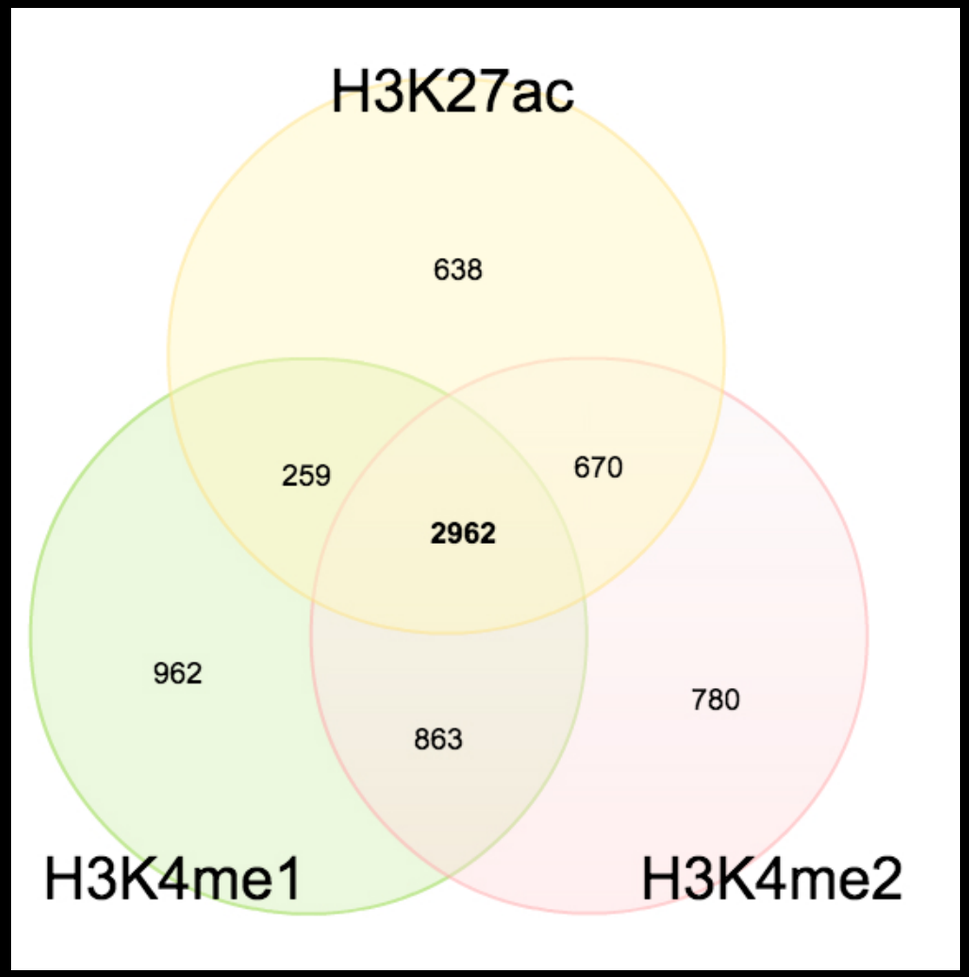
H3K27ac Matched Filter



1964 on promoters

Need to show these promoters are active

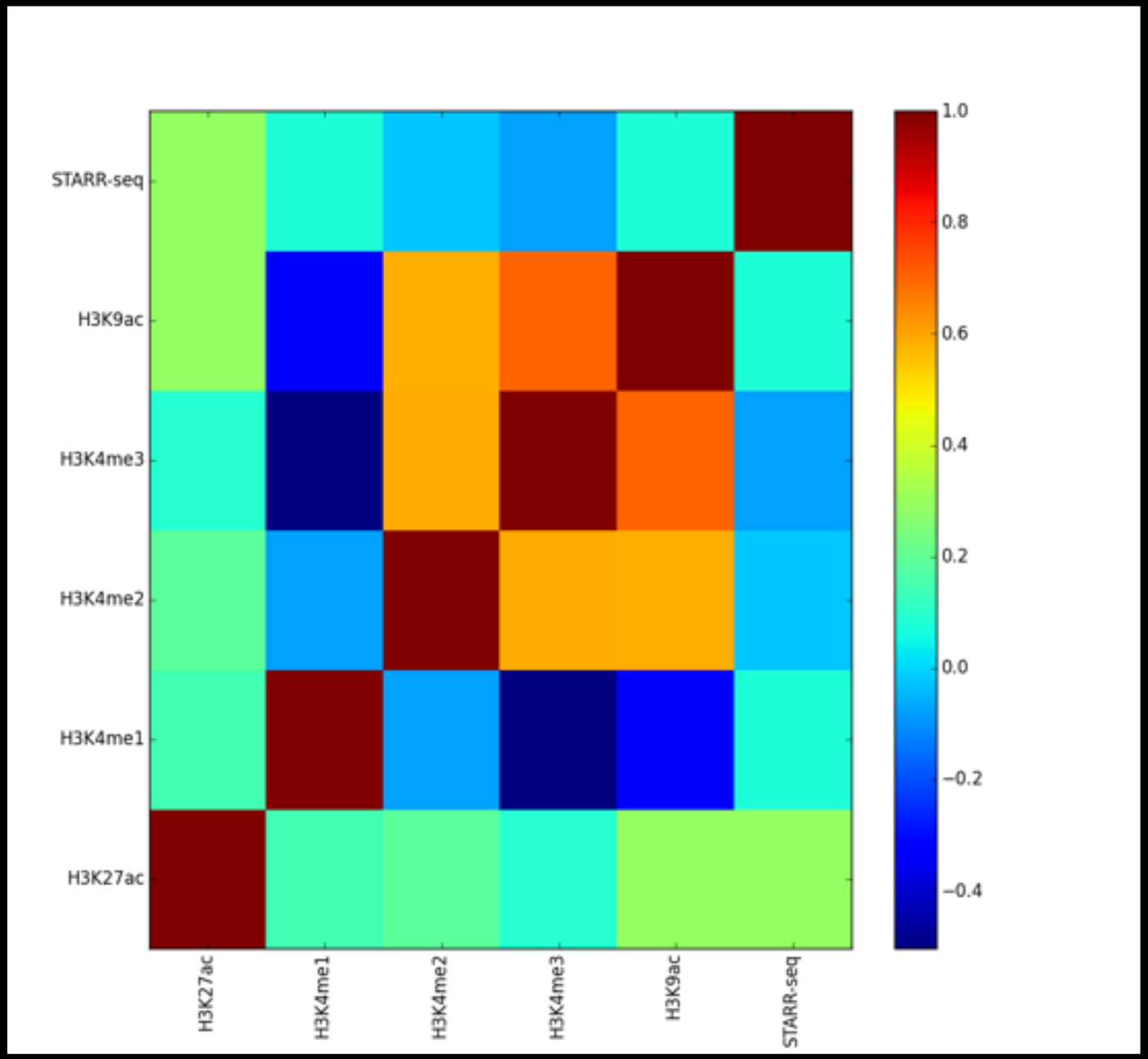
MF comparison



1885 (1490) of STARR-seq peaks are positive in at least two (all three) filters

Most of the matched filter positives are positive on multiple histone marks. H3K9ac and H3K4me3 - most different (promoters).

H3K27ac match filter score most correlated to STARR-seq strength

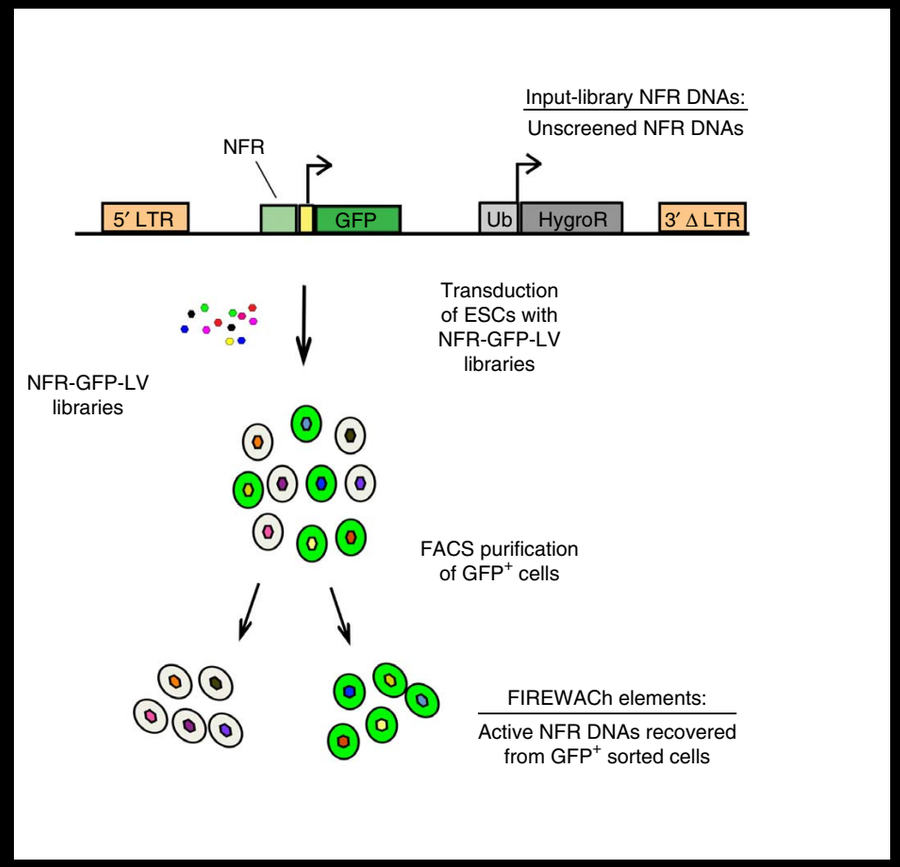


Very little correlation between the matched filter score and STARR-seq assay strength - chromatin regulation versus TF binding strength.
Towards a single enhancer activity score - maybe use TF binding strength!

Moving on to mammals

FIREWACh assay

- Enhancer candidates chosen based on open DNA in cell-line (murine ESC).
- Integrated into virus particles close to a minimal promoter and GFP.
- **Integrated into genome randomly with 1 clone per cell (H1-hESC).**
- **One potential enhancer of length 100-300 bp** per cell.
- FACS to sort cells expressing GFP.
- Small population of cells show positive enhancer activity.
- Amplified positive enhancer sequences with PCR using primers recognizing the flanking sequences.
- Tested enhancer activity using traditional assays.



Pro:
Chromatin context.
Con:
100-300 bp length.

Steps towards a pattern recognition method to predict enhancers based on positives in FIREWACh

ChIP-Seq for histone modification was converted to log enrichment in two steps:

- Adjust for sequencing depth by looking at signals in non-peak regions (H3K27ac) in H3K27ac and in control over 1Mb regions of the genome.
- Calculate log of enrichment of H3K27ac signal after adjusting for sequencing depth over the 1 Mb regions.

Steps towards a pattern recognition method to predict enhancers based on positives in FIREWACh

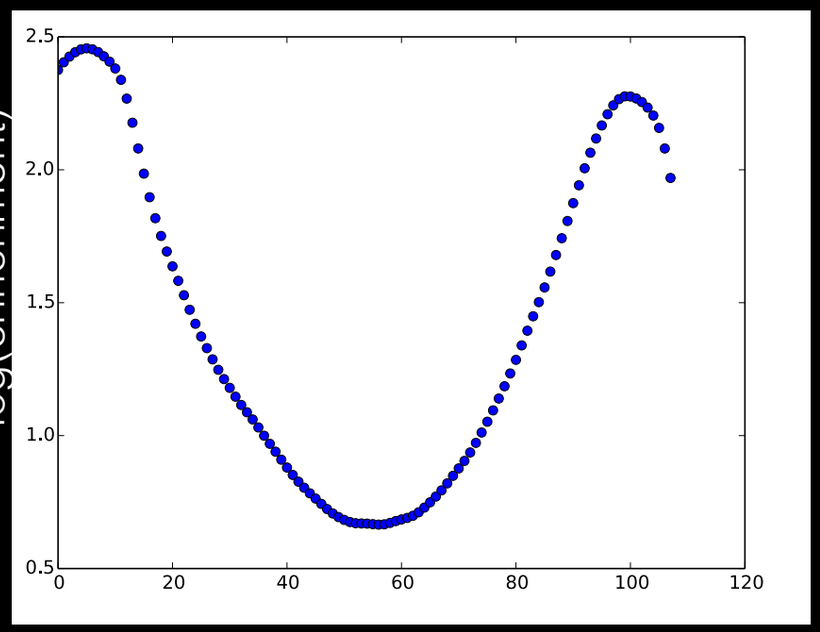
ChIP-Seq for histone modification was converted to log enrichment in two steps:

- Adjust for sequencing depth by looking at signals in non-peak regions (H3K27ac) in H3K27ac and in control over 1Mb regions of the genome.
- Calculate log of enrichment of H3K27ac signal after adjusting for sequencing depth over the 1 Mb regions.

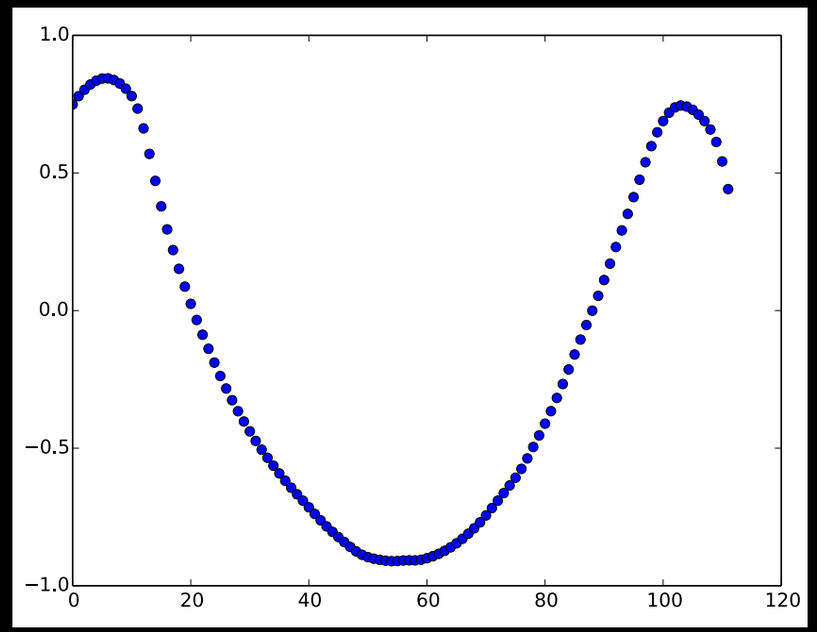
FIREWACh can find positives in repressed or poised enhancers as well (enrichment in H3K27me3 and H3K9me3 signals). Similar to STARR-Seq, I decided to focus on the FIREWACh positives that intersect with H3K27ac/DNase peaks.

Metaprofiles from FIREWACh

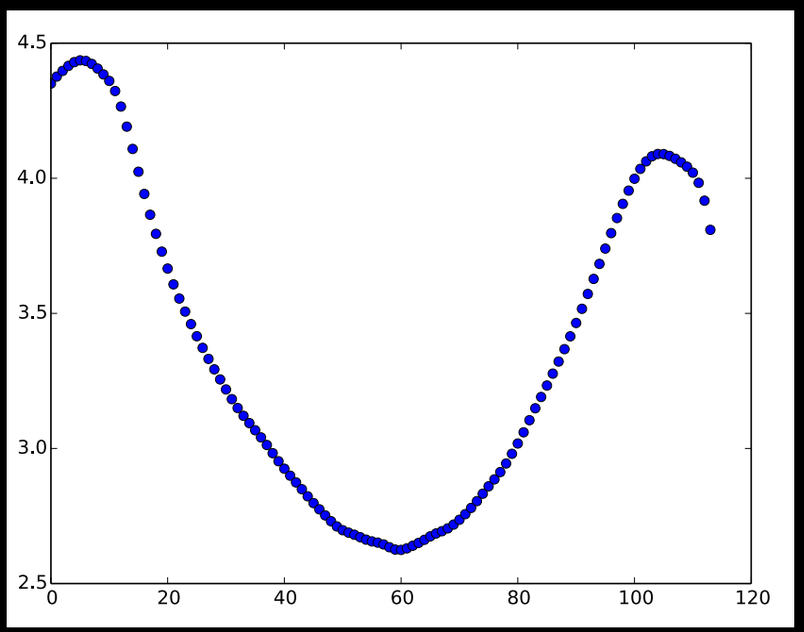
H3K27ac



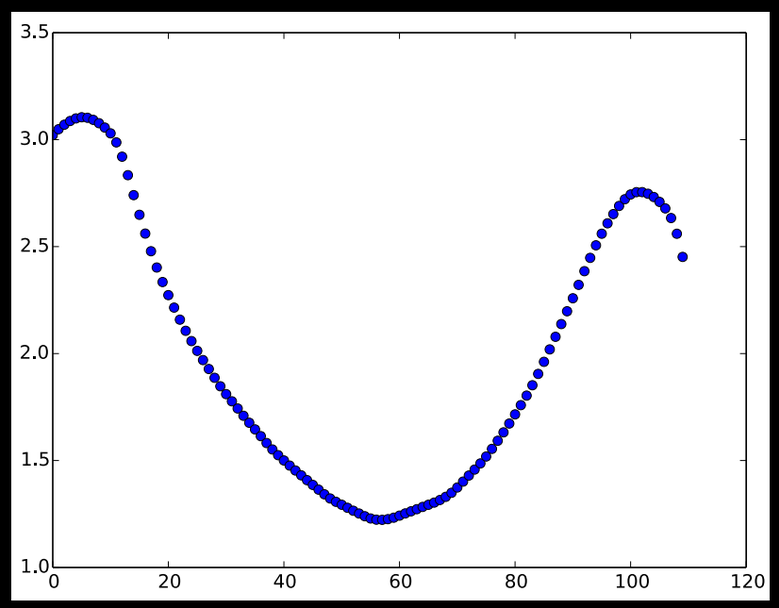
H3K4me1



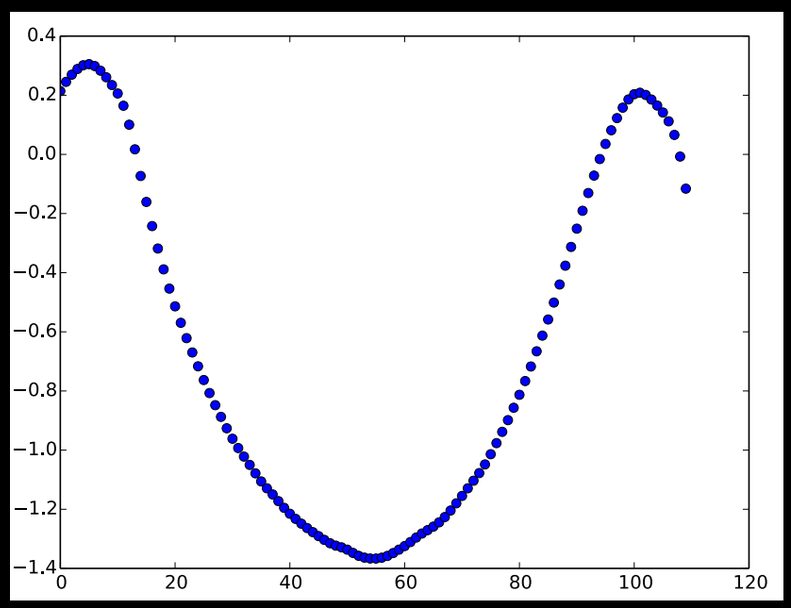
H3K4me3



H3K9ac



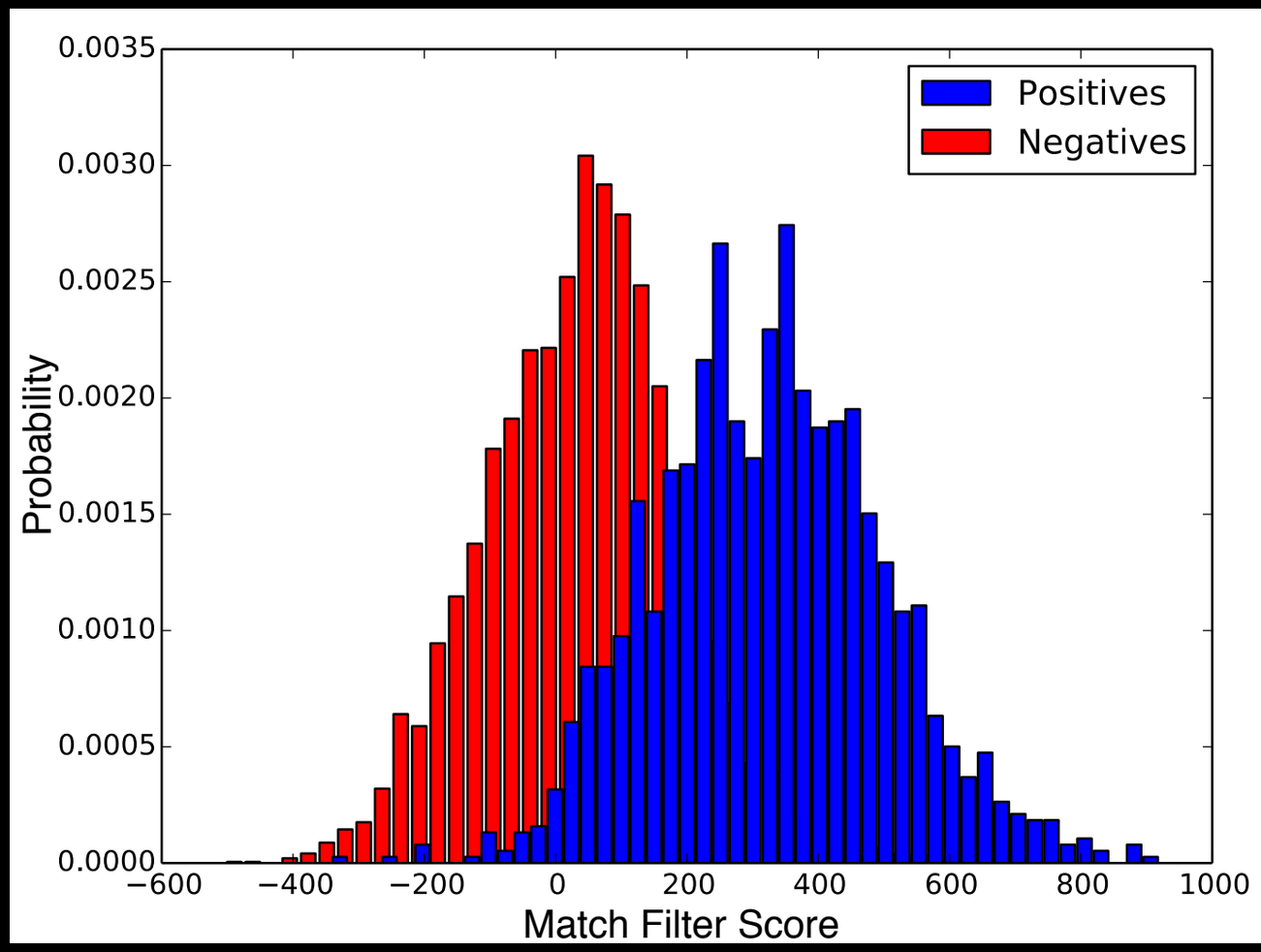
H3K36me3

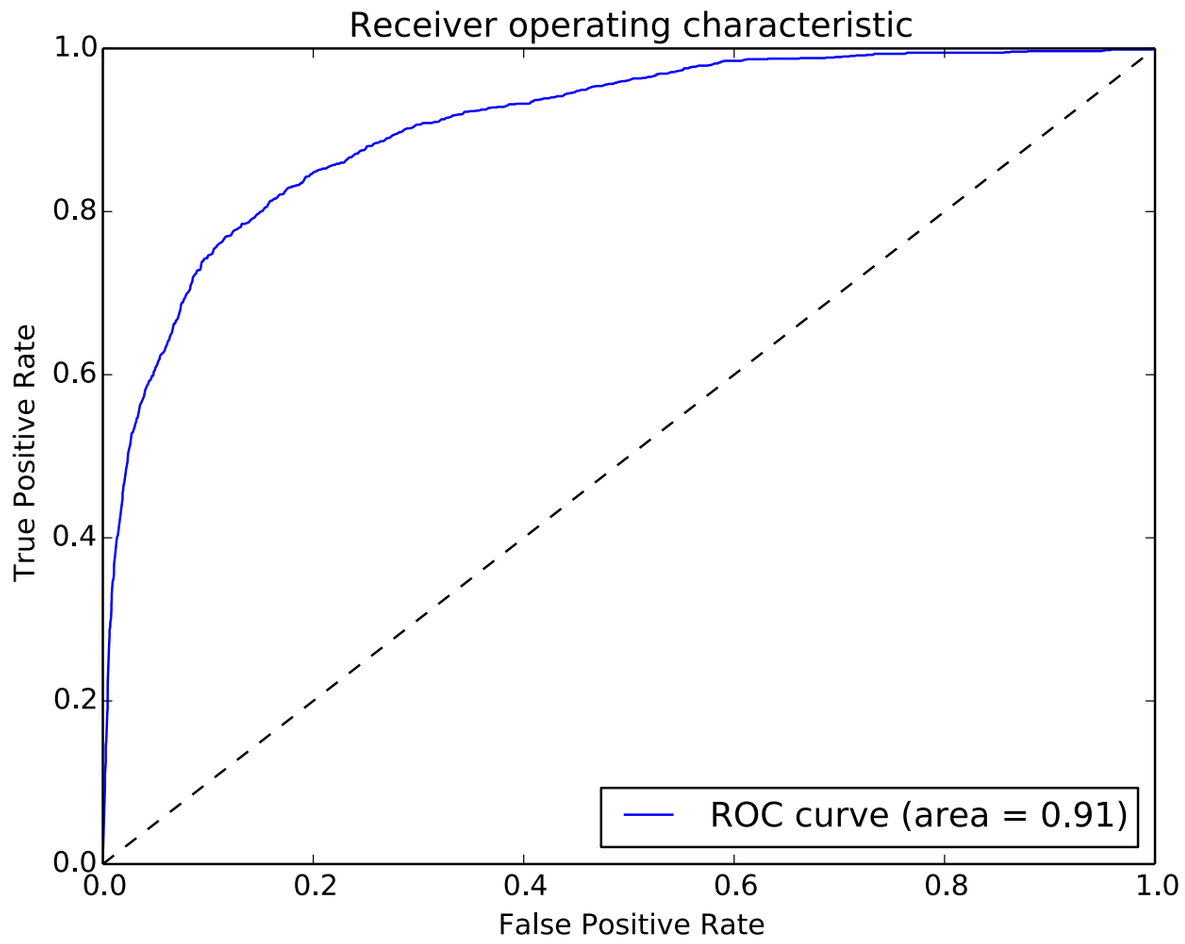


Heterogeneity in the metaprofiles close to regulatory regions - Anshul's paper.

The metaprofile can be used to identify enhancers from random regions in the genome.

Performance of H3K27ac metaprofile





Accuracy of Predictions

Histone Mark (double peak)	**AUC (ROC) mESC cell-line**
H3K27ac	0.91
H3K4me1	0.70!!
H3K4me3	0.87
H3K9ac	0.88
H3K36me3	0.67

Are these metaprofiles conserved across different organisms?

Histone Mark (double peak)	**AUC (ROC) mESC cell-line**	**AUC (ROC) BG3 metaprofile with mESC signal**
H3K27ac	0.91	0.89
H3K4me1	0.70!!	
H3K4me3	0.87!!	
H3K9ac	0.88!!	0.88
H3K36me3	0.67	-

What the future holds?

- Problems with current method - multimappability in ChIP-Seq is causing false positives in chromosome scan (filter development).
- Going to apply these metaprofiles for human enhancer predictions.
- Developing a single enhancer activity score.
- Apply to REMC and make target prediction.
- Look at incidence of rare variants within tissue specific GRN for the regulatory network of each gene - tissue specific variant effects.