文章编号：2095-2163(2020)03-0001-07 中图分类号：TP18 文献标志码：A

生物信息学方法预测增强子及其作用位点综述

章天骄，王亚东
（哈尔滨工业大学 计算机科学与技术学院，哈尔滨 150001）

摘 要：对于多细胞真核生物来说，细胞的特异性功能是十分重要的。这就要求在相同遗传物质的基础上，细胞能够通过不同的基因表达模式来适应环境的变化。基因表达调控的因素有很多，近年来随着对基因组非编码区的研究，发现了一些非编码的 DNA 序列对于基因表达调控具有重要意义。增强子是对基因表达调控具有重要作用的非编码序列元件之一。一些增强子能够通过转录产生具有调控功能的 RNA，也被称为增强子 RNA（enhancer RNA，eRNA）。因此对于增强子的序列特征、作用位点以及在特定时间和特定组织中表达模式的研究成为了基因表达调控领域的一个重要问题。

关键词：增强子；eRNA；基因表达调控

Review of bioinformatics methods for predicting enhancers and their targets

ZHANG Tianjiao, WANG Yadong
(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

【Abstract】 Specific function of cells is very important for multicellular eukaryotes. Despite the same genetic material, cells can adapt to environment changes through different gene expression patterns. There are many factors in gene expression regulation. In recent years, with the study of genomic non-coding region, it has been found that some non-coding DNA sequences are important for gene expression regulation. Enhancers are one of the non-coding sequence components that play an important role in gene expression regulation. Some enhancers can be transcribed into RNA with regulatory function, also known as enhancer RNA (eRNA). Therefore, the study on enhancer sequence features, target genes, and expression patterns in specific time and specific tissues has become an important issue in the field of gene expression regulation.

【Key words】 enhancer; eRNA; gene expression regulation

0 引 言

增强子是 DNA 序列上增强基因表达的顺式调控元件，通常位于转录起始位点较远的位置。与启动子不同，增强子对基因表达的调控作用具有高度明显的组织特异性。增强子对基因表达的调控不是“开关”模式，而是一种可变调控，即影响基因表达量的高低，而不是直接关闭或开启表达的调控方式[1]。

增强子最初于 1981 年由 Banerji 和 Moreau 等人在猿猴空泡病毒 40（SV40）的基因组中被发现[2]。1983 年在小鼠免疫重链基因中发现了第一个非病毒的增强子[3]。哺乳动物中增强子的数目为 50 000 到 100 000 个。大部分增强子位于内含子区和基因间区，少部分位于外显子区[4]。在相近的基因组区域内多个增强子聚集成簇的现象被称为超级增强子或增强子簇。研究发现其横跨很大的基因组区同时富集了大量的转录因子及转录中介复合物[5]。超级增强子经常位于细胞特异性功能基因的附近，并且富含细胞特异性转录因子结合序列模体。虽然超级增强子被广泛应用于多个研究中，但是却没有一个清晰的定义[6]。

1 增强子概述

增强子通过与其目标基因启动子相互作用实现对基因的表达调控。这种相互作用，可能是顺式（in cis）作用，也可能是反式（in trans）作用。顺式作用是指增强子及其作用位点基因在同一条染色体上，反式作用则指增强子及其作用位点基因在不同的染色体上[7]。目前对增强子调控基因表达有 2 种模型，如图 1 所示[8]。第一种是轨道调控模型，即 RNA 聚合酶 II 及其转录复合物沿着 DNA 轨道从增强子到启动子滑动。虽然这种模型在一些例子中被证实是存在的[9]，但过去二三十年的研究更加支持另一种模型。第二种模型是环状调控模型，增强子通过染色质成环现象与其调控基因的启动子区域相互临近[10]。图 2 显示了增强子通过染色质成环与启动子临近调控基因表达的现象[11]。

增强子内部包含多种遗传标记位点，最常见的是转录因子结合位点。转录因子与转录复合物的辅

作者简介：章天骄（1985-），男，博士研究生，主要研究方向：生物信息技术；王亚东（1964-），男，教授，主要研究方向：机器学习、知识工程、生物信息技术等。

收稿日期：2019-04-16

助激活因子 p300/CBP 通过相互作用富集在增强子区域内。这些与转录因子结合的区域内核小体的结合度显著下降,导致容易被脱氧核糖核酸酶 I(Deoxyribonuclease I,简称 DNase I)剪切。这些核小体缺失区域(Nucleosome-depleted Regions,NDRs)两侧被特殊的组蛋白修饰所标记,例如 H3K4me1 和 H3K27ac。H3K4me1 与不活跃和活跃的增强子均相关联,H3K27ac 只与活跃的增强子相关联。图 3 显示了增强子内部常见的遗传标记[1]。

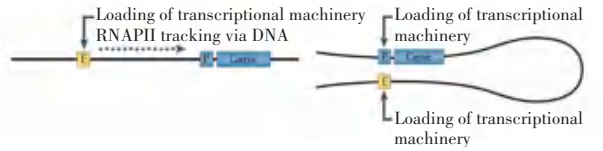

图 1 2 种增强子转录调控模型[8]

Fig. 1 Two models of enhancer transcription regulation[8]

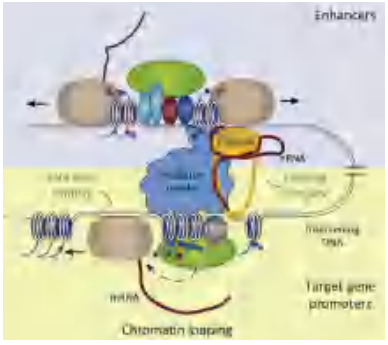

图 2 染色质成环与增强子的转录调控作用[11]

Fig. 2 Chromatin looping and enhancer transcription regulation[11]

图 3 增强子内的遗传标记[1]

Fig. 3 Genetic marks of enhancer[1]

第一次发现增强子的转录现象是在 β-珠蛋白的基因座控制区(Locus Control Region,LCR),随着高通量测序技术的广泛应用,增强子的转录被发现是一种普遍现象。2010 年,Kim 等人[12]对小鼠的中枢神经细胞中剔除 rRNA 后的其它 RNA 测序时,发现了增强子 RNA 的双向转录现象。这种由增强子经过转录过程表达出的 RNA 为 eRNA。这个发现使得人们认识到增强子区域不仅富集转录因子,更是一个转录激活区。与之前的研究一致,增强子区域同样富集转录起始复合物,例如 RNA 聚合酶 II[13]。

eRNA 的长度通常较长,因此 eRNA 经常被认为是长非编码 RNA(Long Noncoding RNA,lncRNA)的子集。然而,一部分 eRNA 却不在 lncRNA 的数据库中[14]。造成这种现象有 2 层原因,其一是 eRNA 通常是通过其转录区域具有增强子遗传标记所发现的[14],而 lncRNA 则是根据其长度大于 200 bp 所定义的[14];其二是 eRNA 由于其不稳定性或者表达量较低没能被构建 lncRNA 数据库的方法所识别[14]。由于这种现象导致 eRNA 和 lncRNA 存在交集,故而将交集部分定义为 lnc-eRNA。

eRNA 作为一种新的转录单元,与其它的转录单元既有相似性也有不同之处。图 4 列出了 4 种常见的转录单元[8]。其中,启动子上游转录区(Promoter Upstream Transcripts,PROMPTs)由于与 eRNA 具有相似的性质而被单独列出来。这四种常见转录单元的部分性质及特征对比见表 1[8]。

图 4 4 种转录单元[8]

Fig. 4 Four transcription units[8]

表 1 4 种转录单元的特征[8]

Tab. 1 Features of four transcription units[8]

Features	eRNA	PROMPT	lncRNA	mRNA
DNase HS	Yes	Yes	Yes	Yes
H3K4me1	High	High	Medium	Low
H3K4me3	Low	Low to medium	Medium	High
H3K36me3	No	No/Low	Yes	Yes/High
H3K27ac	High	High	High	High
RNAP II	Yes	Yes	Yes	Yes
RNAP II Tyr1p	High	High	Unclear	Low
RNAP II Ser2p	No	Yes/low	Yes	Yes/High
RNAP II Ser5p	Yes	Yes	Yes	Yes
RNAP II Ser7p	Yes	Yes	Unclear	Yes
CpG island	Low	High	Medium	High
Splicing	Rare	Rare	Common	Yes
Polyadenylation	Some	Some	Mostly	Mostly
Stability	Low	Low	Low to medium	High
Conservation	Low	Unclear	Medium to high	High
Tissue specificity	Extremely high	Unclear	High	Low

2 增强子预测

对于增强子预测的研究是所有相关增强子研究的基础问题。只有在基因组上识别出增强子所在的位置后,才能对增强子的其它性质及功能进行研究。大量的研究使用了不同的生物数据来预测增强子的位置。这些生物数据主要分为 5 种类别,详述如下。

（1）使用了序列保守性数据和转录因子结合位点数据进行计算学分析。Woolfe 等人[15]、Pennacchio 等人[16]和 Visel 等人[17]对不同物种间非编码元件进行保守性分析来预测增强子。其中，Pennacchio 等人[16]对人类和红鳍东方鲀的基因组进行了序列比对，找到保守的非编码区域。然后对这些区域进行模式序列的搜索，找到 2 个物种共有的具有增强基因表达程度的模式序列。然后，对于所有未知功能的保守非编码区域进行打分，对应数学公式可写为：

$$S_i = \sum_{m \in motifs} x_{mi} \times \log \frac{x_{mi}/n_i}{f_m}. \qquad (1)$$

其中，i 表示第 i 个保守性非编码区域；n 表示区域的长度；m 表示序列模式；$motif$ 表示 2 个物种共有的具有增强基因表达程度的模式序列集合；x_{mi} 表示第 i 个保守性非编码区域中序列模式 m 出现的次数；f_m 表示序列模式 m 在背景区域的频率；S_i 表示对第 i 个保守性非编码区域进行打分得到的最终分数。

这种方法能够有效地识别保守的 DNA 序列，但是也包含了很多非增强子序列元件。同时，不是所有的增强子保守性都很高。

Wasserman 等人[18]将转录因子结合位点和保守性分析相结合，对不同物种的非编码区域进行分析来预测增强子。这对于已知结合序列模式信息的转录因子能够很好地预测其结合的基因组位置，但同时也包含了其它非增强子的与转录因子结合的调控元件序列，因此结果具有很高的假阳性。

（2）使用了调控因子的结合数据，包括转录因子的 ChIP-seq 数据和转录辅激活物 p300 的 ChIP-seq 数据。Chen 等人[19]和 Zinzen 等人[20]使用了转录因子的 ChIP-seq 数据来预测增强子。Chen 等人[19]使用了 13 种转录因子（Nanog，Oct4，STAT3，Smad1，Sox2，Zfx，c-Myc，n-Myc，Klf4，Esrrb，Tcfcp2l1，E2f1 和 CTCF）和 2 种转录调控元件（p300 和 Suz12）对胚胎干（Embryonic Stem，ES）细胞的转录调控网络进行构建。这种方法能够识别那些与已知转录因子结合的增强子。但是这种方法需要知道具体的转录因子来进行实验设计。同时也不能区分增强子和启动子区域，因为这些区域都会结合转录因子。另一方面也不是所有的增强子都与转录因子相结合。

Visel 等人[21]和 May 等人[22]使用了转录辅激活物 p300 的 ChIP-seq 数据来预测增强子。Visel 等人[21]对全基因组 p300 的 ChIP-seq 数据进行分析得到了 p300 的富集位置。然后针对这些 p300 富集的位置是否具有增强子活性进行检测，发现约 88%的位置具有增强子的活性。并且发现这些 p300 富集的位置大多是保守的。这种方法被广泛地应用于增强子位置的预测，但对于活跃的增强子和不活跃的增强子的区分效果不好。

（3）是与染色质的可及性（Chromatin Accessibility）相关的数据。染色质的可及性是指染色质缠绕的核小体从致密变为松散，导致转录调控元件可以顺利结合到其上面起调控作用的性质。应用染色质的可及性来识别增强子主要有以下 3 种技术。Dorschner 等人[23]使用了脱氧核糖核酸酶 I 超敏感位点测序（DNase I Hypersensitive Sites Sequencing，DNase-seq）数据。Giresi 等人[24]使用了甲醛辅助分离调控元件测序（Formaldehyde-assisted Isolation of Regulatory Elements Sequencing，FAIRE-seq）数据。Buenrostro 等人[25]使用了转座酶可及染色质测序（Assay for Transposase-accessible Chromatin Sequencing，ATAC-seq）数据。DNase-seq 需要使用大量的细胞用于实验，而 ATAC-seq 需求的细胞量则很少，同时实验周期也相对较短。但是应用染色质可及性数据进行预测同样也会使结果存在假阳性，即会有其他转录调控单位，例如启动子、隔离子和沉默子等被包含进结果中。

（4）是组蛋白修饰数据。Heintzman 等人[26]应用了 H3K4me1 和 H3K27ac 来预测增强子的位置，这两种组蛋白修饰前者是与增强子特定相关，后者则是与激活的调控区相关，同时应用这两种组蛋白修饰预测得到的基因组调控区域便是激活的增强子区域。此外还有多种组蛋白修饰与 DNA 序列调控元件的关联关系：H3K4me3 与启动子相关联，H3K4me2 与启动子和增强子都相关，H3K9ac 也与激活的调控区相关，H3K36me3 和 H4K20me1 与转录区相关，H3K27me3 与多梳抑制区域相关等。这种预测方法的优点是不同物种间组蛋白修饰数据来源广泛，能够有效地辅助不同需求的研究，缺点是全基因组范围内组蛋白修饰信号十分广泛，不利于高精度预测增强子位置。

（5）是基于 eRNA 数据进行预测。上文中已经提到，增强子会转录出 eRNA，这部分 eRNA 通过测序技术被检测到再映射回原基因组就能得到增强子的位置信息。Kim 等人[12]使用了 RNA-seq 技术，这种技术的优点是 eRNA 和其附近的基因表达水平

可以同时被量化,缺点是低表达水平的 eRNA 不会被检测到。Lai 等人[27],Melgar 等人[28],Mayer 等人[29]分别使用了染色质关联 RNA 测序(Chromatin-associated RNA-seq,ChAR-seq)、GRO-seq 和 NET-seq 技术来检测 eRNA,这三种技术的优势是都可以检测不稳定的 eRNA。Andersson 等人[30]使用了 CAGE 技术来检测 eRNA。这种技术的优点在于可以高精度地确定 eRNA 的转录起始位点,缺点是对于检测表达量较低的 eRNA 需要的样本量较大。同时所有基于 eRNA 数据确定增强子位置的方法都不能用于预测未表达的增强子。

综上,不同的增强子预测方法在本质上是使用了增强子附近不同的生物信号数据。图 5 比较了不同增强子识别方法的差异[31]。

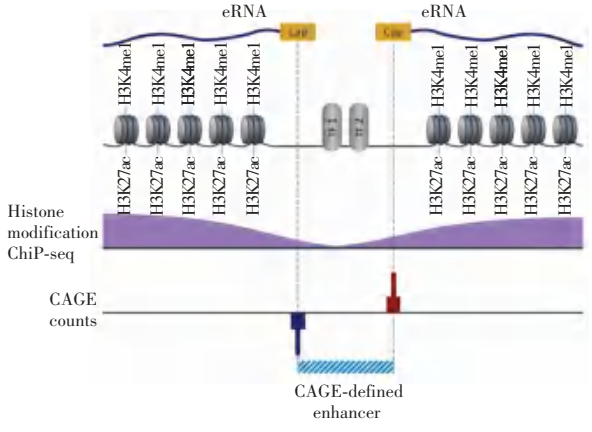

图 5 增强子识别方法的比较[31]

Fig. 5 Comparison of enhancer identification methods[31]

3 增强子作用位点预测

增强子的作用位点指的是增强子对基因表达起增强作用。这种作用体现在增强子与基因的启动子区域相互作用,从而调控基因的表达。通常认为增强子与启动子的相互作用是通过物理上的成环结构实现的。这种物理上的近邻会募集多种转录因子和转录辅助因子结合到增强子和启动子区域。而这些转录因子等则会吸引 RNA 聚合酶从而引起转录的发生。因此要研究一个增强子的生物学功能,确定其作用位点是至关重要的。目前对于增强子作用位点的研究主要分为 2 类方法:基于生物学实验的方法和基于计算学的方法。随着生物实验技术的进步,从生物学实验的角度来研究增强子的作用位点变得准确可靠,但实验成本也随之增高。基于计算学的方法虽然不如实验方法准确,但其高通量的特性和较低的实验成本能够很好地辅助生物学实验的进行。

由于基于生物实验的方法预测增强子作用位点的成本较高,因此需要计算学方法的辅助。近年来由于生物信息学的发展,一系列基于不同增强子特征的计算学预测方法被开发出来。这些计算学方法都需要比较不同细胞类型下增强子附近调控信号的分布模式来预测增强子与基因的关系。

最早被应用于预测增强子与基因调控关系的特征就是基因与增强子间的基因组距离。在定位一个增强子位置后,在基因组上距离其最近的基因被认为是该增强子的调控基因。由于增强子本身具有超远距离调控的性质,这种预测方法的准确率通常不高且变化幅度很大,错误发现率(False Discovery Rate,FDR)约为 40%~73%。由于增强子是对基因表达起增强调控作用,为了提高预测的准确性,研究人员将增强子的调控基因定位为距离其最近的表达基因。这种方案需要用到基因表达数据,其准确性仍然较低,*FDR* 值约为 53%~77%。

Ernst 等人[32]考察了人类 9 种细胞系中增强子附近的组蛋白修饰数据 H3K4me1,H3K4me2 和 H3K27ac。通过对 125 kb 距离内基因的 RNA-seq 表达数据进行关联分析,寻找具有共同变化模式的“增强子-基因”对,来预测增强子的作用位点。这种预测方法使用了增强子的组蛋白修饰数据作为特征与基因表达数据关联,能够一定程度提高预测的准确度。但由于距离的限制只能预测增强子附近 125 kb 范围内的作用位点。

Thurman 等人[33]使用了人类 79 种不同细胞类型的 DHSs 数据来预测增强子的作用位点。研究中通过观察发现不同细胞类型的增强子与其作用位点(基因)的 DHSs 存在很强的关联性。为了挖掘这种关联性,研究时对基因启动子区附近 500 kb 距离内的 DHSs 与基因启动子的 DHSs 计算皮尔森相关系数(Pearson Correlation Coefficient,PCC),数学公式具体如下:

$$r=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}, \tag{2}$$

$$r=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma_X}\right)\left(\frac{Y_i-\bar{Y}}{\sigma_Y}\right). \tag{3}$$

其中,r 表示样本 X 与 Y 的相关系数;$(X_i,\ Y_i)$ 表示容量为 n 的第 i 个样本点;$\bar{X}$ 表示样本 X 的平均值;σ_X 表示样本 X 的标准差。

Thurman 等人[33]识别出相关系数 $r > 0.7$ 的至少与一个启动子相关联的 DHSs 区域。这些区域作为潜在调控基因表达的增强子位点。这种预测方法反映了增强子与启动子的若要行使调控功能则必须在开放的染色质区域内进行这一特点。但并不是所有 DHSs 相关性较高的区域都反映的是增强子与启动子的关联关系，同时 500 kb 的距离也限制了预测的准确度。

Sheffield 等人[34]对 Thurman 等人[33]的方法进行了改进。对 72 种不同细胞类型的基因表达 RNA-seq 数据与 100 kb 距离内的 DHSs 数据进行关联分析，计算皮尔森相关系数。这种预测方法使用了基因的表达数据，能更直观地反映增强子对基因表达的正调控作用。但由于并不是所有增强子调控的基因都处于表达状态，因此对于那些处于沉默状态的基因预测性能不高。同时由于距离限制在 100 kb 范围内，也一定程度影响了预测精度。

Shen 等人[35]将组蛋白修饰数据 H3K4me1，H3K27ac 和 RNA Pol II 数据结合起来预测增强子的作用位点。Shen 等人[35]分别对增强子区域的 H3K4me1 数据和启动子区域的 RNA Pol II 数据，增强子区域的 H3K27ac 数据和启动子区域的 H3K27ac 进行关联分析，计算斯皮尔曼相关系数(Spearman Correlation Coefficient，SCC)，对此会用到如下数学公式：

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}. \tag{4}$$

其中，ρ 表示相关系数；n 表示样本容量；d_i 表示 2 个样本秩次的差值。

这种预测方法使用了 RNA pol II 作为基因启动子区域的标志，使用 H3K4me1 作为增强子区域的标志，同时使用 H3K27ac 作为转录活跃区的标志，通过关联分析，得到增强子与启动子共有的组蛋白修饰变化模式，以此说明二者存在关联关系。使用组蛋白修饰数据能够避免上述依赖表达量方法只能预测处于活跃状态基因的问题。但是由于组蛋白修饰数据本身在基因组中十分广泛，因此预测精度也会受到一定限制。

Andersson 等人[30]使用了 CAGE 数据来预测增强子的作用位点。通过 CAGE 可以测得增强子和基因 TSS 的表达量，选取距离在 500 kb 范围内表达量在 1 TPM 以上的增强子和启动子，计算二者的皮尔森相关系数并进行假设检验。由于上文已经阐述了增强子能够通过转录产生 eRNA 对目标基因起调控作用，因此可以用二者表达量的相关性来预测关联关系。这种基于表达量的方法预测增强子作用位点的优点是能够反映活跃增强子与其调控基因的相关性，但同时有很多增强子的表达量较低、甚至没有表达，这些增强子的作用位点就难以用基于表达量的方法预测，同时 500 kb 的距离也限制了预测的准确度。

Corradin 等人[36]和 Factor 等人[37]分别开发了 PreSTIGE 和 PreSTIGEouse 来预测人类和小鼠中增强子的作用位点。这两种方法都使用了组蛋白修饰数据 H3K4me1 来标识增强子的位置。通过分析与附近基因的表达量相关性确定关联关系。与以往固定距离方法不同，Corradin 等人[36]使用了转录因子 CTCF 的位置数据作为确定增强子与基因间距离的参考。CTCF 是与隔离子活性相关的转录因子，在基因组中起分割作用。增强子只能在隔离子分割的区域内对基因的表达起增强作用。与将最近基因作为增强子作用位点的方法相比，使用 CTCF 数据可以有效降低 *FDR*，其值为 13%～23%。这种方法能够一定程度避免由于距离的不确定性导致预测性能较低的现象。缺点是由于只使用了一种组蛋白修饰数据作为增强子的标识，导致预测结果的准确性仍有一定的不足。

He 等人[38]开发了 IM-PET，应用多种遗传特征的组合来预测增强子的作用位点。研究中根据增强子附近不同遗传特征的类型将所有特征分为 4 个类别，对此可做阐释分述如下。

(1)是反映增强子与启动子表达活跃度的相关性特征。首先通过组蛋白修饰数据 H3K4me1、H3K27ac 和 H3K4me3 估计增强子的表达活性，然后与启动子的 RNA-seq 数据进行关联分析，计算皮尔森相关系数作为第一类特征，简称 EPC。

(2)是反映转录因子与目标启动子关联关系的特征。由于第一类特征只能反映在 DNA 序列层次上的调控，没有反映在转录因子层次上的调控关系，因此需要构建转录因子在增强子区域的结合度与基因表达的关联关系。通过计算二者的皮尔森相关系数作为第二类特征，简称 TPC。

(3)是反映启动子和增强子的保守性的特征。由于启动子和增强子作为调控区域的保守性在序列层次上可能不强，但在同线性的层次上却有着较高的保守性，因此可以分别计算一定距离内启动子和增强子区域在多物种中的保守性得分，将得分标准

化后的乘积作为第三类特征,简称 COEV。

(4)是反映启动子和增强子间距离远近的特征。即转录起始位点到增强子中心的距离作为第四类特征,简称 DIS。

在完成特征集合的构建后,根据已有的训练样本集训练随机森林分类器。然后将此分类器应用到测试样本集上检测性能调整分类器参数。最后对给定的增强子-启动子对数据,应用分类器来预测二者的关联性。图 6 显示了应用 IM-PET 预测增强子与启动子关联关系的流程图。

IM-PET 综合使用了多种特征来分析和预测增强子与启动子间的关联关系。比起前人单纯使用一种或几种特征能够从生物学角度更加全面地描述强子与启动子间的关系。同时由于使用了机器学习的方法对分类器进行训练,使得 *FDR* 大大降低(约 1%),并且可分析的基因组距离大大增加(2 Mb),有效提高了预测精度。但是由于预测方法中对于特征集合的构建使用了较多的特征,并且缺少对这些特征重要程度的描述,使得一些与预测关联性不强的特征也纳入进来,预测结果也会受到一定影响。

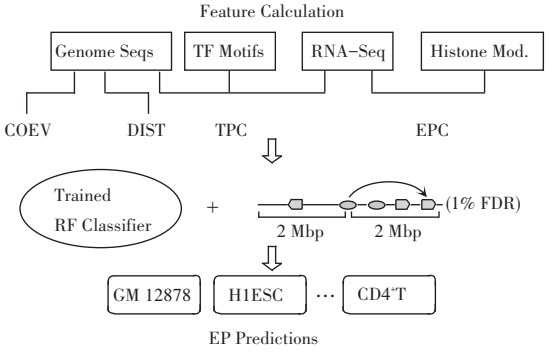

图 6 应用 IM-PET 预测增强子-启动子关联关系流程图[38]

Fig. 6 Schematic diagram for making genome-wide EP predictions using IM-PET[38]

以上从计算学的角度总结了预测增强子作用位点的方法。这些方法大多围绕着提取增强子与启动子附近的相关特征,构建特征值间的关联关系来预测增强子的作用位点。基于计算学的方法能够高效经济地识别潜在的增强子与基因的相互作用关系,但是对于增强子与启动子的反式作用则显得乏力。这些应用特定细胞条件下特征一致性变化的预测方法难以实现对所有类型细胞中均表达的管家基因的预测。因此,这些计算学的方法都给出了生物实验验证增强子与启动子的关联关系。

4 结束语

增强子是基因表达调控的重要元件之一。对于增强子本身的识别及其作用位点的预测一直是相关领域的研究热点问题。近年来生物数据监测技术的不断进步带来了海量的生物数据,同时生物信息技术的发展为研究增强子的生物学功能提供了强大的技术手段。本文总结了目前生物信息领域对增强子相关问题的研究热点,着重总结了增强子及其作用位点预测的研究方法。

参考文献

[1] CORRADIN O, SCACHERI P C. Enhancer variants: Evaluating functions in common disease[J]. Genome Medicine, 2014, 6(10): 85.

[2] BANERJI J, RUSCONI S, SCHAFFNER W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences[J]. Cell, 1981, 27(2 Pt 1): 299.

[3] BANERJI J, OLSON L, SCHAFFNER W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes[J]. Cell, 1983, 33(3): 729.

[4] BIRNBAUM R Y, CLOWNEY E J, AGAMY O, et al. Coding exons function as tissue-specific enhancers of nearby genes[J]. Genome Research, 2012, 22(6): 1059.

[5] HNISZ D, ABRAHAM B J, LEE T I, et al. Super-enhancers in the control of cell identity and disease[J]. Cell, 2013, 155(4): 934.

[6] POTT S, LIEB J D. What are super-enhancers? [J]. Nature Genetics, 2015, 47(1): 8.

[7] SASAKI-IWAOKA H, MARUYAMA K, ENDOH H, et al. A trans-acting enhancer modulates estrogen-mediated transcription of reporter genes in osteoblasts[J]. Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research, 1999, 14(2): 248.

[8] LI W, NOTANI D, ROSENFELD M G. Enhancers as non-coding RNA transcription units: Recent insights and future perspectives [J]. Nature Reviews. Genetics, 2016, 17(4): 207.

[9] HATZIS P, TALIANIDIS I. Dynamics of enhancer-promoter communication during differentiation-induced gene activation[J]. Molecular Cell, 2002, 10(6): 1467.

[10] WANG Qianben, CARROLL J S, BROWN M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking [J]. Molecular Cell, 2005, 19(5): 631.

[11] LAM M T Y, LI Wenbo, ROSENFELD M G, et al. Enhancer RNAs and regulated transcriptional programs [J]. Trends in Biochemical Sciences, 2014, 39(4): 170.

[12] KIM T K, HEMBERG M, GRAY J M, et al. Widespread transcription at neuronal activity-regulated enhancers[J]. Nature, 2010, 465(7295): 182.

[13] KOCH F, FENOUIL R, GUT M, et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters[J]. Nature Structural & Molecular Biology, 2011, 18(8): 956.

[14] DERRIEN T, JOHNSON R, BUSSOTTI G, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression[J]. Genome Research, 2012, 22(9): 1775.

(下转第 13 页)