PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2019,46(4):406~414
www.pibb.ac.cn

染色质互作相关转录因子的挖掘及功能分析*

王 琪 丁砚书 耿宝宝 聂玉敏**
（南京医科大学生物医学工程与信息学院，生物信息学系，南京 210029）

摘要 染色质互作是真核生物基因组组装的基础，并且在调控真核基因细胞特异性表达中发挥重要作用. 染色质互作的发生与特定的蛋白质有关，目前已经发现CTCF（CCCTC binding facor，转录阻抑物）和黏连蛋白与染色质互作相关，然而并不清楚是否还有其他蛋白质参与染色质互作. 我们将整合高通量染色体构象捕获（Hi-C）和染色质免疫沉淀-测序（ChIP-seq）数据，在GM12878和K562细胞系中挖掘与染色质互作相关的转录因子，并对发现的转录因子做功能分析. 我们在频繁发生互作的染色质位点中发现RUNX3、SPI1等转录因子也可能参与染色质互作. 另外，通过FP-growth的数据挖掘方法还发现多个转录因子可能协同作用参与染色质互作. 研究结果将为染色质互作相关实验的开展提供先验知识.

关键词 染色质互作，转录因子，FP-growth，Hi-C
中图分类号 Q5，Q7
DOI：10.16476/j.pibb.2018.0303

在真核细胞中，核小体是染色质的基本结构单位. 真核生物基因组首先被组装成核小体，然后在一些蛋白质的作用下发生互作并逐步折叠形成具有复杂结构的染色质[1]. 染色质互作是真核细胞染色质压缩的基础，同时也是调控真核基因表达的重要因素. 正是由于染色质在三维空间的互作，增强子等远端调控元件才有可能跨越相当长的线性距离与目标基因的启动子发生作用，从而调控目标基因的表达[2-3]. 目前，利用染色体构象捕获（3C）技术及其衍生技术[4-6]，尤其是高通量染色体构象捕获（Hi-C）[7]、配对末端标签测序分析染色质相互作用（chromatin interaction analysis by paired-end tag，ChIA-PET）[8]等高通量测定方法的应用，研究者发现染色质互作还将形成一系列不连续的拓扑相关结构域（topologically associating domains）[9]. 在拓扑相关结构域的内部，染色质互作频繁，增强子-启动子区的互作通常位于某一个特定的拓扑相关结构域内. 染色质互作以及拓扑相关结构域等功能区域的形成对真核基因的精准表达至关重要[10]. 如果特定细胞中染色质互作发生异常，将导致基因的表达异常，甚至导致疾病或者癌症的发生[11-12].

染色质互作的形成与特定的蛋白质相关[13]. 首先，增强子-启动子区的互作由转录因子介导. 转录因子与增强子等远端调控元件结合，并通过与结合到启动子区的转录因子或RNA聚合酶发生相互作用，从而形成染色质环，以实现调控元件的远端调控[14]. 普遍表达的转录因子YY1就是其中一个参与增强子-启动子区互作的蛋白质. YY1与活性增强子和启动子近端调控元件结合并形成二聚体以促进这些DNA元件的相互作用[15]. 其次，CTCF（CCCTC binding facor，转录阻抑物）和黏连蛋白（cohesin）是介导染色质互作的重要结构蛋白质. 其中，黏连蛋白由SMC1和SMC3异二聚体以及RAD21和SCC3亚基组成[16]. CTCF接触结构域和绝缘邻域，将CTCF-CTCF环内的增强子和基因与那些环外的调控元件隔离开来[17]，以这种方式约束CTCF-CTCF环结构内的DNA相互作用，从而促进增强子-启动子接触. 在染色质环形成的挤压模

* 江苏省自然科学基金青年基金项目（BK20161026），江苏省高校自然科学研究面上项目（16KJB180022）和南京医科大学科技发展基金重点项目（2015NJMUZD003）资助.
** 通讯联系人.
Tel：025-86869366，E-mail：yumin_nie@njmu.edu.cn
收稿日期：2018-11-23，接受日期：2019-02-25

型中，CTCF界定了染色质互作的位置，基因组DNA在黏连蛋白中滑动以形成染色质环[18]. 如果黏连蛋白通过NIPBL被加载到活性增强子和启动子上[19]，还可瞬时稳定增强子-启动子相互作用[20-21]. 另外，大多数ZNF143结合位点都与CTCF以及黏连蛋白的结合位点重叠，说明在CTCF和黏连蛋白形成染色质环的过程中，ZNF143辅助CTCF以建立保守的染色质结构[22].

虽然目前对染色质互作的形成有了一定的认识，也发现了一些与染色质互作相关的蛋白质，但对介导染色质互作的蛋白质还是知之甚少，我们的研究希望发现更多与染色质互作相关的蛋白质. 本文中我们将整合Hi-C和ChIP-seq数据，在GM12878和K562细胞系中挖掘与染色质互作相关的转录因子. 染色质互作数据来源于Hi-C实验，并通过Fit-hi-c[23]软件处理得到每对染色质互作发生的显著性. 由于染色质互作具有细胞特异性，因此我们在GM12878细胞系和K562细胞系进行分析，以发现一些相关性和特异性.

1　数据与方法

GM12878和K562细胞系中不同分辨率下的染色质交互数据来源于Rao等[24]的Hi-C实验. 该数据集经过读段比对、过滤以及归一化处理给出了基因组中任何两个位置的染色质互作. 由ChIP-seq实验测定的76（GM12878）和100（K562）个转录因子的结合数据下载于UCSC数据库（http：//genome.ucsc.edu / cgi -bin / hgTrackUi? hgsid=695454653_855MYXXIWS8Oxx2IKbbAPVdbdraB&c=chr21&g=wgEncodeAwgTfbsUniform），通过读段比对和峰值检测，该数据集给出转录因子在全基因组范围内的富集分布.

我们的分析步骤总体来说可以分为5步. 首先从染色质交互数据出发，利用Fit-hi-c在全基因组范围内得到参与互作的染色质位点，其次确定阈值以获得频繁互作的染色质位点，接下来根据染色质位点的互作频繁程度对频繁互作的染色质位点进行划分，并结合转录因子数据构建关联矩阵以挖掘染色质互作相关的转录因子，最后从生物学功能和蛋白质的结构两个方面进行解释（图1）.

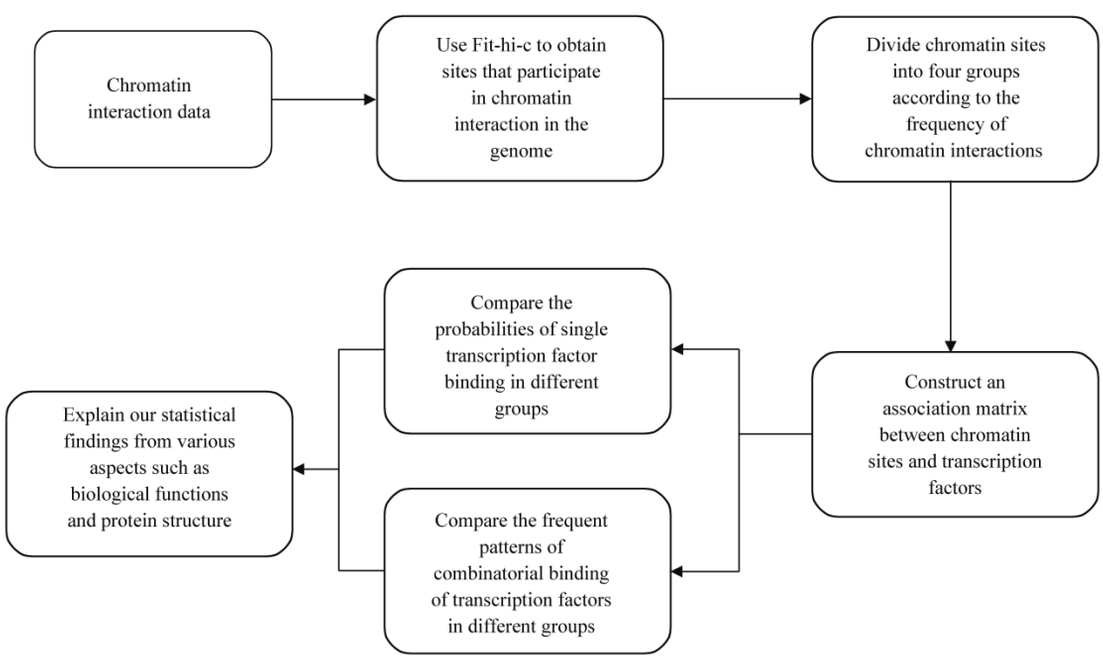

Fig. 1　The overall research design and procedures

1.1　在全基因组范围内获得参与互作的染色质位点

Fit-hi-c处理后的数据会给出每对染色质互作发生的显著性（*P*值）及交互作用矩阵. 在1 Mb、500 kb、250 kb、100 kb、50 kb和25 kb分辨率下，我们选定0、0.1、0.01、0.001、0.0001、0.00001作为显著性的阈值对结果进行筛选，并计算保留下来的有显著意义的互作位点对的个数（图2）. 我们发现当阈值低于0.001之后得到的互作位点对数变化不大，因此我们认为*P*<0.001的互作位点对是显著的、有意义的，并将这些位点定义为参与互作的染色质位点. 进一步统计发现参与互作的染色质位点相对较多，而不参与互作的染色质片段则非常少. GM12878细胞系中不参与互作的染色质片段为2.47%，而K562细胞系中该比例为8.43%. 这说明在全基因组范围内普遍存在染色质互作，相当大一部分的染色质位点都倾向于与其他区域发生物理接触，以调控相关基因的转录或抑制.

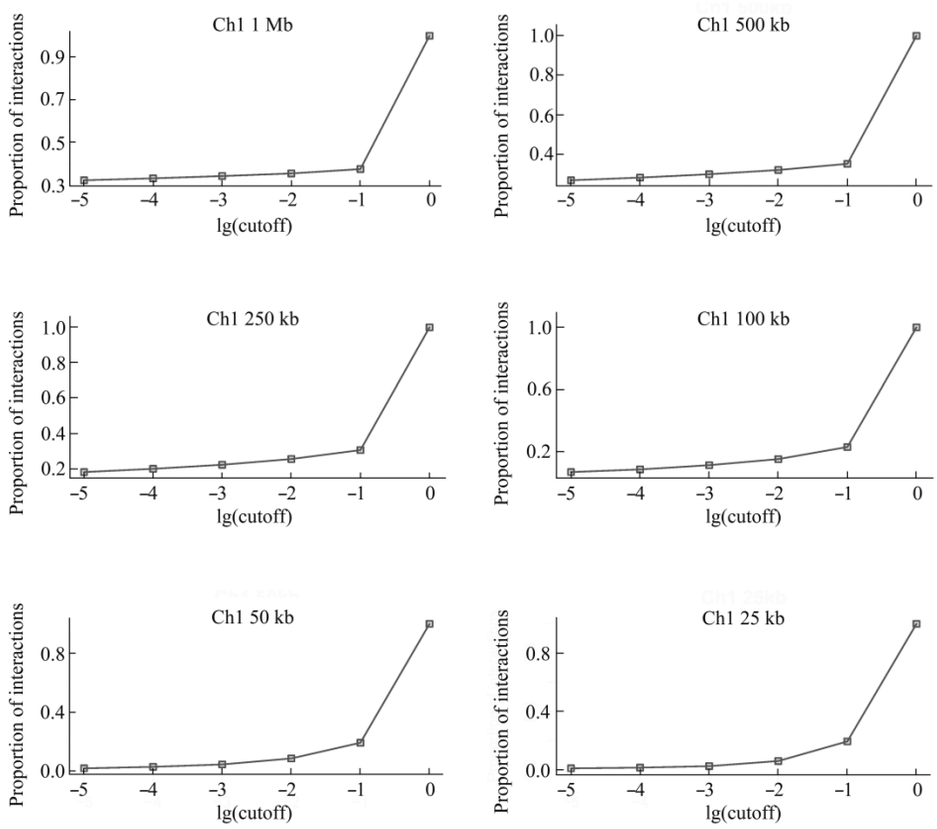

Fig. 2　Defining a threshold to determine the significant chromatin interactions

Take chromosome 1 as an example. At different resolutions， when the threshold was less than 0.001，the proportion of chromatin sites involved in interaction did not change much， so the threshold was set to 0.001.

1.2　根据位点参与互作的频繁程度对染色质位点进行划分

基因组中相当大一部分位点都参与染色质互作，我们接下来在25 kb分辨率下统计每个位点与其他位点发生互作的次数，并根据发生互作次数的四分位点将所有参与互作的染色质位点划分为四个层次（依次为第一、二、三、四层次）. 其中，第一层次的染色质位点互作最频繁，第四层次的染色质位点与其他位点互作较少.

1.3　构建参与互作的染色质位点与转录因子的关联矩阵

以染色质互作位点为行名，转录因子为列名构建关联矩阵，若转录因子结合在某个染色质互作位点区域内，则矩阵中对应的位置为1，否则为0. 如果转录因子与染色质互作位点重合的长度大于20 bp，我们就认为该转录因子结合在该染色质位点内.

1.4　利用统计学方法从关联矩阵中挖掘与染色质互作相关的转录因子

1.4.1　比较不同层次中各转录因子结合概率的差异

在四个层次中统计每个转录因子在该层次的所有互作位点中结合的概率，我们主要比较了差别最大的第一和第四层次，找出了在这两个层次结合概率差别最大（差值的绝对值）的10个转录因子，它们可能与染色质互作存在某种关联. 接下来利用统计学方法在第一和第四层次的位点中比较转录因子的结合是否有显著差异.

每个转录因子在第一和第四层次染色质位点中的分布可以看作是总体分布未知的两个大样本. 依据生物统计学知识，样本量n_1，n_2较大，且$n_1\times P_1>5$，$n_1\times P_1（1-P_1）>5$，$n2\times P_2>5$，$n_2\times P_2（1-P_2）>5$时（P_1，P_2为两个样本的样本率），两个样本率的比较可以用U检验，$U=\dfrac{P_1-P_2}{\sqrt{P_c(1-P_c)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$.

其中，P_c为合并阳性率，即$P_c=\dfrac{X_1+X_2}{n_1+n_2}$. 零假设H0：两个总体无显著差异，即转录因子在两个层次的结合无显著差异. 查表得$U_{0.05}$=1.96，而我们得到的10个转录因子的U值明显高于1.96，因此拒绝零假设，即转录因子在两个层次的染色质位点中的结合有显著差异.

1.4.2　比较不同层次共同发挥作用的转录因子对是否有差异

在上一步中我们已经找出了那些与染色质互作密切相关的单个转录因子，但因为蛋白质经常协同作用共同调节某种生理活动或生理机能，也就是说可能某个转录因子的结合位点在互作位点中不富集，但是某些转录因子对的结合位点在互作位点中富集，这些转录因子很有可能共同出现从而促进或者抑制染色质互作的发生. 为了证实我们的猜测，进一步在关联矩阵中挖掘同时出现的转录因子组合模式. 首先剔除没有转录因子结合的那些染色质位点，然后用FP-growth算法挖掘出各层次中转录因子结合的频繁模式，并在差异最大的第一和第四层次中比较转录因子结合频繁模式的差异. 如果某个频繁模式只出现在其中一个层次的染色质位点中，那么这些频繁模式中的转录因子可能共同发挥作用促进或抑制染色质互作的发生.

1.5　从生物学功能、蛋白质的结构分析等方面解释统计发现

最后，对于发现的可能与染色质互作相关的转录因子，我们从生物学功能、蛋白质结构等方面解释它们的生物学意义，主要应用了基因本体数据库GO的功能分析以及结构域数据库Pfam的结构域分析.

2　结果与讨论

2.1　可能参与染色质互作的单个转录因子

通过比较差距最大的第一和第四层次，我们在GM12878和K562两个细胞系中筛选出了两个层次中单个转录因子结合概率相差最大的10个转录因子（表1）. 其中，GM12878细胞系中有3个是目前已经报道过与染色质互作相关的，它们分别是RAD21、CTCF、SMC3. 这也从一定程度上说明了我们实验结果的可信性，其余的转录因子也很有可能与染色质互作的形成有关，这有待于进一步的实验验证，也为之后的研究提供了一些方向.

Table 1　The top ten transcription factors with the greatest difference in binding at the chromatin interaction sites in the GM12878 and K562 cell lines

GM12878 cell line	K562 cell line
RUNX3	CEBPB
SPI1	REST
BATF	CTCF
CTCF	TEAD4
RAD21	MAX
NFIC	JUND
NR2C2	RCOR1
SMC3	SPI1
EBF1	TAL1
ATF2	MAZ

比较两个细胞系中的转录因子，我们发现RUNX3、NFIC是GM12878细胞系所特有的，在K562细胞系中并不存在这两个转录因子，而TEAD4、TAL1是K562细胞系所特有的，在GM12878细胞系中也不存在. 这说明参与染色质互作形成的蛋白质可能具有细胞特异性，某些蛋白质只在特定的细胞中参与染色质互作的形成. 另外，在两个细胞系中都发现了SPI1、CTCF的存在，说明这2种转录因子在不同细胞系中都参与染色质互作的形成，在介导染色质互作的形成上具有普遍

性. 目前我们的分析只涉及了GM12878和K562两个细胞系，如果将分析扩展到更多细胞系，将得到更可靠的结论.

2.2　共同发挥作用影响染色质互作的转录因子组合模式

我们利用FP-growth算法，并设定置信度和频繁项集中项的个数分别为16和3，在第一、四层次中挖掘转录因子结合的频繁项集，图3和图4分别为GM12878和K562细胞系中转录因子结合的频繁项集的维恩图展示.

在GM12878细胞系中，我们发现YY1、ZNF143等已经报道过的与染色质互作相关的转录因子，在挖掘参与染色质互作的单个转录因子时没有出现，但在几个转录因子协同作用中出现频率很

Fig. 3　The frequent patterns of combinatorial binding of transcription factors in the first and fourth groups in the GM12878 cell line

Combination of transcription factors on the left（green） and right（blue） most frequently bound at chromatin sites in the first and fourth group, respectively. Combination of transcription factors in the middle part most frequently bound at chromatin sites in both groups. All transcription factors except RUNX3 appeared in the middle part， and have been reported as transcription factors related to chromatin interaction. These results suggested that specific transcription factors indeed enriched at chromatin interaction sites.

Fig. 4　The frequent patterns of combinatorial binding of transcription factors in the first and fourth groups in the K562 cell line

Combination of transcription factors on the left（green） and right（blue） most frequently bound at chromatin sites in the first and fourth group, respectively. Combination of transcription factors in the middle part most frequently bound at chromatin sites in both groups.

高. 这从一定程度上证实了我们的猜测，确实存在一些转录因子的组合，可以协同参与染色质互作. YY1通常与其他转录因子一起结合在第一层次的染色质位点中，说明该转录因子并不单独发挥作用，而是与其他转录因子共同作用介导染色质的频繁互作. 另外，我们还发现在维恩图的中间部分，也就是两个层次中共有的出现频率很高的转录因子组合中，几乎都包含SMC3、RAD21、CTCF、YY1、ZNF143这几个已知的参与染色质互作的转录因子，说明这些转录因子的组合在协同介导染色质互作时，与染色质互作的频繁程度无关，它们可能是一些在全基因组范围内普遍介导染色质互作的转录因子. 而右边第四层次有一些新出现的转录因子，如BATF、NR2C2、NFIC、FOXM1，这些转录因子与普遍介导染色质互作的转录因子的组合可能介导少数具有特定功能的染色质互作.

在K562细胞系中我们得到了类似的结果. 发现在左边第一层次中出现的频繁模式，大多为CTCF、RAD21、SMC3、ZNF143、EGR1、MAX、MAZ这些在两者的交集中也就是那些在全基因组范围内普遍介导染色质互作的转录因子的组合. 而右边第四层次则会有一些新的转录因子被发现，如JUND、BHLHE40、TEAD4、CCNT2、TAL1、NR2F2、ZBTB7A，它们可能参与具有特定功能的染色质互作的形成.

综合以上两个细胞系的维恩图结果，我们猜测存在一些普遍介导染色质互作形成的转录因子，它们与不同的转录因子组合以介导染色质互作的形成. 目前已知结构转录因子CTCF与黏连蛋白协同将基因组划分为环状结构域，而YY1形成二聚体以促进染色质互作的形成，其余蛋白质介导染色质互作发生的机制有待于我们进一步的研究和发现. 另外，通过比较发生频繁染色质互作的位点和不频繁发生染色质互作的位点，能找出两者的交集和差集，那么出现在差集中的转录因子对可能协同作用以促进或抑制染色质互作的发生，对这些转录因子对做进一步研究可能发现染色质互作形成的新机制.

2.3　GO功能分析

首先利用metascape对GM12878和K562细胞系中参与染色质互作的单个转录因子进行总体的功能分析（图5），然后在GO数据库中对这些转录因子进行了进一步的分析，我们有一些新的发现，将从三个方面来展示（全部结果的表格见网络版附件）.

2.3.1　细胞组分

在GM12878和K562细胞系中的共18个转录因子中，RAD21和SMC3为核减数分裂黏连蛋白复合物的组成部分，ATF2、CTCF、RAD21、RUNX3、SPI1这5个转录因子为染色体组成部分，CTCF、RAD21、SMC3参与组成了染色体、着丝粒区域，CEBPB、JUND、MAX、MAZ、RCOR1、REST、TEAD4为核质组分，RCOR1、REST为转录阻遏复合物的组成部分，CEBPB、JUND、TAL1为转录因子复合物的组成部分. 从以上结果可以看出，在这18个转录因子中，大部分的转录因子都是染色体或转录蛋白复合物的组成部分，这与我们的研究方向一致.

2.3.2　分子功能

在GM12878和K562细胞系中的共18个转录因子中，ATF2、BATF、CTCF、EBF1、RUNX3、SPI1这6个转录因子具有与RNA聚合酶Ⅱ调节区序列特异性DNA结合的功能，BATF、CTCF、EBF1、SPI1这4个转录因子能与RNA聚合酶Ⅱ核心启动子近端区域序列特异性DNA结合，ATF2、SPI1能与RNA聚合酶Ⅱ末端增强子序列特异性DNA结合. CTCF、JUND、MAX、MAZ、RCOR1、REST、SPI1、TAL1都能与转录调节区DNA结合，CTCF、JUND、MAX、MAZ、REST、SPI1均能与转录调节区序列特异性DNA结合，MAZ、REST、SPI1、TAL1 4个转录因子均能与核心启动子序列特异性DNA结合. 从以上结果可以看出，在这18个转录因子中，大部分的转录因子都具有与RNA聚合酶Ⅱ调节区序列特异性DNA结合的功能，也就是说这些转录因子可能被募集到调控序列，促进染色质互作的产生，从而进一步影响转录调控.

2.3.3　生物学过程

在GM12878和K562细胞系中的共18个转录因子中，BATF、CTCF、EBF1、NR2C2、RAD21、RUNX3、SPI1、JUND、MAX、MAZ、RCOR1、REST、TAL1、TEAD4这些转录因子参与了转录、DNA模板化，RCOR1、REST参与了组蛋白H4的去乙酰化过程. 转录和乙酰化的过程一般伴随着启动子或增强子的活化，这将导致染色质互作的形成或缺失，并最终影响染色质结构.

2.4　Pfam数据库

通过搜寻Pfam数据库，我们发现转录因子

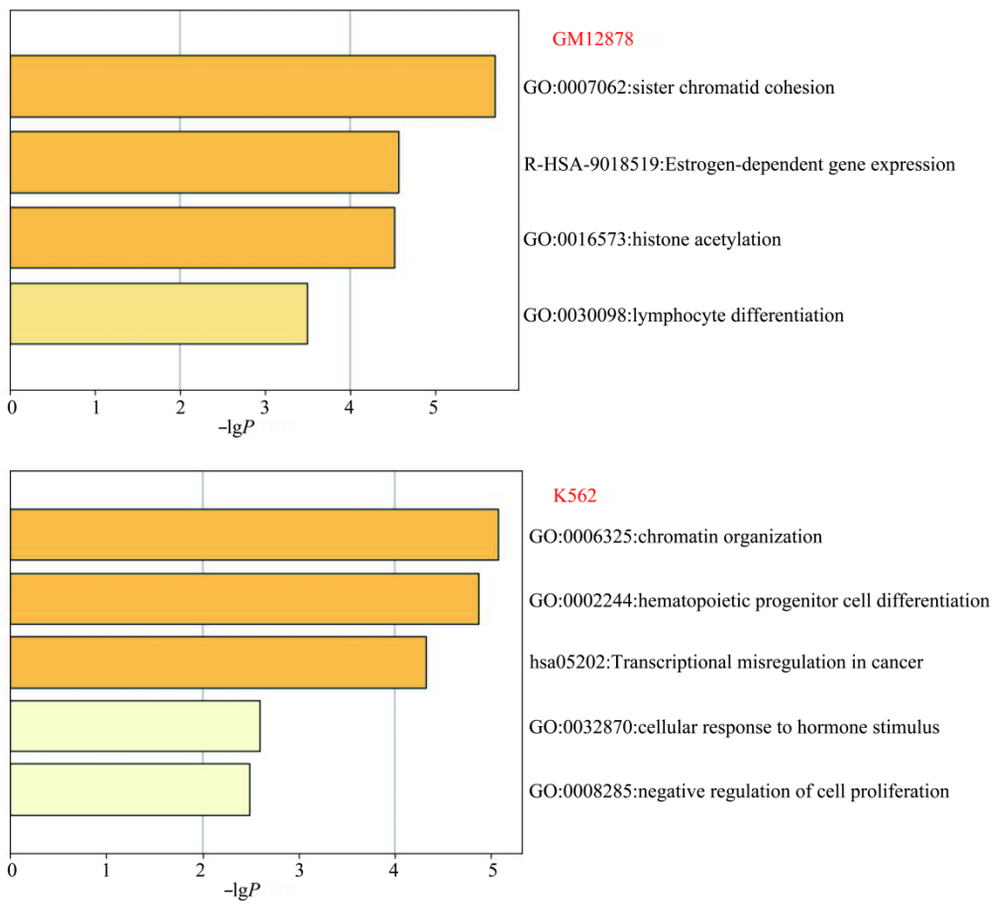

Fig. 5　Functional analysis of transcription factors involved in chromatin interaction in GM12878（above）and K562（below）cell lines

CTCF包含C2H2型的锌指结构域，而在我们发现的转录因子中，MAX、MAZ、REST也包含C2H2锌指结构域，并且EFB1在结构上可以和C2H2锌指结构域结合.这些结构域的存在为它们参与染色质互作提供了基础.

3　总　　结

我们在GM12878和K562两个细胞系中根据染色质位点发生互作的频繁程度对基因组进行划分，并在不同染色质区域对转录因子的结合情况进行富集分析以挖掘与染色质互作相关的转录因子.我们从频繁互作的染色质位点中发现了SMC3、RAD21、CTCF等已知与染色质互作相关的转录因子，还发现了RUNX3、SPI1等可能参与染色质互作的转录因子.此外，我们还发现多个转录因子以很高的概率共定位在染色质频繁发生互作的位点，表明多个转录因子可能共同发挥作用介导染色质互作的形成.

我们的研究发现了一些新的可能与染色质互作有关的转录因子，并发现了一些转录因子的组合与染色质的频繁互作相关.这些转录因子参与染色质互作形成的机制目前还不清楚，我们的结果将为相关研究的开展提供先验知识.

附　件　20180303S1_GM12878_GO.rar和20180303S2_k562_GO.rar见本文网络版（http：//www.ibp.ac.cn）.

参　考　文　献

[1]　Berk A, Lodish H, Zipursky S L, *et al*. Organizing Cellular DNA into Chromosomes. New York: W. H. Freeman, 2000

[2]　Göndör A, Ohlsson R. Enhancer functions in three dimensions: beyond the flat world perspective. F1000 Research, 2018, **7**: 681

[3]　Palstra R J, Grosveld F. Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. Frontiers in

Genetics, 2012, **3**: 195

[4] Roy S S, Mukherjee A K, Chowdhury S. Insights about genome function from spatial organization of the genome. Human Genomics, 2018, **12**(1): 8

[5] Miele A, Gheldof N, Tabuchi T M, *et al*. Mapping chromatin interactions by chromosome conformation capture. Curr Protoc Mol Biol. 2006, chapter 21:Unit 21.11-Unit 21.11

[6] Wei G, Zhao K. 3C-based methods to detect long-range chromatin interactions. Frontiers in Biology, 2011, **6**(1): 76-81

[7] Van Berkum N L , Liebermanaiden E , Williams L , *et al*. Hi-C: a method to study the three-dimensional architecture of genomes. Journal of Visualized Experiments Jove, 2010, **39**(39): 292-296

[8] Tang Z , Luo O J, Li X , *et al*. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell, 2015, **163**(7): 1611-1627

[9] Bonev B , Cavalli G . Organization and function of the 3D genome. Nature Reviews Genetics, 2016, **17**(11):661-678

[10] Liebermanaiden E, van Berkum N L, Williams L, *et al*. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 2009, **326**(5950): 289-293

[11] Corces M R , Corces V G . The three-dimensional cancer genome. Current Opinion in Genetics & Development, 2016, **36**(1): 1-7

[12] Engreitz J M, Agarwala V, Mirny L A. Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. Plos One, 2012, **7**(9): e44196

[13] Guillen-Ahlers H , Shortreed M R , Smith L M , *et al*. Advanced methods for the analysis of chromatin-associated proteins. Physiological Genomics, 2014, **46**(13): 441-447

[14] Spitz F , Furlong E E M . Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics, 2012, **13**(9): 613-626

[15] Weintraub A S , Li C H , Zamudio A V , *et al*. YY1 is a structural regulator of enhancer-promoter loops. Cell, 2017, **171**(7): 1573-1588

[16] Hirano T. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. Genes & Development, 2002, **16**(4): 399-414

[17] Kornberg R D. Mediator and the mechanism of transcriptional activation. Trends in Biochemical Sciences, 2005, **30**(5): 235-239

[18] Sanborn A L , Rao S S P , Huang S C , *et al*. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci USA, 2015, **112**(47): E6456-6465

[19] Allen B L, Taatjes D J. The Mediator complex: a central integrator of transcription. Nature Reviews Molecular Cell Biology, 2015, **16**(3): 155-166

[20] Malik S, Roeder R G. Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends in Biochemical Sciences, 2005, **30**(5): 256-263

[21] Xu H , Balakrishnan K , Malaterre J , *et al*. Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice. Plos One, 2010, **5**(8): e12112

[22] Ye B Y, Shen W L, Wang D, *et al*. ZNF143 is involved in CTCF-mediated chromatin interactions by cooperation with cohesin and other partners. Molecular Biology, 2016, **50**(3): 431-437

[23] Ay F , Bailey T L , Noble W S . Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Research, 2014, **24**(6): 999-1011

[24] Rao S S, Huntley M H, Durand N C, *et al*. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell, 2014, **159**(7): 1665-1680

Mining and Functional Analysis of Chromatin Interaction-related Transcription Factors*

WANG Qi, DING Yan-Shu, GENG Bao-Bao, NIE Yu-Min**

(*Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing* 210029, China)

Abstract Chromatin interaction is the basis of eukaryotic genome assembly and plays an important role in regulating cell-specific expression of eukaryotic genes. The formation of chromatin interactions is associated with specific proteins. It has been suggested that CTCF and adhesion proteins are associated with chromatin interactions, but it is still unclear whether other proteins are involved in chromatin interactions. In this paper, we integrated Hi-C and ChIP-seq data to mine transcription factors related to chromatin interaction in both GM12878 and K562 cell lines, and performed the functional analysis of discovered transcription factors. We found that RUNX3, SPI1 and other transcription factors may also be involved in chromatin interaction in chromatin sites where interaction occurs frequently. In addition, using the data mining method of FP-growth, we found that multiple transcription factors may cooperate to participate in chromatin interaction. Our findings will provide prior knowledge for the development of chromatin interaction experiments.

Key words chromatin interaction, transcription factors, FP-growth, Hi-C

DOI：10.16476/j.pibb.2018.0303

---

* This work was supported by grants from the Natural Science Foundation of Jiangsu Province（BK20161026），Natural Science Foundation of the Higher Education Institutions of Jiangsu Province （16KJB180022），Science and Technology Development Foundation of Nanjing Medical University（2015NJMUZD003）.

** Corresponding author.

Tel：86-25-86869366，E-mail：yumin_nie@njmu.edu.cn

Received：November 23，2018 Accepted：February 25，2019