

DOI: 10.3724/SP.J.1005.2012.00950

哺乳动物转录因子 DNA 结合谱

谷光明, 王进科

东南大学生物电子学国家重点实验室, 南京 210096

摘要: 基因差异表达是生物发育和对刺激作出应答的分子基础, 转录因子在这种基因差异表达中发挥着重要的调控作用。因此, 要弄清楚转录因子调控基因差异表达的机理, 就必须鉴定出它们全部的靶基因并构建其操纵的转录调控网络。对基因组 DNA 的序列特异性结合是转录因子调控基因转录的关键环节, 因此, 要鉴定转录因子的靶基因, 就必须从它们与 DNA 相互作用的分子水平, 鉴定它们能够识别并结合的全部 DNA 序列, 即转录因子 DNA 结合谱。近年来随着 DNA 微阵列芯片和高通量 DNA 测序技术的产生和快速发展, 出现了建立转录因子体内及体外 DNA 结合谱的一系列革命性的新技术, 对该领域的研究带来重大影响。这些新技术主要包括建立转录因子体内 DNA 结合谱的染色质免疫沉淀-芯片技术(ChIP-chip)和染色质免疫沉淀-测序技术(ChIP-Seq), 以及建立转录因子体外 DNA 结合谱的双链 DNA 微阵列芯片技术(dsDNA microarray)、指数富集配体系统进化-系列分析基因表达技术(SELEX-SAGE)、结合-n-测序技术(Bind-n-Seq)、多重大规模并行 SELEX 技术(MMP-SELEX)、凝胶迁移实验-测序技术(EMSA-Seq)和高通量测序-荧光配体互作图谱分析技术(HiTS-FLIP)。文章将对这些新技术做一综述。

关键词: 哺乳动物转录因子; DNA 结合谱; DNA 微阵列芯片; 高通量 DNA 测序

DNA-binding profiles of mammalian transcription factors

GU Guang-Ming, WANG Jin-Ke

The State Key Laboratory of Bioelectronics, Southeast University, Nanjing 210096, China

Abstract: The differential gene expression is the molecular base of development and responses to stimuli of organisms. Transcription factors (TFs) play important regulatory roles in this kind of differential gene expression. Therefore, to elucidate how these TFs regulate the complex differential gene expression, it is necessary to identify all target genes of them and construct the gene transcription regulatory network controlled by them. DNA binding is a key step for TFs regulating gene transcription. Therefore, in order to identify their target genes, it is indispensable to identify all possible DNA sequences that can be recognized and bound by TFs at the molecular level of their interactions with DNA, i.e., construction of the DNA-binding profiles of TFs. In recent years, along with the development of DNA microarray and high-throughput DNA sequencing techniques, there appeared some revolutionary new techniques for constructing DNA-binding profiles of TFs,

收稿日期: 2011-12-20; 修回日期: 2012-05-06

基金项目: 国家自然科学基金项目(编号: 60871014, 61171030)资助

作者简介: 谷光明, 硕士研究生, 专业方向: 生物物理。E-mail: gulong106@126.com

通讯作者: 王进科, 博士, 教授, 博士生导师, 研究方向: 生物医学工程。Tel: 025-83793620; E-mail: wangjinke@seu.edu.cn

网络出版时间: 2012-6-15 09:34

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20120615.0934.010.html>

which greatly promotes studies in this field. These techniques include ChIP-chip and ChIP-Seq for constructing *in vivo* DNA-binding profiles of TFs, dsDNA microarray, SELEX-SAGE, Bind-n-Seq, MMP-SELEX, EMSA-Seq, and HiTS-FLIP for constructing *in vitro* DNA-binding profiles of TFs. This paper reviewed these techniques.

Keywords: mammalian transcription factor; DNA-binding profile; DNA microarray; high-throughput DNA sequencing

转录因子是参与基因表达调控的一类重要的细胞核蛋白质。转录因子分为基础转录因子和调控性转录因子两类,其中基础转录因子与RNA聚合酶在转录起始位点附近的启动子区形成转录机器,实现基因的转录,而调控性转录因子一般与位置多样的增强子序列结合,再通过形成增强体与转录机器相互作用,实现基因转录的调控。调控性转录因子控制复杂多样的生物学过程,在生物对环境和生理刺激作出应答,以及生物的物种差异发育过程中发挥重要作用。例如,少数几个调控性转录因子就可以使高度分化的体细胞实现基因组重编程,逆转为与胚胎干细胞相似的诱导性多能干细胞(Induced pluripotent stem cells, iPS cells)^[1-3]。调控性转录因子基因的突变、活性异常以及其DNA结合靶点序列的多态性等,与人类众多疾病的发生发展、个体疾病的易感性等关系非常密切。如p53与癌症^[4,5],NF- κ B与炎症和肿瘤^[6]。

调控性转录因子一般是序列特异性DNA结合蛋白,在人基因组中已鉴定出的2 000多个转录因子中^[7],有700多个转录因子就属于调控性转录因子。已经证明这些转录因子控制着生物应答及物种发育过程中的基因差异表达,因此要弄清楚这些转录因子调控基因差异表达的分子机理,就必须鉴定出这些转录因子全部的靶基因并构建其操纵的转录调控网络。DNA序列特异性结合是这类转录因子调控基因转录的关键环节,因此,要鉴定它们的靶基因,就必须从它们与DNA相互作用的分子水平,鉴定出它们能够识别并结合的全部DNA序列,即转录因子DNA结合谱。此外,调控性转录因子与不同的DNA结合靶点之间的相互作用具有不同的特异性和亲和性,这些差异形成了对基因表达的一种精密微调控机制。因此,完整地揭示一个转录因子的所有DNA结合靶点,以及转录因子对这些DNA结合靶点的结合

特异性及亲和性,即绘制转录因子DNA结合谱,对于揭示转录因子与不同DNA序列结合的柔性分子机理、弄清转录因子对众多靶基因的精密调控机制是必不可少的。

转录因子DNA结合谱研究的重要价值主要体现在以下两方面:一是可以依据转录因子DNA结合谱信息精确而全面地鉴定转录因子在基因中的所有功能性DNA结合靶点及其调控的靶基因,进而绘制转录因子操纵的转录调控网络;二是依据DNA结合谱及转录因子与DNA结合的柔性分子机理,可以指导设计人工转录因子、小分子转录因子模拟物以及转录因子诱骗子等基因转录治疗药物分子。但截至目前,在已鉴定的700多个人调控性转录因子中,几乎没有一个转录因子的所有DNA靶点及靶基因得到彻底而明确地表征,这成为当前构建基因转录调控网络面临的重大障碍。为解决这一难题,近年来随着DNA微阵列芯片和DNA高通量测序技术的产生和快速发展,出现了建立转录因子体内及体外DNA结合谱的一些革命性新技术,对该领域的研究带来重大影响。

转录因子DNA结合谱根据研究的途径不同,分为转录因子体内DNA结合谱和转录因子体外DNA结合谱。下面就这两种结合谱的基本概念和主要研究方法,以及这些研究方法的优缺点予以概述。

1 体内结合谱

转录因子体内DNA结合谱是指转录因子在细胞内实际结合的所有DNA靶点。绘制转录因子体内DNA结合谱目前主要依赖于染色质免疫沉淀(Chromatin immunoprecipitation, ChIP)及其下游检测技术(图1)。ChIP技术由Orlando等^[8]于1997年创立,该技术的基本流程为:培养细胞、诱导转录因子高表达、甲醛交联(固定处于结合状态的转录因子蛋白与

DNA)、裂解细胞、超声打碎染色质DNA、借助免疫沉淀富集与转录因子结合的DNA片段^[9, 10], 最后再通过下游检测技术分析富集的DNA片段。分析通过染色质免疫沉淀富集的DNA的方法很多, 如传统的狭缝杂交和聚合酶链反应(Polymerase chain reaction, PCR)扩增^[11, 12]; Southern杂交^[8, 13]及ChIP克隆(ChIP-cloning)^[14], 但这些技术的通量低, 只能识别个别DNA靶点及靶基因。2000年Ren等^[15]首次应用ChIP-芯片(ChIP-chip)技术在全基因组范围内鉴定了转录因子Gal4和Ste12的结合位点。2006年Wei等^[16]首次将ChIP-配对端点双标记测序技术(ChIP with paired-end tags, ChIP-PET)用于分析转录因子p53的ChIP富集DNA片段。2007年Johnson等^[17]和Robertson等^[18]首次用ChIP-测序(ChIP-sequencing, ChIP-Seq)技术分析转录因子NRSF和SATA1的ChIP富集DNA片段。ChIP-chip、ChIP-PET及ChIP-Seq的产生, 使ChIP富集DNA片段的分析进入高通量时代, 实现了转录因子结合位点的全基因组分析鉴定。

随着ChIP-PET、ChIP-chip^[19, 20]和ChIP-Seq技术在转录因子体内结合谱研究中的应用, 近年来绘制出许多重要转录因子的体内DNA结合谱。例如, 用ChIP-PET技术研究了细胞重编程及胚胎干细胞多能性相关转录因子Oct4、Nanog及c-Myc的DNA靶点和靶基因^[21]; 用ChIP-chip技术研究了转录因子Sox2、Oct4、Nanog、c-Myc、Klf4、Sall4及STAT3的体内DNA结合靶点和靶基因^[22-26]; 用ChIP-Seq技术研究了转录因子Oct4、Sox2、Nanog、c-Myc、Klf4、STAT3、Smad1及n-Myc的体内DNA结合靶点和靶基因^[27, 28]。这些研究获得了该类转录因子体内DNA结合谱及靶基因的大量信息, 使得构建该类转录因子的靶基因调控线路^[29-31]及网络^[21, 32-35]取得了巨大的进展。依靠这些技术, 也研究了其他重要转录因子的体内DNA结合谱和靶基因, 如p53^[16, 36-40]、p63^[41, 42]、NF- κ B^[43-46]、Sp1^[37]、CREB^[47-49]、E2F1^[50]、ER α ^[51-56]、CRP^[57]、STAT1^[58, 59]、STAT2^[59]、STAT3^[25, 60]、STAT4^[61]、STAT5^[62]、HNF1 α ^[63]、HNF4 α ^[63, 64]、HNF6^[63]、Smad2/3^[65]、PPAR γ ^[66]、PPAR γ /RXR^[67, 68]。与ChIP-chip相比较而言, ChIP-PET和ChIP-Seq是一种对ChIP富集DNA的无偏检测技术, 能够完整显示ChIP富集DNA包含的信息^[48, 69, 70]。并且, ChIP-Seq具有分辨率更高、不受生物芯片上预先设定的基因

组区域的限制、成本不断降低及通量迅速增高等优势, 因而更适合目前高通量绘制DNA结合谱的需求。因此, ChIP-Seq已经基本上取代ChIP-chip和ChIP-PET, 成为研究转录因子、RNA聚合酶、核小体等DNA结合蛋白体内结合靶点的主打技术^[69]。

ChIP技术最大的优点是该技术是一种体内技术, 绘制的转录因子DNA结合谱是一类体内DNA结合谱^[9]。原理上ChIP技术所捕获的DNA是细胞内与转录因子发生结合的DNA, 这些DNA应是具有调控功能的DNA, 但实际情况却并非这样理想。首先, ChIP实验对抗体的品质要求高, 品质较差的抗体以及免疫沉淀环节的低严谨性洗涤往往造成非特异性DNA的沉淀, 产生高背景和假阳性。另外, ChIP技术由于受实验原理的限制, 只能鉴定出某种转录因子在一种细胞中一个很短时间(甲醛交联点)内处于结合态的DNA, 很难反映出一种转录因子在不同组织细胞、不同发育阶段、不同刺激状态下所作用的全部DNA靶点。除此之外, 抗体和转录因子蛋白的结合也受众多因素的干扰, 使得部分蛋白质结合的DNA也难以被ChIP捕获。例如, 增强体中的其他蛋白质可能屏蔽转录因子与抗体的结合, 导致部分DNA信息流失。另外, 从ChIP获得的DNA长片段(一般200~500 bp)中确定转录因子的精确DNA结合靶点(一般6~10 bp)也比较困难。近年来的研究还表明, 有些转录因子常常表现出能够结合非经典的DNA序列, 要从ChIP片段中确定这类DNA靶点, 难度更大。

综上所述, ChIP技术很难反映转录因子在基因组中的所有DNA结合信息^[72], 相应地难以显示出所有靶基因信息。此外ChIP技术绘制的体内DNA结合谱基本上没有反映出转录因子对各DNA靶点结合的亲和性信息, 而转录因子与不同DNA结合靶点相互作用的特异性和亲和性差异, 可能形成转录因子对不同靶基因表达的精密微调。截止目前, 借助ChIP技术鉴定的转录因子靶基因数量还比较有限, 难以绘制出转录因子完整的靶基因转录调控网络。尽管基于ChIP的体内结合谱研究技术还存在上述缺点, 但ChIP技术仍然是目前研究转录因子体内结合靶点的难以替代的核心技术, 因此该技术在转录因子体内结合谱的研究中仍将发挥重要作用。

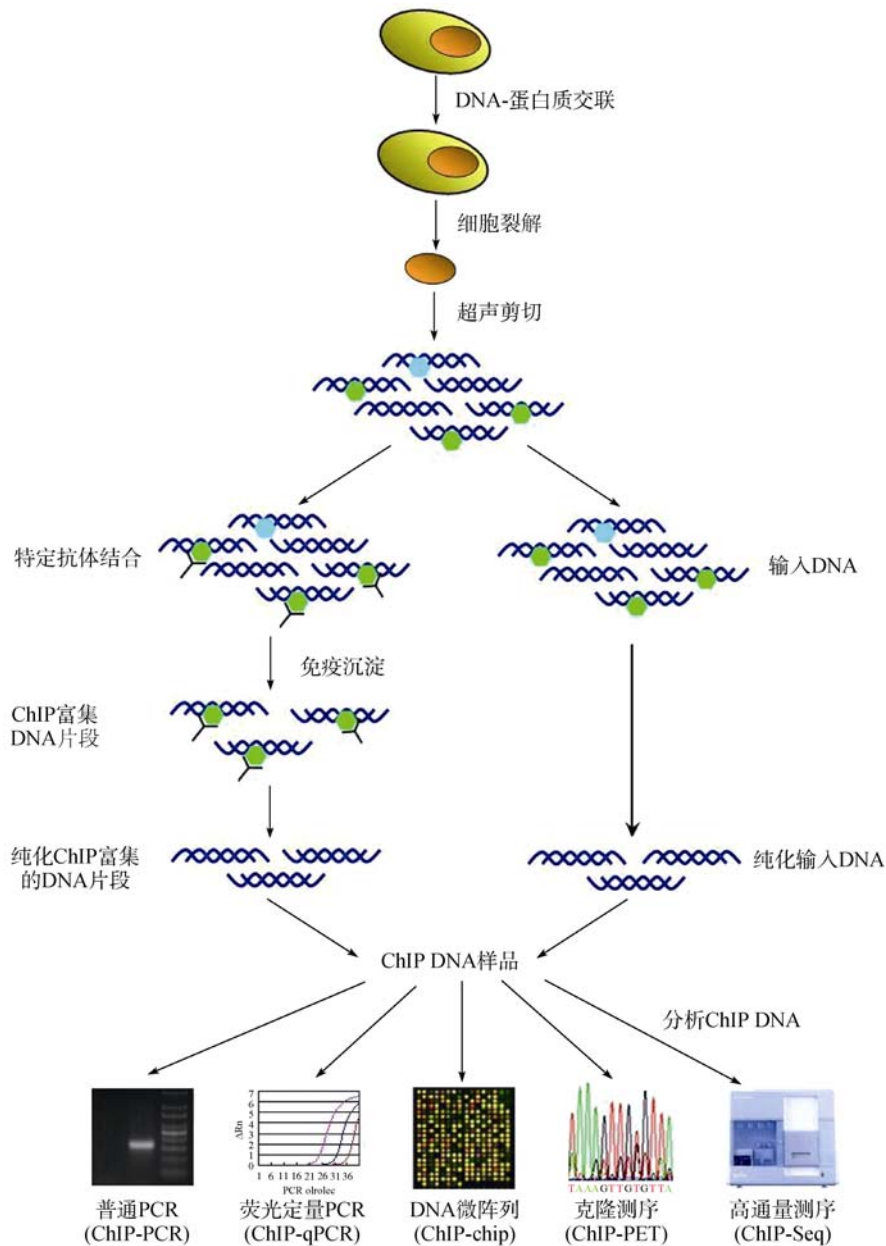


图 1 ChIP 技术流程图

培养细胞、诱导转录因子高表达、甲醛交联(固定处于结合状态的转录因子蛋白与 DNA)、裂解细胞、制备核裂解物、超声打碎染色质 DNA、免疫沉淀富集转录因子结合的 DNA 片段, 最后再通过下游检测技术分析富集的 DNA 片段。下游检测技术包括普通 PCR(检测个别位点, 通量低)、荧光定量 PCR(检测个别位点, 通量低)、DNA 微阵列芯片(如启动子微阵列芯片、CpG 岛微阵列芯片、tiled-DNA 微阵列芯片等)、PET 克隆测序、高通量 DNA 测序(如 Roche/454 FLX、Illumina/Solexa Genome Analyzer、Applied Biosystems SOLiD system, 特别是 Illumina/Solexa Genome Analyzer)等^[21]。

2 体外 DNA 结合谱

鉴于体内 DNA 结合谱存在的缺陷, 近年来转录因子体外 DNA 结合谱的研究得到了越来越多的重视。由于脱离了特定实验材料(如细胞、组织)的限制, 体外结合谱研究的实验过程处于完全人工控制的环

境中, 避免了体内一些复杂的环境因素如其他因子的干扰等, 从而能够完整地显示出转录因子蛋白与 DNA 结合的各种重要信息, 如特异性、亲和性、经典模体(Canonical motif)、非经典模体(Noncanonical motif)、半位点、间隔序列及加和效应等。同时, 通

过使用随机组合 DNA 序列, 体外结合谱研究能够发现一个转录因子结合的所有 DNA 序列以及对各序列的相对亲和性, 因此能绘制出信息完整的高分辨率 DNA 结合谱, 这种结合谱对鉴定功能靶点及靶基因非常重要。目前, 转录因子体外 DNA 结合谱绘制主要依靠双链 DNA 微阵列芯片(dsDNA microarray)技术、与指数富集配体系统进化技术(Systematic evolution of ligands by exponential enrichment, SELEX)相结合的高通量 DNA 测序技术, 以及最新出现的高通量测序-荧光配体互作图谱分析(High-throughput sequencing fluorescent ligand interaction profiling, HiTS-FLIP)技术。

2.1 双链 DNA 微阵列芯片

dsDNA微阵列芯片也称蛋白结合微阵列芯片(Protein-binding microarray, PBM), 最早报道于1999年, 当时美国Harvard大学Bulyk实验室在Affymetrix公

司原位合成的单链DNA微阵列芯片的基础上, 采用引物退火再进行DNA聚合酶延伸的办法制备dsDNA微阵列芯片^[73]。该类芯片在玻片表面固定成千上万的双链DNA探针分子(如含有10碱基随机组合序列的 4^{10} 个探针), 每个探针分子上都含有一个待考察的与某一转录因子蛋白可能结合的DNA位点; 将一种转录因子蛋白与这样的dsDNA微阵列芯片反应, 就可以在一次实验中获得该转录因子蛋白与大量DNA位点的结合信息, 包括结合的特异性及相对亲和性(图2)^[74]。目前, 随机组合序列(k-mer)高密度dsDNA微阵列芯片已经成功地用于绘制转录因子的体外DNA结合谱^[75-77], 使得转录因子及其他DNA结合分子的体外DNA结合谱研究取得了巨大的进展^[78, 79]。截止2009年, Bulyk实验室已经运用PBM技术成功绘制了104个小鼠转录因子蛋白的精细DNA结合谱^[80], 这对认识大量转录因子的

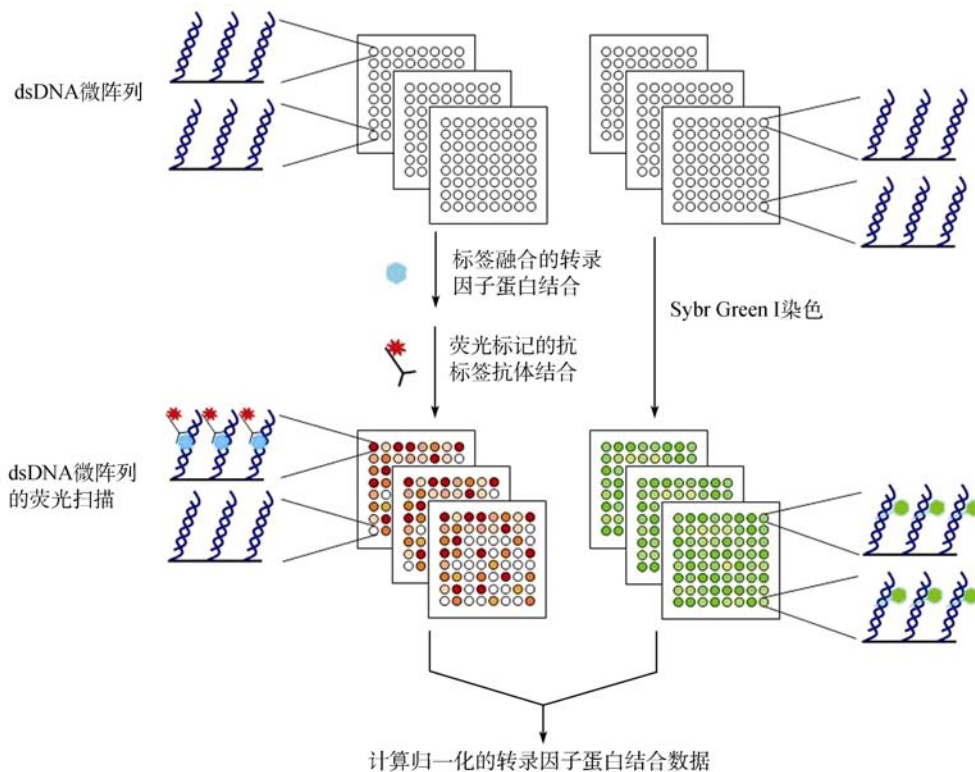


图2 dsDNA微阵列芯片技术

首先原位合成单链DNA微阵列, 再将通用引物与单链DNA上的恒定序列复性, 之后用DNA聚合酶延伸形成dsDNA微阵列。对dsDNA微阵列进行Sybr Green I染色, 再将Tag标记(如6×His、GST)的DNA结合蛋白(转录因子)和红色荧光标记的抗Tag抗体先后与dsDNA微阵列孵育。经过双色荧光扫描, 获取探针密度信号(Sybr Green I染色信号)与DNA结合蛋白结合信号, 以探针密度信号校正DNA结合蛋白结合信号, 获得可相互比较的DNA结合蛋白与大量DNA序列的结合信息, 即相对亲和性, 并筛选出与DNA结合蛋白能够结合的所有DNA序列, 即结合谱。

DNA结合特异性、亲和性等生物学特征发挥了重要作用。这些研究还表明PBM技术可高通量鉴定转录因子蛋白的顺式调控元件和绘制转录调控网络。

与ChIP-chip技术相比较,dsDNA微阵列芯片技术获得的体外结合位点信息能够精确反映ChIP-chip获得的体内结合位点信息。不仅如此,dsDNA微阵列芯片技术还能发现大量ChIP-chip不能够显示的结合位点信息^[82]。此外,dsDNA微阵列芯片技术能够高分辨率反映转录因子与DNA靶点相互作用的规律,如相邻核苷酸间的加和性及核苷酸与氨基酸间的作用关系,可识别出转录因子能够结合的所有DNA序列。利用dsDNA微阵列芯片技术绘制的体外DNA结合谱,可用于提取转录因子的DNA结合模体、预测转录因子的DNA结合靶点、指导设计和筛选人工转录因子、小分子转录因子模拟物以及转录因子诱骗子等。截止2011年,PBM技术已经成功绘制出了415个蛋白及其复合物(几乎全部是转录因子)的体外DNA结合谱,这些数据被系统地储存在专门的数据库UniPROBE(Universal PBM Resource for Oligonucleotide-Binding Evaluation)中^[83,84]。

我们实验室从2003年就开始了基于dsDNA微阵列芯片绘制转录因子DNA结合谱的相关研究。我们提出了3种基于发卡引物的dsDNA微阵列芯片制备技术^[85-87],并以这些技术制备的dsDNA微阵列芯片研究了转录因子NF- κ B与大量DNA结合位点的相互作用^[88]。结果发现NF- κ B对部分突变DNA位点仍具有高亲和力,推测这些位点若在基因组中存在,则可能是NF- κ B的潜在DNA结合靶点,其上下游临近基因有可能是NF- κ B的靶基因。通过对人基因组的全基因组扫描,发现这些位点确实在基因组中有分布,从而将这些位点预测为NF- κ B的假定DNA靶点,同时将这些位点附近的基因预测为NF- κ B的假定靶基因。通过文献检索,发现部分预测的靶基因已经被实验研究鉴定为NF- κ B的真实靶基因(如*NFKB2*、*NFKB1A*、*BCL2*和*VEGFC*等);同时我们从文献未证实的预测靶基因中选择了部分疾病相关基因,并做了ChIP-PCR检测研究,发现这些基因确实受NF- κ B的调控,为其真实靶基因(如*STAT1*、*MIA-53*、*HFE-625*、*LTBP-1*等)(尚未发表)。这些研究提供了一种利用dsDNA微阵列芯片鉴定转录因子靶基因的新策略。

虽然dsDNA微阵列芯片功能强大,但目前使用dsDNA微阵列芯片进行转录因子DNA结合谱的研究还受到一些不利因素的限制。例如,dsDNA微阵列芯片的研究需要设计并制备高密度dsDNA微阵列芯片,而目前能够制备高密度DNA微阵列芯片的公司只有美国Affymetrix和NimbleGen两家,定制一张含有1 048 576个探针(包含所有可能10碱基随机组合序列)的高密度dsDNA微阵列芯片,其价格不低于3万元,加上蛋白、一抗及荧光标记二抗等试剂,一次dsDNA微阵列芯片结合实验的费用在4万元左右。若一个转录因子蛋白做3个重复实验,测定一个转录因子蛋白结合谱的最少费用将达12万元。当DNA结合位点的长度较大时,一张芯片很难容纳所有探针,不得不分片,而这将影响实验中不同探针间的可比性,并且进一步显著增高实验费用。因此,基于dsDNA微阵列芯片绘制转录因子DNA结合谱的研究还存在成本高、实验环节烦琐等不足,依靠dsDNA微阵列芯片绘制大量转录因子的体外DNA结合谱目前还面临一定的困难。

2.2 基于SELEX和DNA克隆测序的传统技术

2.2.1 SELEX-克隆测序技术

SELEX技术是指从随机寡核苷酸文库中筛选出对给定靶分子具高亲和力的寡聚物,从而推论出那些对DNA或RNA结合蛋白具有很高亲和力的配体。这些经SELEX筛选出的高亲和序列可应用于多项研究以及疾病的诊断和治疗等^[89]。图3为该技术的流程图。用于筛选转录因子DNA结合序列时,该技术的基本步骤为:先体外化学合成一个两端带有PCR引物复性位点,中间含有待筛选随机序列的单链寡核苷酸文库,通过单引物复性及聚合酶延伸,形成双链DNA文库,再将文库与转录因子蛋白混合,形成“DNA/蛋白质”复合物,分离出复合物中的DNA,并以其为模板进行PCR扩增,进行下一轮筛选过程;通过重复的筛选与扩增,最后得到高亲和力和高特异性的DNA分子。该过程中,最关键一步就是如何将“DNA/蛋白质”复合物从游离DNA中分离出来,经典的筛选技术有凝胶阻滞法^[90]、亲和层析法^[91]等。

SELEX技术曾被广泛应用于从DNA随机文库中筛选转录因子蛋白结合DNA序列,如Sox2^[92]、Oct4^[93]、Nanog^[94]和c-Myc^[95]。但这些研究中,最终

富集的DNA都采用传统的克隆测序进行分析, 因此分析的通量非常低, 往往克隆数在几十个, 最高也就到几百个, 仅仅反映的是那些亲和性最高的DNA序列, 而排除了中低亲和性的DNA序列, 所以不能全面展示一个转录因子蛋白能够结合的各种DNA序

列及其它们的相对亲和性。可以说这是一种分辨率极低的DNA结合谱。但基于克隆测序的SELEX技术在快速发现大量转录因子的保守模体或基序 (Consensus) 的早期研究中发挥了重要作用。

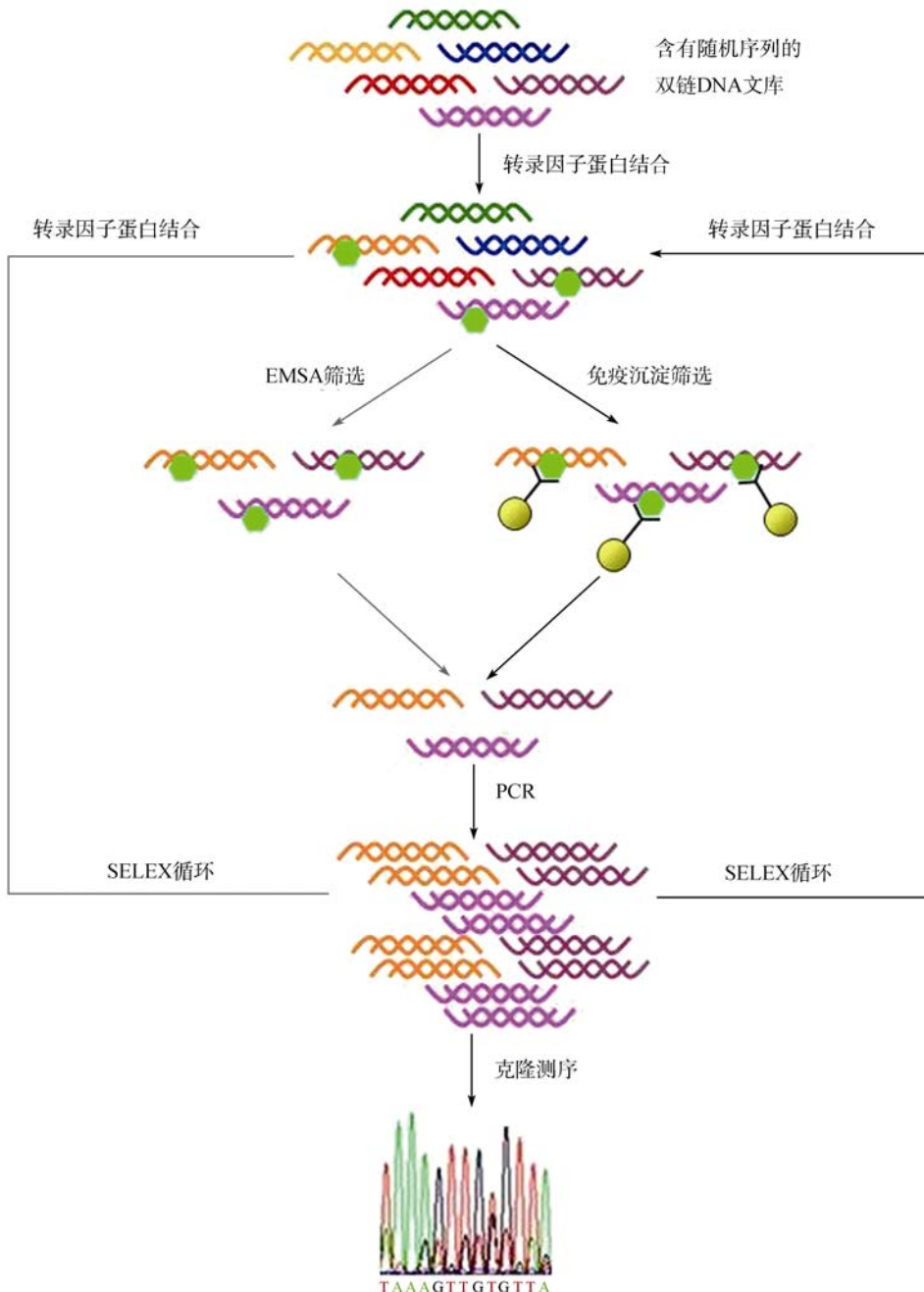


图 3 SELEX 技术流程

合成两端具有PCR引物复性序列、中间具有不同个数随机序列核苷酸的寡核苷酸, 将这种寡核苷酸混合物与DNA结合蛋白(转录因子)结合; 采用适当的分离技术, 如电泳迁移率变动分析(Electrophoretic mobility shift assay, EMSA)、膜结合过滤、磁珠分离等, 分离结合型DNA; 结合型DNA进行PCR扩增; 扩增产物用于下一轮结合筛选; 经过 4~6 轮筛选, 获得与转录因子具有高结合亲和性的DNA^[96]。

2.2.2 SELEX-SAGE

传统SELEX克隆测序的通量非常低, 为了提高SELEX富集DNA的测序通量, 2002 年瑞士生物信息研究所的Bucher等^[97]提出SELEX-基因表达系列分析(SELEX and serial analysis of gene expression, SELEX-SAGE)技术。该技术将SELEX程序与SAGE技术联合起来, 提高了SELEX富集DNA的测序通量。SELEX-SAGE的主要原理见图 4。虽然在目前具有高通量DNA测序技术的情况下, 该技术称不上真正的高通量, 但该技术在高通量DNA测序技术产生之前的数年中, 是研究转录因子DNA结合谱最有效的技术, 产生了大量结合信息, 这些数据被存贮在专门的数据库中, 至今仍可获取使用^[98]。

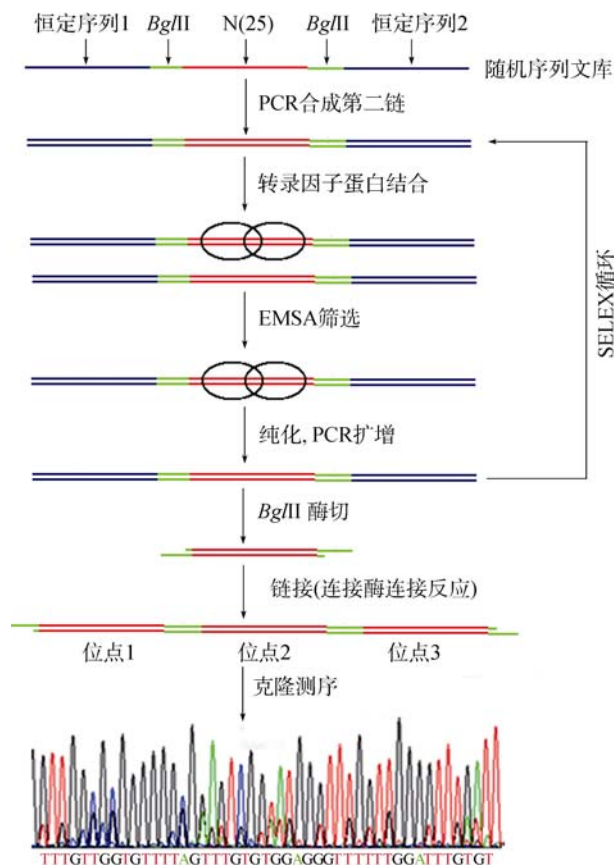


图 4 SELEX-SAGE 原理示意图

SELEX 输入 DNA 文库由大量的双链寡核苷酸构成, 其中每种寡核苷酸的两端为 PCR 引物复性位点的恒定序列, 中间为 25 bp 随机序列。在随机序列和 PCR 引物复性序列间设置限制性内切酶切割位点。完成 SELEX 筛选后, 用限制性内切酶释放中间的随机序列, 再让它们互相链接, 形成大片段, 用大片段进行克隆测序, 提高对 SELEX 富集 DNA 的测序通量。

2.3 基于高通量 DNA 测序技术的转录因子体外结合谱研究技术

2.3.1 高通量 DNA 测序技术

鉴于高通量DNA测序技术对转录因子结合谱研究领域产生的重大影响, 在阐述高通量DNA测序技术在转录因子体外结合谱研究中的应用之前, 有必要简单介绍一下高通量DNA测序技术。最早的DNA测序技术为Sanger等发明的双脱氧核苷酸末端终止法和Gilbert发明的化学降解法^[99, 100], 那时的DNA测序主要用于解码DNA片段的碱基组成信息, 如各种物种基因组序列的测定, 但随着生命科学的发展, DNA测序的目的不仅仅再局限于这一目的, 而更重要的是通过解码序列而识别一个DNA样品中各种DNA分子的身份以及它们的丰度。高通量DNA测序技术的产生, 不但更好地实现第一个目的, 更重要的是真正实现了第二个目的。因此, 高通量DNA测序技术快速而广泛地取代DNA微阵列芯片技术, 成为基因表达和分子相互作用研究的主要工具, 如RNA-测序(RNA-sequencing, RNA-Seq)及ChIP-Seq。

目前, 高通量DNA测序技术平台主要包括Roche/454FLX、Illumina/Solexa Genome Analyzer和Applied Biosystems SOLiD system。这3个技术平台各有优点, 454FLX的测序片段比较长, 高质量的读序(read)能达到400 bp, 因此更多地用于基因组的从头测序和转录组解析; Solexa测序片段较短(70 bp), 但序列解码通量更高, 因此Solexa从一产生就在DNA测序的第二个目的上展示出强大的功能, 目前RNA-Seq和ChIP-Seq研究基本上都采用该技术^[118, 101~104]。

随着高通量DNA测序技术的飞速发展, 该领域技术已经对生命科学的发展产生了巨大的影响, 特别是其对DNA分子的序列丰度的双重解析能力, 已经远远超出了DNA微阵列芯片技术, 因而对DNA微阵列芯片技术的应用领域产生了前所未有的冲击。例如, 检测基因表达的RNA-Seq技术基本上取代了基因芯片技术。同样, 对DNA微阵列芯片技术在转录因子DNA结合谱研究领域产生了深刻的影响, 如ChIP-Seq技术基本上取代了ChIP-chip技术。在过去的3年中, 高通量DNA测序技术对转录因子体外结合谱的研究亦产生深刻影响, 体现了其主导的该领域发生的重大变革和未来走向。

高通量测序技术的发展远未停止,例如 Illumina 公司在 2010 年又推出了 HiSeq2000 测序新技术,虽然其测序原理与 Illumina/Solexa Genome Analyzer 相似,但它融合了最新的光学系统和制造工艺,该光学系统采用 2 个激光源对微流室进行扫描,并使用 4 台照相机对 4 种含有末端阻断基团和不同荧光信号的碱基分别进行记录,降低了不同碱基之间的信号干扰,提高了测序的高精确性和高顺序性。同时,它还使用了新颖的双表面成像技术,增加了微流室的有效面积,从而提高测序产量和降低成本。HiSeq2000 实现了 200 GB 的最高测序产量和最快的数据产生速率。由于 HiSeq2000 测序系统目前尚未有用于转录因子结合谱研究的报道,在此我们只做简单介绍。

2.3.2 Bind-n-Seq 技术

2009 年,美国加州大学基因组中心的 Zykovich 等^[105]首次使用 SELEX 技术联合下游多重平行测序技术实现了对 DNA 结合蛋白 Zif268 和 Aart 结合位点的同时鉴定,此技术被称之为 Bind-n-Seq。图 5 为该技术流程图。该研究首次表明将经典的 SELEX 技术与高通量 DNA 测序技术相结合,可用于研究转录因子高分辨率精细 DNA 结合模体。

为了充分利用高通量 DNA 测序技术巨大的测序通量,该技术对不同转录因子蛋白筛选的 DNA 样品使用了碱基条形码技术。在测序之前将这些样品混合,作为一个测序样品进行测序,这样可极大地降低实验成本。传统 SELEX-克隆测序实验仅能获得几十条到数百条 DNA 序列信息,SELEX-SGAE 技术可获得数千条 DNA 序列信息,而 Bind-n-Seq 技术首次将 SELEX 筛选 DNA 的序列解码能力提高到数百万条,在转录因子体外结合研究中真正实现了结合谱的绘制。例如在该研究中,作者同步分析了 28 个 DNA 样品,每个样品可获得约 10 万个读序(reads)。而且,通过使用条形码技术对不同筛选轮数的 DNA 可同时测序,实现了对高、中、低亲和性序列的全波段扫描,使那些具有中、低亲和性的序列真正进入 DNA 与转录因子相互作用分析的行列,而这些中、低亲和性序列对于建立更精确的结合位点模体是非常必要的^[97]。Zykovich 等^[105]的实验充分证明 Bind-n-Seq 筛选中不同序列被测序的频率反映了蛋白质对不同序列

结合的相对亲和性,因此,通过数字化的中低测序频率,可以发现具有中低亲和性的序列。

与 dsDNA 微阵列芯片技术相比,该技术突破了 dsDNA 微阵列芯片技术难以绘制结合位点大于 10 bp 的转录因子的 DNA 结合谱的限制。对于目前的 DNA 微阵列技术而言,10-mer 的完整目录(1×10^6 个探针点)是其极限^[105]。而很多转录因子的 DNA 结合位点大于 10 bp^[17]。Bind-n-Seq 却不受 10 bp 的限制,例如在研究转录因子 Zif268 和 Aart 的结合位点时,就使用了 21 bp 结合区。也就是说该技术可以用于研究任意长度 DNA 结合位点的转录因子。此外,对于一个结合位点长度未知的新的 DNA 结合蛋白,dsDNA 微阵列芯片可能难以应付,但 Bind-n-Seq 通过使用任意长度的随机寡核苷酸文库,可绘制这种结合序列未知的转录因子蛋白的 DNA 结合谱。

2.3.3 MMP-SELEX 技术

2010 年,芬兰赫尔辛基大学 Taipale 研究组提出多重大规模并行 SELEX (Multiplexed massively parallel SELEX, MMP-SELEX) 技术,用于测定人转录因子结合谱^[106]。他们采用此技术绘制了属于 14 种不同种类的 19 个转录因子的 DNA 结合谱^[106],并用体内技术 ChIP-Seq 实验确证了转录因子 NFATC1 和 RFX3 的特异性。该技术的实验原理及流程见图 6,该技术实质上是基于微孔板偶联 DNA 结合蛋白,进行高效自动化蛋白结合寡核苷酸筛选的 SELEX 新方法,筛选产物使用高通量 DNA 测序技术,如 Illumina Solexa,进行解码测定。该研究建立了一套完整的研究真核转录因子蛋白 DNA 结合特异性的 SELEX 筛选方案和多重测序技术方案,为该类研究提供了经典程序,对该类研究必将产生巨大的推动作用。

2.3.4 EMSA-Seq 技术

2011 年,英国剑桥大学的 Udalova、Ragoussis 和美国哈佛大学的 Bulky 合作研究,提出基于凝胶迁移实验进行 SELEX 筛选的 EMSA-测序 (EMSA-Sequencing, EMSA-Seq) 技术,并将其用于转录因子 NF- κ B DNA 结合谱的大规模表征^[107]。该研究还将 EMSA-Seq 技术与 dsDNA 微阵列芯片技术

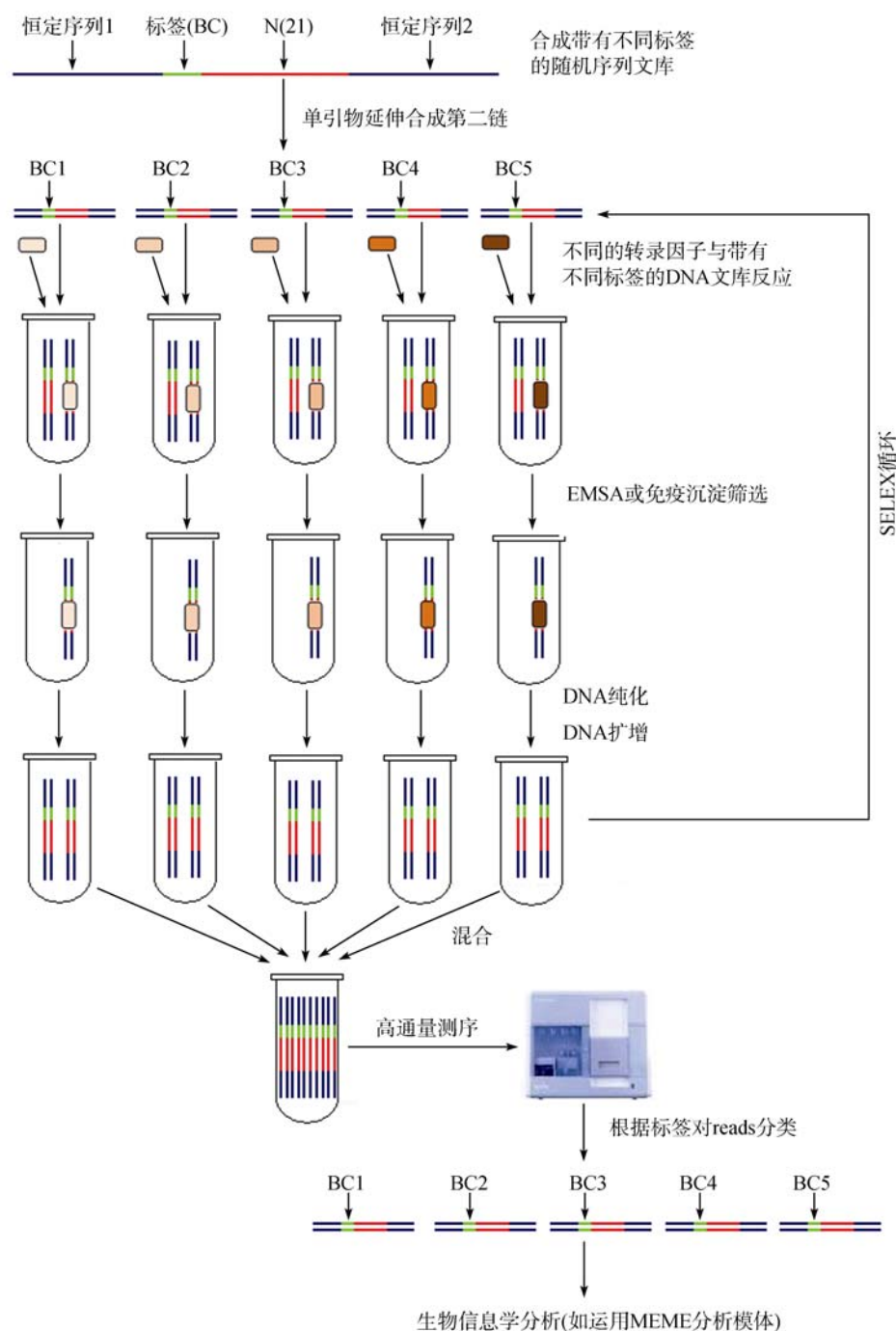


图 5 Bind-n-Seq 技术流程图

Bind-n-Seq 的底物是含有恒定序列区(Primer A 和 Primer B)、3 核苷酸条形码区(Bar-coded, BC)以及 21 bp 随机序列区(Random sequence)的寡核苷酸。打上不同条形码的寡核苷酸与不同的转录因子蛋白结合反应, 分离并收集结合型 DNA, 再用短片段测序技术(如 Solexa)测序。依靠生物信息程序根据条形码对测序片段分类和处理, 产生对应于每种转录因子蛋白的 DNA 结合谱, 并制作结合位点模体。

结合起来, 建立了相互印证的双平台方法(图 7)。

NF- κ B 是炎症、肿瘤相关重要转录因子^[108,109]。NF- κ B/Rel 家族由 RELA(p65)、p50、p52、RELB、C-Rel 等 5 个成员构成, 这些成员间可形成同二聚体

和异二聚体, 通过与基因组中大量不同的 DNA 序列相互作用而控制众多靶基因, 因此, NF- κ B 的 DNA 结合谱研究一直深受重视。著名的转录因子数据库 TRANSFAC 根据实验报道的 NF- κ B 靶基因中的 DNA

位点得出的其DNA结合模体为GGGRNNTYCC (R : G、A; Y : C、T; N : G、A、T、C)^[110]。2002~2004年,英国剑桥大学的Udalova实验室通过EMSA、表面等离子体共振及dsDNA微阵列芯片研究,将其模体从传统的GGGRNNTYCC扩展为GGRNNYYCC^[111~113]。2011年,她们运用EMSA-Seq和dsDNA微阵列芯片研究,又将其模体扩展成RGGRNNHHYY (H : A、T、C),作者将其称为非经典模体。其中运用EMSA-Seq研究了包含RELA的NF- κ B二聚体(RELARELA、RELAp50、RELAp52)同522857个

10-mer 序列的结合,发现用EMSA-Seq产生的结合数据与用dsDNA 微阵列芯片产生的结合数据非常吻合。该研究中还根据新模体RGGRNNHHYY制备了包含803条代表“11-mer”DNA序列探针的dsDNA微阵列芯片,用该种芯片研究了9个NF- κ B二聚体(RELARELA、p50p50、p52p52、RELAp50、RELAp52、RELBp50、RELBp52、C-Relp50、C-Relp52)的结合,证实了EMSA-Seq结果,并且发现NF- κ B家族成员具有不同的结合谱。

将EMSA-Seq体外结合谱与ChIP-Seq体内结合

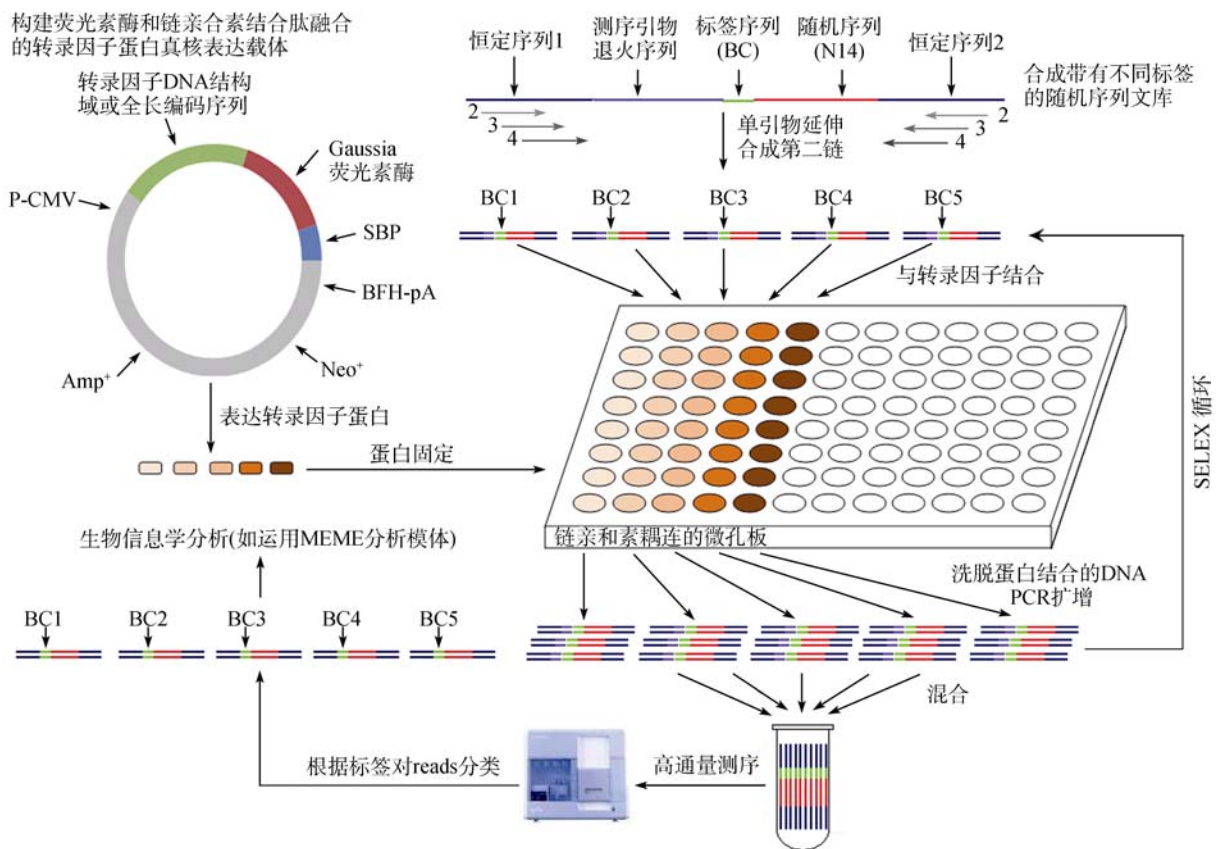


图6 高通量 SELEX 过程示意图

蛋白表达: 构建要研究的转录因子蛋白的真核表达载体,并表达成具有DNA结合域或全长序列转录因子蛋白、链亲和素结合肽(SBP)和荧光素酶的融合蛋白。其中DNA结合域或全长序列转录因子蛋白用于与DNA结合,链亲和素结合肽用于将转录因子蛋白锚定在链亲和素包被的微孔板上,用于筛选和分离结合型DNA;荧光素酶用于筛选阳性细胞、定量转录因子蛋白等。表达载体使用GATEWAY重组克隆系统。配体设计:即待筛选的寡核苷酸,使用条形码技术使多重样品可共同测序。每种DNA配体含有14bp随机序列区和一个5bp条形码(Barcode, BC)。为了增加特异性,每个编码与其他编码至少有2个碱基的差别。这些可变序列的两侧设计恒定序列,该序列包含Illumina Genome Analyzer测序引物位点和PCR扩增引物结合区(恒定序列),这些引物结合序列设计在5'末端,以便容纳部分巢式引物(用恒定序列下的一组箭头显示),用于每个SELEX循环中的PCR扩增。高通量SELEX筛选:含有所有可能14bp序列的双链DNA混合物与固定在96微孔板内的DNA结合蛋白孵育,使DNA与蛋白质结合;之后洗涤并洗脱,获得序列更加特异的DNA群,这些DNA通过PCR扩增进入下一轮SELEX筛选或高通量测序。在PCR扩增环节,不同的SELEX循环使用不同的嵌套引物(图中带数字“2, 3, 4”的一组箭头,数字表示SELEX循环数),以避免污染DNA的扩增。通过重复筛选过程并计算不同循环数后不同序列的丰度构建转录因子的特异性。每个循环中,将多重反应产物混合成单个测序泳道,使用条形码序列识别来自不同转录因子的DNA样品。

谱相比较,发现NF- κ B在基因组中不仅结合经典模体,也结合非经典模体,EMSA-Seq获得的体外结合谱可很好地解释ChIP-Seq数据,并且可以帮助预测NF- κ B靶基因调控序列中单核苷酸多态性的功能效果,识别基因调控序列中的疾病相关单核苷酸多态性^[46]。该研究提供了一个将转录因子体内结合谱与体外结合谱进行结合研究的成功范例,为今后的该类研究提供了一个很好的策略。

2.3.5 HiTS-FLIP 技术

2011年,麻省理工学院的Nutiu等^[114]研究人员利用新一代测序仪定量测定DNA-蛋白质结合亲和力,提出了研究转录因子DNA结合谱的新技术——HiTS-FLIP。该技术的主要原理及实验流程见图8所

示。在该研究中,用HiTS-FLIP技术研究了酵母(*Saccharomyces cerevisiae*)转录激活因子Gcn4与DNA的结合,该蛋白是酵母氨基酸饥饿应答的主要调控者。首先,制备了一个随机25聚体的DNA文库,将其放入Illumina Solexa测序仪中,完成测序。之后变性,剥离测序延伸链,利用新的引物和Klenow酶合成新链。再加入与单体(m)OrBurge结合的GCN4蛋白,并对蛋白结合进行荧光成像。最后,在测序图像上叠加蛋白结合图像,判定蛋白对各种序列的结合亲和性。

该技术能以前所未有的深度识别DNA结合蛋白与海量DNA分子的结合,并能定量测定结合亲和力,获得解离常数。例如,在25聚体的背景下,可获得每个7聚体的蛋白结合亲和力的10万次测定,

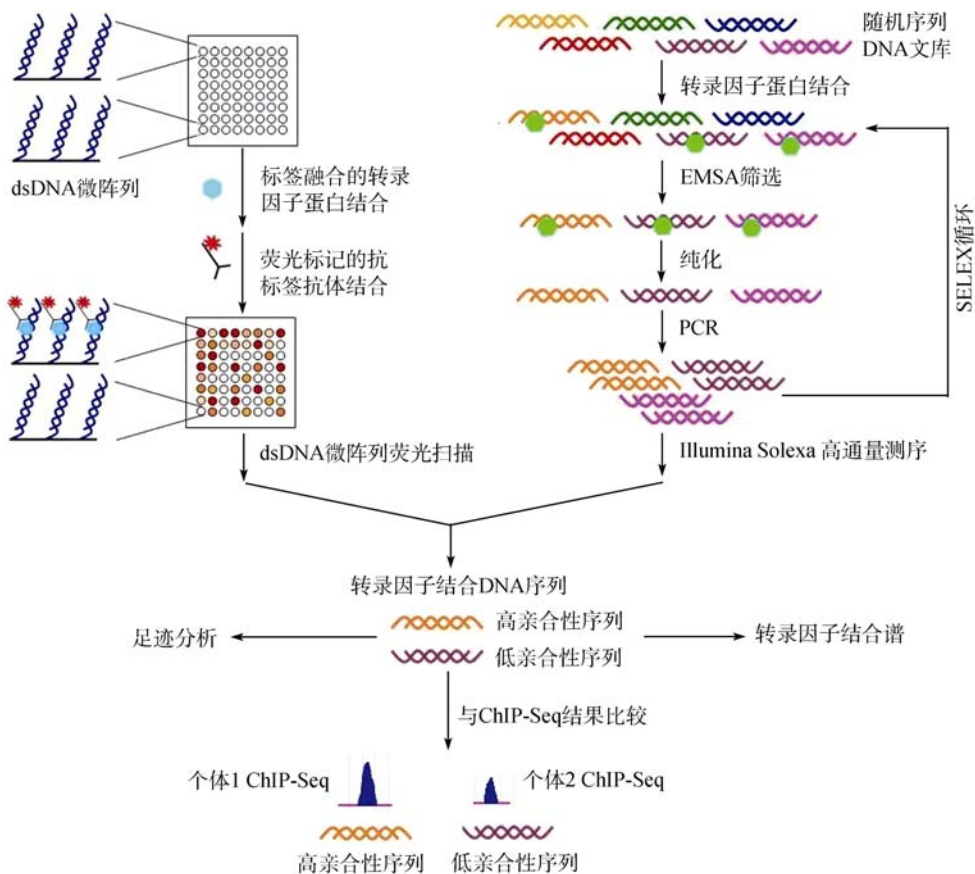


图7 EMSA-Seq及PBM双平台方法用于绘制NF- κ B家族二聚体DNA结合谱的实验流程

在PBM平台中,NF- κ B蛋白与固定大量dsDNA探针的微阵列结合,获取NF- κ B与大量DNA序列的结合亲和性数据。在EMSA-Seq平台中,NF- κ B蛋白与不同序列的DNA混合物结合,再通过双色荧光EMSA分离结合型DNA和非结合型DNA;将结合型DNA切胶回收,进行高通量DNA测序,获得NF- κ B与大量DNA序列的结合亲和性数据。根据双平台数据制作NF- κ B的DNA结合谱。用UV足迹实验验证EMSA-Seq结果。将EMSA-Seq绘制的体内结合谱与ChIP-Seq绘制的体内结合谱相比较,解释ChIP-Seq数据,发现疾病相关单核苷酸多态性。

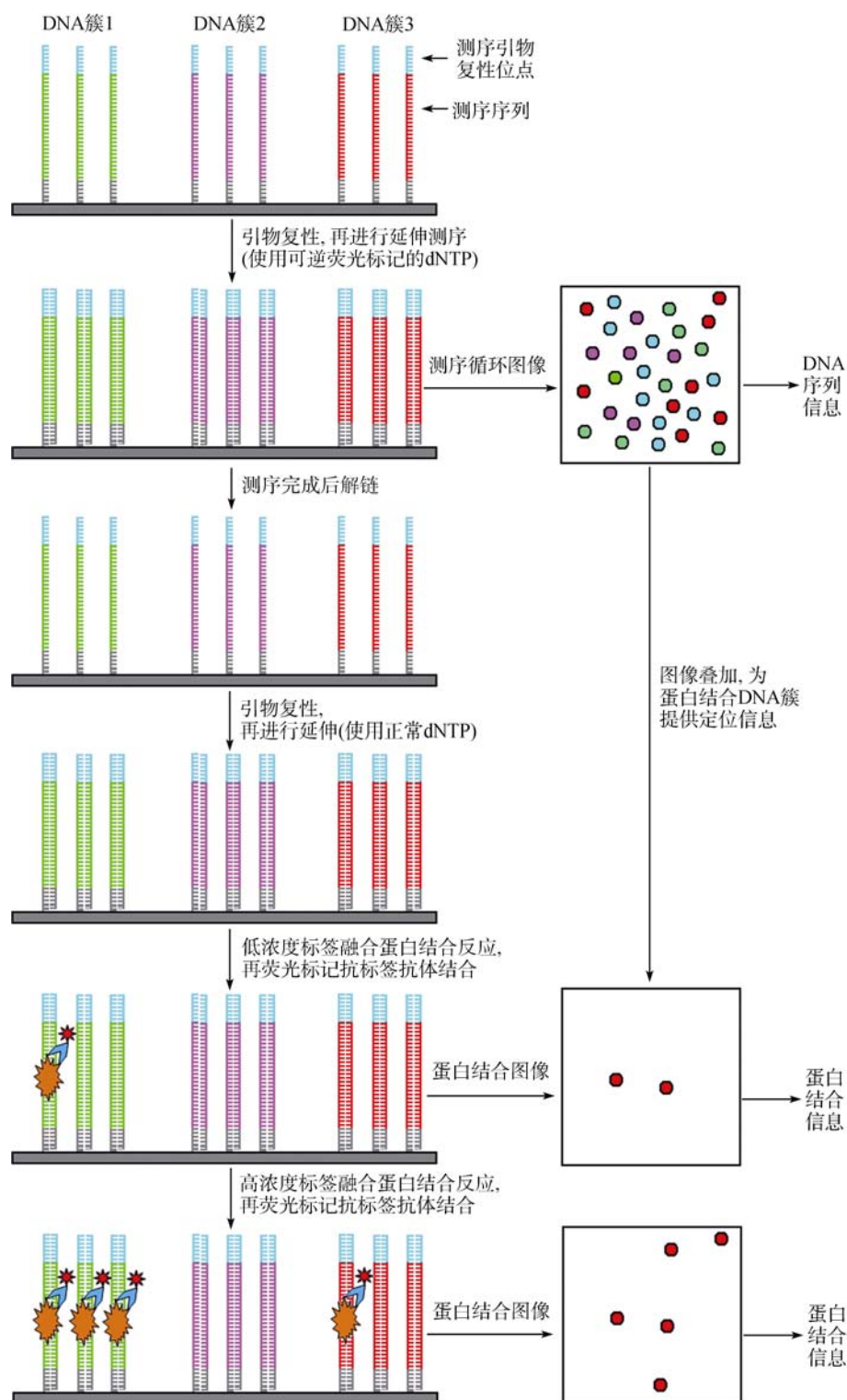


图 8 HiTS-FLIP 方法

HiTS-FLIP 示意图。通过合成测序测定微流室内锚定的单链 DNA 序列。第二链 DNA 剥离并用 *Klenow* 酶和正常 dNTP 重新建立 dsDNA 簇。荧光标记的不同浓度的转录因子蛋白(如 Gcn4p)导入微流室, 结合成像。将测序图像与转录因子蛋白结合图像叠加, 可确定蛋白质结合的 DNA 序列, 以及这些序列的结合常数。

以及每个 12 聚体的 500 次测定。在应用 HiTS-FLIP 研究 Gcn4 蛋白的 DNA 结合中, 能确定该蛋白对具有亚微摩尔亲和性的所有 12 聚体序列的解离常数。因此, 该技术使得高通量 DNA 测序技术能够直接用于绘制转录因子蛋白的 DNA 结合亲和性图谱。所实现的分析深度能够让研究人员了解模体内不同位置间复杂的相互依赖性。通过捕获弱的以及复杂的相互作用, 该数据有助于 GCN4 结合的预测, 并揭示出有着不同表达动力学的 GCN4 调控启动子。这些数据能够更好地区分体内 Gcn4p 结合位点和调控靶点, 并显示与 Gcn4p 有着不同的启动子亲和力的基因有着不同的功能表达动力学。

HiTS-FLIP 方法使得任何研究人员能将测序仪器转换成研究 DNA-蛋白质相互作用的工具, 相比于其他替代工具, 具有几个优势。首先, 它的通量可达数千万到数亿个结合测定, 远远大于使用其他任何技术对于结合的测定, 使得位置间复杂的相互依赖性的分析及较长的或更复杂的结合模体的分析成为可能。在基于 SELEX 筛选结合测序的方法(如 Bind-n-Seq、MMP-SELEX 及 EMSA-Seq)中, 结合亲和性是通过序列的计数所推导的, 而在 HiTS-FLIP 中, 亲和性是直接通过海量的独立的结合直接测定的。该技术测定结合的通量非常高, 可将不同蛋白标记不同荧光, 依靠测序仪测定多重荧光的能力可在微流室中的同一实验中同步测定多个蛋白的结合。此外, 还可以在 8 个测序通道中加入 8 种不同的蛋白, 或在一个测序循环后向某一微流室顺序加入不同的蛋白, 实现蛋白结合的多重化测定。

无疑, HiTS-FLIP 将转录因子体外结合谱、亲和谱的测定通量提到一个史无前例的水平, 该技术将是未来数年中转录因子体外结合谱研究的最前沿、最先进的技术, 将对转录因子体外结合谱的研究带来巨大的影响。但是, 目前该技术还受制于价格极其昂贵的 DNA 测序仪器, 普通实验室很难企及并用于日常化研究。随着 DNA 测序技术的快速发展, 测序仪器的普及和测序成本的下降, 这种技术必将成为研究转录因子体外 DNA 结合谱的主要技术。

3 结 语

综上所述, 可见 DNA 微阵列芯片技术和高通量 DNA 测序技术的产生和发展, 对转录因子 DNA 结

合谱的研究产生了重大影响。随着这些技术在转录因子研究领域的快速应用, 出现了 ChIP-Seq、dsDNA 微阵列芯片、基于 SELEX 结合高通量 DNA 测序(Bind-n-Seq、MMP-SELEX 及 EMSA-Seq), 以及利用高通量 DNA 测序仪直接测定转录因子结合谱(HiTS-FLIP)的许多新技术。这些技术既有各自的优点, 同时也存在自身的缺点, 但就技术的先进性而言, ChIP-Seq 和 HiTS-FLIP 将是研究转录因子体内及体外结合谱的最主要的技术。这两种技术都依赖于高通量 DNA 测序技术, 但 ChIP-Seq 已经在该领域中普及, 原因是 ChIP-Seq 实验的质量主要取决于 ChIP, 而不是 Seq, DNA 测序技术的发展对 ChIP-Seq 实验的主要意义在于实验成本的进一步降低; 而 HiTS-FLIP 则不同, 该实验完全依赖于高通量 DNA 测序技术, DNA 测序技术的发展对于 HiTS-FLIP 研究来说, 不仅仅是实验成本的降低, 更重要的是其普及, 使其成为一种常规实验技术。

参考文献(References):

- [1] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 2007, 131(5): 861–872. [DOI](#)
- [2] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 2006, 126(4): 663–676. [DOI](#)
- [3] Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA. Induced pluripotent stem cell lines derived from human somatic cells. *Science*, 2007, 318(5858): 1917–1920. [DOI](#)
- [4] Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. *Nature*, 2000, 408(6810): 307–310. [DOI](#)
- [5] Vousden KH, Lane DP. p53 in health and disease. *Nat Rev Mol Cell Bio*, 2007, 8(4): 275–283. [DOI](#)
- [6] Dolcet X, Llobet D, Pallares J, Matias-Guiu X. NF- κ B in development and progression of human cancer. *Virchows Arch*, 2005, 446(5): 475–482. [DOI](#)
- [7] Messina DN, Glasscock J, Gish W, Lovett M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res*, 2004, 14(10B): 2041–2047. [DOI](#)
- [8] Orlando V, Strutt H, Paro R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods*,

- 1997, 11(2): 205–214. [DOI](#)
- [9] 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. [DOI](#)
- [10] Das PM, Ramachandran K, van Wert J, Singal R. Chromatin immunoprecipitation assay. *Biotechniques*, 2004, 37(6): 961–969. [DOI](#)
- [11] Peng GH, Chen S. Chromatin immunoprecipitation identifies photoreceptor transcription factor targets in mouse models of retinal degeneration: new findings and challenges. *Vis Neurosci*, 2005, 22(5): 575–586. [DOI](#)
- [12] Goldfarb T, Alani E. Chromatin immunoprecipitation to investigate protein-DNA interactions during genetic recombination. *Methods Mol Biol*, 2004, 262(IV): 223–237. [DOI](#)
- [13] Buchenau P, Hodgson J, Strutt H, Arndt-Jovin DJ. The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. *J Cell Biol*, 1998, 141(2): 469–481. [DOI](#)
- [14] Weinmann AS, Bartley SM, Zhang T, Zhang MQ, Farnham PJ. Use of chromatin immunoprecipitation to clone novel E2F target promoters. *Mol Cell Biol*, 2001, 21(20): 6820–6832. [DOI](#)
- [15] Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. Genome-wide location and function of DNA binding proteins. *Science*, 2000, 290(5500): 2306–2309. [DOI](#)
- [16] Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, Liu J, Zhao XD, Chew JL, Lee YL, Kuznetsov VA, Sung WK, Miller LD, Lim B, Liu ET, Yu Q, Ng HH, Ruan Y. A global map of p53 transcription-factor binding sites in the human genome. *Cell*, 2006, 124(1): 207–219. [DOI](#)
- [17] Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, 2007, 316(5830): 1497–1502. [DOI](#)
- [18] Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao YJ, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*, 2007, 4(8): 651–657. [DOI](#)
- [19] Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. A high-resolution map of active promoters in the human genome. *Nature*, 2005, 436(7052): 876–880. [DOI](#)
- [20] Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*, 2006, 125(2): 301–313. [DOI](#)
- [21] Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet*, 2006, 38(4): 431–440. [DOI](#)
- [22] Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell*, 2008, 132(6): 1049–1061. [DOI](#)
- [23] Sharov AA, Masui S, Sharova LV, Piao Y, Aiba K, Matoba R, Xin L, Niwa H, Ko MS. Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics*, 2008, 9(1): 269. [DOI](#)
- [24] Yang JC, Chai L, Fowles TC, Alipio Z, Xu D, Fink LM, Ward DC, Ma YP. Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. *Proc Natl Acad Sci USA*, 2008, 105(50): 19756–19761. [DOI](#)
- [25] Kidder BL, Yang J, Palmer S. Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. *Plos One*, 2008, 3(12): e3932. [DOI](#)
- [26] Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K. Role of the murine reprogramming factors in the induction of pluripotency. *Cell*, 2009, 136(2): 364–377. [DOI](#)
- [27] Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 2008, 133(6): 1106–1117. [DOI](#)
- [28] Ouyang Z, Zhou Q, Wong WH. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc Natl Acad Sci USA*, 2009, 106(51): 21521–21526. [DOI](#)
- [29] Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS,

- Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 2005, 122(6): 947–956. [DOI](#)
- [30] Cole MF, Young RA. Mapping key features of transcriptional regulatory circuitry in embryonic stem cells. *Cold Spring Harb Symp Quant Biol*, 2008, 73: 183–193. [DOI](#)
- [31] Kashyap V, Rezende NC, Scotland KB, Shaffer SM, Persson JL, Gudas LJ, Mongan NP. Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. *Stem Cells Dev*, 2009, 18(7): 1093–1108. [DOI](#)
- [32] Mathur D, Danford TW, Boyer LA, Young RA, Gifford DK, Jaenisch R. Analysis of the mouse embryonic stem cell regulatory networks obtained by ChIP-chip and ChIP-PET. *Genome Biol*, 2008, 9(8): R126. [DOI](#)
- [33] Pan G, Thomson JA. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res*, 2007, 17(1): 42–49. [DOI](#)
- [34] Chen X, Vega VB, Ng HH. Transcriptional regulatory networks in embryonic stem cells. *Cold Spring Harb Symp Quant Biol*, 2008, 73: 203–209. [DOI](#)
- [35] Orkin SH, Wang J, Kim J, Chu J, Rao S, Theunissen TW, Shen X, Levasseur DN. The transcriptional network controlling pluripotency in ES cells. *Cold Spring Harb Symp Quant Biol*, 2008, 73: 195–202. [DOI](#)
- [36] Miled C, Pontoglio M, Garbay S, Yaniv M, Weitzman JB. A genomic map of p53 binding sites identifies novel p53 targets involved in an apoptotic network. *Cancer Res*, 2005, 65(12): 5096–5104. [DOI](#)
- [37] Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammanna H, Helt G, Struhl K, Gingeras TR. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, 2004, 116(4): 499–509. [DOI](#)
- [38] Mirza A, Wu Q, Wang L, McClanahan T, Bishop WR, Gheyas F, Ding W, Hutchins B, Hockenberry T, Kirschmeier P, Greene JR, Liu S. Global transcriptional program of p53 target genes during the process of apoptosis and cell cycle progression. *Oncogene*, 2003, 22(23): 3645–3654. [DOI](#)
- [39] Kaneshiro K, Tsutsumi S, Tsuji S, Shirahige K, Aburatani H. An integrated map of p53-binding sites and histone modification in the human ENCODE regions. *Genomics*, 2007, 89(2): 178–188. [DOI](#)
- [40] Smeenk L, van Heeringen SJ, Koeppel M, van Driel MA, Bartels SJJ, Akkers RC, Denissov S, Stunnenberg HG, Lohrum M. Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res*, 2008, 36(11): 3639–3654. [DOI](#)
- [41] Vigano MA, Lamartine J, Testoni B, Merico D, Alotto D, Castagnoli C, Robert A, Candi E, Melino G, Gidrol X, Mantovani R. New p63 targets in keratinocytes identified by a genome-wide approach. *EMBO J*, 2006, 25(21): 5105–5116. [DOI](#)
- [42] Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol Cell*, 2006, 24(4): 593–602. [DOI](#)
- [43] Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M. Distribution of NF- κ B-binding sites across human chromosome 22. *Proc Natl Acad Sci USA*, 2003, 100(21): 12247–12252. [DOI](#)
- [44] Lim CA, Yao F, Wong JJY, George J, Xu H, Chiu KP, Sung WK, Lipovich L, Vega VB, Chen J, Shahab A, Zhao XD, Hibberd M, Wei CL, Lim B, Ng HH, Ruan YJ, Chin KC. Genome-wide mapping of RELA(p65) binding identifies E2F1 as a transcriptional activator recruited by NF- κ B upon TLR4 activation. *Mol Cell*, 2007, 27(4): 622–635. [DOI](#)
- [45] Schreiber J, Jenner RG, Murray HL, Gerber GK, Gifford DK, Young RA. Coordinated binding of NF- κ B family members in the response of human cells to lipopolysaccharide. *Proc Natl Acad Sci USA*, 2006, 103(15): 5899–5904. [DOI](#)
- [46] Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, Hong MY, Karczewski KJ, Huber W, Weissman SM, Gerstein MB, Korbel JO, Snyder M. Variation in transcription factor binding among humans. *Science*, 2010, 328(5975): 232–235. [DOI](#)
- [47] Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH. Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell*, 2004, 119(7): 1041–1054. [DOI](#)
- [48] Euskirchen G, Royce TE, Bertone P, Martone R, Rinn JL, Nelson FK, Sayward F, Luscombe NM, Miller P, Gerstein M, Weissman S, Snyder M. CREB binds to multiple loci

- on human chromosome 22. *Mol Cell Biol*, 2004, 24(9): 3804–3814. [DOI](#)
- [49] Zhang X, Odom DT, Koo SH, Conkright MD, Canetti G, Best J, Chen H, Jenner R, Herbolsheimer E, Jacobsen E, Kadam S, Ecker JR, Emerson B, Hogenesch JB, Unterman T, Young RA, Montminy M. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc Natl Acad Sci USA*, 2005, 102(12): 4459–4464. [DOI](#)
- [50] Bieda M, Xu X, Singer MA, Green R, Farnham PJ. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res*, 2006, 16(5): 595–605. [DOI](#)
- [51] Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 2005, 122(1): 33–43. [DOI](#)
- [52] Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoutte J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS, Brown M. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet*, 2006, 38(11): 1289–1297. [DOI](#)
- [53] Laganier J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguere V. Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc Natl Acad Sci USA*, 2005, 102(33): 11651–11656. [DOI](#)
- [54] Liu Y, Gao H, Marstrand TT, Strom A, Valen E, Sandelin A, Gustafsson JÅ, Dahlman-Wright K. The genome landscape of ERalpha- and ERbeta-binding DNA regions. *Proc Natl Acad Sci USA*, 2008, 105(7): 2604–2609. [DOI](#)
- [55] Kwon YS, Garcia-Bassets I, Hutt KR, Cheng CS, Jin M, Liu D, Benner C, Wang D, Ye Z, Bibikova M, Fan JB, Duan L, Glass CK, Rosenfeld MG, Fu XD. Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. *Proc Natl Acad Sci USA*, 2007, 104(12): 4852–4857. [DOI](#)
- [56] Lin CY, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, Yeo A, George J, Kuznetsov VA, Lee YK, Charn TH, Palanisamy N, Miller LD, Cheung E, Katzenellenbogen BS, Ruan Y, Bourque G, Wei CL, Liu ET. Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet*, 2007, 3(6): e87. [DOI](#)
- [57] Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJW. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E-coli* chromosome. *Proc Natl Acad Sci USA*, 2005, 102(49): 17693–17698. [DOI](#)
- [58] Bhinge AA, Kim J, Euskirchen GM, Snyder M, Iyer VR. Mapping the chromosomal targets of STAT1 by Sequence Tag Analysis of Genomic Enrichment (STAGE). *Genome Res*, 2007, 17(6): 910–916. [DOI](#)
- [59] Hartman SE, Bertone P, Nath AK, Royce TE, Gerstein M, Weissman S, Snyder M. Global changes in STAT target selection and transcription regulation upon interferon treatments. *Gene Dev*, 2005, 19(24): 2953–2968. [DOI](#)
- [60] Durant L, Watford WT, Ramos HL, Laurence A, Vahedi G, Wei L, Takahashi H, Sun HW, Kanno Y, Powrie F, O'Shea JJ. Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. *Immunity*, 2010, 32(5): 605–615. [DOI](#)
- [61] Good SR, Thieu VT, Mathur AN, Yu Q, Stritesky GL, Yeh N, O'Malley JT, Perumal NB, Kaplan MH. Temporal induction pattern of STAT4 target genes defines potential for Th1 lineage-specific programming. *J Immunol*, 2009, 183(6): 3839–3847. [DOI](#)
- [62] Nelson EA, Walker SR, Alvarez JV, Frank DA. Isolation of unique STAT5 targets by chromatin immunoprecipitation-based gene identification. *J Biol Chem*, 2004, 279(52): 54724–54730. [DOI](#)
- [63] Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, Fraenkel E, Bell GI, Young RA. Control of pancreas and liver gene expression by HNF transcription factors. *Science*, 2004, 303(5662): 1378–1381. [DOI](#)
- [64] Rada-Iglesias A, Wallerman O, Koch C, Ameur A, Enroth S, Clelland G, Wester K, Wilcox S, Dovey OM, Ellis PD, Wraight VL, James K, Andrews R, Langford C, Dhami P, Carter N, Vetrie D, Ponten F, Komorowski J, Dunham I, Wadelius C. Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. *Hum Mol Genet*, 2005, 14(22): 3435–3447. [DOI](#)
- [65] Koinuma D, Tsutsumi S, Kamimura N, Taniguchi H, Miyazawa K, Sunamura M, Imamura T, Miyazono K, Aburatani H. Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. *Mol Cell Biol*, 2009, 29(1): 172–186. [DOI](#)
- [66] Lefterova MI, Zhang Y, Steger DJ, Schupp M, Schug J, Cristancho A, Feng D, Zhuo D, Stoeckert CJ, Jr, Liu XS,

- Lazar MA. PPARgamma and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev*, 2008, 22(21): 2941–2952. [DOI](#)
- [67] Nielsen R, Pedersen TA, Hagenbeek D, Moulos P, Siersbaek R, Megens E, Denissov S, Borgesen M, Francoijs KJ, Mandrup S, Stunnenberg HG. Genome-wide profiling of PPAR gamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Gene Dev*, 2008, 22(21): 2953–2967. [DOI](#)
- [68] Hamza MS, Pott S, Vega VB, Thomsen JS, Kandhadayar GS, Ng PWP, Chiu KP, Pettersson S, Wei CL, Ruan YJ, Liu ET. *De-Novo* identification of PPAR gamma/RXR binding sites and direct targets during adipogenesis. *Plos One*, 2009, 4(3): E4907. [DOI](#)
- [69] Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 2009, 10(10): 669–680. [DOI](#)
- [70] Hoffman BG, Jones SJ. Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing. *J Endocrinol*, 2009, 201(1): 1–13. [DOI](#)
- [71] Collas P, Dahl JA. Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. *Front Biosci*, 2008, 13(17): 929–943. [DOI](#)
- [72] Chambers I, Tomlinson SR. The transcriptional foundation of pluripotency. *Development*, 2009, 136(14): 2311–2322. [DOI](#)
- [73] Bulyk ML, Gentalen E, Lockhart DJ, Church GM. Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nat Biotechnol*, 1999, 17(6): 573–577. [DOI](#)
- [74] Bulyk ML, Huang X, Choo Y, Church GM. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci USA*, 2001, 98(13): 7158–7163. [DOI](#)
- [75] Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, III, Bulyk ML. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol*, 2006, 24(11): 1429–1435. [DOI](#)
- [76] Alleyne TM, Peña-Castillo L, Badis G, Talukder S, Berger MF, Gehrke AR, Philippakis AA, Bulyk ML, Morris QD, Hughes TR. Predicting the binding preference of transcription factors to individual DNA k-mers. *Bioinformatics*, 2009, 25(8): 1012–1018. [DOI](#)
- [77] Zhu C, Byers KJ, McCord RP, Shi Z, Berger MF, Newburger DE, Saulrieta K, Smith Z, Shah MV, Radhakrishnan M, Philippakis AA, Hu Y, De Masi F, Pacek M, Rolfs A, Murthy T, Labaer J, Bulyk ML. High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res*, 2009, 19(4): 556–566. [DOI](#)
- [78] Puckett JW, Muzikar KA, Tietjen J, Warren CL, Ansari AZ, Dervan PB. Quantitative microarray profiling of DNA-binding molecules. *J Am Chem Soc*, 2007, 129(40): 12310–12319. [DOI](#)
- [79] Carlson CD, Warren CL, Hauschild KE, Ozers MS, Qadir N, Bhimsaria D, Lee Y, Cerrina F, Ansari AZ. Specificity landscapes of DNA binding molecules elucidate biological function. *Proc Natl Acad Sci USA*, 2010, 107(10): 4544–4549. [DOI](#)
- [80] Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen XY, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML. Diversity and complexity in DNA recognition by transcription factors. *Science*, 2009, 324(5935): 1720–1723. [DOI](#)
- [81] Berger MF, Bulyk ML. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc*, 2009, 4(3): 393–411. [DOI](#)
- [82] Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulyk ML. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat Genet*, 2004, 36(12): 1331–1339. [DOI](#)
- [83] Newburger DE, Bulyk ML. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res*, 2009, 37(Database issue): D77–82. [DOI](#)
- [84] Robasky K, Bulyk ML. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res*, 2011, 39(Database issue): D124–D128. [DOI](#)
- [85] Wang J, Bai Y, Li T, Lu Z. DNA microarrays with unimolecular hairpin double-stranded DNA probes: fabrication and exploration of sequence-specific DNA/protein interactions. *J Biochem Biophys Methods*, 2003, 55(3): 215–232. [DOI](#)
- [86] Wang JK, Li TX, Bai YF, Zhu Y, Lu ZH. Fabrication of unimolecular double-stranded DNA microarrays on solid surfaces for probing DNA-protein/drug interactions. *Molecules*, 2003, 8(1): 153–168. [DOI](#)
- [87] Wang JK, Li TX, Lu ZH. A method for fabricating uni-dsDNA microarray chip for analyzing DNA-binding proteins. *J Biochem Biophys Meth*, 2005, 63(2): 100–110.

- [DOI](#)
- [88] Wang JK, Li TX, Bai YF, Lu ZH. Evaluating the binding affinities of NF-kappaB p50 homodimer to the wild-type and single-nucleotide mutant Ig-kappaB sites by the unimolecular dsDNA microarray. *Anal Biochem*, 2003, 316(2): 192–201. [DOI](#)
- [89] Djordjevic M. SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. *Biomol Eng*, 2007, 24(2): 179–189. [DOI](#)
- [90] Tsai RY, Reed RR. Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz. *Mol Cell Biol*, 1998, 18(11): 6447–6456. [DOI](#)
- [91] Liu J, Stormo GD. Combining SELEX with quantitative assays to rapidly obtain accurate models of protein-DNA interactions. *Nucleic Acids Res*, 2005, 33(17): e141. [DOI](#)
- [92] Maruyama M, Ichisaka T, Nakagawa M, Yamanaka S. Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *J Biol Chem*, 2005, 280(26): 24371–24379. [DOI](#)
- [93] Tantin D, Gemberling M, Callister C, Fairbrother W. High-throughput biochemical analysis of in vivo location data reveals novel distinct classes of POU5F1(Oct4)/DNA complexes. *Genome Res*, 2008, 18(4): 631–639. [DOI](#)
- [94] Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 2003, 113(5): 631–642. [DOI](#)
- [95] Papoulas O, Williams NG, Kingston RE. DNA binding activities of c-Myc purified from eukaryotic cells. *J Biol Chem*, 1992, 267(15): 10470–10480. [DOI](#)
- [96] Stoltenburg R, Reinemann C, Strehlitz B. SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng*, 2007, 24(4): 381–403. [DOI](#)
- [97] Roulet E, Busso S, Camargo AA, Simpson AJ, Mermod N, Bucher P. High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. *Nat Biotechnol*, 2002, 20(8): 831–835. [DOI](#)
- [98] Jagannathan V, Roulet E, Delorenzi M, Bucher P. HTPSELEX--a database of high-throughput SELEX libraries for transcription factor binding sites. *Nucleic Acids Res*, 2006, 34(Database issue): D90–D94. [DOI](#)
- [99] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*, 1977, 74(12): 5463–5467. [DOI](#)
- [100] Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA*, 1977, 74(2): 560–564. [DOI](#)
- [101] Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res*, 2008, 36(16): 5221–5231. [DOI](#)
- [102] Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods*, 2008, 5(9): 829–834. [DOI](#)
- [103] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 2007, 448(7153): 553–560. [DOI](#)
- [104] Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*, 2009, 138(5): 1019–1031. [DOI](#)
- [105] Zykovich A, Korf I, Segal DJ. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res*, 2009, 37(22): e151. [DOI](#)
- [106] Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpaa MJ, Bonke M, Palin K, Talukder S, Hughes TR, Luscombe NM, Ukkonen E, Taipale J. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res*, 2010, 20(6): 861–873. [DOI](#)
- [107] Wong D, Teixeira A, Oikonomopoulos S, Humburg P, Lone IN, Saliba D, Siggers T, Bulyk M, Angelov D, Dimitrov S, Udalovala IA, Ragoussis J. Extensive characterization of NF-kappaB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. *Genome Biol*, 2011, 12(7): R70. [DOI](#)
- [108] Sen R, Baltimore D. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell*, 1986, 46(5): 705–716. [DOI](#)
- [109] Sen R, Smale ST. Selectivity of the NF-kB response. *Cold Spring Harb Perspect Biol*, 2010, 2(4): a000257. [DOI](#)
- [110] Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, 2003, 31(1): 374–378. [DOI](#)
- [111] Nijnik A, Mott R, Kwiatkowski DP, Udalovala IA. Comparing the fine specificity of DNA binding by NF-kB p50 and p52 using principal coordinates analysis. *Nucleic Acids*

- Res*, 2003, 31(5): 1497–1501. [DOI](#)
- [112] Udalova IA, Mott R, Field D, Kwiatkowski D. Quantitative prediction of NF-kappa B DNA-protein interactions. *Proc Natl Acad Sci USA*, 2002, 99(12): 8167–8172. [DOI](#)
- [113] Linnell J, Mott R, Field S, Kwiatkowski DP, Ragoussis J, Udalova IA. Quantitative high-throughput analysis of transcription factor binding specificities. *Nucleic Acids Res*, 2004, 32(4): e44. [DOI](#)
- [114] Nutiu R, Friedman RC, Luo S, Khrebtukova I, Silva D, Li R, Zhang L, Schroth GP, Burge CB. Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat Biotechnol*, 2011, 29(7): 659–664. [DOI](#)