a) Tag cluster identification upstream known pre-miRNAs

- 33 RNA libraries of CAGE tags
- Ensembl66 annotated transcripts

b) Tag cluster filtering and background creation

- TSS1 and TSS2 with pre-miRNA
- TSS of another transcript
- Protein-coding Ensembl transcript

- Background noise
- Exon spanning


c) Modeling of putative promoters and background region features

- CpG density
- Conservation score
- TATA box affinity
- Normalized tag counts

- Mixture model for promoters and background sequences

- 1000 bp


d) Analysis of promoter sequences - Transcription factor binding site enrichment

- Promoter
- pre-miRNA