

一种检测小鼠转录因子活性的微阵列芯片的构建及应用*

孙义民^{1,2,3,4)**} 曾令琴^{1,2)} 张 岩^{1,2)} 魏 丽^{3,4)} K R MITCHELSON^{1,2,3)}

张 亮^{3,4)} 程 京^{1,2,3,4)}

⁽¹⁾清华大学医学系统生物学研究中心, 北京 100084; ⁽²⁾清华大学医学院, 北京 100084;

⁽³⁾生物芯片北京国家工程研究中心, 北京 102206; ⁽⁴⁾博奥生物有限公司, 北京 102206)

摘要 转录因子是一类具有序列特异性的双链 DNA 结合蛋白。在哺乳动物细胞中, 一个转录因子可以调控多个靶基因, 一个基因也同时受到多个转录因子的调控, 从而形成复杂的基因表达网络调控机制。在蛋白质水平检测特定状态下组织或细胞中的转录因子活性对深入研究基因表达调控的分子机制具有重要的意义。针对 TRANSFAC 数据库提供的 226 个小鼠转录因子的 PSSM, 设计了 240 个转录因子结合探针, 构建了可以在蛋白质水平同时检测约 200 个小鼠转录因子的微阵列芯片, 实验结果显示: a. 针对含有不同 DNA 结合结构域的 7 个转录因子, 进行依赖于抗体的转录因子活性检测, 结果证明该芯片具有良好的特异性; b. 针对 NF κ B 转录因子纯品, 芯片的检测灵敏度可以达到 0.5 nmol/L, 线性范围为 0.5~20 nmol/L, 表明芯片具有较高的灵敏度和良好的定量能力; c. 针对经典的 IKK, JNK 信号通路和 JAK/STAT 信号通路, 该芯片可以检测到其关键转录因子的活性变化, 证明芯片的可靠性; d. 针对小鼠前脂肪细胞系 3T3-L1 的分化, 该转录因子活性芯片不但检测到了已经报道的活性发生变化的转录因子, 如 PPAR 家族和 C/EBP 家族外, 还发现了以前没有报道过的 SRF 等, 充分显示了芯片技术的高通量优势和发现新数据的能力。

关键词 基因表达调控, 转录因子, 微阵列芯片, TF-ELISA

学科分类号 Q81

DOI: 10.3724/SP.J.1206.2008.00399

随着人类第一号染色体的序列公布^[1], 标志着“人类基因组计划(HGP)”正式全部完成, 在 HGP 的带动下, 包括小鼠, 水稻, 酵母等几十种模式生物的基因组序列也公布于世^[2], 生物学家们的研究重点逐渐地由基因组学向功能基因组学过渡。包含转录调控, 转录后加工, 翻译调控, 翻译后修饰等几个核心层次的基因表达调控, 无疑在功能基因组研究领域占据了重要的位置。近年来, 高通量 DNA 甲基化检测^[3]和 microRNA 研究^[4]的兴起使基因表达调控变得更加神秘和复杂。

转录因子介导的转录调控掌控着基因表达的第一道关卡。转录因子是细胞内一类重要的调控蛋白, 其特异地结合基因组 DNA 的顺式作用元件以调节基因的转录过程。在哺乳动物细胞中, 如人和小鼠, 编码基因不到全基因组 10% 的转录因子通过复杂的网络调控机制, 严谨地控制超过 20 000 个基因的表达情况, 一个转录因子调控多个靶基

因, 同时一个基因也会受到多个转录因子的组合调控。转录因子以重要的角色参与了包括肿瘤发生, 细胞分化等几乎所有的生物学过程^[5], 尤其是日本京都大学 Yamanaka 教授^[6,7]最近发表的 Oct4、Sox2、Myc 和 Klf4 可以逆转体细胞成为干细胞的科研成果, 使生物学家对转录因子的兴趣又上升到了一个新的高度。

转录因子的检测可以分为 4 个层次: mRNA 水平, 可以通过表达谱芯片或 RNA 印迹进行分析; 蛋白质的量, 可以通过 Western blot 法进行分析; 结合活性水平, 可以通过凝胶滞留实验(EMSA)进行分析; 激活活性水平, 通常用报告系

* 国家高技术研究发展计划(863)(2006AA020701)资助项目。

** 通讯联系人。

Tel: 010-80727291, E-mail: ymsun@capitalbio.com

收稿日期: 2008-06-03, 接受日期: 2008-06-30

统进行研究. 由于基因表达调控网络的存在, 使转录因子研究的各个层次间没有必然的相关性, 所以针对每个层次开发高通量的检测手段, 对全方位研究转录因子有着重要的意义. 本文构建的转录因子活性芯片就是在结合活性水平开发的高通量分析方法. EMSA 是经典的在体外研究转录因子的方法^[8], 其结果可靠, 为大家所公认, 但却存在通量低, 耗时费力的缺点. 近年来出现的通过 ELISA 技术来检测转录因子的方法^[9]在一定程度上弥补了 EMSA 的劣势, 可以较方便地检测在某生物学过程中的转录因子活性变化, 但其依赖于转录因子的抗体, 部分地限制了该技术的应用. 目前, 通过结合位点分析, 可以对共调控一簇基因的转录因子进行生物信息学预测^[10], 对深入研究转录调控的机制提供了一定的线索, 但其预测的准确率有待于进一步提高. 综上所述, 开发一种检测转录因子活性的高通量实验技术对洞察基因表达调控的网络机制至关重要^[11].

本课题组的前期实验采用 OATFA (Oligonucleotide array-based transcription factor assay) 技术实现了在微阵列芯片上检测 AP1, NF κ B 和 SP1 三个转录因子的活性^[12]. 本研究将 OATFA 技术应用在小鼠模型上, 构建了可以检测约 200 个转录因子的小鼠转录因子活性谱芯片(MOUSE OATFA), 实验结果表明, 我们成功地开发了一种在蛋白质水平高通量检测转录因子活性的稳定、可靠的微阵列芯片平台.

1 材料与方法

1.1 材料

小鼠 RAW264.7 和 3T3-L1 细胞株为清华大学医学系统生物学研究中心冻存; LPS(Salmonella Minnesota)购自美国 Sigma 公司(Cat#. L6261); 小鼠干扰素 γ (IFN γ , 10^7 U/mg)购自美国 PeproTech 公司(Cat#. 315-05); 重组转录因子纯品 NF κ B(p50)购自美国 Promega 公司(Cat#. E3770); 转录因子抗体 AP1(sc-1694x)、C/EBP α (sc-61)、HNF1(sc-8986)、HNF4 (sc-8987)、NF1 (sc-5567)、NF κ B (sc-372x)、PPAR γ (sc-7196)、SRF (sc-335)和 STAT1 (sc-346)购自美国 Santa Cruz 公司; SmartArrayTM 芯片点样仪、LuxScanTM 10K 激光共聚焦芯片扫描仪、芯片清洗仪、芯片杂交盒、和 DNA 点样液均为博奥生物有限公司产品; 数据分析用软件 LuxScanTM 3.0 为博奥生物有限公司产品.

1.2 实验方法

1.2.1 细胞培养及处理. 小鼠巨噬细胞系 RAW264.7 用含 10%胎牛血清, 100 mg/L 链霉素和 100 U/ml 青霉素的 RPMI 1640 培养基进行培养, 当细胞长至 80%左右时, 用终浓度为 500 μ g/L 的 LPS 进行刺激, 收集刺激 2 h 的细胞样品进行实验. 小鼠前脂肪细胞 3T3-L1 用含 10%小牛血清, 100 mg/L 链霉素和 100 U/ml 青霉素的 DMEM 培养基进行培养, 当细胞长至 80%左右时, 用终浓度为 50 U/ml 的 IFN γ 进行刺激, 收集刺激 20 min 的细胞样品进行实验. 在分化研究时, 当 3T3-L1 细胞长满 48 h 后, 转到含 10%胎牛血清, 0.5 mmol/L 3- 异丁基 -1- 甲基黄嘌呤 (IBMX), 1 μ mol/L 地塞米松和 1.7 μ mol/L 胰岛素的 DMEM 培养基中进行诱导分化, 48h 后用含 10%胎牛血清的 DMEM 培养基继续培养 24 h, 收集细胞进行实验. 所有细胞均在 37 $^{\circ}$ C, CO $_2$ 浓度为 5%的培养箱中培养.

1.2.2 芯片探针的设计及芯片的点制.

芯片探针的设计模板来源于商品化的 TRANSFAC[®] Professional r8.2 数据库中的转录因子结合位置权重矩阵 PSSM^[13](<http://www.biobase.de>). 探针设计过程采用模拟退火的算法^[14, 15], 对探针的退火温度(T_m)、二级结构形成、灵敏度和特异性进行综合评分, 最终确定探针的序列. 针对每个转录因子设计一个“探针组”, 其中包括 3 条探针. “探针 1”为固定在芯片上的探针, 长度为 30 nt, 其 5'端进行氨基修饰; “探针 2”与“探针 1”序列完全互补; “探针 3”中除了 30 nt 与“探针 1”完全相同的序列外, 还在 3'端增加了 T7 启动子通用引物序列(5' CCCTATAGTGAGTCGTA-TTACCCC 3'). “探针 2”和“探针 3”经过体外退火形成转录因子结合探针, 其中包含了 1~2 个转录因子核心结合序列. 除了针对转录因子设计的 240 个“探针组”外, 还设计了作为质量控制的 4 个 HNC 探针(只包含“探针 1”)和 4 个 BNC 探针组. 用于芯片实验的“探针库”包括了 240 个转录因子结合探针和 4 个 BNC 双链探针(“探针 2”和“探针 3”的退火产物).

所有探针都在上海生工生物技术公司合成, 然后以 20 μ mol/L 的浓度溶于 DNA 点样液中. 探针经 SmartArrayTM 芯片点样仪点制在经过醛基修饰的光学基片上, 每个探针重复 2 个点, 点间距为 200 μ m, 每张芯片共包含 22 行、24 列. 点制好的

芯片通过探针上的氨基和芯片表面的醛基发生的希夫碱反应进行固定化处理, 然后室温干燥保存备用。

1.2.3 细胞核蛋白的制备. 采用美国 Pierce 公司 NE-PER® Nuclear and Cytoplasmic Extraction Reagents 试剂盒, 按照试剂盒说明书进行核蛋白的提取. 收集细胞后, 加入适量的含有蛋白酶抑制剂的 CER I, 混匀后在冰上放置 10 min, 加入 CER II, 在冰上放置 1 min 后, 16 000 *g* 离心 5 min, 取上清转到预冷的 Eppendorf 管内. 在离心沉淀下来的细胞核中加入含有蛋白酶抑制剂的 NER, 充分混匀后, 冰上放置 10 min, 再混匀, 重复 4 次后, 16 000 *g* 离心 10 min, 收集上清. 提取出的细胞核蛋白和胞浆蛋白采用美国 Pierce 公司 BCA™ Protein Assay Kit 试剂盒进行浓度测定, 分装后保存在 -80℃.

1.2.4 转录因子活性芯片的检测流程. a. 结合: 含 240 个转录因子结合探针的“探针库”与细胞核蛋白(或转录因子纯品)在含有 polydIdC 的缓冲液中进行结合反应, 反应体系为 20 μ l, 室温反应 1 h. b. 分离: 将上述的结合混合物进行 1.5% 琼脂糖凝胶电泳分离, 电泳条件为 12 V/cm, 室温电泳 30 min, 电泳缓冲液为 0.5 $\times$ TBE. 将含有 DNA- 转录因子复合物的凝胶部分切下, 通过 QIAGEN 公司 QIAEX II® Gel Extraction Kit 试剂盒进行 DNA 的回收纯化. c. 单引物扩增 (single primer amplification, SPA): 以回收的 DNA 作为模板, 利用荧光(Cy5 和 Cy3)标记的引物(5'GGGGTAAT-ACGACTCACTATAGGG 3')进行单引物扩增. d. 杂交: 将分别标记不同荧光的配对样品的扩增产物同时加入含有 3 $\times$ SSC、0.2%SDS、5 $\times$ Denhart's 和 25%甲酰胺的杂交缓冲液中, 42℃ 杂交 12~16 h. 然后在芯片清洗仪中用洗液 I (2 $\times$ SSC、0.2%SDS)、洗液 II (0.2 $\times$ SSC)分别洗 4 min, 离心甩干芯片. e. 图像获得和数据处理: 芯片扫描采用 LuxScan™ 10K 激光共聚焦芯片扫描仪, 芯片上的信号强度使用 LuxScan™ 3.0 软件进行提取. 针对芯片上每个探针的 2 个重复点计算平均值, 代表该探针的信号强度. 每张芯片中两个通道的信号值采用中位值线性归一化方法, 得到每个探针的比值(ratio). 多次实验的比值采用 *t* 检验进行显著性分析, 当多次实验比值的平均值大于(或等于)1.5 并且 *P* 值小于(或等于)0.05 时, 该探针对应的转录因子被认为其活性发生了显著的变化.

1.2.5 基于抗体的转录因子结合特异性检测. 结合反应按照 1.2.4 中步骤进行. 然后向反应体系中加入某特定转录因子的抗体, 4℃ 反应 2 h, 同时制备一份只加正常 IgG(不加抗体)的样品作为阴性对照. 通过 Protein A(或 G)包被的磁珠将抗体 - 转录因子 - 结合探针的复合物收集回来. 回收复合物中的探针进行后续的 SPA 扩增和杂交, 扩增和杂交按照 1.2.4 中步骤进行.

1.2.6 转录因子的 ELISA 检测(TF-ELISA). 具体操作步骤参考本研究小组已经发表的文献[12]. 针对每个转录因子, 包被在 96 孔板上的序列分别为: AP1, CTCGACCCGAATGAGTCAGCATGAGTAGTG; C/EBP α , AATATTGCCCAACCAACATTGCGTAAGGAC; NF κ B, GGAGGGACTTTCCAAGAGG-GACTTTCCAAG; PPAR γ , AGCGTACAGGGTCA-AAGGTCAAGGCTCAAG; SRF, CTCGGTATGCC-CATATATGGTAGTACCGGA; STAT1, GATCTTCCGTAAAGTCCTTCCCGGAACGTC.

2 结 果

2.1 转录因子活性芯片的构建

针对 TRANSFAC 数据库提供的 226 个小鼠转录因子结合 PSSM, 采用了模拟退火的算法, 设计了 260 个“探针组”, 每个探针组包含 3 条探针, 同时还设计了作为杂交对照的 HNC 和作为结合对照的 BNC. 为了进行探针的优化筛选, 把这些对照探针和 260 个转录因子结合探针点制成微阵列芯片, 然后将每个探针进行扩增标记, 与芯片进行单独杂交分析, 去掉有交叉杂交和杂交信号强度弱的探针, 最后保留 240 个转录因子结合探针, 可以检测约 200 个小鼠的转录因子.

小鼠转录因子活性芯片进行检测时主要分为 4 个步骤(图 1): a. 结合. 把 240 个转录因子结合双链探针与 4 个 BNC 双链探针混合在一起作为“探针库”, 将这个探针库与 5 μ g 从细胞或组织中提取的核蛋白在结合缓冲液中进行孵育, 同时加入适量的为减少非特异结合的 polydIdC, 使核蛋白中的转录因子与其对应的双链探针充分结合. b. 分离. 采用琼脂糖凝胶电泳的方式将混合物中未结合的游离探针去除, 将若干转录因子 -DNA 的结合复合物进行试剂盒回收, 将复合物中的 DNA 纯化出来. c. 扩增和标记. 利用通用的荧光标记引物, 采用单引物扩增(SPA)的方法, 对回收的 DNA 进行扩增标记. 虽然 SPA 技术的扩增倍数不高, 但其在

扩增前后可以保持良好的定量关系，使实验结果更加准确可靠。d. 芯片杂交和数据分析。将分别标记不同荧光的来自于两个样品的核酸探针混合在一起，与芯片进行杂交。

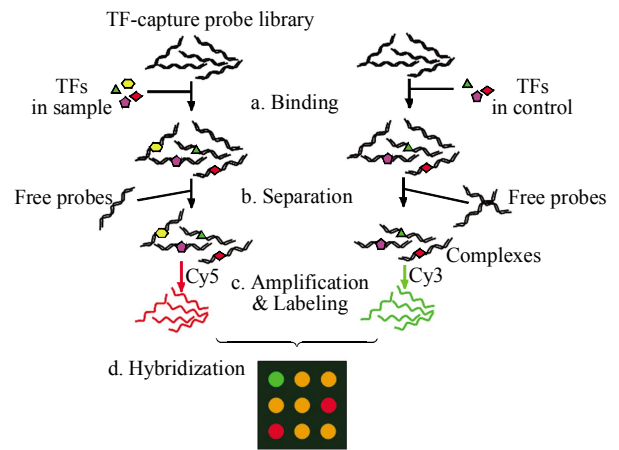


Fig. 1 The MOUSE OATFA direct comparison assay (DCA) procedure

2.2 转录因子活性芯片的结合特异性检测

在 TRANSFAC 数据库中，根据转录因子的 DNA 结合结构域的不同，将转录因子主要分为 4 个大类(Superclass). 为了检测芯片上的探针与其对应的转录因子之间的结合特异性，我们从 4 大类转录因子中随机选择了 AP1, HNF1 等 7 个转录因子(表 1)，在核蛋白与“探针库”的结合反应结束后，通过各自的抗体将转录因子-DNA 复合物免疫沉淀下来，然后回收复合物中的 DNA 进行扩增、标记和杂交，通过芯片上的信号来判断芯片的结合特异性(图 2)。

在哺乳动物细胞中，转录因子和其结合序列之间存在交叉结合的现象，尤其在具有同样 DNA 结合结构域的转录因子家族中，即一个转录因子可以结合序列相似的多个 DNA 片段，同一个 DNA 序列有时也可以被多个转录因子结合。对于转录因子

Table 1 Classification of seven selected TFs detecting the binding specificity

	AP1	NF1	NFκB	STAT1	HNF1	HNF4	PPARγ
Subfamily	JUN/FOS	—	—	—	HNF1	HNF4	PPAR
Family	AP1-like	NF1	Rel	STAT	Homeo domain only	Thyroid hormone receptor-like	
Class	bZIP	SMAD/NF1	RHR	STAT	Homeo domain		Cys4
Superclass	Basic domains		Beta-scaffold		Helix-turn-helix		Zinc-finger

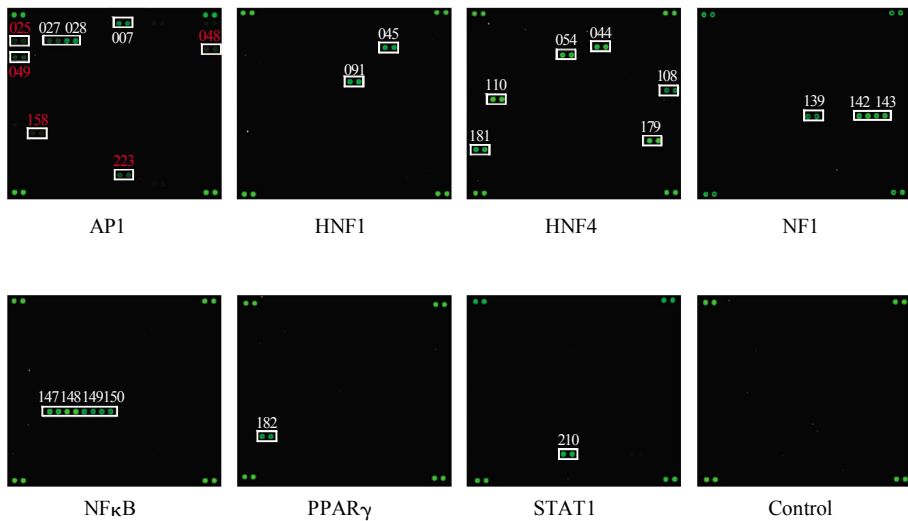


Fig. 2 Antibody-based individual TF assay in nuclear extracts

To test the binding specificity of physiological TFs in cellular nuclear extracts, we constructed the antibody-based individual assay. Seven transcription factors including AP1, HNF1, HNF4, NF1, NFκB, PPARγ and STAT1 covering all the four superclasses were chosen. The probes with white letters were all specially designed for the corresponding TFs. For AP1, the probes with red letters were for some other bZIP TFs whose binding sequences also included the whole or half AP1-site.

HNF1、NFκB 和 NF1，芯片上所有为其设计的结合探针都得到了明显的富集(图 2)，而在没有加任

何抗体的阴性对照杂交图片中，看不到非特异的杂交信号。对于 STAT(signal transducer and activator

of transcription)家族, 其 DNA 结合序列非常接近, 通过转录因子活性芯片无法将各成员之间有效地区分, 针对 STAT 家族, 芯片上共设计了 4 条探针, 由于各探针与 STAT1 转录因子之间结合能力的差异, 只有其中一个探针被显著地富集出来. HNF4 和 PPAR γ 都属于 Cys4 锌指蛋白家族, 与 COUP-TF 三者之间的 DNA 结合序列非常相似, 但通过 HNF4 和 PPAR γ 的抗体富集出来的探针却不尽相同, 说明各探针与转录因子的结合能力并不一致. 在转录因子 AP1 的检测结果中, 富集的探针中除了 mmu-007、mmu-027 和 mmu-028 三个特异的探针外(白色标注), 还包括了 mmu-025(红色标注)等 5 个探针, 序列分析显示, 在这 5 个探针中都含有一个(5' TGAC/GTCA 3')或半个(5' TGAC 3') AP1 的结合位点, 可以被 AP1 蛋白中的 JUN 或 FOS 亚基单独结合, 这 5 个探针对应的转录因

子, 如 MAF 和 NRF, 与 AP1 一样, 同属于 bZIP 家族.

2.3 转录因子活性芯片的灵敏度、定量能力和重复性

通过商品化的转录因子纯品 NF κ B(p50), 我们对小鼠转录因子活性芯片的灵敏度和定量能力进行了检测, 实验结果如图 3 所示. 通过 NF κ B 转录因子的纯品实验, 芯片上 4 个为其设计的探针都得到了明显的富集. 通过纯品实验中富集探针的 DNA 序列, 利用 MDscan 和 WebLogo 软件, 绘制了 NF κ B 的结合矩阵, 与经典的 NF κ B 结合序列(5' GGGAATTTC 3')得到了很好的吻合. 利用 NF κ B 纯品蛋白, 我们进行了浓度依赖的定量检测, 发现芯片检测 NF κ B 的灵敏度为 0.5 nmol/L, 比本课题组的前期研究^[12]提高了一个数量级, 线性区间为 0.5~20 nmol/L.

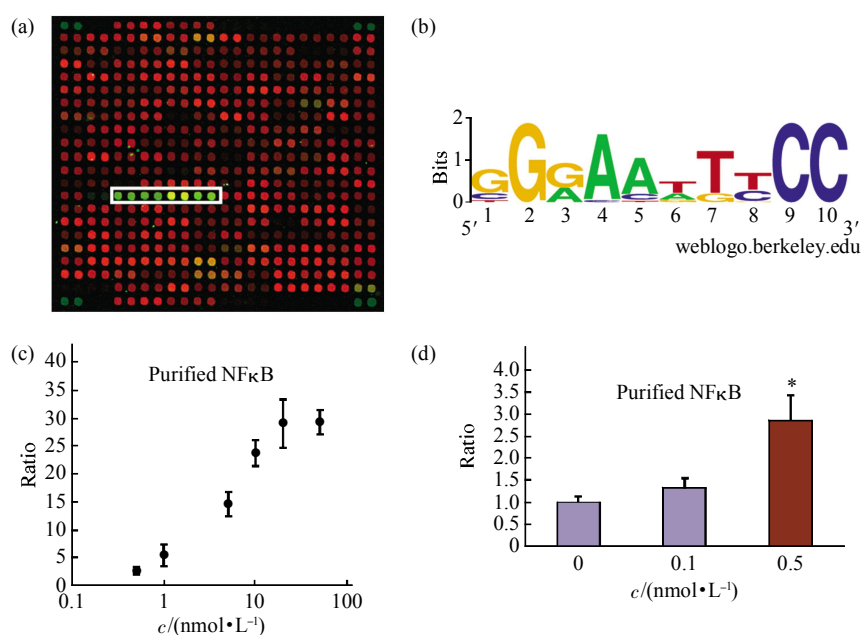


Fig. 3 Detection of sensitivity and quantitative ability by purified NF κ B (p50)

(a) Hybridization picture. The probes enriched by purified NF κ B were labeled as Cy3 (green fluorescence) and TF-capture probe library was directly labeled as Cy5 (red fluorescence). The probes within the white frame were designed for NF κ B. (b) Binding matrix of NF κ B based on the experimental data with the aid of MDscan (<http://mdscan.stanford.edu>) and WebLogo (<http://weblogo.berkeley.edu>). (c) Quantitative curve of purified NF κ B in which y axis was the enriched ratio. (d) Sensitivity of microarray using purified NF κ B. Brown column stood for the concentration whose enriched ratio was significantly (*) higher than the before.

针对小鼠巨噬细胞系 RAW264.7, 我们对不同含量的核蛋白提取物进行了芯片测试. 利用 5 μ g 核蛋白进行实验, 得到的各探针的比值与 1 μ g, 2 μ g, 10 μ g 核蛋白之间的相关系数分别为 0.95, 0.98 和 0.98, 并发现利用 5 μ g 核蛋白可以得到最

佳的实验效果. 使用 5 μ g 核蛋白提取物, 我们进行了 3 次独立的重复实验, 3 次实验结果的聚类分析非常相似, 并且每 2 次实验间的相关系数可以达到 0.99(图 4), 说明小鼠转录因子活性芯片具有良好的重复性.

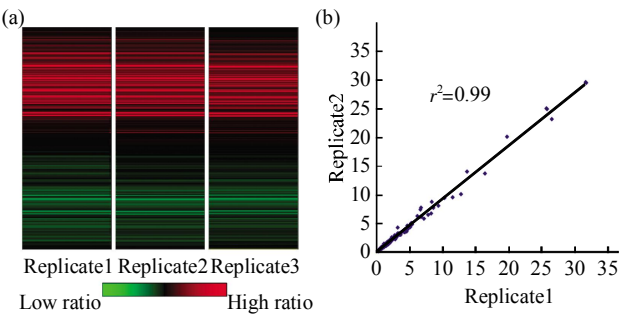


Fig. 4 Reproducibility of microarray using nuclear extracts of RAW264.7

(a) Clustering analysis of ratio from 3 independent replicates. (b) Correlative analysis of ratio from 2 independent replicates.

2.4 转录因子活性芯片在 IKK, JNK 信号通路和 JAK/STAT 信号通路中的应用

LPS 是革兰氏阴性细菌的主要成分，它可以通过糖脂结构中的脂肪酸链激活细胞的免疫应答引起内毒素效应，如革兰氏阴性脓血症。当利用 LPS 对细胞进行刺激时，LPS 首先与游离在血清中的 LBP 蛋白结合，然后通过细胞表面的 CD14 分子介导 LPS 效应。LPS-LBP-CD14 三分子复合物与结合

在细胞表面的 TLR 受体相互作用后，可以激活下游的 I κ B 蛋白激酶(IKK)，IKK 磷酸化位于细胞质中的 I κ B，导致 I κ B 被泛素化降解，使定位在胞浆中的 NF κ B 进入到细胞核中起始下游靶基因的转录，从而引起一些促炎症因子如 IL1、IL6、IL8 和 TNF α 的合成和释放^[16]。同时，在 LPS 的刺激下，JNK 通过 MAPK 信号通路被激活，进而磷酸化 AP1 使其活性升高^[17]。在本研究中，我们利用终浓度为 500 μ g/L 的 LPS 刺激小鼠巨噬细胞 RAW264.7 2 h 后，收集刺激前后的核蛋白进行转录因子的活性变化研究，实验结果如图 5 所示。在独立的 3 次重复性实验中，共有 4 个探针在刺激前后发生了显著的变化(ratio > 1.5, $P < 0.05$)，其中 mmu-007 和 mmu-012 对应的是转录因子 AP1，而 mmu-148 和 mmu-150 对应的是 NF κ B。在考察芯片平台重复性的自身 - 自身的杂交实验中，没有任何一个探针的比值发生了显著的变化，从另外一个角度上证明了实验结果的可靠性。通过 TF-ELISA 技术，我们对 LPS 刺激模型中 AP1 和 NF κ B 转录因子的活性升高进行了验证，与芯片的结果完全符合。

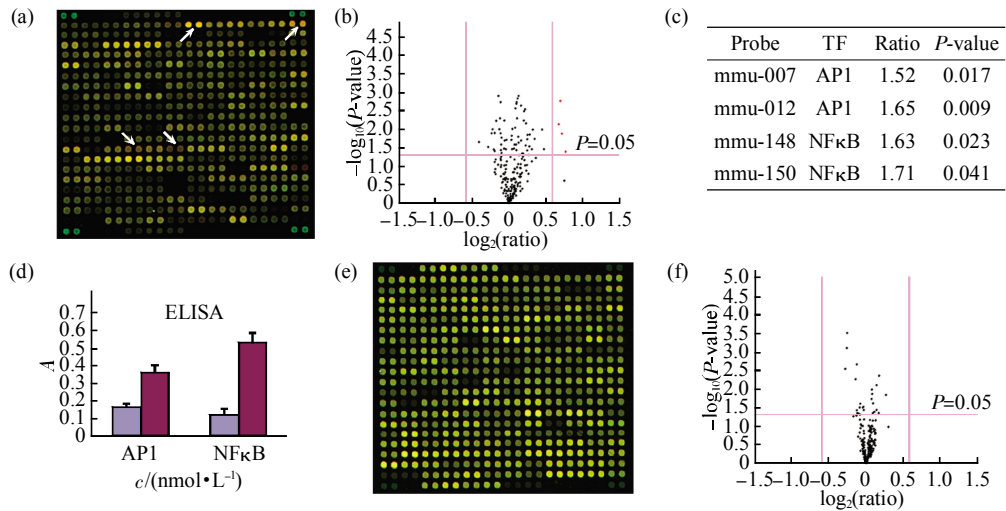


Fig. 5 Detection of changed TFs in LPS-stimulated RAW264.7 cells

After 2 h of stimulation by LPS (500 μ g/L), nuclear proteins were extracted from treated and untreated cells and used to perform OATFA analysis. (a) Hybridization picture in which the DNA enriched from treated cells was labeled with Cy5. The probes oriented by arrows were the changed ones. (b) Volcano plots depicting fold change ($\log_2$ ratio, x-axis) and statistical significance ($-\log_{10} P$, y-axis). The red spots on the top right region stood for the up-regulated TFs (ratio > 1.5, $P < 0.05$, $n=3$). (c) List of changed probes. ■: LPS(-); ■: LPS(+). (d) Result from antibody-based TF-ELISA validation assay. (e) Hybridization picture of self-self experiment in which the identical DNA was labeled and hybridized together. (f) Volcano plots from self-self experiment.

另外，一个研究得很清楚的模型是干扰素 γ (IFN γ)介导的 JAK/STAT 信号通路。IFN γ 与细胞表面的受体结合后，激活在细胞内表面与其相互作用的 JAK1 和 JAK2，然后招募转录因子 STAT1 与

IFN γ 受体的胞内部分结合，这时 JAK 会磷酸化 STAT1 的酪氨酸残基(Y701)导致 STAT1 发生蛋白质二聚化，二聚化后的 STAT1 进入到细胞核中调节下游靶基因的表达^[18]。在本研究中，我们利用终

浓度为 50 U/ml 的 $\text{IFN}\gamma$ 刺激小鼠前脂肪细胞系 3T3-L1, 为了评估转录因子活性芯片对活性发生快速变化的转录因子的检测能力, 我们选择刺激 15 min, 然后收集刺激前后的核蛋白进行转录因子的活性变化研究, 实验结果如图 6 所示. 在独立的

3 次重复性实验中, 共有 2 个探针在刺激前后发生了显著的变化($\text{ratio} > 1.5$, $P < 0.05$), 2 个探针对应的都是 STAT 家族的转录因子. 通过 TF-ELISA 技术, 我们验证了 STAT1 的活性变化, 与芯片结果吻合.

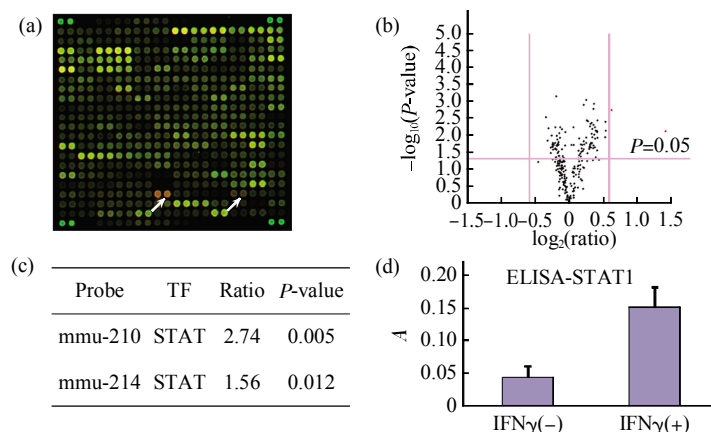


Fig. 6 Detection of changed TFs in $\text{IFN}\gamma$ -stimulated 3T3-L1 cells

After 15min of stimulation by $\text{IFN}\gamma$ (50 U/mL), nuclear proteins were extracted from treated and untreated cells and used to perform OATFA analysis. (a) Hybridization picture in which the DNA enriched from treated cells was labeled with Cy5. The probes oriented by arrows were the changed ones. (b) Volcano plots depicting fold change ($\log_2$ ratio, x-axis) and statistical significance ($-\log_{10}$ P-value, y-axis). The red spots on the top right region stood for the up-regulated TFs ($\text{ratio} > 1.5$, $P < 0.05$, $n=3$). (c) List of changed probes. (d) Result from antibody-based TF-ELISA validation assay.

2.5 转录因子活性芯片在小鼠前脂肪细胞3T3-L1分化中的应用

肥胖是威胁人类健康的国际性难题, 可以导致糖尿病、肿瘤、心脏病等多种并发症, 小鼠前脂肪细胞 3T3-L1 向脂肪细胞的分化是研究肥胖最常用的细胞模型. 目前, 已经有文章详细研究了

3T3-L1 细胞分化过程中基因表达随分化时间的变化情况^[9], 但对于这些表达发生变化的基因调控机制还不是十分清楚. 在本研究中, 我们对 3T3-L1 细胞分化 3 天时的转录因子活性变化进行了检测, 实验结果如图 7 所示. 活性上调的转录因子包括 NFAT, PPAR, C/EBP 和 GR, 下调的转录因子包

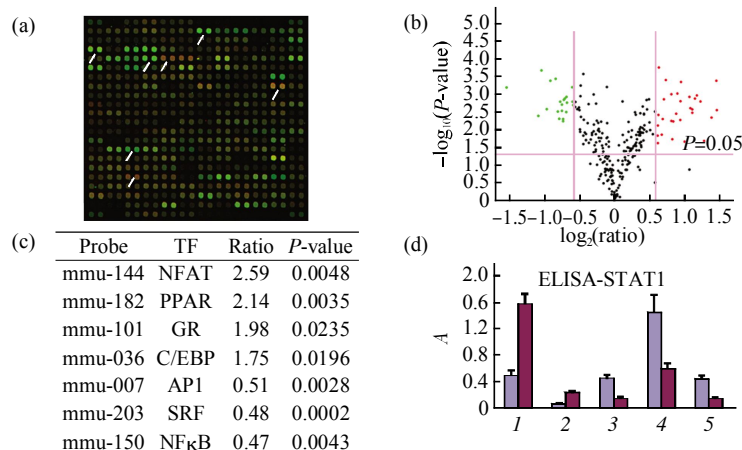


Fig. 7 Application of MOUSE OATFA in preadipocyte 3T3-L1 differentiation

After 2 days of postconfluence in DMEM with 10% bovine serum, 3T3-L1 cells were induced to differentiate by changing the medium to DMEM containing 10% fetal bovine serum, 0.5 mmol/L 3-isobutyl-1-methylxanthine, 1 $\mu\text{mol/L}$ dexamethasone, and 1.7 $\mu\text{mol/L}$ insulin. After 48 h this medium was replaced with DMEM supplemented with 10% fetal bovine serum, and cells were maintained in this medium for another 24 h before experimentation. (a) Hybridization pictures in which the DNA enriched from differentiated cells was labeled with Cy5. The spots presenting green and red were the changed ones. (b) Volcano plots depicting fold change ($\log_2$ ratio, x-axis) and statistical significance ($-\log_{10}$ P-value, y-axis). The red spots on the top right region stood for the up-regulated TFs ($\text{ratio} > 1.5$, $P < 0.05$, $n=3$) and green ones on the top left region stood for the down-regulated TFs ($\text{ratio} < 0.66$, $P < 0.05$, $n=3$) after differentiation. (c) Part of changed probes. (d) Result from antibody-based TF-ELISA validation assay. 1: PPAR γ ; 2: C/EBP α ; 3: AP1; 4: SRF; 5: NF κ B. ■: 0d; ■: 3d.

括 AP1, NF κ B 和 SRF. 其中, C/EBP 家族和 PPAR 家族是最早被发现也是研究最清楚的调控 3T3-L1 细胞分化的转录因子^[20]. AP1, NF κ B 和 GR 是糖皮质激素响应的转录因子^[21], 诱导细胞分化的地塞米松就是糖皮质激素的一种, 对于转录因子 NFAT 和 SRF, 目前尚未见相关报道.

3 讨 论

TRANSFAC 数据库^[13]和 JASPAR^[22]是两个国际公认的转录因子结合位点数据库, 其中更以 TRANSFAC 应用更加广泛. 本研究中, 小鼠转录因子活性芯片的探针设计部分就是基于 TRANSFAC 数据库进行的, 芯片探针对于转录因子的检测能力与数据库质量密切相关. 随着数据库的不断更新, 我们的转录因子活性芯片也会随之更新.

基于转录因子结合序列的活性检测, 无论是传统的 EMSA、报告系统, 还是本研究中的芯片技术, 对于检测某些哺乳动物细胞的转录因子都有一个缺陷, 就是从实验结果不能精确给出到底是哪一个转录因子的活性发生了变化, 这些技术的直观结果就是所有能结合在这个序列上的转录因子都有可能发生了活性变化, 例如 HNF4、PPAR 和 COUP-TF, GATA 家族, STAT 家族等, 所以对于芯片上检测出来的活性发生变化的转录因子, 要通过依赖于抗体的其他技术进行验证, 如 TF-ELISA^[9]技术或基于 EMSA 技术的 Supershift 实验. 对本研究中研制的转录因子活性芯片, 比较精确的定位是: 一种在蛋白质水平针对转录因子结合活性进行检测的指导性的高通量筛选技术, 相当于高通量的 EMSA.

与本课题组的前期研究^[12]相比, 在本研究中构建的小鼠转录因子活性芯片的主要革新在于: a. 通过严谨的探针设计和实验条件优化, 将能够检测的转录因子由 3 个增加到约 200 个, 真正的使转录因子的高通量筛选成为可能; b. 引入单引物扩增 (SPA)^[23]技术, 在保持原有定量关系的同时, 使检测 NF κ B 纯品的灵敏度提高了一个数量级; c. 引入引物末端荧光标记方法, 实现了芯片的双通道检测, 使成对样品间的比较更加直接和准确.

目前, 研究转录调控水平的生物芯片平台有 mRNA 表达谱芯片和 ChIP-chip 技术^[24], 如果能将这两种芯片与本研究中的转录因子活性芯片联合使用, 将对深入理解基因表达调控的网路机制产生巨大的帮助. 以小鼠前脂肪细胞 3T3L1 分化模型为

例, 通过转录因子活性芯片的分析, 不但检测到了已经报道过的 PPAR 和 C/EBP 家族, 同时也发现了从未研究过的 SRF、NFAT 等转录因子, 如果利用 ChIP-chip 技术在全基因组范围内检测 SRF 的靶基因, 然后再联合分化前后的 mRNA 表达谱芯片结果, 就能精确找出 SRF 在脂肪细胞分化中调控的表达发生变化的功能基因, 这将有助于深入研究 SRF 在脂肪分化中所发挥的生物学功能, 目前本课题组正在进行相关的工作.

疾病(如肿瘤、老年痴呆症等)发生过程中的生物标志物筛选对深入了解分子机制, 疾病的分子分型与个体化诊疗、药物开发等领域有着重要的意义. 传统的生物标志物研究一直停留在“量”上, 如哪些基因的 mRNA 表达“量”发生了变化^[25], 哪些 microRNA 的表达“量”发生了变化^[26], 随着生物学研究领域的拓宽和技术的进步, 与遗传相关的 SNP(单核苷酸多态性)检测^[27]以及表观遗传学中的 DNA 甲基化逐渐成为了生物标志物筛选的焦点之一^[3], 并具有大规模推广的潜力. 随着转录因子活性检测芯片技术的成熟和推广, 可以预见, 代表生物分子生命力的转录因子“活性”检测将成为生物标志物筛选的又一重要领域.

参 考 文 献

- 1 Gregory S G, Barlow K F, McLay K E, *et al.* The DNA sequence and biological annotation of human chromosome 1. *Nature*, 2006, **441** (7091): 315~321
- 2 Waterston R H, Lindblad-Toh K, Birney E, *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002, **420** (6915): 520~562
- 3 Gebhard C, Schwarzfischer L, Pham T H, *et al.* Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res*, 2006, **66** (12): 6118~6128
- 4 Thomson J M, Parker J, Charles M P, *et al.* A custom microarray platform for analysis of microRNA gene expression. *Nat Methods*, 2004, **1** (1): 47~53
- 5 Boyadjiev S A, Jabs E W. Developmental biology: Frontiers for clinical genetics. *Clin Genet*, 2000, **57** (4): 253~266
- 6 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
- 7 Takahashi K, Tanabe K, Ohnuki M, *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 2007, **131** (5): 861~872
- 8 Garner M M, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory

- system. *Nucleic Acids Res*, 1981, **9** (13): 3047~3060
- 9 Renard P, Ernest I, Houbion A, *et al.* Development of a sensitive multi-well colorimetric assay for active NF κ B. *Nucleic Acids Res*, 2001, **29**(4): e21
- 10 Ho Sui S J, Mortimer J R, Arenillas D J, *et al.* oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res*, 2005, **33** (10): 3154~3164
- 11 Walhout A J. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res*, 2006, **16** (12): 1445~1454
- 12 Shao W, Wei H J, Qiao J Y, *et al.* Parallel profiling of active transcription factors using an oligonucleotide array-based transcription factor assay (OATFA). *J Proteome Res*, 2005, **4** (4): 1451~1456
- 13 Wingender E, Chen X, Hehl R, *et al.* TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res*, 2000, **28** (1): 316~319
- 14 Metropolis N, Rosenbluth A W, Rosenbluth M N, *et al.* Equation of state calculations by fast computing machines. *J Chem Phys*, 1953, **21** (6): 1087~1092
- 15 Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by simulated annealing. *Science*, 1983, **220** (4598): 671~680
- 16 Schletter J, Heine H, Ulmer A J, *et al.* Molecular mechanisms of endotoxin activity. *Arch Microbiol*, 1995, **164** (6): 383~389
- 17 Swantek J L, Cobb M H, Geppert T D. Jun N-terminal kinase/stress-activated protein kinase (JNK/SAPK) is required for lipopolysaccharide stimulation of tumor necrosis factor α (TNF- α) translation: glucocorticoids inhibit TNF- α translation by blocking JNK/SAPK. *Mol Cell Biol*, 1997, **17** (11): 6274~6282
- 18 Waite K J, Floyd Z E, Reily P A, *et al.* Interferon- γ -induced regulation of peroxisome proliferator-activated receptor γ and STATs in adipocytes. *J Biol Chem*, 2001, **276** (10): 7062~7068
- 19 Burton G R, Nagarajan R, Peterson C A, *et al.* Microarray analysis of differentiation-specific gene expression during 3T3-L1 adipogenesis. *Gene*, 2004, **329**: 167~185
- 20 MacDougald O A, Lane M D. Transcriptional regulation of gene expression during adipocyte differentiation. *Annu Rev Biochem*, 1995, **64**: 345~373
- 21 Heck S, Bender K, Kullmann N, *et al.* I κ B-independent downregulation of NF κ B activity by glucocorticoid receptor. *EMBO J*, 1997, **16** (15): 4698~4707
- 22 Sandelin A, Alkema W, Engstrom P, *et al.* JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res*, 2004, **32**: D91~D94
- 23 Smith L, Underhill P, Pritchard C, *et al.* Single primer amplification (SPA) of cDNA for microarray expression analysis. *Nucleic Acids Res*, 2003, **31** (3): e9
- 24 Ren B, Robert F, Wyrick J J, *et al.* Genome-wide location and function of DNA binding proteins. *Science*, 2000, **290** (5500): 2306~2309
- 25 van de Vijver M J, He Y D, van't Veer L J, *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*, 2002, **347** (25): 1999~2009
- 26 Guo Y, Chen Z, Zhang L, *et al.* Distinctive microRNA profiles relating to patient survival in esophageal squamous cell carcinoma. *Cancer Res*, 2008, **68** (1): 26~33
- 27 Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3 000 shared controls. *Nature*, 2007, **447** (7145): 661~678

Construction and Application of a Microarray for Profiling Mouse Transcription Factor Activities*

SUN Yi-Min^{1,2,3,4)**}, ZENG Ling-Qin^{1,2)}, ZHANG Yan^{1,2)}, WEI Li^{3,4)},
K R MITCHELSON^{1,2,3)}, ZHANG Liang^{3,4)}, CHENG Jing^{1,2,3,4)}

¹⁾ Medical Systems Biology Research Center, Tsinghua University, Beijing 100084, China;

²⁾ School of Medicine, Tsinghua University, Beijing 100084, China;

³⁾ National Engineering Research Center for Beijing Biochip Technology, Beijing 102206, China;

⁴⁾ CapitalBio Corporation, Beijing 102206, China)

Abstract Differential gene expression under particular conditions is tightly controlled at the transcriptional and post-transcriptional level. Transcription factors (TFs), key regulators at the transcriptional level, constitute a large and diverse group of double-strand DNA binding proteins which bind close to a target gene to activate or repress its transcription. Large-scale profiling of active transcription factors is essential to assist investigation into complicated gene regulatory networks as well as to help global pathway mapping. Here, a reproducible and reliable high-throughput mouse oligonucleotide array-based platform which can simultaneously explore the activities of nearly 200 different transcription factors is presented. The array comprises 240 synthetic probes which contain TF binding sequences based on 226 PSSMs provided by the TRANSFAC® database. The binding specificity between TF and its binding sequence was validated using antibody-based individual TF assay for AP1, HNF1, HNF4, NF1, NFκB, PPARγ and STAT1. Then purified NFκB was used to test the sensitivity and quantitative ability of the microarray. It could detect as little as 0.5 nmol/L of NFκB protein and maintain linear increase from 0.5 nmol/L to 20 nmol/L. Reproducibility was also explored using nuclear extracts from mouse macrophage RAW264.7 cells. The correlative coefficient between two independent experiments was 0.99 when using the identical nuclear protein. The reliability of the platform was then validated using two well studied cell models including LPS-stimulated MAPK pathway and IFNγ-stimulated JAK/STAT pathway. And finally, the TF profiling microarray was applied into the study of mouse preadipocyte 3T3-L1 differentiation. In addition to the well studied C/EBP family and PPAR family, two unknown transcription factors, SRF and NFAT, were also found changed in the process of adipocyte differentiation. Totally, a reliable microarray platform to profile the mouse transcription factor activities was successfully constructed.

Key words gene regulation, transcription factor, microarray, TF-ELISA

DOI: 10.3724/SP.J.1206.2008.00399

*This work was supported by a grant from Hi-Tech Research and Development Program of China (006AA020701).

**Corresponding author.

Tel: 86-10-80727291, E-mail: ymsun@capitalbio.com

Received: June 3, 2008 Accepted: June 30, 2008