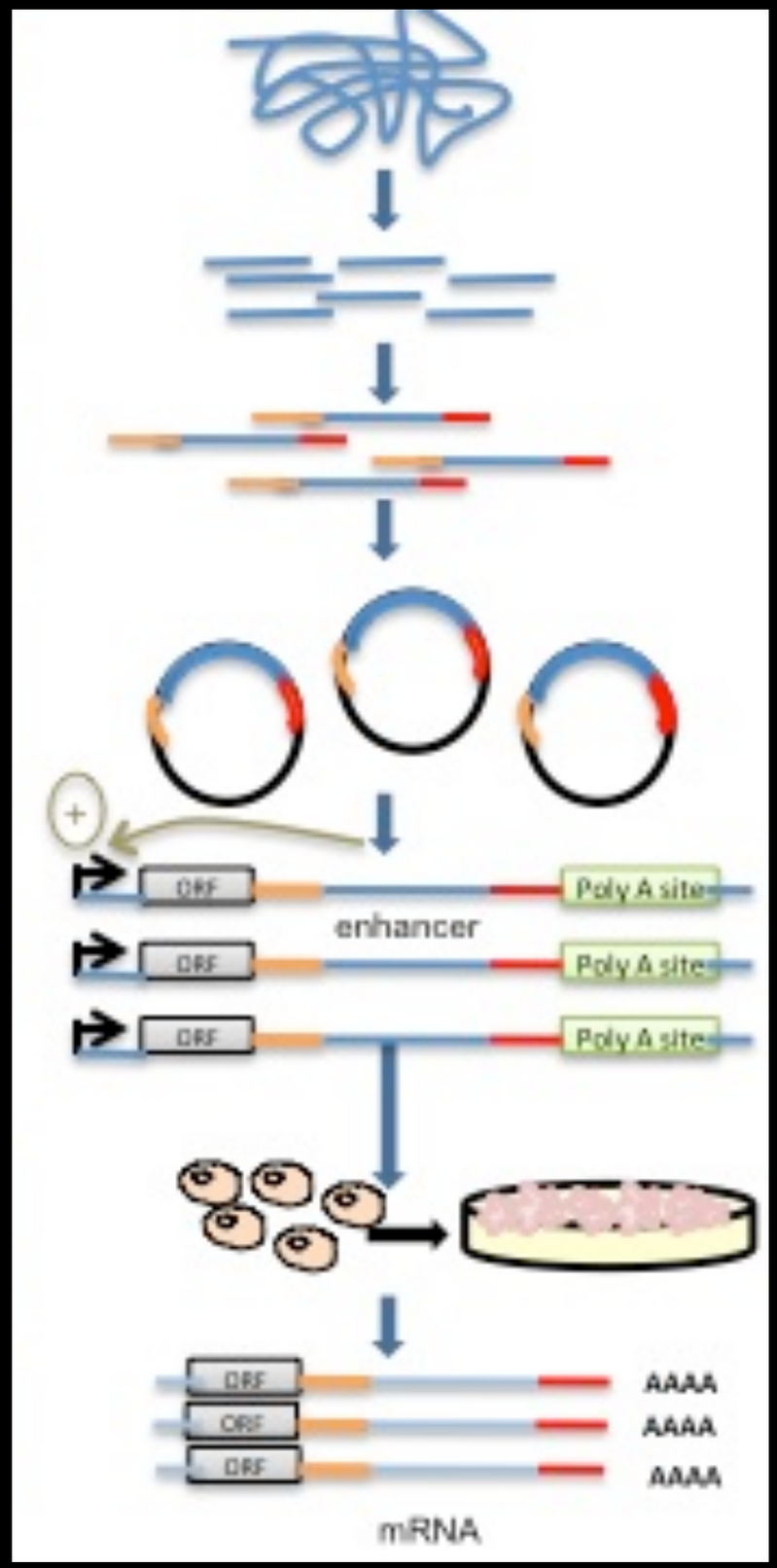
Matched Filter for Enhancer Predictions

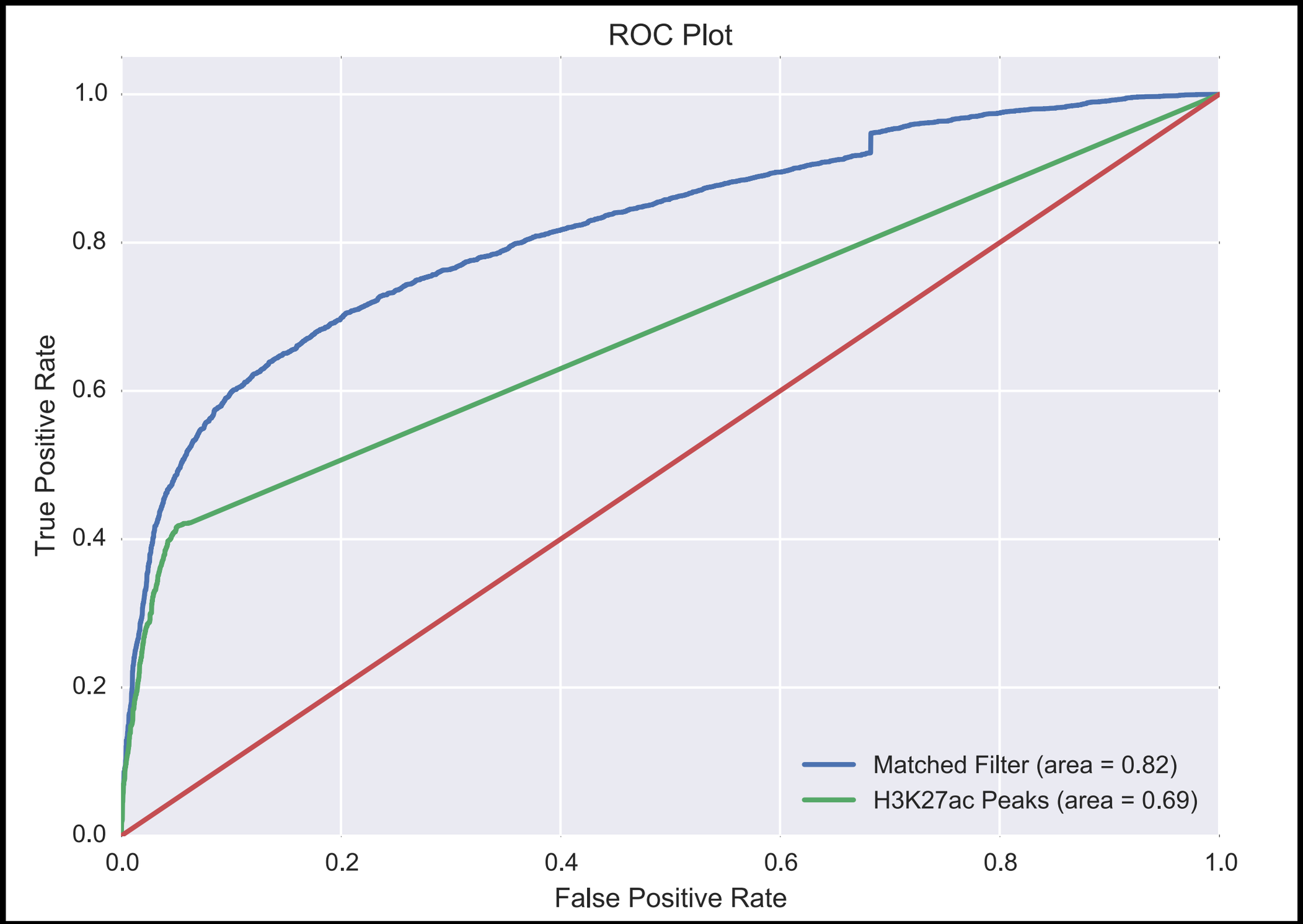
Anurag Sethi

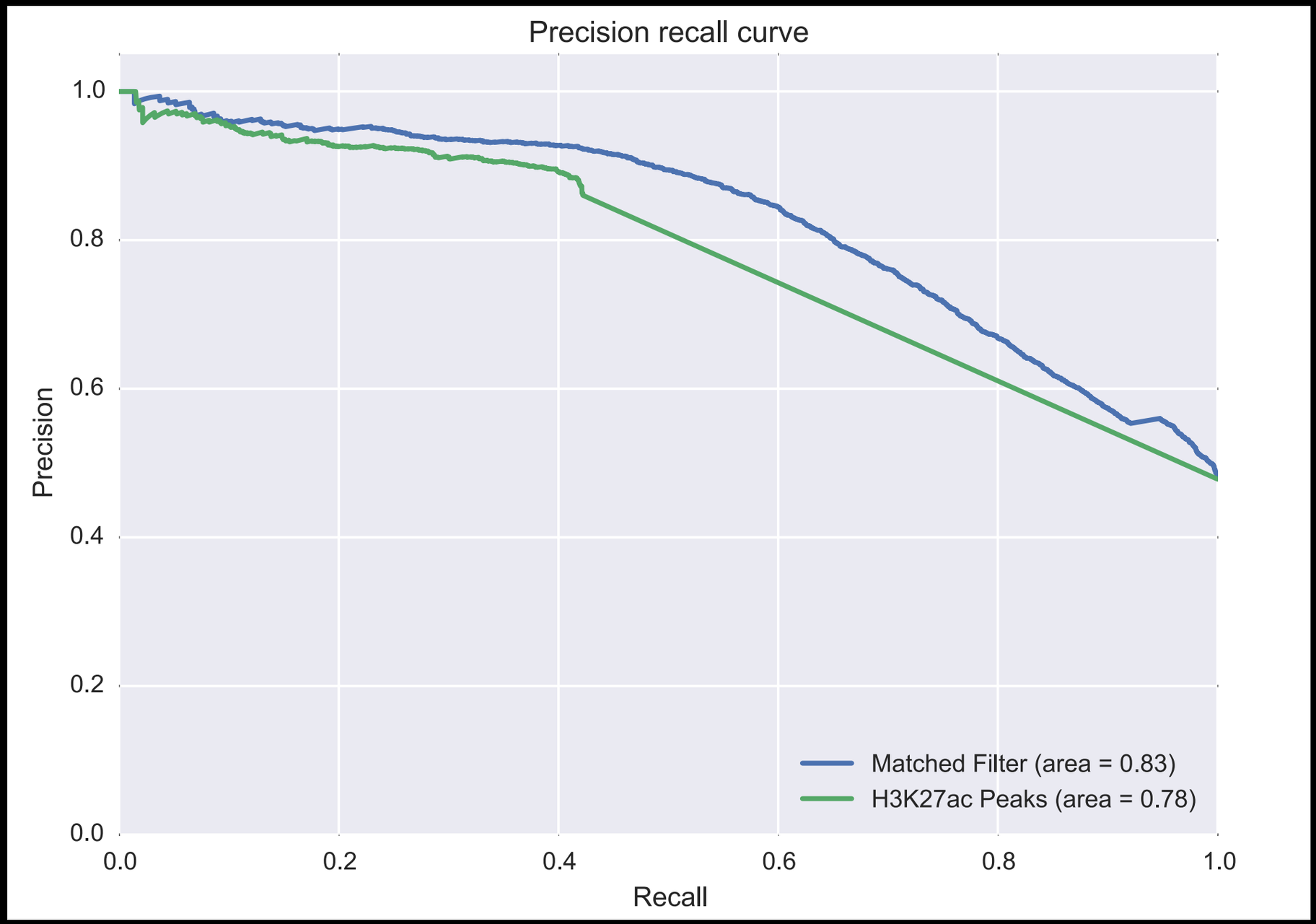
STARR-Seq is performed in plasmid environment



Minimal chromatin marks
Test for TF binding (TFs associated with appropriate cell type and promoter)

STARR-Seq (all peaks)

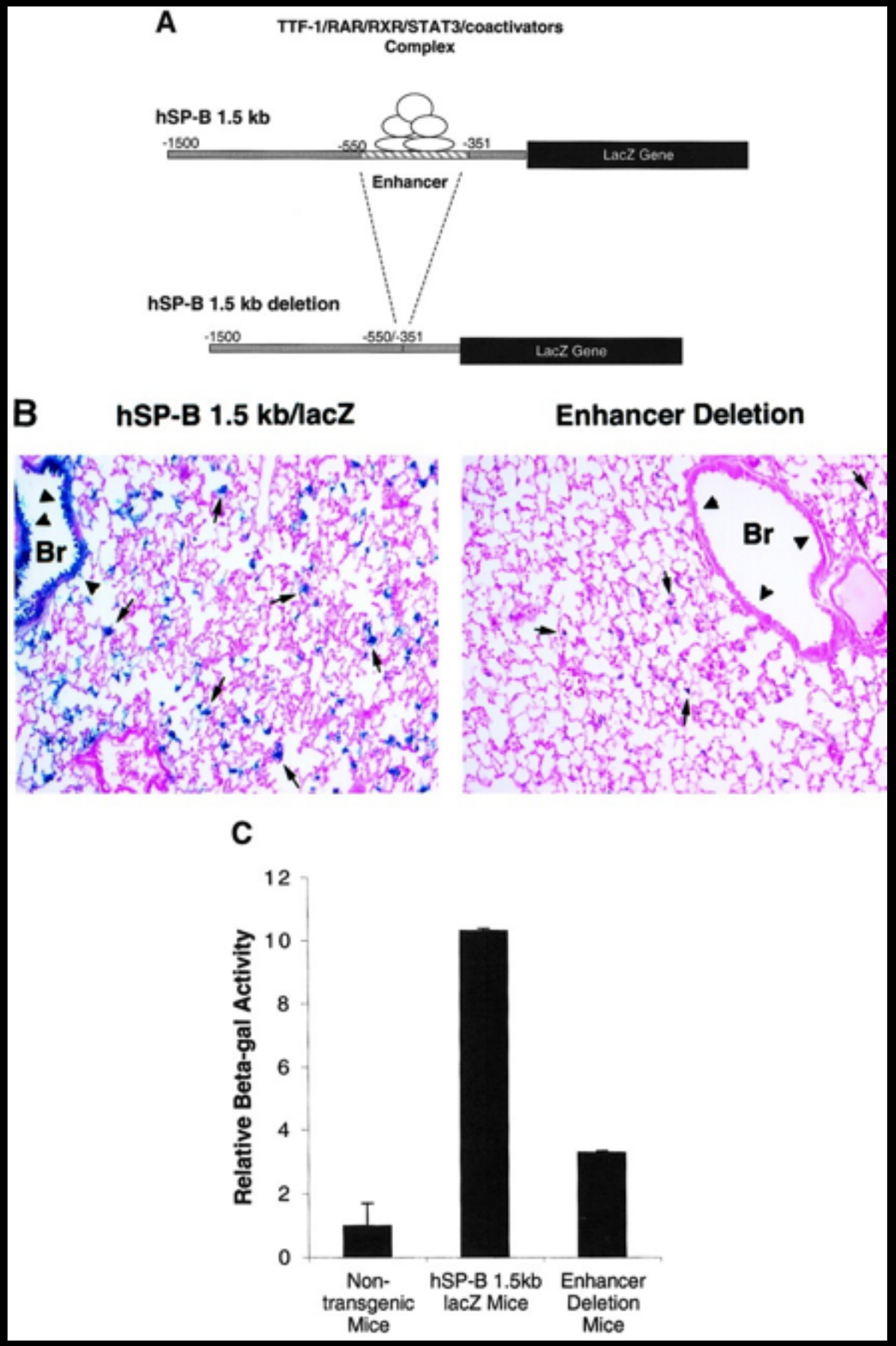



Precision recall curve
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Matched Filter (area = 0.83)
H3K27ac Peaks (area = 0.78)

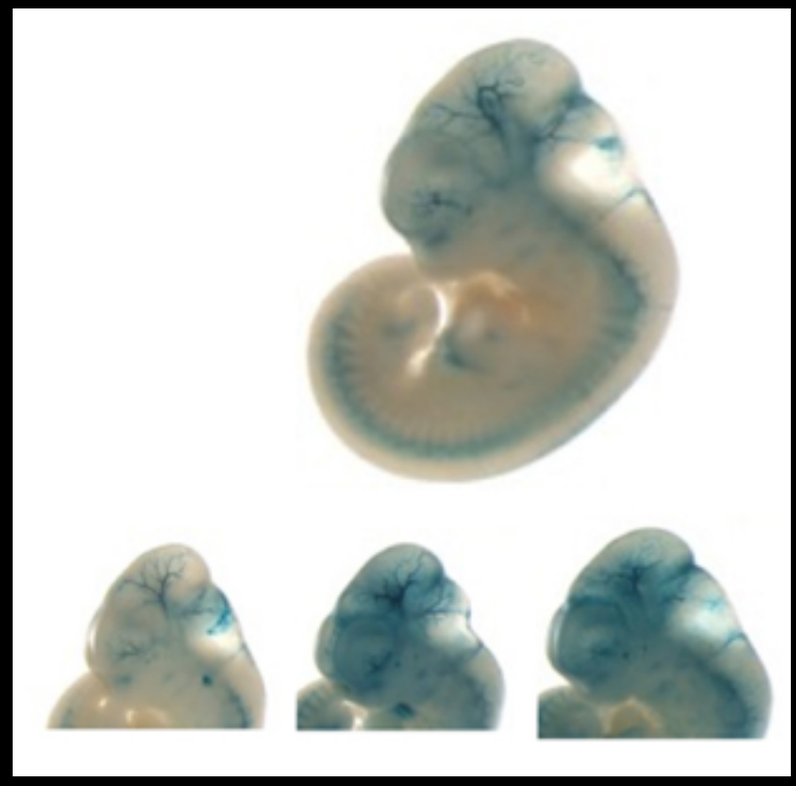
Experiments Currently being done

- Do match filter score around H3K27ac/DHS peaks and calculate ROC/PR curves.
- Extend H3K27ac/DHS peaks to 2 k and do same calculations

Low throughput assay - Len's assay

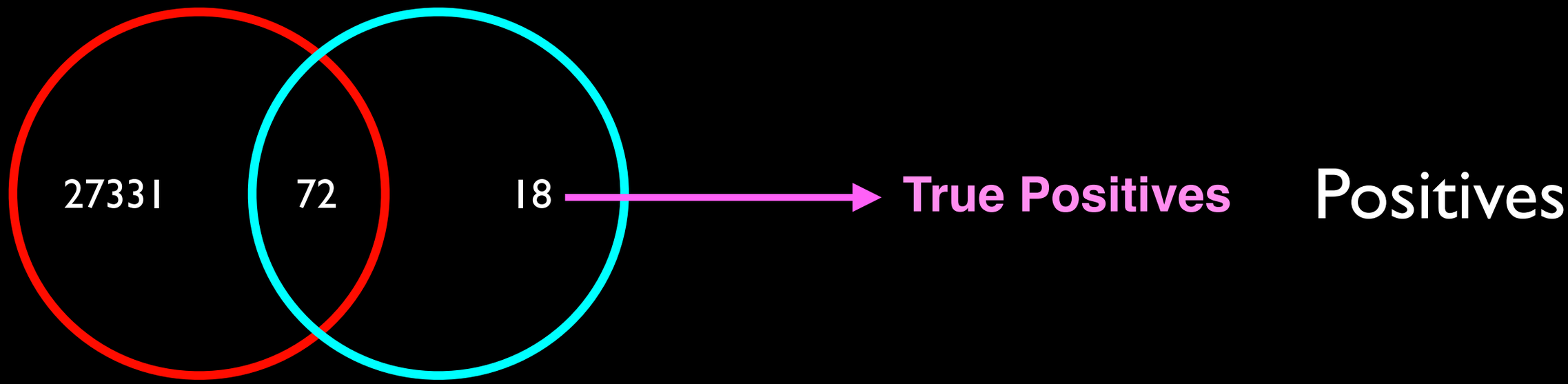


Clone enhancer sequence + minimal promoter + LacZ in to mice.
Strongest test for enhancer prediction.

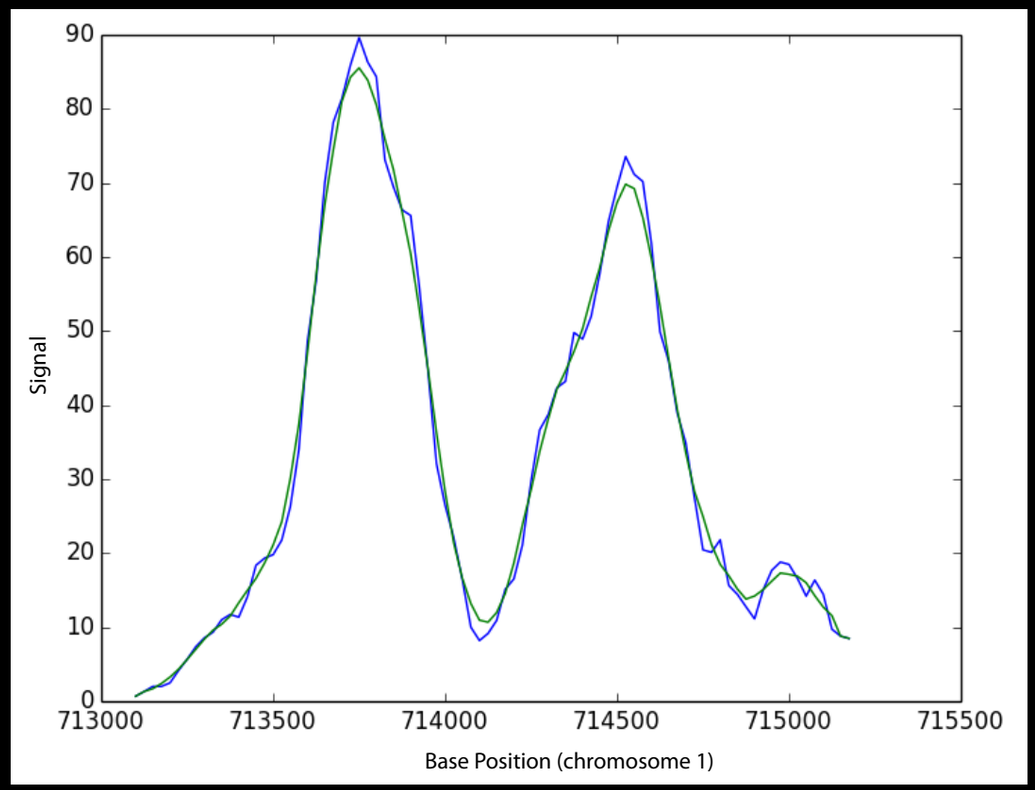


Very few tissue specific enhancers known
Region being tested is placed in chromosome environment and should include the chromatin signals as well.

The H3K27ac peaks are not sufficient to get the positives in the genome



Having a double peak requirement reduces the false positive rate



H3K27ac

True Positives

27331 | 72 | 18

H3K27ac peaks

True Negatives

9878 | 78 | 199

H3K27ac double peaks

Reduction in false positive rate