

综述与专论

细胞筛选平台在人类功能基因组研究中的应用*

武春晓¹⁾ 石太平²⁾ 马大龙^{1,2)**}¹⁾北京大学人类疾病基因研究中心, 北京 100083; ²⁾国家人类基因组北方研究中心, 北京 100176)

摘要 人类基因组全序列的精细图已完成, 当前生命科学面临的重要任务就是如何将基因组序列信息转化为基因的功能信息, 了解生命活动的分子机理, 改善人类健康, 为生物技术发展提供动力. 在一系列功能基因组研究新技术中, 高通量 (high-throughput) 和高内涵 (high-content) 的细胞筛选技术平台已经显示出巨大潜力, 发挥着越来越重要的作用. 通过在体外培养的哺乳动物细胞中基因过表达或抑制基因表达, 分析所产生信号传导通路和 / 或细胞表型改变, 可以直接发现基因功能. 近年来一些技术上的进展, 使细胞筛选平台具有微量、自动、高效、高通量, 以及可以系统研究的特点, 已经成为功能基因组研究的核心方法之一. 近 2~3 年来已经出现一批成功应用细胞筛选平台进行大规模功能基因组研究的报道. 我国在这一领域的研究也开始起步, 将对我国生物技术的源头创新研究产生深远的影响.

关键词 功能基因组, 高通量, 高内涵, 细胞筛选技术

学科分类号 Q2

人类基因组计划在 2004 年完成了精细的全序列图, 许多模式生物基因组序列图也已完成, 当前生命科学面临的重要任务就是如何将基因组序列信息转化为基因的功能信息, 了解生命活动的分子机理, 改善人类健康, 为生物技术发展提供动力. 相应的研究模式也从以对单个基因水平的功能研究为重点的传统分子生物学, 转向以基因组功能通路 - 网络, 乃至细胞 - 组织 - 器官 - 个体等更高水平为目标的系统生物学研究模式^[1]. 研究基因编码的蛋白质在生理或病理状态下的功能变化、与其他生物大分子的相互作用以及对信号传导途径的影响, 有助于深入认识各种疾病发生发展的分子机制, 发现新的药物作用靶点. 而在功能基因组研究中, 需要整合一系列的新技术, 如生物芯片、生物信息学、蛋白质组学、结构基因组学、酵母双杂交、基因敲除、RNAi 等, 而其中一个重要的关键技术是细胞水平的功能筛选技术 (cell-based screening technology). 这一技术发挥了承上启下的作用, 可以看作是功能基因组研究的枢纽环节, 特别是新药开发中应用的高通量 (high-throughput) 和高内涵 (high-content) 细胞筛选技术将在功能基因组研究中发挥越来越重要的作用^[2]. 通过在体外培养的哺乳动物细胞中基因过表达或抑制基因表达, 分析所产生信号传导通路和 / 或细胞表型改变, 可以直接发

现基因功能. 近年来一些技术上的进展, 使细胞筛选平台具有微量、自动、高效、高通量、以及可以系统研究的特点, 已经成为功能基因组研究的核心方法. 本篇综述重点阐述这一领域的一些新进展.

1 人类基因的 ORFeome 计划

所谓人 ORFeome, 就是获得全部人类基因所有的转录体完整的开放读码框 ORF 的独立克隆, 测序验证, 并系统性构建到各种表达载体中 (真核 / 原核, 瞬时 / 稳定 / 调控表达, 融合蛋白 / 标签等). 建立 ORFeome, 是从基因组水平研究基因编码蛋白质的结构、功能、相互作用及细胞定位等系统生物学研究开展之前的一个必要的基础工作. 建立 ORFeome 工作充满挑战, 这是因为基因数目相对巨大, 人类基因组中蛋白质编码基因总数在 20 000 ~ 25 000 个之间. 并且由于在转录水平, 受可变剪接、替代启动子使用、RNA 编辑机制调节, 在 DNA 水平, 由于单核苷酸多态性 (SNP) 和基因

* 科技部“功能基因组与生物芯片”重大专项支持项目 (2002BA711A01-01).

** 通讯联系人.

Tel/Fax: 010-82801149, E-mail: madl@bjmu.edu.cn, dlma@263.net

收稿日期: 2005-07-28, 接受日期: 2005-08-31

重排作用,使得这些基因转录产物更加丰富多样。此外,低丰度基因,某些组织、发育特异性表达基因,以及部分长度大于 5 kb 基因的克隆工作难度相对更大。目前建立 ORFeome 的策略主要有两种:

1.1 全长 cDNA 文库随机测序法

建立人 cDNA 质粒文库,特别是全长 cDNA 文库,通过大规模随机测序获得全长人 cDNA 克隆。早期大规模 EST 测序计划产生极为丰富的 cDNA 克隆文库,如 IMAGE collection 包含数量巨大的人 EST 克隆,其中部分克隆包含完整的 ORF,可以从中筛选而有效获得部分完整的 ORF。但这远远不能覆盖整个 ORFeome。为获得人类基因全长 cDNA 克隆,多个国家启动实施了大规模的全长 cDNA 质粒文库克隆测序计划,其中包括美国 NIH 的 MGC 计划^[3](Mammalian Gene Collection),日本的 FLJ 计划^[4](Full-length Long Japan)和 KIAA 计划等。它们通过建立数百个人不同组织或细胞系的全长 cDNA 文库,或富集全长 cDNA 的质粒文库,进行大规模 5'端随机测序,并经过与人基因组序列比对分析验证,从中筛选可能的全长 cDNA 独立克隆,并将其全长测序验证,最终高通量地获得基因全长 cDNA 克隆。其中日本的 FLJ 计划采用了独创的 oligo-capping 的方法,可以保证高效率地获得全长人 cDNA,而且规模巨大,尤其引人注目。所有这些计划的成果已经收录于 H-Invitational Database(<http://www.h-invitational.jp>)。

需要指出,现在 MGC 的 ORF 克隆没有构建到表达载体上,而 FLJ 的全长 cDNA 克隆虽然构建在真核表达载体上,但不能高效灵活地克隆到其他表达载体上。而且,全长 cDNA 克隆 ORF 两端有长短不等的 5'和 3'非翻译区 UTR,以及天然终止码,许多基因 5' UTR 带有 non-canonical ORF,体外可以翻译表达出非天然蛋白质,5' UTR 的终止码会干扰 ORF 表达效率,天然终止码影响融合蛋白质载体构建。这些诸多潜在缺陷导致它们尚不能马上适用于进一步研究工作。

在上述工作基础上,如果再获得余下的人类基因全长 cDNA 克隆,其效率必然逐渐降低,需要有其他更有效的方法和技术替代补充,或者选择新的特殊组织器官 cDNA 文库。

1.2 生物信息学指导的系统性 RT-PCR 扩增克隆靶基因的方法

随着人类基因组数据库的不断更新完善,目前已经可以根据其中的 Refseq 数据库,并结合 EST

比对拼接等生物信息学方法,准确预测靶基因全长 ORF,设计引物,通过高通量 PCR 扩增,靶向性克隆的方法获得人类基因全长 ORF。目前在国际上很多实验室的 cDNA 克隆即采用这一技术获得新的人类基因全长 ORF。我国国家人类基因组北方研究中心在科技部“功能基因组与生物芯片”重大专项的支持下,采用此种方法成功扩增了超过 700 个全长人基因 ORF 及其剪接异构体,并将其克隆到 pcDNA3.1 真核表达载体上,完成测序验证,其中大部分基因为功能未知基因。

由此可见,生物信息学指导的系统性 RT-PCR 扩增克隆靶基因的方法是 ORFeome 建立后期一个重要的补充方法。它可以选择性获得目的基因 ORF 序列克隆,有效排除 5'和 3'非翻译区 UTR 对后续蛋白质表达干扰,这种方法在发现新的剪切异构体方面也有独到的优势。

2 人类基因 RNAi 表达文库建立

RNA 干扰(RNA interference, RNAi)是一种保护宿主细胞免受外来 DNA 入侵的防御机制。经典理论认为,它通过一定机制将双链 RNA(dsRNA)剪切成短的干扰 RNA (small interfering RNA, siRNA),siRNA 介导特异性同源 RNA 的降解。在哺乳动物细胞,通过载体表达短的发夹样 RNA (short hairpin RNA, shRNA)来发挥类似 siRNA 机制作用,稳定地抑制靶基因的表达(knockdown),使基因沉默(gene silencing),由此可以通过 loss-of-function 方法研究基因功能。

在绝大多数人类基因序列已知条件下,理论上每个基因都可以用 RNAi 技术使基因沉默。为了从基因组水平使用 RNAi 技术为基础的功能筛选研究,需要先构建大规模人 RNAi 表达文库。2004 年 Paddison^[5]和 Berns^[6]两个研究小组各自独立构建了逆转录病毒载体 shRNA 表达文库。Paddison 小组构建的 RNAi 文库内含 28 000 个独立测序验证的克隆,针对 9 610 个人类基因和 5 563 个鼠基因,并且可以通过细菌交配辅助的 DNA 重组(mating-assisted genetically integrated cloning, MAGIC)方法很便捷地将 shRNA 转移到其他商品化的表达载体上。Berns 小组构建的 RNAi 文库含 23 742 个独立克隆,针对 7 914 个人类基因,经检测它们平均对 70%的靶基因有 70%的抑制效果。为提高筛选效率,两个文库都采用了条形码标记系统。我国的国家基因组北方研究中心梁子材小组利

用新的相向表达载体, 构建了大于 5×10^7 容量的 siRNA 文库, 用于细胞表型筛选, 成功获得了一批与细胞增殖相关的基因^[7]. 可以预计, 针对全部人类基因的 RNAi 表达文库将在不久后完成, 在细胞筛选技术中发挥越来越重要的作用.

3 报告基因检测技术

以各种外源性荧光蛋白或荧光素酶为代表的报告基因在细胞筛选中有广泛的用途. 一是在细胞筛选大多需要将待测基因 DNA 转染到细胞中, 而现在的转染效率不可能达到 100%, 因此报告基因可以标记出被转染的细胞, 或作为内对照, 对检测结果作正态化处理, 减少实验误差; 二是报告基因编码的蛋白质产物活性变化, 可以用于灵敏地反映细胞内某种特定生化或信号转导通路的改变等. 如报告蛋白绿色荧光蛋白 GFP 及其各种衍生物应用于检测细胞内蛋白质稳定性^[8]、蛋白质间相互作用、各细胞器 pH 变化^[9]、凋亡诱导^[10]、融合蛋白细胞内定位^[11]、转录因子活化^[12]等.

高通量细胞筛选要求快速、灵敏、简便、可靠地读取实验检测结果(read-out). 荧光素酶报告基因在此领域有得天独厚的优势. 它们的检测灵敏度高达 fg 级, 线性范围在 7~8 个数量级, 反应后只需几秒钟即可读取检测结果, 此外, 荧光素酶还具有没有细胞内源性酶活性干扰、细胞毒性小等特点. 新近开发的双荧光素酶检测系统即是充分发挥其优势的代表: 它们组合了萤火虫(*Photinus pyralis*)和海肾(*Renilla reniformis*)两种荧光素酶, 在单管中检测两个荧光素酶报告基因, 这两个荧光素酶报告基因具有兼容的化学特性、操作条件、速度、灵敏度、线性范围和仪器需求. 使用标准的荧光发光计可在 30 s 内手动完成检测. 而且, 使用经特别设计的被动裂解液, 可快速、高效地将培养在多孔板中的细胞裂解, 不必要对每一个样品分别进行单独操作, 并使萤火虫和海肾荧光素酶具有最佳检测灵敏度和稳定性. 此技术解决了过去采用氯霉素乙酰转移酶(CAT)、 β -半乳糖苷酶(β -Gal)或葡萄糖醛糖苷酶(GUS)等作为辅助报告基因时的缺陷——灵敏度低、线性范围窄、受内源性酶活性干扰、操作时间长、两个报告基因不兼容、需要分别检测等, 使其尤其适用于高通量细胞筛选. 国际上多个研究组, 以及我国的国家人类基因组北方研究中心均采用这一检测体系建立了多种细胞筛选平台.

4 高内涵细胞表型筛选技术

如果一次筛选可以读取多个参数而非仅仅一个参数, 则筛选的效率将大大提高. 这就是起源于药物筛选的高内涵筛选技术(HCS)的目标. 高内涵技术建立在自动高解析度光学/荧光显微图像获取技术(硬件)和后续的自动数字图像分析处理储存技术(软件)整合基础之上. 这种技术尤其适用于各种细胞表型的筛选, 如增殖、凋亡、分化、衰老、细胞周期、黏附等细胞表型的变化等. 随着显微镜分辨率和图片解析度的提高, 各种细胞染色技术和荧光蛋白等报告基因技术也应用到高内涵筛选技术当中, 用于观测各种细胞器的变化, 并且可以自动实时动态监测如钙流、受体内化、核转位等动态变化. 细胞图片(动态/静态)自身蕴涵丰富的信息量, 并且可以永久保存, 可以为以后各种图像分析提供原始数据, 这些都是其他筛选技术无法比拟的. 高内涵细胞表型筛选技术潜力巨大, 是未来人类功能基因组细胞筛选技术的主要发展方向.

5 细胞芯片技术

功能基因组研究是要发现每个基因的功能, 因此细胞筛选也应采用逐个基因 gene-by-gene 的研究方式, 细胞芯片技术提供了这样的高通量平台. 细胞芯片(cell microarray)技术最初由 Ziauddin 等^[13]在 2001 年发明, 他们将表达质粒、凝胶、和转染试剂混合物以序列点阵形式点样到经过包被的玻片上, 再把细胞以爬片方法覆盖在点样的玻片上. 细胞可以摄取并表达点在玻片上的质粒, 结果一个质粒 DNA 点仅仅转染覆盖在它上面的细胞, 使得被转染的细胞和 DNA 序列点阵以同样形式排列在玻片上, 一个点上细胞的表型变化与此位点的质粒对应. 一个 DNA 点的直径大约为 150 μm , 可以转染约 30~80 个细胞, 这样一个芯片能够检测几百个基因. 细胞芯片的转染方式与常规的转染方式不同(图 1): 它是首先将待测质粒和转染试剂混合物固定在玻片特定位置上, 然后再铺细胞, 只有生长在质粒/转染试剂混合物位点上的细胞被转染. 而常规的方法是先铺细胞, 再转染, 细胞被随机转染. 由于与常规转染方法的顺序相反, 细胞芯片的转染技术又称为反式转染技术(reverse transfection).

细胞芯片技术还可以与 RNAi 技术完全兼容^[14]. 筛选结果(如细胞形态等表型变化)可以在显微镜下

观察. 这一技术与高解析度显微镜自动化数字影像分析系统和信息管理系统结合, 可以离线(off-line)自动分析亚细胞水平 - 单个细胞水平 - 单个照片水平 - 点阵水平的各种细胞表型变化, 一次筛选可以获得多种结果, 适用于高内涵筛选系统^[15], 具有强大的功能和广泛的用途, 是目前最具发展潜力的细胞筛选技术. 此外, 细胞芯片技术对细胞表面受体蛋白研究似乎有独到的优势. 细胞芯片技术的局限是它只能研究贴壁细胞. 并且由于细胞之间没有分隔, 不能进行分泌蛋白质和培养上清的多种酶学检测. 细胞芯片技术适合于自动化操作, 其中自动化单 cDNA 研究 RISC^[16] (robotic of single cDNA

investigation) 技术就是采用全自动的机械臂进行单个质粒小量制备和微孔细胞转染技术. 质粒转化的细菌在 96 孔板中扩增, 机械臂自动小量提取纯化质粒, 再将质粒转染 96 或 384 孔板培养的哺乳动物细胞. 由于一个孔中只有一种质粒被提取, 细胞只被一种质粒转染, 96 孔板中其他质粒 / 细胞隔离, 因此它适用于逐个基因 gene-by-gene 的研究方式, 可以发现传统的大批量基因混合转染方法遗漏的功能相对较弱的基因, 还可用于分泌蛋白质和酶学检测方面. 在结果检测方面, 也出现了大批微量、灵敏、自动、快速、可靠检测方法, 如荧光素酶报告基因的检测等.

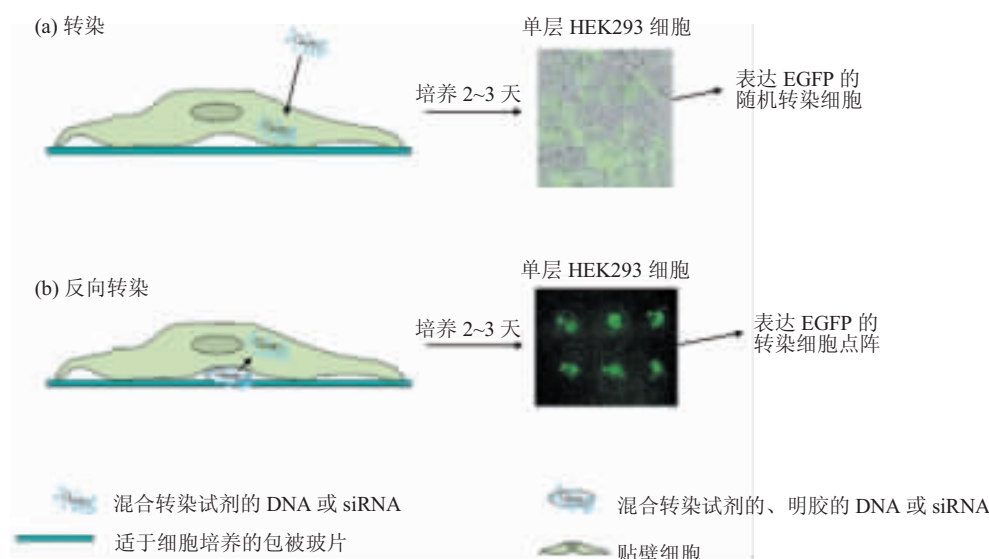


Fig. 1 Principle of reverse transfection^[17]

图 1 转染与反向转染^[17]

6 细胞筛选相关技术的应用

在上述技术进展基础上, 近 2~3 年来已经集中出现一批成功应用细胞筛选平台进行大规模功能基因组研究的报道.

6.1 基因过表达——信号传导通路细胞筛选

Novartis 生物医学研究所利用 MGC 人全长 cDNA 文库构建了含有 20 704 个独立序列已知的质粒, 代表约 12 000 个基因的真核表达质粒文库, 约占人类基因总数的 50%. 他们通过 RISC^[16] 技术, 以逐个基因普筛方式, 将这些质粒与信号传导通路

荧光素酶报告质粒共转到 384 孔板培养的细胞中, 系统性地发现了数以百计的已知或未知的参与几个重要疾病相关信号转导通路的调节分子, 如 AP-1、CRE、P53、WNT、EPO 等, 并对其中作用最强的新调节基因进行了进一步的实验验证^[18-22]. 以其中 AP-1/P53/EPO 信号转导通路筛选为例, 他们总共获得了 129 个 AP-1 阳性克隆, 阳性率为 0.5%, 其中包括许多已知 AP-1 通路调节基因, 如 c-FOS、c-JUN、GADD45、TRAF6、erbB2、FOSL1、ATF3、MAP3K11、RICK 以及 PKC 等, 其激活 AP-1 作用水平都排在前 0.5%, 证明他们筛选方法

的可靠性,更有意义的是他们在各个通路都发现了几十个新的调控基因. p53 和 EPO 的筛选结果阳性率也比较类似. 日本研究小组从几个人 cDNA 文库中随机选取了 150 000 个独立的表达克隆,采用类似方法筛选了 NF κ B 通路和 MAPK 通路的调节分子,他们一共找到 83 个可以激活 NF κ B 报告基因的阳性克隆,其中 28 个是已知的 NF κ B 通路调节基因,其余是未知基因和已知基因的新功能. MAPK 通路筛选也发现了 54 个激活基因,其中 27 个与 NF κ B 通路调节基因重叠,提示它们可能参与两个通路之间的交叉作用^[23]. 这些研究证明,我们可以通过高通量细胞筛选高效率地从功能基因组水平系统性地阐释人信号传导通路机制,这种后基因组时代研究模式,其效率和系统性是以往单基因研究为重点的传统分子生物学研究模式所无法企及的. 当然,高通量筛选提供出候选基因之后,后期必须要经过二次筛选和传统研究模式的精细深入的研究验证.

我国的国家人类基因组北方研究中心也在近年利用双荧光素酶报告基因建立了相似的 NF κ B、AP-1、NFAT、CRE、STAT 等十多个筛选通路,并且以此为基础建立了对外服务平台,为多个课题组提供技术服务,对国内实验室的基因如 LAPTM4B、PDCD10、VCY2IP1 等基因发现了有价值的功能线索. 其中部分通路分别对自己克隆的 700 条新基因进行了筛选,至少发现 20 条基因经重复实验验证对上述通路有调节作用.

6.2 基因过表达-细胞表型变化细胞筛选

Michiels 小组^[24]构建逆转录病毒表达载体上的 13 000 多个独立 cDNA 质粒,通过细胞筛选检测了其中可以诱导成骨细胞分化、上皮细胞形态改变等多种表型变化的基因. 上海顾健人小组将 29 910 和 22 926 多个人独立 (非全长) cDNA 表达质粒克隆分别转染了人肝癌细胞系 SMMC7721 和鼠 NIH3T3 细胞系,通过克隆形成和结节形成等细胞增殖有关的表型变化寻找与肿瘤发生发展相关的基因,他们总共找到 3 806 个与细胞增殖有关的基因,其中有 2 836 个已知基因,372 个未知基因,598 个结构未知的 Unigenes^[25]. 国家人类基因组北方研究中心利用基因过表达-报告基因系统-细胞表型变化-荧光染料观察的细胞筛选体系,对自己克隆的近 600 条新基因进行了功能筛选(图 2),并利用传统方法流式细胞术 (FACS) 和 MTT 法加以验证,发现了 14 条新基因可以诱导细胞凋亡,8

条新基因与细胞增殖相关.

6.3 基因过表达-多通路协同筛选

最近 Cecil 等^[26]为寻找具有糖尿病相关的分泌蛋白质,用生物信息学方法从人 EST 数据库中筛选出 8 000 个独立的、预测可能编码分泌蛋白的人 cDNA 表达质粒,收集转染细胞上清,根据糖摄取、糖原生成、胰岛素信号传导等糖代谢关键通路机制建立了大于 100 个高通量细胞筛选体系,通过协同多重细胞筛选,结果发现了一个已知的分泌蛋白质 BMP-9,并经动物试验验证了它在调节糖稳态的新的生理功能. Michele 等^[27]应用同一文库筛选了其中可能有 T 细胞免疫调节功能的基因,他们从 3 个方面建立 3 组共 50 多个相互协同的筛选体系: a. 原代 T 细胞活化过程中参与或被调控的细胞表面标记检测; b. T 细胞分泌的各种细胞因子的检测; c. T 细胞活化过程中参与调控基因表达信号转导通路的检测. 从总共获得的 400 000 个筛选结果中找到 68 个阳性结果,并从中发现一个新的具有免疫调节功能的分泌蛋白 TIP, TIP 可以刺激原代人和鼠 T 细胞表达 IFN- γ 、TNF- α 和 IL-10,动物体内试验显示 TIP 可以在小鼠急性移植植物抗宿主反应疾病模型中起保护作用. 这两个报道突出体现了细胞筛选技术在寻找药物靶点方面的巨大潜力: 将人类 ORFeome 和 RNAi 表达文库,通过系统地建立大量的高通量细胞筛选通路反复筛查,可以快速准确地发现并优选新的药物靶点.

6.4 基因敲减 (gene-knockdown) -RNAi 细胞筛选

RNAi 介导的 loss-of-function 细胞筛选技术在低等生物如线虫和果蝇等已成功应用. 在建立了人 RNAi 表达文库之后,Westbrook 和 Kolfshoten 分别对影响细胞表型的基因进行了筛选,在 2005 年 6 月的 Cell 杂志报道了新发现的 2 个肿瘤抑制基因 REST 和 PITX1^[28, 29],初步证实 RNAi 在人细胞筛选中的有效应用. 此外,如前所述, RNAi 技术与细胞芯片技术兼容,与自动数据影像处理系统结合,在高通量高内涵细胞筛选方面有巨大优势. Kittler^[30]研究组采用了一种更简便有效的 esiRNA 技术: 先利用已测序的人 cDNA 质粒克隆逆转录制备长的双链 RNA (long dsRNA),再用核糖核酸内切酶 (endoribonuclease) 将其剪切成短的双链 RNA,称为 esiRNA. 使用 esiRNA 可以确保有效并特异地使靶基因 mRNA 沉默,从而无须像 siRNA 那样验证沉默子的有效性. 他们用此技术制备了针对 15 497 个人类基因的 esiRNA 文库,选取其中的

5 305 个 esiRNA，采用初步高通量细胞活性筛选，随后高内涵显微照相细胞筛选的两步筛选法，在初筛发现的 275 个影响细胞活性的基因当中找到 37 个与 HeLa 细胞分裂有关的基因，其中 10 个基因

可以导致细胞有丝分裂停滞的表型异常，使细胞有丝分裂停滞在双细胞核阶段. 而北京大学人类疾病基因研究中心此前克隆的功能未知基因 CKLFSF4 就是这 10 个基因之一^[31].

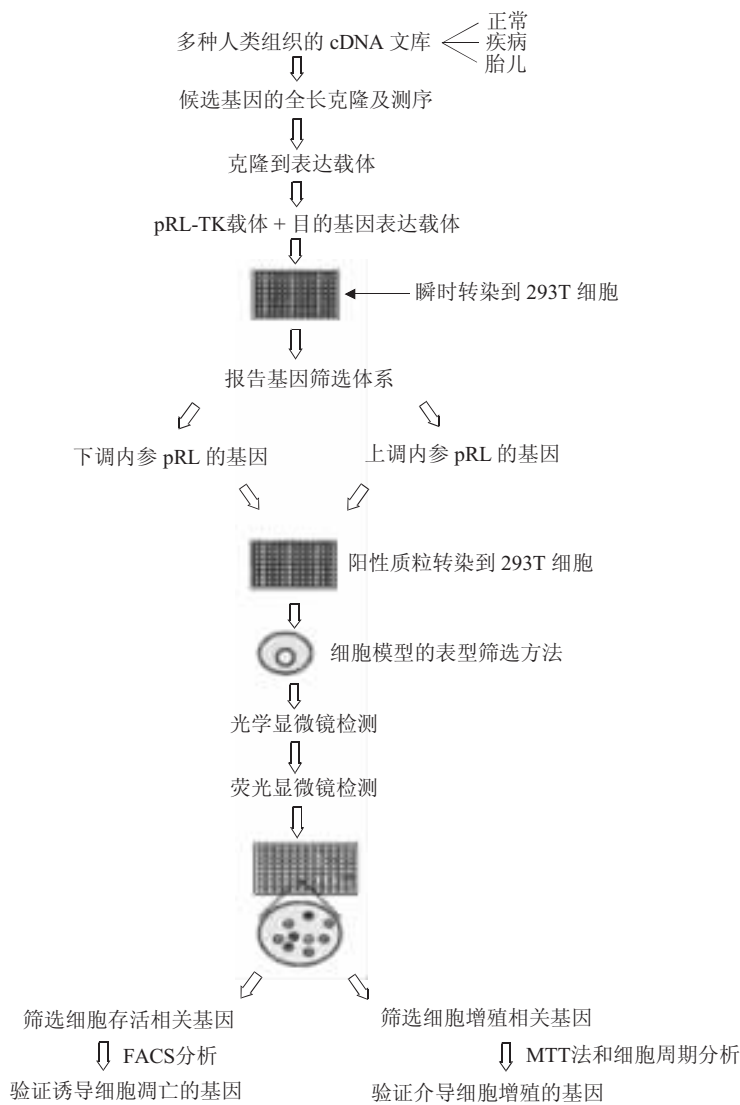


Fig. 2 Cell-based screening system integrating of gene over-expression, reporter gene system, cellular phenotype assay and fluoressence stain

图 2 基因过表达-报告基因系统-细胞表型变化-荧光染料观察的细胞筛选体系

7 展 望

以上这些报告显示，细胞筛选在人类功能基因组研究中的巨大效用和潜力. 由于哺乳动物细胞种类繁多，可以建立的各种分子通路和细胞表型模型筛选体系更是灵活多样，因此我们可以将人 ORFeome 和 RNAi 表达文库通过各种筛选体系反

复进行功能研究，发现各个基因在各种条件下在各种分子通路或网络中的作用和联系，由此将获得丰富的数据，为数据库中基因提供详尽功能注释. 结合其他功能基因组研究方法获得的数据，必将为我们精确选择药物靶标，全面认识自身生理和疾病规律奠定坚实基础.

参 考 文 献

- 1 Butcher E C, Berg E L, Kunkel E J. Systems biology in drug discovery. *Nat Biotechnol*, 2004, **22** (10): 1253~1259
- 2 Kramer R, Cohen D. Functional genomics to new drug targets. *Nature Rev Drug Discov*, 2004, **3** (11): 965~972
- 3 Mammalian Gene Collection (MGC) Program Team, Strausberg R L, Feingold E A, *et al.* Generation and initial analysis of more than 15 000 full-length human and mouse cDNA sequence. *Proc Natl Acad Sci USA*, 2002, **99** (26): 16899~16903
- 4 Ota T, Suzuki Y, Nishikawa T, *et al.* Complete sequencing and characterization of 21 243 full-length human cDNAs. *Nature Genetics*, 2004, **36** (1): 40~45
- 5 Paddison P J, Silva J M, Conklin D S, *et al.* A resource for large-scale RNA-interference-based screens in mammalian cells. *Nature*, 2004, **428** (6981): 427~431
- 6 Berns K, Hijmans E M, Mullenders J, *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*, 2004, **428** (6981): 431~434
- 7 Chen M, Zhang L, Zhang H Y, *et al.* A universal plasmid library encoding all permutations of small interfering RNA. *Proc Natl Acad Sci USA*, 2005, **102** (7): 2356~2361
- 8 Dantuma N P, Lindsten K, Glas R, *et al.* Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nature Biotechnol*, 2000, **18** (5): 538~543
- 9 Xu X, Gerard A L, Huang B C, *et al.* Detection of programmed cell death using fluorescence energy transfer. *Nucleic Acids Res*, 1998, **26** (8): 2034~2035
- 10 Llopis J, McCaffery J M, Miyawaki A, *et al.* Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc Natl Acad Sci USA*, 1998, **95**(12): 6803~6808
- 11 Simpson J C, Wellenreuther R, Poustka A, *et al.* Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep*, 2000, **1** (3): 287~292
- 12 Misawa K, Nosaka T, Morita S, *et al.* A method to identify cDNAs based on localization of green fluorescent protein fusion products. *Proc Natl Acad Sci USA*, 2000, **97** (7): 3062~3066
- 13 Ziauddin J, Sabatini D M. Microarrays of cells expressing defined cDNAs. *Nature*, 2001, **411** (6833): 107~110
- 14 Conrad C, Erfle H, Wornat P, *et al.* Automatic identification of subcellular phenotypes on human cell arrays. *Genome Research*, 2004, **14** (6): 1130~1136
- 15 Mousses S, Caplen N J, Cornelison R, *et al.* RNAi microarray analysis in mammalian cultured cells. *Genome Res*, 2003, **13**(10): 2341~2347
- 16 Stefan G. The art and design of genetic screens: Mammalian culture cells. *Nature Rev Genetics*, 2004, **5** (3): 179~189
- 17 Vanhecke D, Janitz M. Functional genomics using high-throughput RNA interference. *Drug Discovery Today: Targets*, 2005, **10**(3): 205~212
- 18 Jun Liu, Bang A G, Kintner C, *et al.* Identification of the Wnt signaling activator leucine-rich repeat in Flightless interaction protein 2 by a genome-wide functional analysis. *Proc Natl Acad Sci USA*, 2005, **102** (6): 1927~1932
- 19 Conkright M D, Canettieri G, Screaton R, *et al.* TORCs: Transducers of regulated CREB activity. *Mol Cell*, 2003, **12** (2): 413~423
- 20 Chanda S K, White S, Orth AP, *et al.* Genome-scale functional profiling of the mammalian AP-1 signaling pathways. *Proc Natl Acad Sci USA*, 2003, **100** (21): 12153~12158
- 21 Iourgenko V, Zhang W J, Mickanin C, *et al.* Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc Natl Acad Sci USA*, 2003, **100** (21): 12147~12152
- 22 Huang Q, Raya A, DeJesus P, *et al.* Identification of p53 regulators by genome-wide functional analysis. *Proc Natl Acad Sci USA*, 2004, **101** (10): 3456~3461
- 23 Matsuda A, Suzuki Y, Honda G, *et al.* Large-scale identification and characterization of human genes that activate NFκB and MAPK signaling pathways. *Oncogene*, 2003, **22** (21): 3307~3318
- 24 Michiels F, van Es H, van Rompaey L, *et al.* Arrayed adenoviral expression libraries for functional screening. *Nature Biotechnol*, 2002, **20** (11): 1154~1157
- 25 Wan D, Gong Y, Qin W, *et al.* Large-scale cDNA transfection screening for genes related to cancer development and progression. *Proc Natl Acad Sci USA*, 2004, **101** (44): 15724~15729
- 26 Chen C, Grzegorzewski K J, Barash S, *et al.* An integrated functional genomics screening program reveals a role for BMP-9 in glucose homeostasis. *Nature Biotechnol*, 2003, **21** (3): 294~301
- 27 Fiscella M, Perry J W, Teng B, *et al.* TIP, a T-cell factor identified using high-throughput screening increases survival in a graft-versus-host disease model. *Nature Biotechnol*, 2003, **21**(3): 302~307
- 28 Westbrook T F, Martin E S, Schlabach M R, *et al.* A genetic screen for candidate tumor suppressors identifies REST. *Cell*, 2005, **121**(6): 837~848
- 29 Kolfschoten IGM, van Leeuwen B, Berns K, *et al.* A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell*, 2005, **121** (6): 849~858
- 30 Kittler R, Putz G, Pelletier L, *et al.* An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature*, 2004, **432** (7020): 1036~1040
- 31 Han W L, Ding P G, Xu M X, *et al.* Identification of eight genes encoding chemokine-like factor super family members 1~8 (CKLFSF1~8) by in silico cloning and experimental validation. *Genomics*, 2003, **81** (6): 609~617

Applications of Cell-based Screening Assays in Human Functional Genomics Research*

WU Chun-Xiao¹⁾, SHI Tai-Ping²⁾, MA Da-Long^{1,2)**}

¹⁾Human Disease Genomics Research Center, Peking University, Beijing 100083, China;

²⁾Chinese National Human Genome Center, Beijing 100176, China)

Abstract After the full sequence map of human genome has been completed, the great challenges for the life scientists is how to transform the genomic sequence into gene function information, to elucidate the molecular mechanism of biological process, to improve human health and supply impetus powers to the progression of biotechnology. Among the series of novel functional genomic technologies, high-throughput and high-content cell-based screen platforms have showed enormous potential and will play more and more important roles in human functional genomics research. By overexpression or knock-down of genes in cultured mammalian cells *in vitro* and analysing the consequent changes in signal transduction pathways and/or cell phenotypes, gene functions can be identified directly. With the recent technique progresses, the cell-based screening assays can possess the characteristics of miniaturization, automation, high efficiency, high-throughput and feasibility, and have become one of the pivotal methods in functional genomics. In the last 2~3 years, the successful applications of cell-based screen technology in large-scale functional genomics research have been reported in the literatures. In China, investigation in this field has been set out, and it will have profound influences on boosting the R & D of Chinese biotechnology.

Key words functional genome, high throughput screening, high content screening, cell-based screening technology

*This work was supported by a grant from The National Key Technologies R&D Program "Functional Genome and Biochips" of Chinese Ministry of Science and Technology (2002BA711A01-01).

**Corresponding author . Tel/Fax: 86-10-82801149, E-mail: madl@bjmu.edu.cn, dlma@263.net

Received: July 28, 2005 Accepted: August 31, 2005