

Review of NGS for chromatin biology

- overview:

研究染色质开放性的常用方法：ATAC-Seq, FAIRE-Seq, DNase-Seq, MNase-Seq

研究三位基因组的常用方法：Hi-C, CHIP-PET

- 相关链接或资料：

- paper

- Meyer C A, Liu X S. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology[J]. Nature Reviews Genetics, 2014, 15(11): 709-721.
- Li Z, Schulz M H, Look T, et al. Identification of transcription factor binding sites using ATAC-seq[J]. Genome biology, 2019, 20(1): 1-21.
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627383/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3149993/>
- Lieberman-Aiden E, Van Berkum N L, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome[J]. science, 2009, 326(5950): 289-293.
- Fullwood M J, Liu M H, Pan Y F, et al. An oestrogen-receptor- α -bound human chromatin interactome[J]. Nature, 2009, 462(7269): 58-64.
- <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-S12-S11>
- <http://journals.im.ac.cn/html/cjbcn/2020/12/gc20122791.htm>

- Some blogs

- <https://cmb.i-learn.unito.it/mod/wiki/prettyview.php?pageid=135>
- <https://www.activemotif.com/blog-atac-seq>
- <https://zhuanlan.zhihu.com/p/349659624>
- <https://zhuanlan.zhihu.com/p/47467408>

- Some video

- <https://youtu.be/nkWGmaYRues>
- <https://youtu.be/uD2uTC5hXIg>
- https://youtu.be/VYCT_9I_jzY

CHIP-Seq

Description

- 全称: Chromatin immunoprecipitation followed by next-generation DNA sequencing
- 目的: 特异性转录因子(TFs)的结合位点
- 过程:
 - a. The first stage is to stabilize the link between proteins and DNA (DNA-protein crosslinking) thanks to formaldehyde (甲醛)
 - b. The second stage is to fragment the DNA (with the bound proteins) thanks to a process known as "sonication" (超声波) in lysis buffer
 - c. The third stage consists in the addition of a specific antibody for the protein of interest, the antibody is linked to a beads(sepharose or magnetic beads) which, thanks to its weight, deposits the antibody-protein-DNA complex on the bottom of the eppendorf, the complexes are then collected and purified from non-specific proteins
 - d. The last stage is the removal of the DNA protein bond thanks to the protease K, the extracted DNA fractions can then be sequenced with NGS
- Control experiment
 - From the original CHIP-seq experiment
 - Without using an antibody to enrich any particular protein

Pros

- 捕获任何生物体整个基因组的转录因子或组蛋白修饰的DNA靶点
- 定义转录因子结合位点
- 结合RNA测序和甲基化分析揭示基因调控网络

Cons

- 必须为每个TF产生抗体
- TF要已知
- ChIP实验不能区分不同TF亚型(蛋白质亚型)

Application

- 寻找已知的TF结合位点
- Motif in all peaks, 可以寻找TF结合的motif,
- 比较不同细胞同一个蛋白的peaks

ATAC-Seq

Description

- 全称： Assay for Transposase-Accessible Chromatin with high-throughput sequencing
- 目的：检测染色质的开放区域
- 过程：
 - a. 它利用了细菌Tn5转座酶（预先加载了sequencing adapters）的过度活跃，插入到染色质的可访问区域
 - b. 利用NGS技术对这些片段进行PCR扩增和测序
 - c. 测序峰对应于开放的染色质

Pros

- 快速、简单、灵敏的方法(3小时即可完成制备)
- 适用于多种细胞类型和物种
- 不需要事先知道序列或结合蛋白
- 不需要超声（ChIP-Seq），苯酚-氯藻提取(FAIRE)，或抗体(ChIP-Seq)
- 可适配单细胞测序分析
- 与DNase-seq相比，有明显的实验优势，它需要更少的细胞，步骤简单，在18年的统计数据中，ATAC-seq的研究数量是DNase-seq的12倍；

Cons

- 细胞数量必须从一开始就进行优化:细胞过少导致低转位，细胞过多导致过转位(对于人体细胞研究，推荐500-50000个细胞，但根据细胞类型和种类，最佳数量可能会有所不同)
- Tn5转座酶的切割机制比较复杂，需要Tn5二聚体起作用，Tn5二聚体的大尺寸使得切割事件依赖于邻近蛋白质(TF或组蛋白)的结构特征和可接近DNA的大小，导致有一些小片段被切割的可能性很低。除此之外，酶的DNA结合偏好也导致了序列特异性的切割偏差。

Application

- 核小体定位:识别核小体在分化或实验条件之间的位置变化
- 与Dnase-Seq输出的信息互补
- 可以用于生成高分辨率的核小体位置图以及转录因子结合谱图

FAIRE-Seq

Description

- 全称：Formaldehyde（甲醛）-assisted isolation of regulatory elements followed by sequencing
- 目的：鉴定无核小体区域（与ATAC-seq结果类似）
- 过程：DNA -蛋白质复合物在体内用甲醛进行短暂的交联，超声剪切，苯酚-氯仿萃取。在水相中回收的DNA被荧光标记并杂交到一个DNA微阵列。这种方法提取非交联DNA，只有这些核小体缺失的区域将被纯化、富集和测序。

Pros

- Protocol 简单，便宜
- 不需要抗体，不需要酶
- 不需要单细胞悬液或细胞核分离，所以它很容易适应于组织样本的使用

Cons

- 对交联的依赖
- 所需最少100,000个细胞数

DNase-seq

Description

- 全称：DNase I digestion of chromatin is combined with next-generation sequencing
- 目的：定位基因组中的活性基因调控元件，包括增强子和启动子
- 过程：
 - a. 用DNase I处理DNA -蛋白复合物，切割DNase I 的超敏感位点，而包裹在核小体和更高阶结构中的DNA区域对其更具抗性；
 - b. 调控蛋白结合的序列不受DNase I酶切的影响；
 - c. 进行DNA提取和测序；

Pros

- 能检测开放的染色质
- 不需要事先知道序列或结合蛋白
- 不需要超声（ChIP-Seq），苯酚-氯藻提取(FAIRE)，或抗体(ChIP-Seq)
- 与甲醛辅助分离调控元件和测序(FAIRE-seq)相比，对启动子有更大的敏感性

- 与ATAC-Seq相比，DNase-seq的footprints更准确（footprints, where read coverage suddenly drops within peak regions of high coverage）

Cons

- DNase I 是序列特异性的，超敏感位点可能不能解释整个基因组
- 多重纯化步骤造成的DNA损失限制了灵敏度

Application

- 定位基因组中的活性基因调控元件，包括增强子和启动子
- DNase I 与ChIP数据的整合可以帮助识别和区分类似的蛋白结合位点

MNase-seq

Description

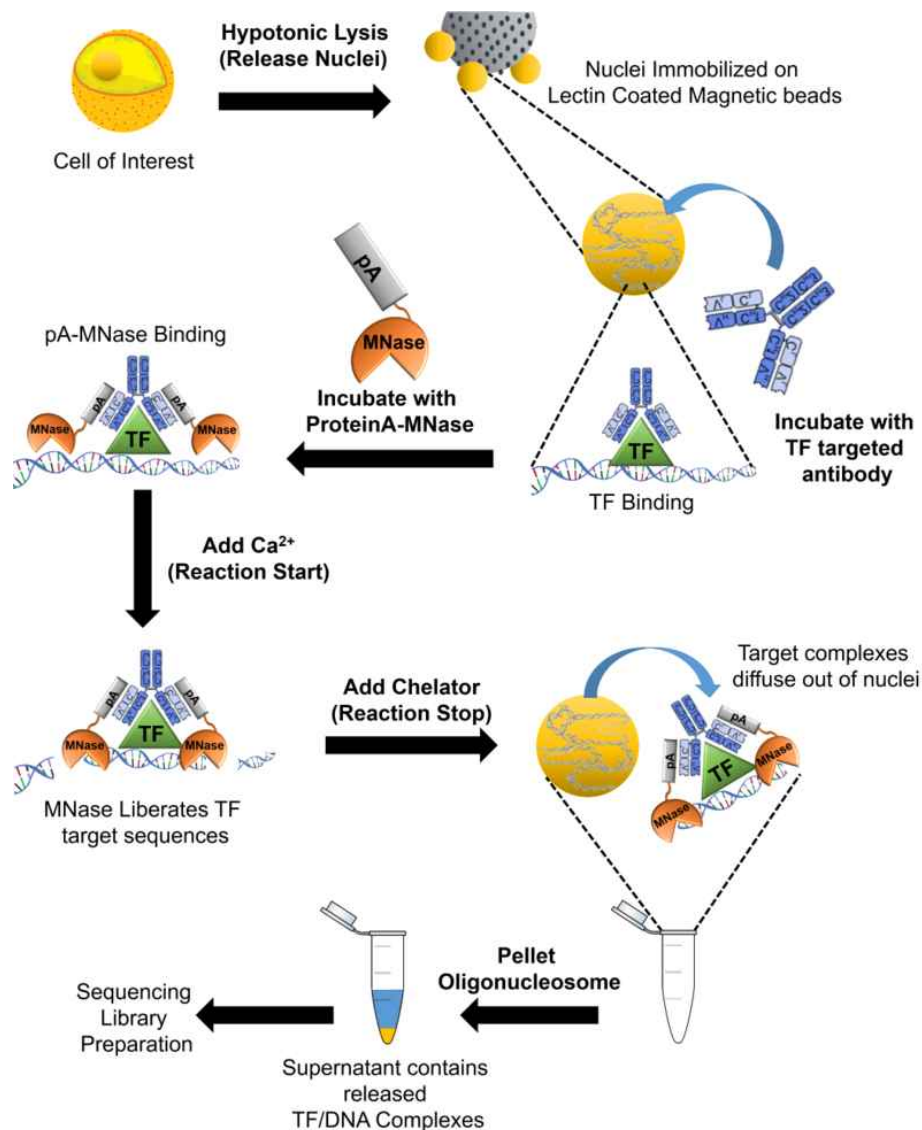
- 全称：micrococcal nuclease (MNase, 微球菌核酸酶) digestion of chromatin is followed by next-generation sequencing
- 目的：鉴定核小体占据位点
- 过程：
 - a. 首先从感兴趣的组织中分离出真核细胞的细胞核
 - b. MNase-seq使用微球菌核酸酶结合和切割真核染色质DNA的蛋白非结合区域
 - c. 核小体或DNA-蛋白质复合物可以从样品中纯化，结合的DNA随后可以通过凝胶电泳和提取纯化

Pros

- chip-seq的超声波处理，需要更高的温度和去污剂，可能导致转录因子的损失；
- 由于染色质区域彼此之间紧密
- 技术扩展： cut&run seq，sc MNase-seq

CUT&RUN

- CUT&RUN是一种新型的基于 MNase 的免疫沉淀，使用标记有抗体的MNase来特异性结合目标蛋白周围的DNA；
- 过程如图：



- 与ChIP-Seq相比，ChIP-Seq protocol中的交联步骤会产生假阳性结合位点，并且超声波处理具有很强的背景噪音；CUT&RUN测序由于低背景，更低的测序深度就可以满足需求。
- 在CUT&RUN中，细胞或细胞核在整个过程中保持完整，只有目标结合位点被释放到溶液中，CUT&RUN方法极大地减少了非特异性背景。
- CUT&RUN-seq 的主要限制是由于钙依赖性 MNase 反应的时间不当而导致 DNA 过度消化的可能性。

CONCLUSION

- 检测染色质开放性四种方法：ATAC-Seq, DNase-seq, MNase-seq, FAIRE-seq
- MNase-seq与 DNase-seq相比：
 - 相同点：都属于核酸酶，用于测定染色质开放区域，可用于核小体在生物体的定位信息，都对单细胞测序进行了优化；
 - 不同点：足够浓度下，DNase I 能够将核小体结合的DNA消化至10bp，而微球菌核酸酶不能；DNase I 可以识别DHS，DHS 是对 DNase 处理过敏的 DNA 区域，通常指示调节区域（例如启

动子，增强子），而MNase不能；

- MNase-seq与 FAIRE-seq相比：

- 相对于其他三类，FAIRE-seq 的两个主要缺点是: 1. 所需的细胞数最少为 100,000 个。2. 对交联的依赖。FAIRE-seq 依赖于目标蛋白质与DNA交联，交联可能会结合其他蛋白，有些蛋白质与DNA瞬时相互作用，限制了可以从水相中回收和分析的数量，因此，FAIRE-seq获得的整体分辨率比较低。

- MNase-seq与 ATAC-seq相比：

- ATAC-seq优点：

它不依赖于微球菌核酸酶的可变消化，也不依赖于交联或苯酚-氯仿提取；ATAC-seq可在几个小时内完成

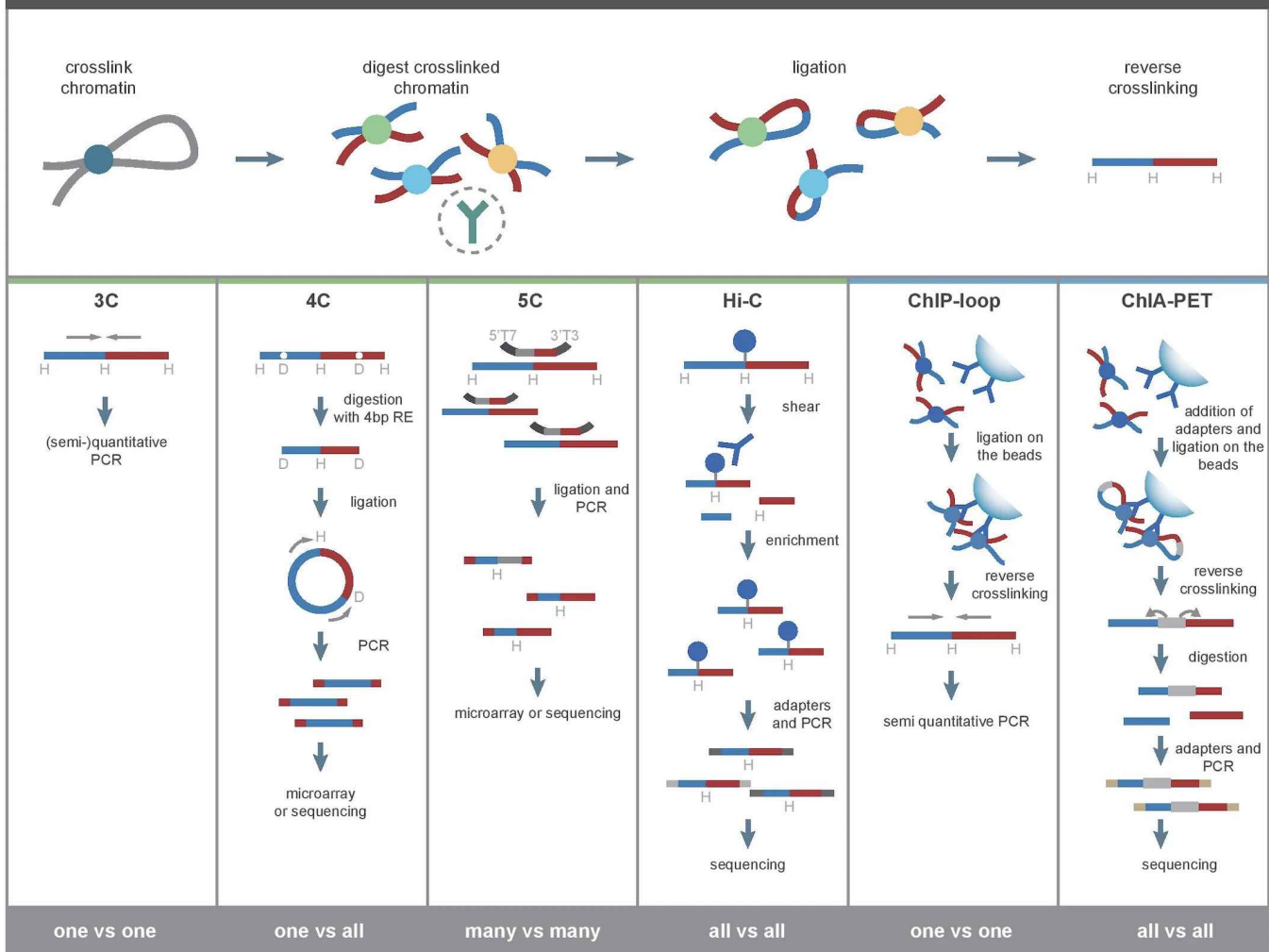
- ATAC-seq缺点：它需要更高的测序覆盖率和由于DNA非特异性插入线粒体DNA，导致线粒体污染

Hi-C

Background

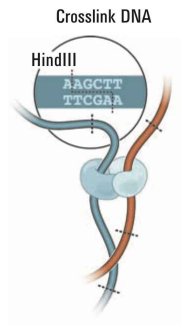
- 了解Hi-C，要先从**chromosome conformation capture（染色质构象捕获技术，3C）**开始了解
- 3C（one vs one）：基因组捕获技术（Chromosome conformation capture, 3C）是最早研究三维基因组的技术，需要提前知道互相作用区域，才能量化一对基因组基因座之间的互相作用，主要捕捉点对点的染色体交联。
- 4C（one vs all）：染色体构象捕获芯片（Chromosome conformation capture-on-chip, 4C），主要捕获一个特定位点对其他所有位置的染色体交联。该技术不需要知道作用区域的先验知识就可以使用。
- 5C（many vs many）：染色体构象捕获碳拷贝（Chromosome conformation capture carbon copy, 5C），可以检测某段区域内所有的互作，但是该区域一般<1 Mb。覆盖度的问题也就造成该技术不适用于全基因组测序。
- Hi-C（all vs all）：高通量基因组捕获技术，基本解决了上述技术的缺点，可以实现全基因组覆盖检测全部未知互作区域。
- ChIA-PET：该技术将 HiC 与 ChIP-seq 结合，可以检测目的蛋白质的所有互相作用。

Chromosome Conformation Technologies



Description

- 全称：chromosome conformation capture
 - 目的：了解在染色体组中空间距离相近的DNA片段，并构建相应的3D基因组数据信息。
 - 过程：
 - 甲醛固定
- 先加入甲醛，导致空间相邻的染色质片段之间进行共价连接。虚线标记的为限制点。



- 酶切序列

用限制性内切酶切割基因组，一般有两种酶可供选择：6bp 的限制性内切酶，4bp 的限制性内切酶。EcoR1 或 HindIII 用于每4000bp切割一次基因组，在人类基因组中产生约100万个片段。这里是HindIII。

Cut with
restriction
enzyme



- 末端修复

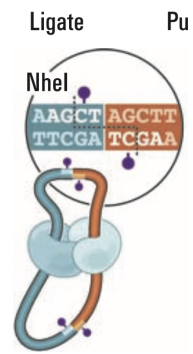
得到的片段具有粘性末端，产生的粘性末端被补平修复，加入生物素，其中的一个核苷酸被生物素标记。

Fill ends
and mark
with biotin



- 连接及解交联

使用DNA连接酶连接互作片段，形成环状。将连接DNA片段的蛋白质消化掉，得到交联片段



- 序列打断

使用超声波或其他方式，再次打断片段

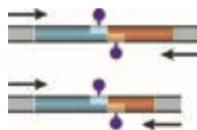
Purify and shear DNA;
pull down biotin



- 上机测序

用磁珠将带有生物素的捕获，制作文库，上机测序

Sequence using
paired-ends



Pros:

- a. 主要用于解释基因组的三维结构
- b. 可以帮助深入研究染色体结构变异
- c. 可以解析所有的染色质构象

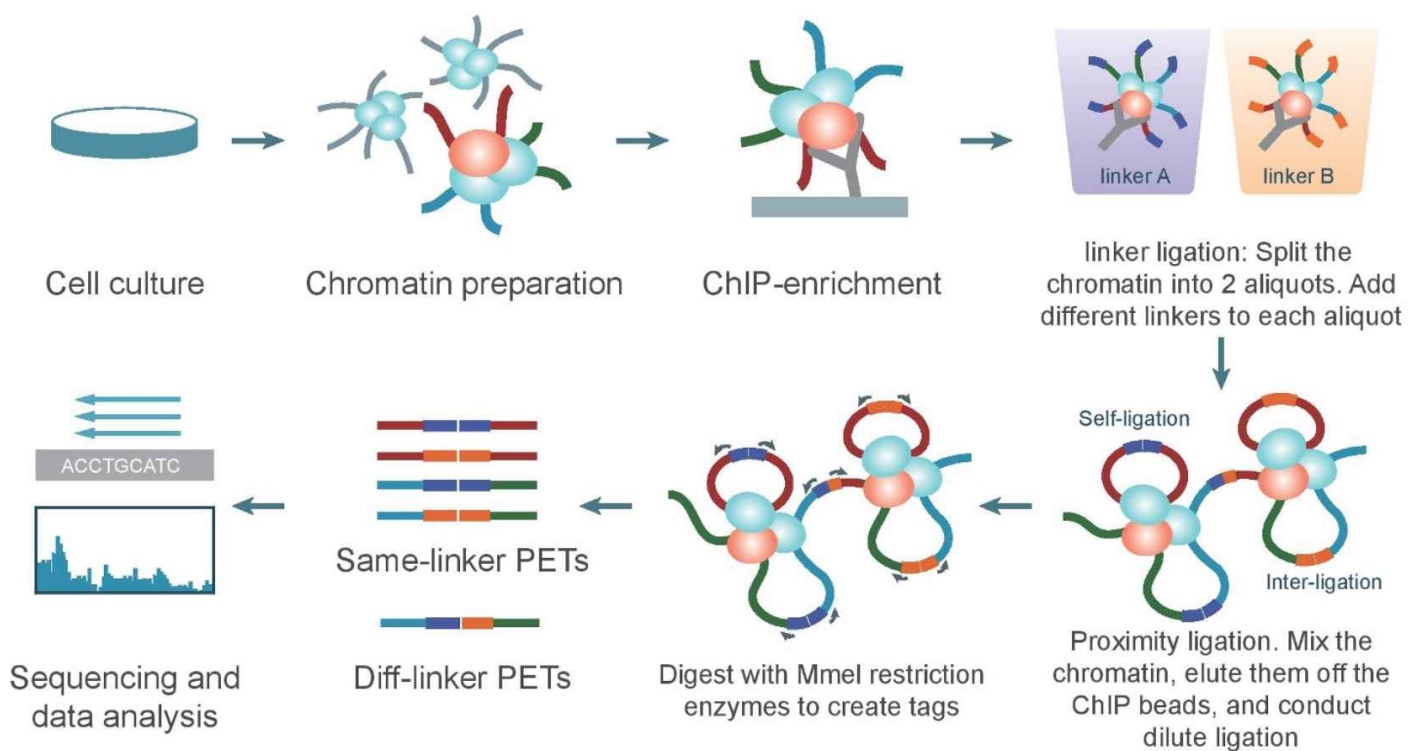
Cons:

- a. 数据量庞大
- b. 数据特异性不好

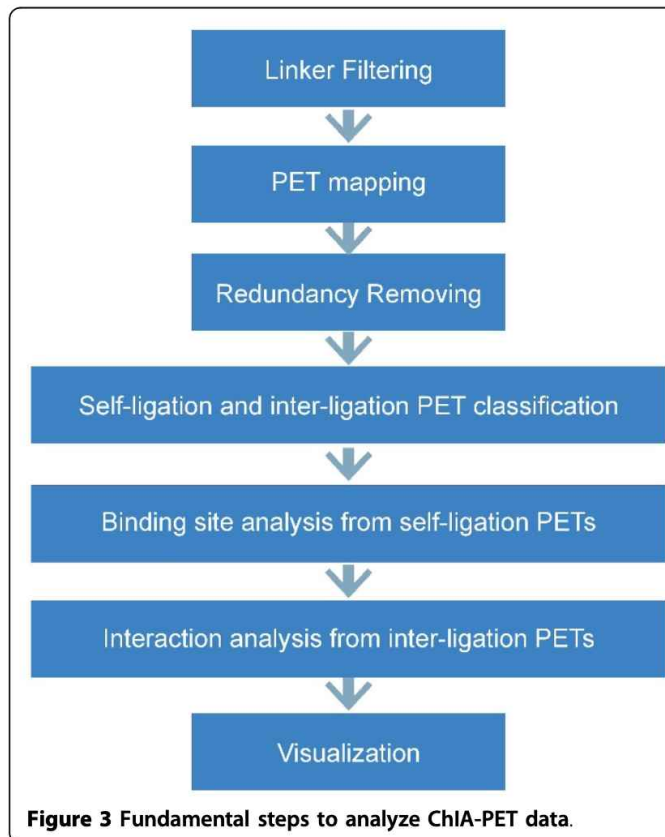
CHIA-PET

Description

- 全称： Chromatin Interaction Analysis with Paired-End-Tag sequencing
- 过程：
 - 湿实验过程



- 干实验过程



Pros:

- a. 与Hi-C相比，有更高的分辨率
- b. 具有比Hi-C更高的分辨率，可以与Hi-C一起重建基因组的三维结构
- c. 可以研究蛋白质的功能性，比如转录因子的调控关系

Cons:

- a. ChIA-PET实验需要数千万个细胞。在许多实际应用中，比如病人的组织，细胞数量是有限的。
- b. 在目前的ChIA-PET协议中，使用限制性内切酶MmeI对连接的DNA进行消化，制备DNA结构，从DNA片段中只保留20个核苷酸用于进一步分析，这是相当短的，特别是对于高重复序列的基因组。我们需要使DNA片段更长，以便更具体地映射到参考基因组。
- c. 目前，ChIA-PET和Hi-C生成的交互数据主要以二维图形的形式呈现，远未实现三维可视化。我们需要重建细胞核内基因组的三维结构，以真实的空间方式展示染色质相互作用
- d. 使用DNA-FISH对结构的验证率很低，DNA-FISH的分辨率需要提高，数以万计的染色质相互作用已经由ChIA-PET报告，但只有几十到几百已经被DNA-FISH验证。

Application:

- a. 转录因子介导的DNA元件间的相互作用
- b. 染色质交互网络
- c. 染色质相互作用的功能研究
- d. 染色质三维结构重建