Computational identification of putative lincRNAs in mouse embryonic stem cell

Published: 07 October 2016

周润

2017 1 10

introduction

- mouse embryonic stem cells (ESCs) are pluripotent stem cells derived from the preimplantation
- lncRNAs：length longer than 200nt、no ORF、lower expression abundance and evolutionary conservation
- genomic features: Pol II transcripts with a poly-A tail and 5’ capping、exons and introns
- influence:gene expression regulatory networks、ESC differentiation

Problem:5’end incompleteness of lincRNA

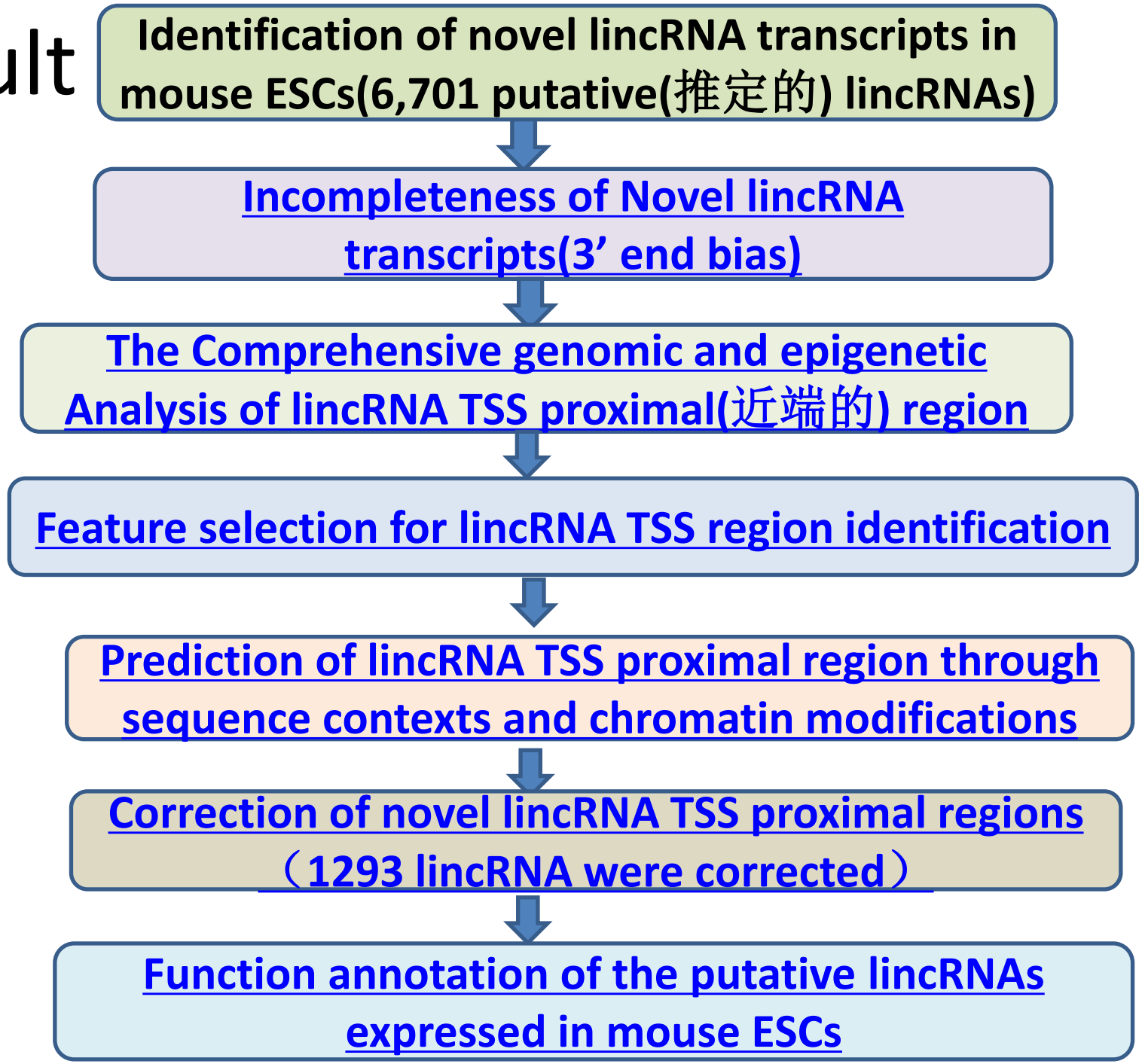
- RNA-seq data (polyA+ RNA by the first-strand with oligo-dT(primer) cause biased enrichment of the 3’ ends relative to 5’ ends)
- low expression abundance
- CG-rich(low read coverage by resisting the DNA denaturation(变性))

Solution: prediction model of lincRNA TSS proximal regions based on the machine learning method RBF SVM

Datasets

- 14 RNA-Seq data in mouse ESCs(fasta)
- mESC ChIP-Seq data of H3k4me3, H3k9ac, H3k27ac and PolII (fasta)
- CGIs and repeat elements annotations

Result


Identification of novel lincRNA transcripts in mouse ESCs(6,701 putative(推定的) lincRNAs)
Incompleteness of Novel lincRNA transcripts(3’ end bias)
The Comprehensive genomic and epigenetic Analysis of lincRNA TSS proximal(近端的) region
Feature selection for lincRNA TSS region identification
Prediction of lincRNA TSS proximal region through sequence contexts and chromatin modifications
Correction of novel lincRNA TSS proximal regions （1293 lincRNA were corrected）
Function annotation of the putative lincRNAs expressed in mouse ESCs

Conclusion

- a novel catalog of mouse ESCs-expressed lincRNAs with relatively complete transcript length

- putative lincRNAs expressed in mouse ESCs might be useful for the investigation of the transcriptional regulation and posttranscriptional regulation of the lincRNAs in mouse ESCs and even the mammalian development

启示：

严格的lincRNA的筛选标准

解决lincRNA5’端不完整性问题,为后续研究奠定基础

缺点：

不具有普遍性，未在其他模式生物中测试

thank you!