

# 期末报告

## 小鼠胚胎发育过程关键基因的挖掘

姓名：李群

老师：吴飞珍

学号：20111510021

邮箱：[20111510021@fudan.edu.cn](mailto:20111510021@fudan.edu.cn)

时间：2021年1月11日

题目：小鼠胚胎发育过程关键基因的挖掘

## 摘要以及问题

组蛋白修饰是控制许多关键细胞过程的基本表观遗传方式之一。组蛋白修饰是否能够从亲代传递到子代，以及如何传递仍是表观遗传学领域长久以来悬而未决的问题。清华大学颀伟组课题组通过优化ChIP技术开发出能够利用低细胞量的测序方法，在小鼠胚胎发育过程中成功节食组蛋白修饰的传递问题。

2016年，清华大学生命科学学院颀伟和医学院那洁研究组在Nature上发表研究论文[1]，报道了哺乳动物组蛋白修饰从如何从亲代传递到子代的，以及早期胚胎发育中组蛋白修饰遗传和重编程的模式和分子机制。本期末报告数据来自该篇论文，分析内容也是参考该论文，但是原文工作量很大，且着眼点是胚胎发育全过程的全局性，本报告则是希望找到在早期发育阶段ZGA中重要的生物过程，与原文内容并不重复。

## 数据介绍

数据来自GEO数据库（GSE71434）[1]，选取35个样本进行分析，具体如下。

## RNA-Seq（17个样本）

时期	重复数
Growing_oocyte_10days	1
Growing_oocyte_14days	1
GV_8weeks	1
MII_oocyte	2
zygote_PN5	2
2cell_early	2
2cell_late	2
4cell	2
8cell	2
ICM	2

## 方法

### RNA-Seq

RNA-seq分析采用的是HISAT2+StringTie+Ballgown流程[2]。将所有RNA-seq比对到小鼠参考基因组（mm9），基因表达水平由R包Ballgown进行计算。

### GO分析

采用的是R包clusterProfiler和org.Mm.eg.db，数据可视化采用的是R包ggplot2和pheatmap。另外其他R包见附录

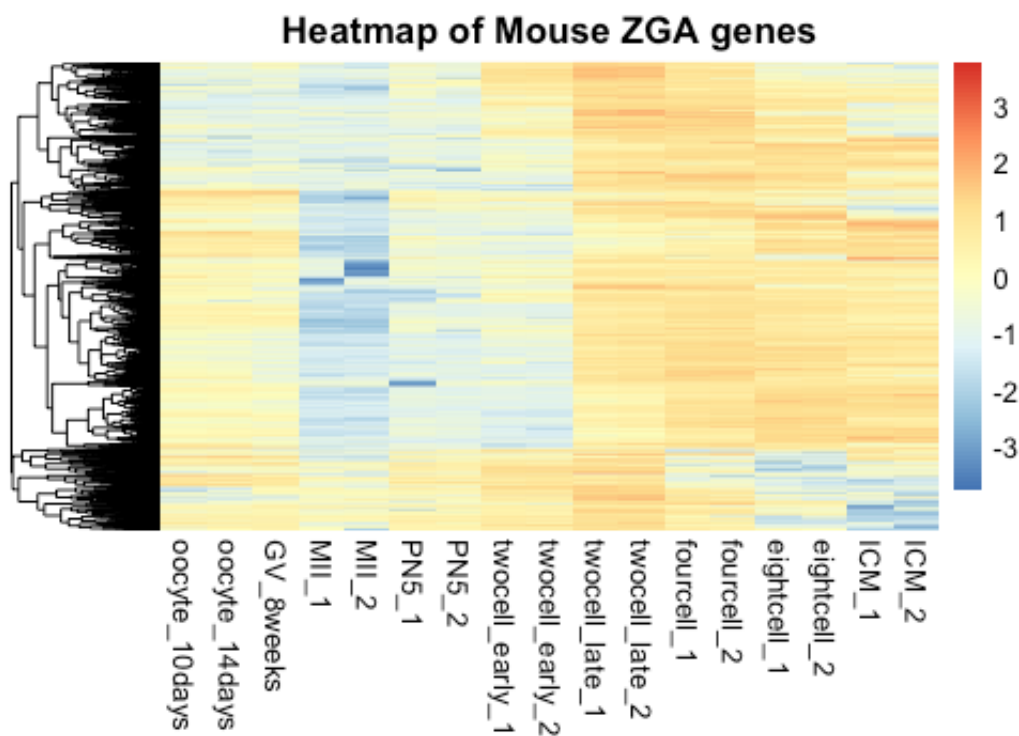
### 合子激活基因的筛选

通过选择在两细胞晚期高表达 (FPKM > 5) 和在MII中低表达的基因 (FPKM < 5) 来鉴定ZGA（合子激活）基因[3]。

另有参考：Genes with FPKM(one-cell) > 2 and FPKM(MII) < 0.5 in total RNA-seq data were defined as one-cell minor ZGA genes (the different cut-offs were in part due to the different data ranges between total RNA-seq and mRNA-seq data). Similarly, genes with FPKM(early two-cell) > 2, FPKM(MII) < 0.5 in total RNA-seq data were defined as early two-cell minor ZGA genes[3].

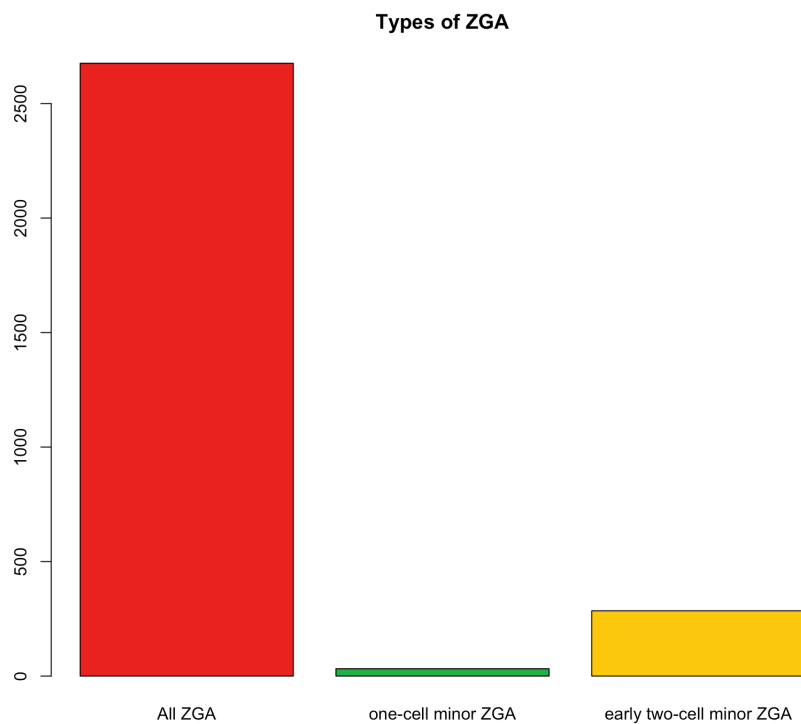
## 结果

## 1 小鼠ZGA基因的表达情况



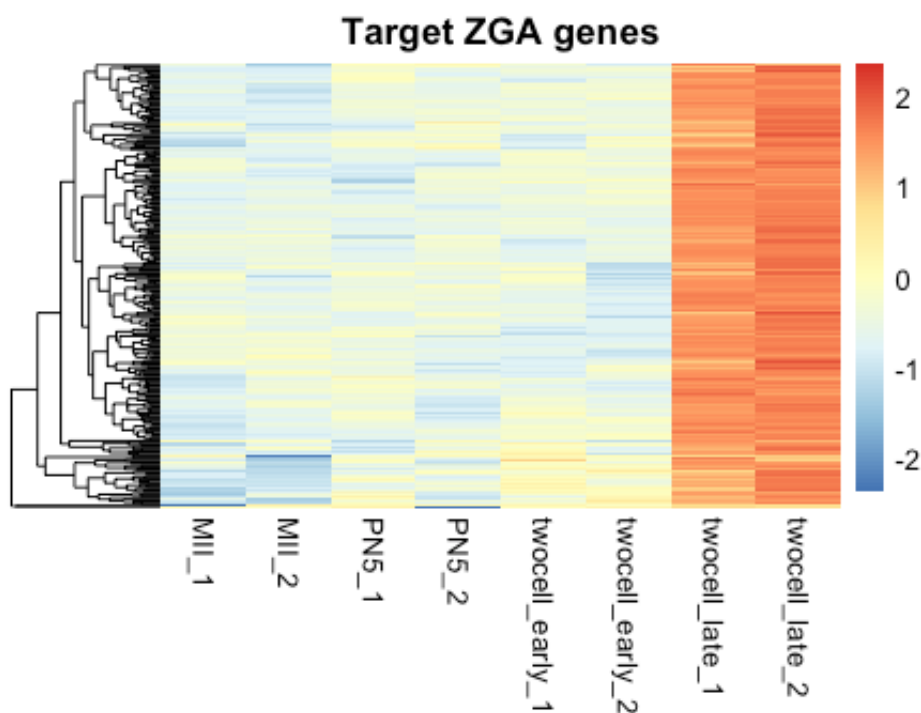
大部分与ZGA相关的基因在MII，PN5以及2cell早期处于低表达状态，在2cell晚期发生激活。

## 2. ZGA相关基因统计



颜伟课题组将ZGA的过程划分的更加细致，包含多个minor ZGA过程。

### 3. 关键基因的筛选



数值为log2FPKM

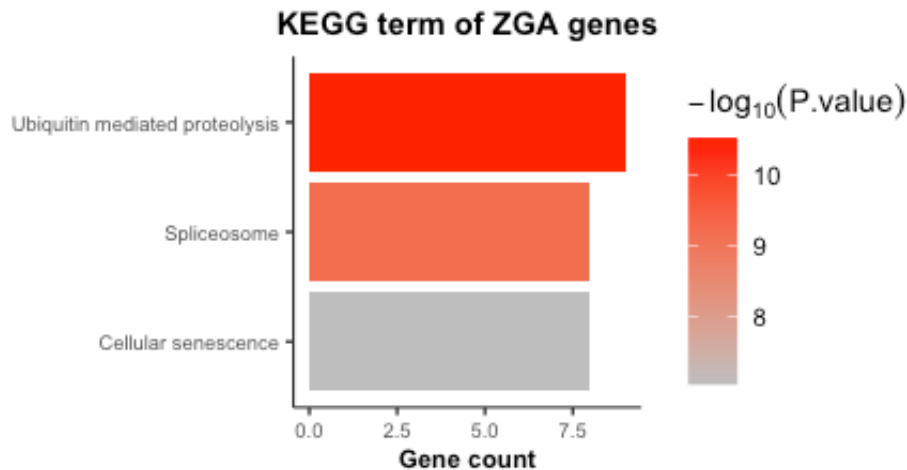
筛选出在MII、PN5和2cell早期均低表达，在2cell晚期高表达的基因（FoldChange $\geq$ 2），作为本次分析报告的目标基因。

### 4. GO



颜色表示p值，红色表示显著，灰色表示不显著。纵坐标表示GO注释，横坐标表示基因数

## 5. KEGG



颜色表示p值，红色表示显著，灰色表示不显著。纵坐标表示KEGG注释，横坐标表示基因数

经过注释发现，大部分基因与剪切相关。（GO中有多个GO注释条目，KEGG中存在一个与剪接体相关的条目），推测该过程（从MII到2细胞晚期的合子激活过程）与剪接因子相关，该过程产生多样性的RNA，从而翻译成多种蛋白，进一步实现功能。

## 讨论

合子基因组激活(ZGA)是生命中的第一个转录事件[4]。在ZGA之前，大部分的转录本都是母表达的且研究的不多，其转录过程中的剪切比较复杂[5]。在2008年，研究发现RNA聚合酶II转录终止过程涉及poly(A)位点的剪切[6]。2020年颀伟课题组发现RNA聚合酶II在ZGA中发挥重要作用[3]，另外作者团队还利用ChIP-Seq实验进一步进行研究发现，ZGA的过程是一个渐进的过程（包含多个minor ZGA过程）。

经过本次报告分析，可以判断与ZGA相关的基因可能参与到与剪切相关的生物过程。剪切在产生蛋白质的多样性上发挥了重大的作用，为接下来的发育过程提供多样的基础保障[7]。

## 参考

- [1] Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, Ming J, Wu X, Zhang Y, Xu Q, Liu W, Kou X, Zhao Y, He W, Li C, Chen B, Li Y, Wang Q, Ma J, Yin Q, Kee K, Meng A, Gao S, Xu F, Na J, Xie W. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*, 537(7621):553-557 (2016).
- [2] Pertea, M., Kim, D., Pertea, G. *et al.* Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*, 11, 1650–1667 (2016).
- [3] Liu, B., Xu, Q., Wang, Q. *et al.* The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature*, 587, 139–144 (2020).
- [4] Aoki, F., Worrall, D. M. & Schultz, R. M. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev. Biol* 181, 296–307 (1997).

- [5] Abe, K. et al. The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing. EMBO J 34, 1523–1537 (2015).
- [6] West S, Proudfoot NJ, Dye MJ. Molecular dissection of mammalian RNA polymerase II transcriptional termination. Mol Cell, 14;29(5):600-10 (2008)
- [7] Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. Function of alternative splicing. Gene, 1,514(1):1-30 (2013)

## 附件

---

1. [R 脚本](#)  
<https://github.com/leequn/biowufz/blob/main/report4/reportRscripts.R>
2. [表达数据](#)  
<https://github.com/leequn/biowufz/blob/main/report4/tmp.txt>

## 致谢

---

感谢吴老师、顾老师及其课题组成员的讲解与帮助

感谢复旦大学生物医学研究院，生物信息平台提供的服务器