

# 超声处理对 PRC2 相关蛋白染色质免疫共沉淀测序结果的影响

杨根生<sup>1</sup>, 焦芳芳<sup>1</sup>, 吕瑞途<sup>1</sup>, 郭睿<sup>1,2</sup>

1 复旦大学 生物医学研究院, 上海 200031

2 复旦大学 附属浦东医院肿瘤科, 上海 201399

杨根生, 焦芳芳, 吕瑞途, 等. 超声处理对 PRC2 相关蛋白染色质免疫共沉淀测序结果的影响. 生物工程学报, 2020, 36(2): 341–352.

Yang GS, Jiao FF, Lü RT, et al. Effect of sonication on results of ChIP-seq experiments involving PRC2 related proteins. Chin J Biotech, 2019, 36(2): 341–352.

**摘要:** 超声处理是染色质免疫共沉淀 (Chromatin immunoprecipitation, ChIP) 实验过程中染色质片段化的重要手段之一。选用多梳抑制复合物 2 (Polycomb repressive complex 2, PRC2) 相关蛋白组蛋白甲基转移酶 EZH2 及其催化产物 H3K27me3 为代表, 研究不同分子量大小的蛋白在不同超声处理时间下对染色质免疫共沉淀测序 (Chromatin immunoprecipitation sequencing, ChIP-seq) 实验的影响, 结果表明在启动子区或非启动子区, 不同超声时间下小分子量组蛋白 H3K27me3 结合位点的注释基因均无明显差异, 说明超声时间对组蛋白 ChIP-seq 数据影响不大。与组蛋白不同, 超声时间从 10 min 延长至 20 min 后启动子区 EZH2 新增结合位点的注释基因能够显著聚类在与肌动蛋白丝组装等相关通路上。超声 20 min 相比 10 min, 非启动子区 EZH2 新增基因的 GO (Gene ontology) 聚类通路要远多于丢失基因, 且超声 30 min 相比 20 min, 非启动子区丢失基因的 GO 聚类通路要远多于新增基因, 这些通路大多与 RNA 聚合酶 II (RNA polymerase II, RNAPII)、器官发育、细胞形态发生相关。这表明超声时间不足或过长均会导致 EZH2 的基因组定位信息的不全。另外, 超声主要影响启动子区中的 PRC2 非结合区域及二价启动子区域的 EZH2 结合位点, 还影响非启动子区中 PRC2 结合区域、PRC2 不结合区域以及活化态增强子区域的 EZH2 结合位点。综上, 建议对大分子量的染色质修饰相关蛋白优化超声处理时间, 使形成的染色质片段聚集在 100–500 bp 可获得比较全面的基因组信息。对小分子量的组蛋白来说, 超声时间对 ChIP-seq 结果影响不大。

**关键词:** 超声时间, 染色质片段化, EZH2, PRC2, 启动子, 增强子

Received: June 11, 2019; Accepted: August 30, 2019

Supported by: National Natural Science Foundation of China (No. 31771450).

Corresponding author: Rui Guo. Tel: +86-21-54237598; E-mail: guorui@fudan.edu.cn

国家自然科学基金 (No. 31771450) 资助。

网络出版时间: 2019-09-05

网络出版地址: <http://kns.cnki.net/kcms/detail/11.1998.Q.20190904.1746.002.html>

# Effect of sonication on results of ChIP-seq experiments involving PRC2 related proteins

Gensheng Yang<sup>1</sup>, Fangfang Jiao<sup>1</sup>, Ruitu Lü<sup>1</sup>, and Rui Guo<sup>1,2</sup>

*1 Institutes of Biomedical Sciences, Fudan University, Shanghai 200031, China*

*2 Department of Oncology, Pudong Medical Center, Fudan University, Shanghai 201399, China*

**Abstract:** Sonication is one of the essential strategies of chromatin fragmentation in Chromatin immunoprecipitation (ChIP) assay. The impact of proteins with different molecular weights generated under different duration of sonication on the results of Chromatin immunoprecipitation sequencing (ChIP-seq) experiments involving the Polycomb Repressive Complex 2 (PRC2) related protein EZH2, which is a histone methyltransferase, and its product H3K27me3 was investigated. The results indicate that in the promoter region or nonpromoter region, there were hardly any differences among the H3K27me3 peaks annotated genes from different duration of sonication, which suggesting that the duration of sonication had little effect on histone ChIP-seq results. In contrast, in the promoter region newly gained EZH2 peaks annotated genes at 20 min sonication time were significantly clustered in the gene ontology (GO) pathways related to actin filament bundle compared with 10 min. In the nonpromoter region, compared with 10 min, the GO pathways of newly gained EZH2 peaks annotated genes at 20 min sonication time is much more than that of lost EZH2 peaks annotated genes. And in the nonpromoter region, compared with 20 min, the GO pathways of lost EZH2 peaks annotated genes at 30 min sonication time is much more than that of newly gained EZH2 peaks annotated genes. Most of these pathways are associated with RNA polymerase II (RNAPII), organ development and cell morphogenesis. These suggest that the genomic information of EZH2 will be lost if the duration of sonication is not enough or too long. Different duration of sonication mainly affect the EZH2 peaks in PRC2 unoccupied region and the bivalent promoter in the promoter region, as well as the PRC2 occupied region, PRC2 unoccupied region and the activated enhancer in the nonpromoter region. Therefore, the sonication for the chromatin related proteins with high molecular weights need to be optimized to make chromatin fragments size vary from 100 bp to 500 bp, which will yield relatively comprehensive genomic information of the target protein. For histones, which are of small molecular weights, duration of sonication has little effect on them.

**Keywords:** duration of sonication, fragmentation of chromatin, EZH2, PRC2, promoter, enhancer

染色质免疫共沉淀是目前表观遗传学及分子生物学的研究手段之一,可以分析蛋白质和 DNA 在天然染色质环境中的相互作用<sup>[1]</sup>。将 ChIP 实验与二代高通量测序结合,探究目的蛋白在基因组上分布情况的技术称为染色质免疫共沉淀测序 (Chromatin immunoprecipitation sequencing, ChIP-seq)<sup>[2]</sup>。ChIP-seq 技术主要用于研究目的蛋白 (例如表观遗传修饰的组蛋白、染色质修饰蛋白、转录因子、辅因子及其他染色质相关蛋白) 在基因组上的定位及富集分析。ChIP-seq 实验的主要流程包括: 1) 交联,通过甲醛处理等方法将染色质上的 DNA 与目的蛋白交联; 2) 染色质片段化,运用超声或酶解的方法将染色质片段化; 3) 蛋白质免疫沉淀,将结合了目的蛋白的染色质片段通过

抗体进行富集纯化; 4) 解交联,通过加热等方式解除 DNA 与目的蛋白的交联; 5) DNA 纯化,添加蛋白酶 (Protease) 和核糖核酸酶 (RNase) 酶解消化去除解交联产生的蛋白和 RNA,用 PCR 回收试剂盒纯化与目的蛋白结合的 DNA<sup>[3]</sup>; 6) 建库并进行二代测序。对 ChIP-seq 影响最大的环节为染色质片段化和蛋白质免疫沉淀,这两步都需要根据具体研究的蛋白去调整实验条件,本文就针对染色质片段化环节进行讨论。

染色质片段化主要有超声和酶解两种方式。对于酶解法来说,目前主要采用微球菌核酸酶 (Micrococcal nuclease, MNase) 或商业试剂盒中的混合酶。酶解法的优点在于比超声更加温和,对蛋白与 DNA 以及蛋白与蛋白之间的相互作用影响更

小,但仍存在很多问题,例如酶解主要消化核小体之间的连接 DNA,所以只能形成整数倍核小体长度的染色质片段,且大多数为一个核小体长度,这样不仅会丢失一些结合在连接 DNA 区域的转录因子的信息,而且还会降低一些横跨多个核小体长度的染色质结合蛋白的富集度。另外较小的消化片段会使 qPCR 验证更加困难,酶切温度 37 °C 会影响蛋白表位,从而降低免疫沉淀效率等<sup>[4]</sup>。因此,目前 ChIP 实验采用的最主流的染色质片段化方式仍然为超声破碎。超声主要受到细胞类型、细胞密度、ChIP 裂解缓冲液、超声时间以及目的蛋白种类 5 个因素影响。一般使细胞密度低于  $10^7$  个/mL 进行超声,对于第一次进行 ChIP 实验的蛋白,一般会使用较温和的裂解缓冲液(含 0.1% SDS),实验过程中控制染色质片段化程度最常用方法是调节超声时间的长短。超声时间越长,染色质片段化程度越高,但时间过长超声波可能会破坏目的蛋白-DNA 之间的相互作用以及蛋白的稳定性,反之超声时间越短,染色质片段化会不彻底,染色质中不易打开的部分会形成较长的染色质片段,这会导致距离目的蛋白实际基因组定位很远的位点也被认为有目的蛋白的富集,从而导致 ChIP-seq 的分辨率大大降低。目前一般认为超声片段分布在 100–1 000 bp 范围内为宜,最好集中分布在 500 bp 以内,即大约 3 个核小体长度以内<sup>[2,4-5]</sup>。过长的染色质片段会在建库的过程中通过切胶回收的方式去除。

为了研究不同分子量蛋白在不同超声时间对 ChIP-seq 数据的影响,选取表观遗传领域已研究较多的 PRC2 相关蛋白 EZH2(分子量为 85 kDa) 和其催化生成的 K27 位具有三甲基化修饰的组蛋白 H3K27me3(分子量为 17 kDa) 为例进行探讨。EZH2 作为 PRC2 的催化亚基,是一个含有 SET 结构域的组蛋白甲基转移酶(Histone methyltransferase),能够甲基化组蛋白 H3K27<sup>[6-8]</sup>。PRC2 能够与组蛋白去乙酰化酶 HDACs 互作将乙酰化转变为甲基化从而抑制基因表达<sup>[9]</sup>,PRC2 在一些发育调节基因的启动子上形成 H3K27me3 起到转录抑制作

用<sup>[10-14]</sup>。PRC2 催化生成 H3K27me3 后还可招募多梳抑制复合物 1 (Polycomb repressive complex 1, PRC1) 泛素化 H2AK119 从而保持基因的抑制<sup>[15]</sup>。EZH2 参与胚胎发育、干细胞干性维持以及造血细胞增殖并起到抑制或沉默基因表达的作用。在慢性淋巴细胞性白血病、前列腺癌、胃癌、胶质瘤、乳腺癌等多种癌症中 EZH2 呈现高表达,而且 EZH2 的高表达对肿瘤的发生、增殖、侵袭和转移均有促进作用<sup>[16-17]</sup>。有研究表明 EZH2 可不依赖于 PRC2 独立行使转录激活因子的功能,例如在乳腺癌中 EZH2 不依赖于 PRC2 作为转录激活因子激活 NF- $\kappa$ B 靶基因从而促进乳腺癌的发展<sup>[9,18]</sup>。由此可见,目前对 EZH2 和 H3K27me3 染色质调节机制研究比较广泛,可供参考的 ChIP-seq 数据和信息较多,因此可选择这两个蛋白为代表来量化不同程度的染色质片段化对目的蛋白 ChIP-seq 数据的影响。使用相同的细胞数、不同的超声时间对以 H3K27me3 为代表的组蛋白和以 EZH2 为代表的染色质修饰蛋白进行 ChIP-seq 实验,在相同的测序深度下进行数据分析,探究能够产生最全面 ChIP-seq 数据信息的超声条件。

## 1 材料与方法

### 1.1 材料

#### 1.1.1 耗材及试剂

胎牛血清 FBS、MEM 非必需氨基酸 NEAA、 $\beta$ -巯基乙醇购自 Gibco 公司;DMEM 高糖培养基、双抗(氨苄青霉素和链霉素)、磷酸盐缓冲液(PBS)、0.25%胰酶为 Hyclone 公司产品;明胶(Gelatin)、脱氧胆酸钠(Na-deoxycholate)、二硫苏糖醇(DTT)、苯甲基磺酰氟(PMSF)、多聚甲醛、乙基苯基聚乙二醇(NP-40)、乙二胺四乙酸(EDTA)、无 EDTA 的蛋白酶抑制剂混合物(EDTA-free protease inhibitor cocktail)为 Sigma 公司产品;LIF 购自 Millipore 公司;Hepes、Tris、Triton X-100、SDS 购自 Amresco 公司;PCR 产物回收试剂盒购自 Qiagen 公司;NaCl、LiCl 购自国

药集团; RNase A 购自 Thermo 公司; Proteinase K 为 TaKaRa 公司产品; 免疫磁珠 Dynabeads protein A/G 购自 Invitrogen 公司; KAPA Hyper 文库构建试剂盒购自罗氏公司; H3K27me3 兔源单抗、Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb 购自 CST 公司; EZH2 兔源单抗、Ezh2 (D2C9) Rabbit mAb 购自 CST 公司。

### 1.1.2 仪器设备

细胞超净台、细胞培养箱购自 Thermo 公司; QSONICA 超声波破碎仪为 Misonix 公司产品; 脱色摇床购自其林贝尔仪器制造有限公司; 恒温振荡仪 ThermoMixer 为 Eppendorf 公司产品; DynaMag-2 磁力架购自 Life Technologies 公司。

## 1.2 方法

### 1.2.1 小鼠胚胎干细胞培养

用适量含 0.1% 明胶 (Gelatin) 的灭菌水溶液铺满细胞培养皿, 室温放置 30 min 以上。吸除盛有细胞的培养基, PBS 清洗。0.05% 胰酶消化约 3 min 后加入 3 mL 新鲜的小鼠胚胎干细胞培养基 (DMEM 高糖培养基, 10% FBS, 1% 双抗, 1% NEAA, 0.1%  $\beta$ -ME, 1 000 U/mL LIF) 重悬, 离心去上清。吸除 Gelatin, 按需要的细胞量加入胰酶消化后重选的细胞悬液及新鲜培养基, 轻轻摇匀, 放入 37 °C、5% CO<sub>2</sub> 浓度的细胞培养箱中。小鼠胚胎干细胞生长速度很快, 一天传两代, 故一般按照 1 : 10 隔天进行传代培养。

### 1.2.2 交联

向盛有约  $6 \times 10^6$  细胞数量的 6 cm 细胞培养皿中加入终浓度为 1% 的多聚甲醛 (4 mL 培养基加入 37% 多聚甲醛 108  $\mu$ L), 在室温脱色摇床上缓慢转动进行蛋白质与 DNA 的固定, 10 min 后加入 2.5 mol/L 的甘氨酸溶液 200  $\mu$ L 终止交联, 反应 5 min 后吸除上清并用预冷的 PBS 清洗 2 次, 加入 PBS 刮下细胞。

### 1.2.3 超声

用 1 mL ChIP 裂解缓冲液 (ChIP lysis buffer: 50 mmol/L Hepes (pH 7.5), 500 mmol/L NaCl,

1 mmol/L EDTA (pH 8.0), 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, 0.1% DTT, 1 mmol/L PMSF, 2% EDTA-free Protease Inhibitor Cocktail) 重悬细胞放入 QSONICA 超声仪中进行超声处理。将细胞超声破碎液 13 000 r/min 离心 10 min, 取 2% 体积 (20  $\mu$ L) 上清作为 Input 检查超声情况, 剩下的上清冻于 -80 °C 或放置冰上待超声情况检测结束之后使用。

### 1.2.4 解交联

向 Input 中加入 5 倍体积 (100  $\mu$ L) 的 ChIP 洗脱缓冲液 (ChIP Wash Buffer: 50 mmol/L Hepes (pH 7.5), 500 mmol/L NaCl, 1 mmol/L EDTA (pH 8.0), 1% Triton X-100, 0.1% Na-deoxycholate), 放入 65 °C 的恒温振荡仪 (ThermoMixer) 4 h 以上解交联。

### 1.2.5 消化 RNA 和蛋白质

加入等体积 (120  $\mu$ L) 无 NaCl 的 ChIP TE 缓冲液 (ChIP TE Buffer without NaCl: 50 mmol/L Tris-HCl (pH 8.0), 10 mmol/L EDTA (pH 8.0)) 以及 4  $\mu$ L RNase A, 4  $\mu$ L ChIP 蛋白酶 K 消化缓冲液 (ChIP proteinase K digest buffer: 300 mmol/L CaCl<sub>2</sub>, 100 mmol/L Tris-HCl (pH 8.0)) 以及 4  $\mu$ L Proteinase K 放入 50 °C 的恒温振荡仪上消化 1 h。

### 1.2.6 DNA 纯化与超声检测

用 Qiagen PCR 产物回收试剂盒纯化 Input 中的 DNA, 取 1% Input 跑胶 (2% 的琼脂糖凝胶) 检测超声情况, 剩下 1% Input 作为 ChIP 的阴性对照。

### 1.2.7 免疫沉淀

检测超声片段处于预期范围内后, 向剩余的超声破碎液中加入等体积的 ChIP 稀释缓冲液 (ChIP Dilution Buffer: 50 mmol/L Hepes (pH 7.5), 500 mmol/L NaCl, 1 mmol/L EDTA (pH 8.0)), 目的是将 SDS 等去垢剂浓度降低 1 倍以创造温和的结合条件, 再加入目的蛋白的抗体进行蛋白质免疫沉淀, 放入 4 °C 旋转培养器中孵育 4 h 以上或过夜。再加入免疫磁珠 Dynabeads protein A/G 各 30  $\mu$ L 继续 4 °C 孵育 2 h。在 DynaMag-2 磁力架上

用预冷的 ChIP 洗涤缓冲液 (ChIP Wash Buffer: 50 mmol/L Hepes (pH 7.5), 500 mmol/L NaCl, 1 mmol/L EDTA (pH 8.0), 1% Triton X-100, 0.1% Na-deoxycholate), ChIP RIPA 缓冲液 (ChIP RIPA buffer: 50 mmol/L Hepes (pH 7.5), 300 mmol/L LiCl, 1 mmol/L EDTA, 0.5% NP-40, 0.7% Na-deoxycholate), 以及 ChIP TE 缓冲液 (ChIP TE buffer: 50 mmol/L Tris-HCl (pH 8.0), 10 mmol/L EDTA (pH 8.0), 50 mmol/L NaCl) 洗磁珠 (Beads), 次数分别为 3, 3, 1。加入 100  $\mu$ L ChIP Elution Buffer, 后续解交联、消化 RNA 和蛋白、DNA 纯化步骤参考前面对 Input 操作的流程。

### 1.2.8 qPCR 验证并进行建库

检测 ChIP 的成功主要手段并非测量 ChIP 得到的 DNA 总量的大小, 而是需要经过 qPCR 计算目的蛋白在基因组阳性区域相对阴性区域的富集度来衡量。ChIP 得到的 DNA 总量中有很多属于背景 DNA 故不应作为判断 ChIP 好坏的标准, 一般目的蛋白阳性区域相比阴性区域的富集度低于 4 倍认为 ChIP 没有成功。将纯化得到的 DNA 按照 KAPA Hyper 文库构建试剂盒建库并进行二代测序。

### 1.2.9 ChIP-seq 生物信息学分析

用 Trim galore (version 0.5.0) 进行数据清洗; 用 Bowtie2 (version 2.4.3.3) 进行比对; 用 Samtools (version 1.6) 进行排序和去重; 用 macs2 (version 2.1.1) 识别峰 (callpeak); 用 Deeptools (version 3.2.0) 绘制目的蛋白基因组平均密度分布图以及相关散点图; 用 R 包 ChIPseeker (version 1.16.1) 对 peaks 注释和可视化<sup>[19]</sup>; 用 R 包 Eulerr (version 5.1.0) 进行 venn 图绘制; 用 R 包 ClusterProfiler (version 3.8.1) 进行 GO 富集分析<sup>[20]</sup>。

## 2 结果与分析

### 2.1 超声时间对染色质片段化的影响

使用细胞数为  $6 \times 10^6$  的小鼠胚胎干细胞进行甲醛交联并裂解, 分别超声 10 min、20 min、30 min 处理。超声后纯化染色质得到 DNA 片段大小随

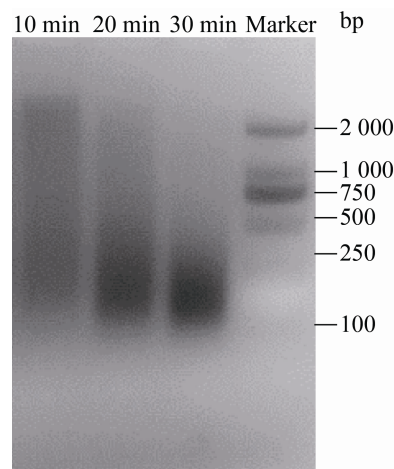


图 1 不同超声时间下染色质片段凝胶电泳图

Fig. 1 Agarose gel electrophoresis for chromatin fragment from different duration of sonication.

超声时间延长而变化(图 1)。超声 10 min 所得 DNA 在 100–2 000 bp 之间均匀分布; 超声 20 min 所得 DNA 在 100–2 000 bp 之间分布, 但在 100–500 bp 富集较多; 30 min 超声所得 DNA 主要富集在 100–300 bp 之间。可以看出超声时间对染色质片段化后的 DNA 大小影响较大。

### 2.2 超声对组蛋白修饰染色质富集的影响

对  $2 \times 10^6$  个小鼠胚胎干细胞在超声 10 min、20 min 和 30 min 时分别进行 H3K27me3 的 ChIP-seq 实验, 得到 H3K27me3 基因组结合位点分别为 16 348、16 837、19 443 个, 结合位点随超声时间增加而增加。对 H3K27me3 结合位点在基因组各类区域分布情况进行分析, 结果 (图 2A) 显示 3 种超声时间条件得到的基因组结合位点在基因组各区域的分布相差不大, 超声 30 min 相比 20 min 在启动子区 H3K27me3 的结合位点分布占比略有降低, 基因间隔区略有升高。H3K27me3 结合位点在全基因组上的平均密度分布分析显示, 超声时间对 H3K27me3 平均密度分布无显著影响 (图 2B), 且可看出 H3K27me3 主要富集在启动子区域, 这提示 H3K27me3 在启动子区和非启动子区的功能具有差异, 故后续对启动子区和非启动子区分别进行分析。

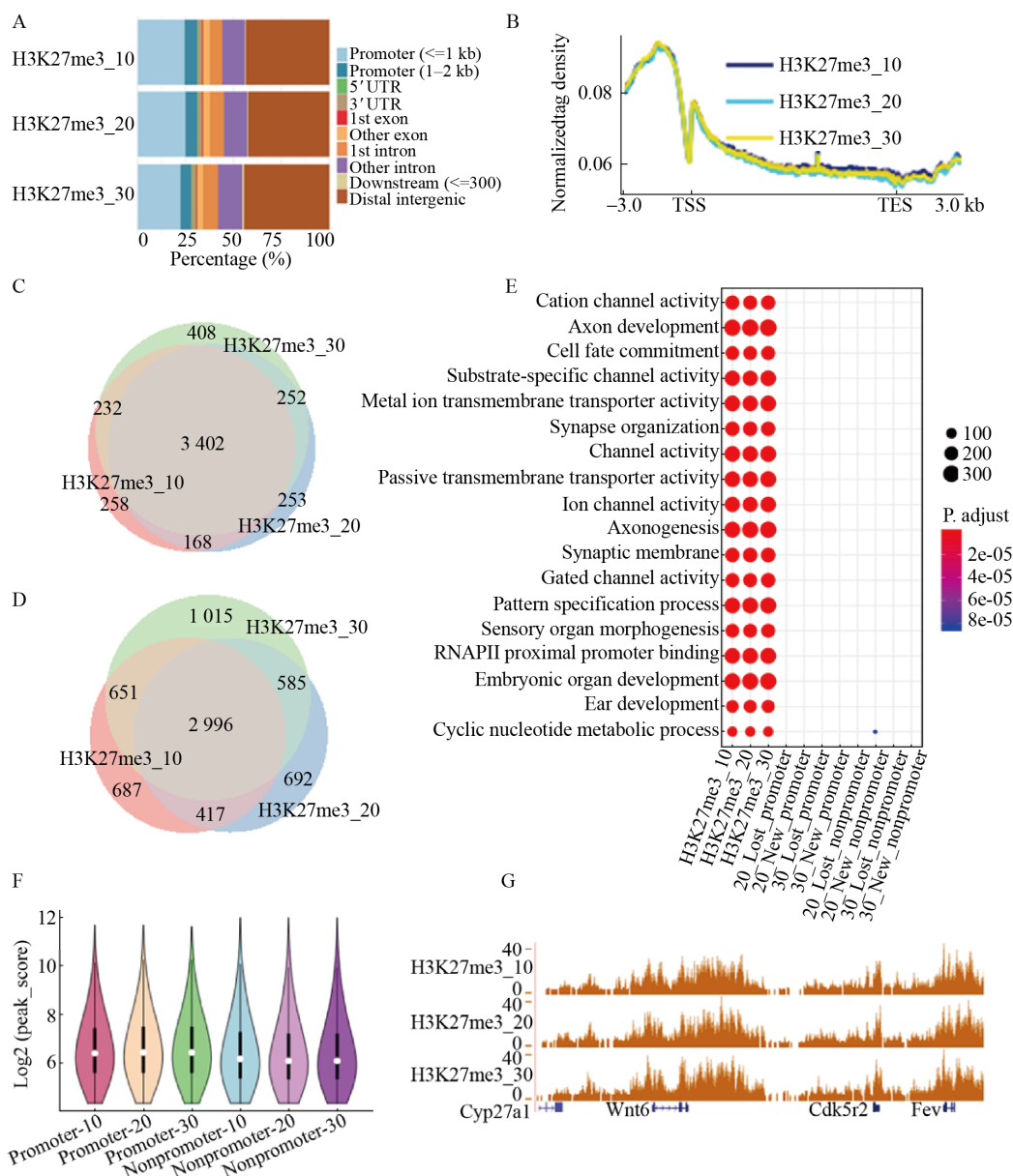


图 2 超声对 H3K27me3 ChIP-seq 的影响

Fig. 2 The effect of sonication on H3K27me3 ChIP-seq. (A) Genomic distribution of H3K27me3 ChIP-seq peaks in mESCs. (B) Average genome-wide occupancies of H3K27me3 along the transcription unit. The gene body length is aligned by percentage from the TSS to TTS. Three kilobases upstream of TSS and 3 kb downstream of TTS are also included. (C) Venn diagram showing H3K27me3 peaks annotated genes from different duration of sonication in promoter region. (D) Venn diagram showing H3K27me3 peaks annotated genes from different duration of sonication in nonpromoter region. (E) GO enrichment analysis comparison ( $P < 0.01$ ): H3K27me3\_10, H3K27me3\_20, H3K27me3\_30 represents the H3K27me3 peaks annotated genes at 10 min, 20 min and 30 min sonication time; 20\_Lost\_promoter means the lost H3K27me3 peaks annotated genes in promoter region after prolonged sonication from 10 min to 20 min; 20\_New\_promoter means the newly gained H3K27me3 peaks annotated genes in promoter region after prolonged sonication from 10 min to 20 min; 30\_Lost\_promoter means the lost H3K27me3 peaks annotated genes in promoter region after prolonged sonication from 20 min to 30 min; 30\_New\_promoter means the newly gained H3K27me3 peaks annotated genes in promoter region after prolonged sonication from 20 min to 30 min. (F) Chromatin accessibility of H3K27me3 peaks from different duration of sonication. (G) Genome-browser view of H3K27me3 ChIP-seq peaks from different time duration of sonication.

由于不重叠但位置相近的基因组结合位点的注释基因可能相同,所以结合位点之间的重叠情况不能准确反映不同超声时间下的真实差异。因此将 H3K27me3 基因组结合位点注释到最邻近基因上,得到启动子区不同超声时间下 H3K27me3 结合位点注释基因数为 4 060、4 075、4 294 个(图 2C),非启动子区基因数为 4 751、4 690、5 247 个(图 2D)。由不同超声时间下注释基因的韦恩重叠分析可发现启动子区基因重叠比例高于非启动子区,这提示超声对启动子区域的影响相比非启动子区要小(图 2C-D)。将不同超声时间下 H3K27me3 的基因组结合位点注释基因进行 GO 聚类分析,结果显示超声时间 10 min、20 min、30 min 得到的 GO 通路数量分别为 1 479、1 539、1 506 ( $P < 0.01$ ),且最显著富集的前 10 条通路与已报道文献一致能够显著富集在轴突发育 (Axon development)、细胞命运决定 (Cell fate commitment)、胚胎器官发育 (Embryonic organ development) 等 H3K27me3 主要相关功能的通路上(图 2E)<sup>[21-23]</sup>。随着超声时间延长,除了非启动子区域 20 min 相比 10 min 的新增基因有一条显著富集的通路外,其他区域均未显著富集。图 2F 表示不同超声时间得到的 H3K27me3 基因组结合位点的染色质开放程度的比较,由小提琴箱式图可看出启动子区与非启动子区中不同超声时间对 H3K27me3 基因组结合位点的染色质开放程度几乎没有影响,同时可看到启动子区的 H3K27me3 的开放程度大于非启动子区。图 2G 可直观地看到不同超声时间下 H3K27me3 的基因组结合位点几乎没有差异。

综上所述,在启动子区或非启动子区,不同时间超声处理得到组蛋白 H3K27me3 结合位点的注释基因及染色质开放程度均无明显差异,且超声时间对 H3K27me3 在基因组各区域分布情况以及平均分布密度影响不大。

### 2.3 超声对非组蛋白染色质富集的影响

为了研究超声时间对不同分子量蛋白的影响,使用不同超声时间处理的  $6 \times 10^6$  个小鼠胚胎干细胞对较大分子量的非组蛋白 EZH2 进行 ChIP-seq 实验,与较小分子量的组蛋白 H3K27me3 进行对比分析。EZH2 作为 PRC2 的组分,是 H3K27 的甲基转移酶,分子量约 85 kDa。ChIP-seq 分析结果显示超声 10 min、20 min 和 30 min 分别得到 EZH2 基因组结合位点数为 7 334、7 890、7 789。随着超声时间延长,分布在基因组各区域的 EZH2 结合位点的比例波动不大(图 3A)。EZH2 结合位点在全基因组上的平均密度分布分析显示,超声时间对平均密度分布无明显影响(图 3B)。

启动子区不同超声时间下 EZH2 结合位点的注释基因数为 4 320、4 624、4 545 个(图 3C),非启动子区注释基因数为 2 127、2 255、2 207 个(图 3D)。由韦恩重叠分析可发现启动子区基因重叠比例高于非启动子区,这提示与 H3K27me3 类似,超声对 EZH2 启动子区域的影响相比非启动子区也小(图 3C-D)。将不同超声时间条件下 EZH2 基因组结合位点注释基因进行 GO 聚类分析,结果表明超声时间 10 min、20 min、30 min 得到的 GO 通路数分别为 2 604、2 727、2 672 ( $P < 0.01$ ),且最显著富集的前 10 条通路与已报道文献一致,能够显著富集在分化发育等 EZH2 主要相关功能的通路上,例如轴突发育 (Axon development)、胚胎器官发育 (Embryonic organ development)、细胞命运决定 (Cell fate commitment)<sup>[24-25]</sup>。在启动子区超声 20 min 相比 10 min 新增的 EZH2 结合位点注释基因富集在 4 条 GO 通路上,例如肌动蛋白丝成束 (Actin filament bundle)、应力纤维 (Stress fiber) 等通路上,启动子区其他部分均无显著富集。

对于非启动子区,超声 20 min 相比 10 min 新增基因的富集通路 (113 条) 远多于丢失基因富集的通路 (8 条),30 min 比 20 min 丢失基因富



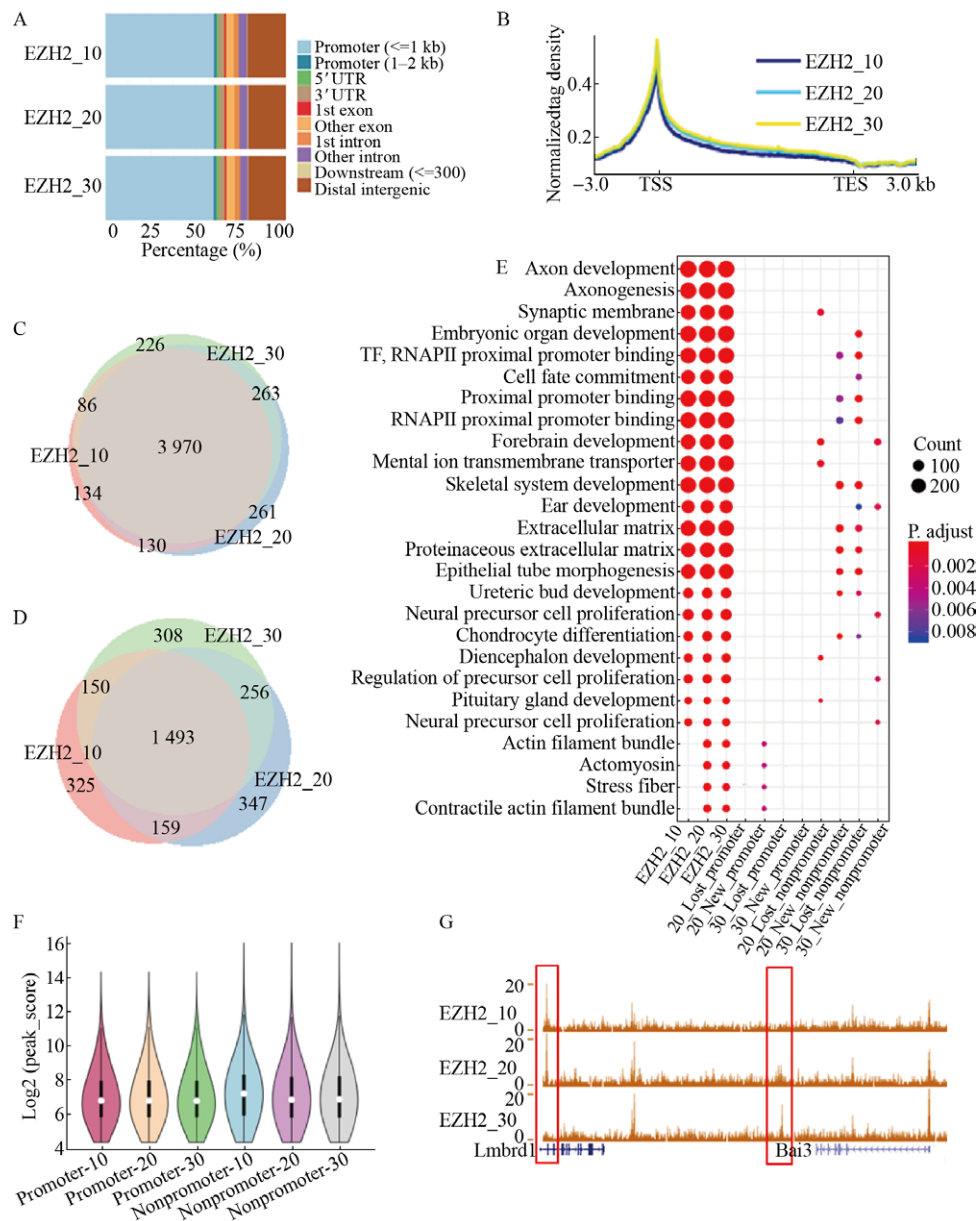


图 3 超声对 EZH2 ChIP-seq 的影响

Fig. 3 The effect of sonication on EZH2 ChIP-seq. (A) Genomic distribution of EZH2 ChIP-seq peaks in mESCs. (B) Average genome-wide occupancies of EZH2 along the transcription unit. The gene body length is aligned by percentage from the TSS to TTS. Three kilobases upstream of TSS and 3 kb downstream of TTS are also included. (C) Venn diagram showing EZH2 peaks annotated genes from different duration of sonication in promoter region. (D) Venn diagram showing EZH2 peaks annotated genes from different duration of sonication in nonpromoter region. (E) GO enrichment analysis comparison ( $P < 0.01$ ): EZH2\_10, EZH2\_20, EZH2\_30 represents the EZH2 peaks annotated genes at 10 min, 20 min and 30 min sonication time; 20\_Lost\_promoter means the lost EZH2 peaks annotated genes in promoter region after prolonged sonication from 10 min to 20 min; 20\_New\_promoter means the newly gained EZH2 peaks annotated genes in promoter region after prolonged sonication from 10 min to 20 min; 30\_Lost\_promoter means the lost EZH2 peaks annotated genes in promoter region after prolonged sonication from 20 min to 30 min; 30\_New\_promoter means the newly gained EZH2 peaks annotated genes in promoter region after prolonged sonication from 20 min to 30 min. (F) Chromatin accessibility of EZH2 peaks from different duration of sonication. (G) Genome-browser view of EZH2 ChIP-seq peaks from different time duration of sonication.



集的通路 (149 条) 也远多于新增基因富集的通路 (17 条), 在前 20 条最显著富集的通路中仍然是非启动子区 20 min 相比 10 min 新增的通路多于丢失, 非启动子区 30 min 丢失的通路多于新增, 且这些通路主要与 RNAPII、器官发育、细胞形态发生和细胞分化相关 (图 3E)。这表明超声时间不足 (10 min) 或超声时间过长 (30 min) 均会造成得到的 EZH2 基因组结合信息不全。图 3F 表示不同超声时间得到的 EZH2 基因组结合位点的染色质开放程度的比较, 由小提琴箱式图可看出启动子区不同超声时间对 EZH2 基因组结合位点的染色质开放程度几乎没有影响, 但超声 20 min 和 30 min 相比 10 min, EZH2 基因组结合位点的染色质开放程度均发生下降, 这提示超声时间延长可能打开了更多的染色质开放程度低的区域, 即染色质致密不易被超声破碎的部分。图 3G 可直观地看到不同超声时间下 EZH2 的基因组结合位点存在差异。

综上, 虽然超声时间对非组蛋白 EZH2 在基因组各区域分布情况以及平均分布密度影响不大, 但相比 H3K27me3, EZH2 ChIP-seq 数据受超声时间影响较大。而且超声对非启动子区的影响大于启动子区。相比 10 min 和 30 min, 超声 20 min 即超声片段集中分布在 100–500 bp 时, 能够拿到更全面的 EZH2 基因组结合信息。

#### 2.4 EZH2 随超声时间延长丢失和新增基因组结合位点的具体定位

由上一节的分析结果可知 20 min 为最优的超声条件, 选取超声 20 min 相比 10 min 启动子区新增的 EZH2 基因组结合位点、超声 20 min 相比 10 min 非启动子区新增的 EZH2 基因组结合位点以及 30 min 相比 20 min 非启动子区丢失的 EZH2 基因组结合位点这 3 部分为代表探究超声时间具体影响了基因组的哪些区域。

SUZ12 和 EED 是 EZH2 行使催化功能所必需的两个非催化亚基<sup>[9]</sup>, EZH2 和 SUZ12 及 EED 三

者共定位的基因组结合位点可认为是 PRC2 的结合区域。EZH2 除了作为表观遗传书写器 (Writer), 依赖 PRC2 起到转录抑制的作用外, 还被报道在多种癌症中能够不依赖于 PRC2 作为转录激活因子独立行使转录激活功能<sup>[9,18]</sup>。故可分析超声时间影响了基因组上 PRC2 结合区域还是 PRC2 不结合区域的 EZH2 结合位点。另外对于启动子区和非启动子区还可进一步细分, 启动子区依据组蛋白修饰可划分为 H3K4me3 和 H3K27me3 共定位的二价启动子 (Bivalent promoter)、只有 H3K4me3 而没有 H3K27me3 的活跃启动子 (Active promoter)、只有 H3K27me3 而没有 H3K4me3 的抑制启动子区 (Repressed promoter)<sup>[26]</sup>, 非启动子区最主要的特征就是散布着增强子, 根据组蛋白修饰可分为具有 H3K4me1 和 H3K27ac 的活化态增强子 (Active enhancer)、只有 H3K4me1 的待发态增强子 (Primed enhancer) 以及具有 H3K4me1 和 H3K27me3 的准备态增强子 (Poised enhancer), 也可称为静态增强子<sup>[27]</sup>。故还可以分析不同超声时间影响了哪一类启动子或增强子区域的 EZH2 结合位点。

对启动子区中的 PRC2 结合区域、PRC2 不结合区域、二价启动子区、活跃启动子区、抑制启动子区的 20 min 新增 EZH2 结合位点进行 spearman 相关分析和 GO 聚类比较分析 ( $P < 0.05$ ), 发现启动子区 20 min 新增的 EZH2 基因组结合位点主要分布在 PRC2 不结合区域及二价启动子区域 (图 4A–B)。对非启动子区中 PRC2 结合区域、PRC2 不结合区域、活化态增强子、待发态增强子、准备态增强子区的 20 min 新增 EZH2 结合位点和 30 min 丢失 EZH2 结合位点的进行 spearman 相关分析和 GO 聚类比较分析 ( $P < 0.05$ ) 发现, 非启动子区随超声时间延长丢失和新增的 EZH2 基因组结合位点主要不仅分布在 PRC2 结合区域、而且也分布在 PRC2 不结合区域以及活化态增强子上 (图 4C–E)。

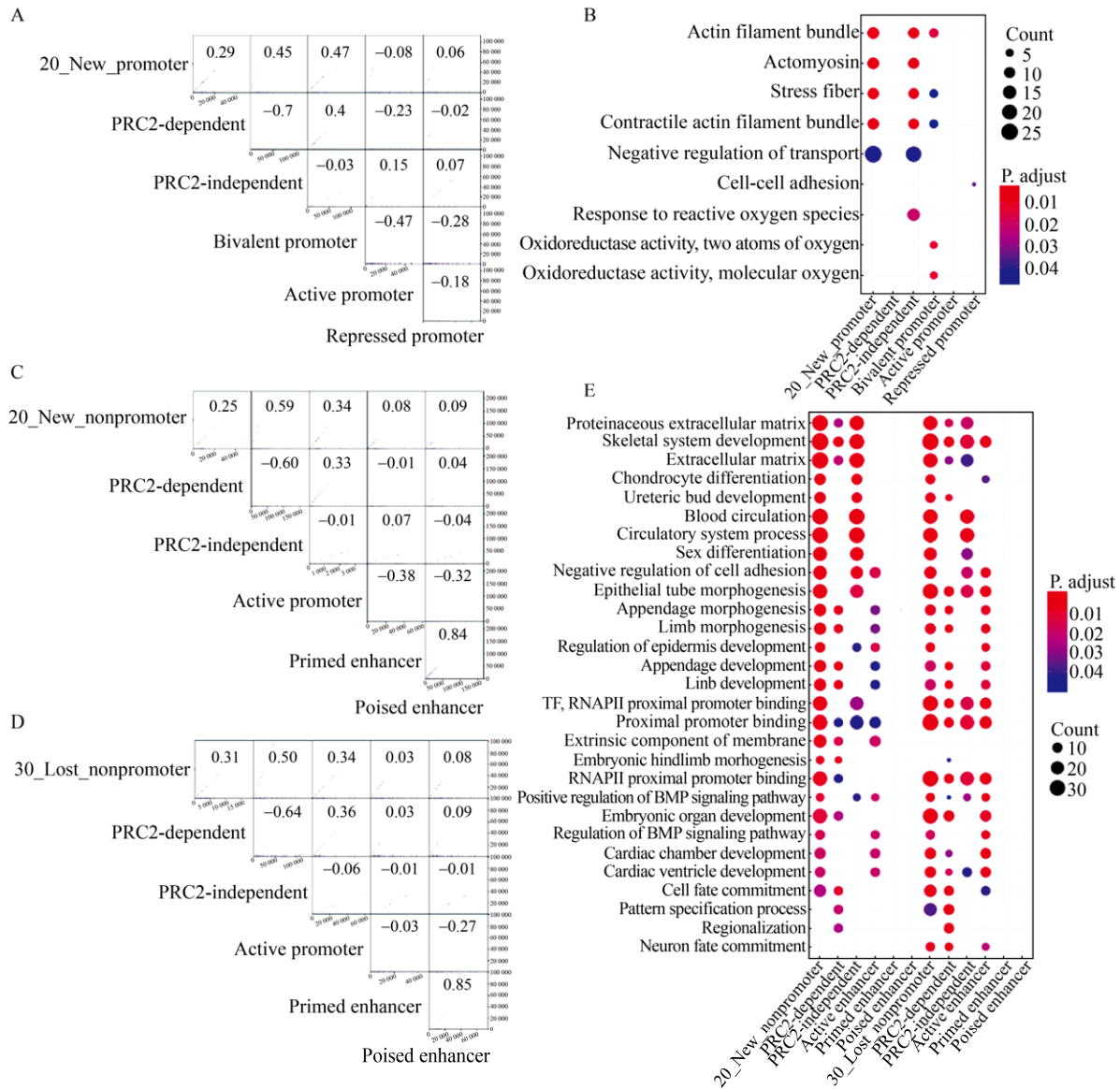


图 4 EZH2 随超声时间延长丢失和新增基因组结合位点的具体定位区域

Fig. 4 EZH2 lost and newly gained peaks in specific genomic regions after prolonged sonication. (A–B) Scatterplot of correlations and GO enrichment analysis ( $P < 0.05$ ) of the EZH2 peaks in 20\_New\_promoter, PRC2 occupied region, PRC2 unoccupied region, Bivalent promoter, active promoter, repressed promoter. (C–E) Scatterplot of correlations and GO enrichment analysis ( $P < 0.05$ ) of the EZH2 peaks in 20\_New\_nonpromoter, 30\_Lost\_nonpromoter, PRC2 occupied region, PRC2 unoccupied region, Active enhancer, Primed enhancer, Poised enhancer.

### 3 讨论

ChIP-seq 实验中超声时间延长会得到更多小分子量的染色质 DNA 片段，但是也会伴随产热，对蛋白质稳定性有影响。本研究以组蛋白 H3K27me3 和其甲基化转移酶 EZH2 为例，研究

不同超声时间对不同分子量大小蛋白 ChIP-seq 的影响。对小分子量组蛋白 H3K27me3，在启动子区或非启动子区不同超声时间得到的基因组结合位点的注释基因无明显差异。较短超声时间虽然产生 100–500 bp DNA 片段比例较少，但是对于小分子量组蛋白，足够得到比较全面的基因组结

合位点信息, 延长时间超声并不能显著增加基因组结合位点信息。

与组蛋白不同, 将超声时间从 10 min 延长至 20 min 后, 启动子区新增 EZH2 的基因组结合位点注释基因能够显著 ( $P < 0.01$ ) 聚类在与肌动蛋白丝组装的相关通路上 (图 3E), 而且新增的 EZH2 基因组结合位点主要分布在 PRC2 不结合区域及二价启动子区域, 因为这两个区域的 EZH2 结合位点注释基因也显著 ( $P < 0.05$ ) 聚类在肌动蛋白丝组装等相关通路上 (图 4E)。有研究表明在 T 细胞和纤维母细胞中敲除 EZH2 会导致聚合的 F-actin 降低, 提示 EZH2 可以促进 G-actin 组装聚合为 F-actin, 即 EZH2 对肌动蛋白丝的组装起到促进作用<sup>[28-29]</sup>。而基因组上 PRC2 不结合区域的 EZH2 可能具有转录激活作用, 这提示 EZH2 可能以不依赖 PRC2 的方式作为转录激活因子促进胚胎干细胞中肌动蛋白丝的组装。若超声时间不足, 则会丢失 EZH2 调节肌动蛋白丝这部分信息。另外非启动子区 20 min 新增基因的聚类 (113 条) 要远多于丢失基因 (8 条), 非启动子区 30 min 丢失基因的聚类 (149 条) 要远多于新增基因 (17 条), 且这些区域大多显著 ( $P < 0.01$ ) 聚类在 RNAPII、器官发育、细胞形态发生相关通路上 (图 3E)。这表明超声时间不足或过长均会丢失 EZH2 的基因组定位信息且超声对非启动子区的影响要大于启动子区。进一步分析还发现超声主要影响非启动子区中 PRC2 结合区域、PRC2 不结合区域以及活化态增强子区域的 EZH2 结合位点, 因为这些区域的 EZH2 结合位点注释基因也能显著 ( $P < 0.05$ ) 富集在 RNAPII、器官发育、细胞形态发生等相关通路上 (图 4E)。

综上所述, 超声步骤对于染色质免疫沉淀测序实验非常重要, 建议对大分子量的染色质修饰相关蛋白优化超声处理时间, 使形成的染色质片段聚集在 100–500 bp 可获得比较全面的基因组信息。对小分子量的组蛋白来说, 超声时间对其影响不大。

## REFERENCES

- [1] Kuo MH, Allis CD. *In vivo* cross-linking and immunoprecipitation for studying dynamic protein: DNA associations in a chromatin environment. *Methods*, 1999, 19(3): 425–433.
- [2] Mardis ER. ChIP-seq: welcome to the new frontier. *Nat Methods*, 2007, 4(8): 613–614.
- [3] Das PM, Ramachandran K, Vanwert J, et al. Chromatin immunoprecipitation assay. *Biotechniques*, 2004, 37(6): 961–969.
- [4] Skene PJ, Henikoff S. A simple method for generating high-resolution maps of genome-wide protein binding. *eLife*, 2015, 4: e09225.
- [5] Schmid CD, Bucher P. ChIP-seq data reveal nucleosome architecture of human promoters. *Cell*, 2007, 131(5): 831–832.
- [6] Nekrasov M, Wild B, Müller J. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *Embo Rep*, 2005, 6(4): 348–353.
- [7] Tie F, Stratton CA, Kurzhals RL, et al. The N terminus of Drosophila ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27. *Mol Cell Biol*, 2007, 27(6): 2014–2026.
- [8] Højfeldt JW, Laugesen A, Willumsen BM, et al. Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat Struct Mol Biol*, 2018, 25(3): 225–232.
- [9] Tan JZ, Yan Y, Wang XX, et al. EZH2: biology, disease, and structure-based drug discovery. *Acta Pharmacol Sin*, 2014, 35(2): 161–174.
- [10] Margueron R, Reinberg D. The polycomb complex PRC2 and its mark in life. *Nature*, 2011, 469(7330): 343–349.
- [11] Simon JA, Kingston RE. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell*, 2013, 49(5): 808–824.
- [12] Chamberlain SJ, Yee D, Magnuson T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells*, 2008, 26(6): 1496–1505.
- [13] Leeb M, Pasini D, Novatchkova M, et al. Polycomb complexes act redundantly to repress genomic

- repeats and genes. *Gene Dev*, 2010, 24(3): 265–276.
- [14] Riising EM, Comet I, Leblanc B, et al. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell*, 2014, 55(3): 347–360.
- [15] Kahn TG, Dorafshan E, Schultheis D, et al. Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic Acids Res*, 2016, 44(21): 10132–10149.
- [16] Burdach S, Plehm S, Unland R, et al. Epigenetic maintenance of stemness and malignancy in peripheral neuroectodermal tumors by EZH2. *Cell Cycle*, 2009, 8(13): 1991–1996.
- [17] Jiang HM, Gupta R, Somma J. EZH2, a unique marker of malignancy in effusion cytology. *Diagn Cytopathol*, 2014, 42(2): 111–116.
- [18] Lee ST, Li ZM, Wu ZL, et al. Context-specific regulation of NF- $\kappa$ B target gene expression by EZH2 in breast cancers. *Mol Cell*, 2011, 43(5): 798–810.
- [19] Yu GC, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 2015, 31(14): 2382–2383.
- [20] Yu GC, Wang LG, Han YY, et al. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, 2012, 16(5): 284–287.
- [21] Juan AH, Wang S, Ko KD, et al. Roles of H3K27me2 and H3K27me3 examined during fate specification of embryonic stem cells. *Cell Rep*, 2016, 17(5): 1369–1382.
- [22] Liu JC, Wu XW, Zhang HY, et al. Dynamics of RNA polymerase II pausing and bivalent histone H3 methylation during neuronal differentiation in brain development. *Cell Rep*, 2017, 20(6): 1307–1318.
- [23] Khan AA, Ham SJ, Yen LN, et al. A novel role of metal response element binding transcription factor 2 at the hox gene cluster in the regulation of H3K27me3 by polycomb repressive complex 2. *Oncotarget*, 2018, 9(41): 26572–26585.
- [24] Collinson A, Collier AJ, Morgan NP, et al. Deletion of the polycomb-group protein EZH2 leads to compromised self-renewal and differentiation defects in human embryonic stem cells. *Cell Rep*, 2016, 17(10): 2700–2714.
- [25] Thornton SR, Butty VL, Levine SS, et al. Polycomb repressive complex 2 regulates lineage fidelity during embryonic stem cell differentiation. *PLoS ONE*, 2014, 9(10): e110498.
- [26] Du Z, Li H, Wei Q, et al. Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. Japonica. *Mol Plant*, 2013, 6(5): 1463–1472.
- [27] Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell*, 2013, 49(5): 825–837.
- [28] Bryant RJ, Winder SJ, Cross SS, et al. The polycomb group protein EZH2 regulates actin polymerization in human prostate cancer cells. *Prostate*, 2008, 68(3): 255–263.
- [29] Su IH, Dobenecker MW, Dickinson E, et al. Polycomb group protein Ezh2 controls actin polymerization and cell signaling. *Cell*, 2005, 121(3): 425–436.

(本文责编 陈宏宇)