

基因芯片与高通量 DNA 测序技术前景分析

滕晓坤^①, 肖华胜^{①②*}

① 生物芯片上海国家工程研究中心, 上海 201203;

② 国家人类基因组南方研究中心, 上海市-科技部共建疾病与健康基因组学重点实验室, 上海 201203

* 联系人, E-mail: huasheng_xiao@shbiochip.com

收稿日期: 2008-10-03; 接受日期: 2008-10-08

国家高技术研究发展计划(批准号: 2006AA020704)和上海市科学技术委员会(批准号: 05DZ22201)资助项目

摘要 基因芯片与第二代 DNA 测序是两种重要的高通量基因组学研究技术, 对于揭示基因组的结构与功能已经并正在发挥重要的推动作用. 基因芯片技术建立了 10 多年, 技术日渐成熟, 在功能基因组、系统生物学、药物基因组的研究中已经得到了广泛的应用. 2003 年, 454 公司首先建立了高通量的第二代测序技术, 其他公司相继推出了 Solexa 和 Solid 测序技术. 虽然第二代测序技术建立的时间不长, 但发展非常快, 已经应用于基因组, 包括测序和表观基因组学以及功能基因组学研究的许多方面. 本文简要综述了基因芯片和第二代测序技术及其应用进展, 并分析了这两种高通量基因组学技术的前景.

关键词

基因芯片
第二代高通量测序技术
基因组

20 世纪 80 年代启动的由多个国家参加的人类基因组计划, 被称为是继曼哈顿原子计划、阿波罗登月计划之后的第三大科学计划, 这个计划的完成对人类认识自身, 提高健康水平, 推动生命科学、医学、生物技术、制药业、农业等的发展具有极其重要的意义. 2000 年 6 月 26 日由美国、英国、法国、德国、日本和中国科学家同时向世界宣布人类基因组工作草图已基本完成, 2002 年 2 月又公布了“精细图”. 这是基因组学研究领域的一个里程碑式的工作. 随着人类基因组计划的实施, 也推动了模式生物基因组的测序, 在这期间还完成了大鼠、小鼠、水稻等物种的测序. 众多物种的基因组序列被解开, 大量的基因通过基因组测序被识别, 接下来的工作是要对这些基因的结构、功能、表达和分布进行深入的研究, 急需高通量、大规模的分析手段和方法来解决这些需求. 20 世纪 90 年代建立起来的基因芯片技术和最近发展起来的第二代 DNA 测序技术是高通量研究基因的结构

和功能的两种比较重要的技术, 推动了功能基因组和系统生物学研究的发展. 本文简要介绍了这两种技术和应用, 并对这两种技术在生命科学中的应用前景进行了分析.

1 基因芯片技术及其发展

生物芯片的概念由 Fodor 等人在 1991 年提出^[1], 是指能够快速并行处理多个样品并对其所包含的各种生物信息进行解剖的微型器件, 它的加工运用了微电子工业和微机电系统加工中所采用的一些方法, 只是由于其所处理和分析的对象是生物样品, 所以叫生物芯片(Biochip).

在生物芯片技术中, 基因芯片技术建立最早, 也最为成熟. 基因芯片, 又称 DNA 微阵列(DNA microarray), 是把大量已知序列探针集成在同一个基片(如玻片、膜)上^[2-7], 经过标记的若干靶核苷酸序列与芯片特定定位点上的探针杂交, 通过检测杂交信号, 对生

物细胞或组织中大量的基因信息进行分析。其突出特点在于高度并行性、多样性、微型化和自动化。高度的并行性不仅可以大大提高实验的进程,而且有利于DNA芯片技术所展示图谱的快速对照和阅读。多样性可以在单个芯片中同时进行样品的多参数分析,从而避免因不同实验条件产生的误差,大大提高分析的精确性。微型化可以减少试剂用量和减小反应液体积,降低实验费用。高度自动化则可以降低制造芯片的成本和保证芯片的制造质量。1995年*Science*杂志首次报道了Schena等人^[8]用DNA微阵列技术并行检测拟南芥多个基因的表达水平。1994年第一张商业化基因芯片由Affymetrix公司推出。

基因芯片的制备方法主要有两种:原位合成法和点样法。从目前应用的情况来看,原位合成的基因芯片的密度高,重复性好,制备过程中的质量控制比较容易,但是成本较高。而点样技术主要应用在部分没有商业化芯片的物种的基因芯片的制备,制备的成本较低。原位合成的芯片是今后的一个发展和应用方向。

基因芯片的原位合成法是基于组合化学的合成原理^[9],通过一组定位模板来决定基片表面上不同化学单体的偶联位点和次序,把腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)、胸腺嘧啶(T)四种不同碱基的核苷酸按不同次序化学偶联在相应的位点,原位合成序列不同的寡核苷酸探针,形成DNA芯片。这一技术是由Affymetrix公司的Fodor及其同事最先发明的^[1],他们使用含光敏化学保护基的DNA合成试剂,用光脱保护法直接在基片上合成寡核苷酸探针,即光导向原位合成法。该方法的优点在于精确性高,缺点是制造光掩蔽剂既费时又昂贵。

美国Wisconsin大学的Singh-Easson等人利用计算机设计的虚拟掩模来制备DNA芯片,他们使用一种包括480000个微小铝镜阵列的光学数字微镜设备,通过计算机精确地控制光线在微镜上的反射方向,将紫外光束直接照射到玻璃基片上,从而在基片的特定区域上选择性地脱除保护基。采用虚拟掩模技术不用耗资费时去加工光掩模,降低了基因芯片的制备时间和成本^[10]。Nimblegen公司利用这一技术进行了高密度芯片的商业化生产。

国内的东南大学生物芯片研究小组,结合组合

化学合成与软光刻微印刷技术的原理,用一套组合的分子印章印刷的方式——压印的方式分配DNA合成试剂到基片的特定位点,多次压印直至指定的探针微阵列形成而制得DNA芯片。该技术利用预先制作的分子印章实现空间寻址原位合成,设计适当的合成顺序与设计凹凸位点不同的印章即可在基片上原位合成出位置和序列预定的寡核苷酸或寡肽阵列。东南大学利用软光刻原位合成法,成功得到特征位点尺寸为100 μm与30 μm的寡核苷酸微阵列^[11]。

点样法制备基因芯片首先按常规方法制备cDNA(或寡核苷酸)探针库,然后通过特殊的微喷头,分别把不同的探针溶液按照预先排定的顺序逐点在片基表面,并通过物理和化学作用使探针固定于基片表面。探针片段除了可使用寡聚核苷酸探针,也可使用较长的基因片段以及核酸类似物探针(如PNA等)。点样法的优越性在于可以充分利用原有的合成寡核苷酸或cDNA探针库,探针的长度可以任意选择,灵活性大,可根据需要自行制备。最早的点样装置是由Stanford大学的Shalon和Brown等研究人员自己搭建研制的,美国的Cartesian Technologies, TeleChem, Biodot, PerkinElmer, GE Healthcare等公司,英国的Genetix公司,中国的博奥生物公司等都有商业成型的点样机。在基因芯片制备方面除了Affymetrix, Agilent和Nimblegen等几个公司外,大多公司普遍采用点样技术制作DNA芯片。

20世纪90年代末,中国政府和科学家开始启动生物芯片的研究与开发。我国最早的一张DNA微阵列膜,是在中国科学院的人类基因组项目的支持下,利用了国家人类基因组南方研究中心所积累的大量人类cDNA克隆模板点制成功的。这张“基因芯片”成功地运用于首例肝癌及癌旁组织的比较转录组研究,其成果于2001年发表于*PNAS*^[12]。“九五”末期,国家高技术研究发展(863)计划又连续拨出专款,支持北京、上海和南京若干个课题组开始实施包括基因芯片的“生物芯片”研制工作,成立上海的“华冠”公司专门从事诊断用生物芯片的研发。国内若干民营企业也开始投入对生物芯片的研发。比较突出的包括广州“益生堂”、上海“博星”和浙江湖州的“江南生物”等企业。

从 2000 年开始, 政府陆续投入近 5 亿元人民币, 建立了北京、上海两个生物芯片国家工程研究中心, 国家高技术研究发展(863)计划在“十五”和“十一五”期间对生物芯片技术的研究与产品的开发进行了重点资助, 已经取得了一批重要的研究成果和开发了一批生物芯片产品. 如生物芯片北京国家工程研究中心, 开发了基因芯片的扫描仪, 基因芯片的点样仪, 杂交仪以及样品处理的仪器, 形成了基因芯片的制备和应用的一套仪器设备. 同时, 利用自主研发的基因芯片技术平台, 开发了几十种基因表达谱芯片、miRNA 芯片、CGH 芯片、DNA 甲基化芯片以及应用于食品安全检测, 临床诊断的基因芯片. 生物芯片上海国家工程研究中心, 建立了基于点样的基因芯片技术平台, 开发了 20 余种基因表达谱芯片和 Pathway 芯片, 主要的物种有人、大鼠、小鼠以及微生物全基因组芯片. 同时还引进了国际上用得最多, 技术也最为成熟的 Affymetrix, Agilent, Illumina 和 Nimblegen 的系统和体系, 建立了严格的质量控制和应用体系, 提供基因芯片的全面解决方案. 承担了国家重点基础研究发展计划(973)、国家高技术研究发展计划(863)和上海市科委的项目, 为国家重点基础研究发展计划(973)、国家高技术研究发展计划(863)、自然科学基金和地方政府的项目提供了技术支持和技术服务, 为完成这些项目提供了技术保障. 通过服务和合作在 *PNAS*^[13,14], *Genomics*^[15,16], *BBRC*^[17,18], *JBC*^[19], *FEBS letter*^[20,21] 等杂志上发表了论文.

2 基因芯片技术的应用和发展趋势

随着基因芯片技术的日渐成熟, 在功能基因组、疾病基因组、系统生物学等领域中得到了广泛的应用, 已经发表了上万篇研究论文, 每年发表的论文呈现增长的趋势.

芯片制备技术极大地推进了生物芯片的发展, 从实验室手工或机械点制芯片到工业化原位合成制备, 从几百个点的芯片到几百万点的高密度芯片, 生物芯片从一项科学成为一项技术, 被越来越多的研究者广泛运用. 各个实验室不断产生海量的杂交数据, 相同领域的研究者需要比较不同实验平台产生的数据, 作为基于分子杂交原理的高通量技术, 芯片实验的标准化、可信度、重现性和芯片结果是否能作

为定量数据等问题成为所有的芯片使用者关心的课题. 迈阿密原则和微阵列质量控制系列研究回答了这两个问题.

迈阿密原则(Minimum Information About a Microarray Experiment, MIAME, 微阵列实验最小信息量)提出了生物芯片标准化的概念^[22], 该原则的制定使世界各地实验室的芯片实验数据可以为所有的研究者共享. 同时, 美国国家生物信息学中心(NCBI)和位于英国的欧洲生物信息学研究所(EBI)也建立了 GEO^[23] (<http://www.ncbi.nlm.nih.gov/geo/>) 和 ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) 公共数据库, 接受和储存全球研究者根据迈阿密原则提交的生物芯片数据, 对某项研究感兴趣的研究人员可以下载到相关课题的芯片原始数据进行分析.

2006 年美国 FDA 联合多个独立实验室进行了 MAQC 系列实验(micro array quality control, MAQC), 旨在研究目前所使用的芯片平台的质量控制. 该研究的 12 篇系列文章发表在 2006 年 9 月份的 *Nature Biotechnology* 上, 用严格的实验分析了目前主流芯片平台数据质量^[24-26], 芯片数据和定量 PCR 结果之间的相关性^[27], 芯片数据均一化方法^[28], 不同芯片平台之间的可重现性^[29,30]. 证明了不同芯片平台产生的数据具有可比性和可重现性, 各种芯片平台之间的系统误差远远小于人为操作和生物学样品之间本身的差异, 肯定了芯片数据的可信性, 打消了以往对芯片数据的种种猜疑, 明确了基于杂交原理的芯片同样可以作为一种定量的手段. 推动了生物芯片技术在分子生物学领域更广泛的应用.

生物信息学和统计学是在处理基因芯片产生的海量数据中必不可少的工具^[31]. 随着芯片应用的推进, 芯片数据分析的新理论和新算法不断地被开发出来, 这些方法帮助生物学家从海量的数据里面快速筛选出差异表达的基因. 一次芯片实验获得的是成千上万个基因的表达信息, 任何一种单一的分析方法都很难将所有蕴含在数据中的生物学信息全部提取出来^[32], 从近年来生物信息学研究的趋势来看, 目前研究的重点开始转向芯片数据储存、管理、共享和深度信息挖掘, 旨在从芯片数据中获得更多的生物学解释, 而不再停留在单纯的差异表达基因筛选上^[32,33].

目前基因芯片的制备向两个主要方向发展. 第

一, 高密度化, 具体表现为芯片密度的增加, 目前原位合成的芯片密度已经达到了每平方厘米上千万个探针. 一张芯片上足以分析一个物种的基因组信息. 第二, 微量化, 芯片检测样品的微量化, 目前芯片检测下限已经能达到纳克级总 RNA 水平, 这为干细胞研究中特别是 IPS 干细胞对单个细胞的表达谱研究提供了可能. 另一方面, 微量化也体现芯片矩阵面积的微量化, 即在同一个芯片载体上平行的进行多个矩阵的杂交, 大大减少系统和批次可能带来的差异, 同时削减实验费用.

微阵列技术改变了生物学研究的方法, 使得微量样品快速高通量的分析成为可能, 从单个基因的研究迅速扩展到全基因组的系统生物学研究. 微阵列技术帮助生物学研究进入后基因组时代, 研究成果层出不穷^[34].

2001 年国家人类基因组南方研究中心韩泽广博士研究小组利用 cDNA 芯片对肝癌和正常组织中的 12393 个基因和 EST 序列进行了表达谱筛查, 其中发现了 2253 个基因和 EST 在肝癌中发生了差异表达, 并对这些差异基因的信号通路进行了分析, 发现 WNT 信号通路在肝癌的发生中出现了表达异常^[12]. 2002 年中国科学院神经科学研究所张旭博士研究组利用表达谱芯片对大鼠外周神经损伤模型背根神经节的基因表达进行了研究, 通过对 7523 个基因进行表达谱筛查, 发现一批与神经损伤和疼痛相关的基因和潜在的药物靶点, 并解释了临床上使用的 Gabapentin 治疗疼痛的分子机制^[35]. 同年 *Nature* 发表的荷兰癌症研究所的研究小组利用表达谱芯片对乳腺癌患者 5 年内的转移情况进行分子水平的分型工作^[36], 研究结果显示可以利用基因表达谱的差异来预测肿瘤的预后情况. 这项工作成为用表达谱对疾病进行分子分型和预后研究的经典案例. 2006 年根据这项研究成果生产的表达谱芯片成为第一张通过 FDA 批准进入临床使用的表达谱芯片, 迈出了芯片从实验室走向临床的第一步.

2003 年启动的人类基因组单体型计划(International Hapmap Project)采用了 5 个不同的高通量平台进行基因分型工作, 其中芯片完成了超过 50% 的工作量^[37]. 2005 年 Gunderson 等人用 DNA 芯片检测人类基因组中的 SNP 位点^[38], 芯片的检测结果与传统的

基于 PCR 技术的基因分型具有非常好的相关性. Roch 的 Amplichip CYP450 检测芯片在 2005 年通过了 FDA 批准进入临床使用, 用以检测患者体内决定细胞色素氧化酶活性的多态性位点, 预测患者药物代谢水平的高低^[39], 这也是第一张进入临床检查的 SNP 芯片. 2007 年 Burton 等人利用微阵列检测了 1000 例四种疾病患者和对照组的 1500 例健康人的 14000 多个 SNP 位点, 发现了患者与健康人自身免疫方面存在的基因组 SNP 位点差异性^[40].

2006 年国家人类基因组南方研究中心和生物芯片上海国家工程研究中心合作, 利用比较基因组杂交(aCGH)和表达谱芯片联合分析了肝癌样品中基因组拷贝数变异和基因表达量变化的相关性, 为从基因组结构变异角度研究肝癌的发生机制提供了研究基础^[20]. 2007 年 Carter 等人报道利用 SNP 芯片检测基因组拷贝数变化(copy number variation), 发现人类基因组中存在约 12% 的拷贝数变异^[41], 这一发现远远出乎以前人们对人类基因组多样性的预测. 2008 年 Lee 等人^[42]报道了用比较基因组杂交(aCGH)对遗传疾病进行分型, 发现了存在于精神发育迟缓病人染色体 15q13.3 区域的一个缺失, *Nature Genetics* 刊登了这一结果.

小 RNA 或非编码 RNA(ncRNA) 是新兴的一个研究领域. 这类小分子 RNA 在生理过程中扮演着调控分子的角色, 在发育过程和人类疾病尤其是肿瘤发生发展中起到非常重要的作用^[43,44]. 这些小分子核酸高度同源, 序列差异往往就是一个碱基, 且长度在 21~35 碱基之间, 这对芯片检测是一个很大的挑战, 要求既能检测到低分子量的目的片段, 又能区分一个碱基的差异, 芯片技术的发展已经部分解决这些问题, 利用芯片检测小分子 RNA 的表达谱成为肿瘤分型的另一种途径^[45], 在乳腺癌^[46]、肺癌^[47]、前列腺癌^[48]等疾病的分子分型和分子标志物寻找中取得非常好的结果.

DNA 修饰以及 DNA-蛋白质相互作用是细胞对基因表达的调控方式. 2007 年 Meier 等人^[49]采用 ChIP-chip 实验方法, 检测到了发生 DNA 损伤时, 相应的修复因子在 DNA 上的分布. 同年, Guenther 等人^[50]用 ChIP-chip 实验发现基因的转录起始可能不是特异的, 大多数的人类基因包括原来被证明转录失活

的编码基因都能启动转录, 细胞通过对基因组的修饰来控制转录的延伸过程, 从而达到调控基因转录的目的. 2008 年 Lupien 等人^[51]用 ChIP-chip 和一系列的实验, 证实了基因组表观遗传学修饰能调控 FoxA1, 通过 FoxA1 改变染色质的结构, 从而控制基因的组织特异性表达.

基因芯片的应用加速了生命科学研究的进程, 微量化和并行化的分析帮助科学家从海量数据中发掘有用的信息.

3 高通量测序技术

高通量测序技术是对传统测序一次革命性的改变, 一次对几十万到几百万条 DNA 分子进行序列测定, 因此在有些文献中称其为下一代测序技术(next generation sequencing)足见其划时代的改变^[52], 同时高通量测序使得对一个物种的转录组和基因组进行细致全貌的分析成为可能, 所以又被称为深度测序(deep sequencing)^[53]. 高通量测序平台的代表是罗氏公司(Roche)的 454 测序仪(Roch GS FLX sequencer), Illumina 公司的 Solexa 基因组分析仪(Illumina Genome Analyzer)和 ABI 的 SOLiD 测序仪(ABI SOLiD sequencer). 2008 年 4 月 Helico BioScience 公司的 Timothy 等人在 *Science* 上报道了他们开发的真正的单分子测序技术, 并利用该技术对一个 M13 病毒基因组进行重测序^[54]. 这项技术之所以被称为真正的单分子测序, 是因为它完全跨过了上述 3 种高通量测序依赖的基于 PCR 扩增的信号放大过程, 真正达到了读取单个荧光分子的能力, 向 1000 美元测定一个人类基因组的目标迈出了一大步.

这些平台共同的特点是极高的测序通量, 相对于传统测序的 96 道毛细管测序, 高通量测序一次实验可以读取 40 万到 400 万条序列. 读取长度根据平台不同从 25 碱基到 450 碱基, 不同的测序平台在一次实验中, 可以读取 1G 到 14G 不等的碱基数^[55], 这样庞大的测序能力是传统测序仪所不能比拟的.

4 高通量测序的应用

高通量测序可以帮助研究者跨过文库构建这一实验步骤, 避免了亚克隆过程中引入的偏差. 依靠后期强大的生物信息学分析能力, 对照一个参比基因

组(reference genome)高通量测序技术可以非常轻松地完成基因组重测序(re-sequence), 2007 年 van Orsouw 等人^[56]结合改进的 AFLP 技术和 454 测序技术对玉米基因组进行了重测序, 该重测序实验发现的超过 75% 的 SNP 位点能够用 SNPWave 技术验证, 提供了一条对复杂基因组特别是含有高度重复序列的植物基因组进行多态性分析的技术路线. 2008 年 Hillier^[57]对线虫 CB4858 品系进行 Solexa 重测序, 寻找线虫基因组中的 SNP 位点和单位点的缺失或扩增. 但是也应该看到, 由于高通量测序读取长度的限制, 使其在对未知基因组进行从头测序(*de novo sequencing*)的应用受到限制, 这部分工作仍然需要传统测序(读取长度达到 850 碱基)的协助. 但是这并不影响高通量测序技术在全基因组 mRNA 表达谱, microRNA 表达谱, ChIP-chip 以及 DNA 甲基化等方面的应用.

2008 年 Mortazavi 等人^[58]对小鼠的大脑、肝脏和骨骼肌进行了 RNA 深度测序, 这项工作展示了深度测序在转录组研究上的两大进展, 表达计数和序列分析. 对测得的每条序列进行计数获得每个特定转录本的表达量, 是一种数码化的表达谱检测, 能检测到丰度非常低的转录本. 分析测得的序列, 有大于 90% 的数据显示落在已知的外显子中, 而那些在已知序列之外的信息通过数据分析展示的是从未被报道过的 RNA 剪切形式, 3' 末端非翻译区, 变动的启动子区域以及潜在的小 RNA 前体, 发现至少有 3500 个基因拥有不止一种剪切形式. 而这些信息无论使用芯片技术还是 SAGE 文库测序都是无法被发现的. 同年 Sugarbaker 利用 mRNA 深度测序对恶性胸膜瘤和对照样品进行比较, 发现了肿瘤中存在的 15 个不同的点突变^[59].

高通量测序另一个被广泛应用的领域是小分子 RNA 或非编码 RNA(ncRNA)研究. 测序方法能轻易的解决芯片技术在检测小分子时遇到的技术难题(短序列, 高度同源), 而且小分子 RNA 的短序列正好配合了高通量测序的长度, 使得数据“不浪费”, 同时测序方法还能在实验中发现新的小分子 RNA. 在衣藻^[60]、斑马鱼^[61]、果蝇^[62]、线虫^[63]、人^[64]和黑猩猩^[65]中都已经成功地找到了新的小分子 RNA. 在线虫中获得了 40 万个序列, 通过分析发现了 18 个新的小

RNA分子和一类全新的小分子RNA^[63], 通过对人胚胎干细胞发育前后的分析, 获得了 334 个小RNA的表达谱带, 包括新发现的 104 个小RNA^[64].

在DNA-蛋白质相互作用的研究上, 染色质免疫沉淀-深度测序(ChIP-seq)实验也展示了其非常大的潜力. 染色质免疫沉淀以后的DNA直接进行测序, 对比ref-seq可以直接获得蛋白与DNA结合的位点信息, 相比ChIP-chip, ChIP-seq可以检测更小的结合区段、未知的结合位点、结合位点内的突变情况和蛋白亲和力较低的区段. 2007 年Johnson等人^[66]用ChIP-seq 对转录因子NRSF在DNA上的结合位点进行了全基因组的筛查, 获得了 1946 个结合位点, 最小能分辨的结合位点为 50 个碱基, 这些高质量的ChIP-seq结果提供了研究新的DNA-蛋白质相互作用的内容, 其中包括了胰岛发育调控网络中的重要转录因子. 同年Robertson等人^[67]用同样的方法检测转录因子和基因组DNA的结合情况. 这两项研究同时验证了以往用ChIP-chip实验检测到的结合位点, 同时发现新的结合位点, Robertson等人^[67]发现, ChIP-seq的分辨率可达 40 碱基. 2008 年Chen等人^[68]在*Cell*上发表论文, 用ChIP-seq检测了Nanog, Oct4, STAT3, Smad1, Sox2 等 13 个序列特异性的转录因子与基因组DNA的结合情况, 这些转录因子都是LIF和BMP途径的重要调控分子. 这些转录因子在ES细胞里结合位点为我们揭示了ES细胞内决定ES细胞发育方向的调控网络.

5 基因芯片和高通量测序技术的应用前景

高通量测序技术虽然建立的时间不长, 但是在基因组的各个研究领域都显示出其非凡的魅力, 而且日益显示出其对基因芯片“取而代之”的咄咄态势. 那么, 基因芯片向何处去呢?

基因芯片技术经过近 15 年的发展已经形成了一个系统的平台, 从样品制备、芯片制作、芯片杂交、数据扫描到后期的数据管理, 储存以及深度数据挖掘都有了标准化的流程、坚实的理论和实验的支持, 成为一个非常稳定可信的实验技术, 为广大的研究者所运用, 同时也积累了庞大的公共数据库. 深度测序要建立这样的体系同样需要若干年的完善. 芯片杂交结果直观, 分析快速, 适合对大量生物学样品进行已知信息的检测, 同时芯片数据分析有成熟

完整的理论, 为后期数据分析提供强大的支持.

基因芯片的缺点, 就在于它是一个“封闭系统”, 它只能检测人们已知序列的特征(或有限的变异). 而深度测序的强项, 就在于它是一个“开放系统”, 它的发现能力和寻找新的信息的能力, 从本质上高于芯片技术. 研究者可以充分享受这两个平台的比较优势, 在获取新信息的基础上, 利用芯片的强项, 即对已知信息的高通量、低成本(相对)的检测能力, 对大量样品进行快速检测, 短时间内获得有大量有效的数据.

作为两个高通量的基因组学研究技术, 在应用的某些方面存在重叠和竞争, 但是在更多方面是优势互补, 两种方法联合使用, 将解决以前的单种技术难以解决的问题. 如Euskirchen等人^[69]同时用ChIP-chip和ChIP-seq对STAT1 的结合位点进行了检测, 结果非常有趣, 两种技术对于强阳性的区段具有非常好的相关性, 而对于一些弱的结合位点, ChIP-chip和ChIP-seq都会丢失部分信息, 而一种方法丢失的信息又恰好能被另一种方法所检出, 完整的数据是来自两部分的整合. 同样的情况也发生在mRNA表达谱检测上, 一种技术能弥补另一种技术遗漏的部分^[70,71]. 因此对一个生物学问题的回答需要不同实验技术的协同配合. 例如目前新兴的Target sequencing 或者叫做序列捕获, Sequence Capture, 技术, 就是结合了芯片和深度测序, 利用芯片探针捕获待测片段, 再用深度测序技术分析核酸序列^[72], 利用高密度芯片和 454 测序仪曾成功的捕获了 6726 个 500 碱基长度的外显子和 200 kb到 5 Mb的DNA区段, 测序结果显示大多数的捕获DNA是符合设计要求的目的片段, 该实验验证了序列捕获的特异性和可行性, 芯片的序列捕获技术将来有可能在对基因组区段测序的研究中取代多重PCR过程^[73]. 芯片这种高通量技术显示出其在样品选择和富集方面的优势和潜力.

随着科学技术的进步, 能不断地给一项技术带来新的增长点, 基因芯片和深度测序是点杂交技术和测序的高通量革命, 两大分子生物学经典实验技术都发展到了高通量的时代, 正如他们以前对生命科学研究所做出的贡献一样, 今后这两大技术必将继续协同配合推动生命科学研究进入新的纪元.

参考文献

- 1 Fodor S P, Read J L, Pirrung M C, et al. Light-directed, spatially addressable parallel chemical synthesis. *Science*, 1991, 251(4995): 767—773[DOI]
- 2 Pease A C, Solas D, Sullivan E J, et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci USA*, 1994, 91(11): 5022—5026[DOI]
- 3 Service R F. Microchip arrays put DNA on the spot. *Science*, 1998, 282(5388): 396—399[DOI]
- 4 Beattie W G, Meng L, Turner S L, et al. Hybridization of DNA targets to glass-tethered oligonucleotide probes. *Mol Biotechnol*, 1995, 4(3): 213—225[DOI]
- 5 Cheng J, Shoffner, M A, Hvhichia G E, et al. Chip PCR. II . Investigation of different PCR amplification systems in microfabricated silicon-glass chips. *Nucleic Acids Res*, 1996, 24(2): 380—385[DOI]
- 6 McCall G, Labadie J, Brock P. et al. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc Natl Acad Sci USA*, 1996, 93(24): 13555—13560[DOI]
- 7 Gilles P N, Wu D J, Foster C B, et al. Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nat Biotechnol*, 1999, 17(4): 365—370[DOI]
- 8 Schena M, Shalon D, Davis R W, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 1995, 270(5235): 467—470[DOI]
- 9 Pirrung M C. Spatially Addressable Combinatorial Libraries. *Chem Rev*, 1997, 97(2): 473—488
- 10 Singh-Gasson S, Green R D, Yue Y, et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol*, 1999, 17(10): 974—978[DOI]
- 11 Xiao P F, He N Y, Liu Z C, et al. *In situ* synthesis of oligonucleotide arrays by using soft lithography. *Nanotechnology*, 2002, 13: 756—762[DOI]
- 12 Xu X R, Huang J, Xu Z G, et al. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc Natl Acad Sci USA*, 2001, 98(26): 5[DOI]
- 13 Zheng P Z, Wang K K, Zhang Q Y, et al. Systems analysis of transcriptome and proteome in retinoic acid/arsenic trioxide-induced cell differentiation/apoptosis of promyelocytic leukemia. *Proc Natl Acad Sci USA*, 2005, 102(21): 7653—7658[DOI]
- 14 Huang W, He Y, Wang H, et al. Linkage disequilibrium sharing and haplotype-tagged SNP portability between populations. *Proc Natl Acad Sci USA*, 2006, 103(5): 1418—1421[DOI]
- 15 Hua Y J, Tu K, Tang Z Y, et al. Comparison of normalization methods with microRNA microarray. *Genomics*, 2008, 92(2): 122—128[DOI]
- 16 Zhang Y, Huang J, Jia S, et al. SAGE tag based cDNA microarray analysis during larval to pupal development and isolation of novel cDNAs in *Bombyx mori*. *Genomics*, 2007, 90(3): 372—379[DOI]
- 17 Zhao B, Liang R, Ge L, et al. Identification of drought-induced microRNAs in rice. *Biochem Biophys Res Commun*, 2007, 354(2): 585—590[DOI]
- 18 Li R Y, Zhang Q H, Liu Z, et al. Effect of short-term and long-term fasting on transcriptional regulation of metabolic genes in rat tissues. *Biochem Biophys Res Commun*, 2006, 344(2): 562—570[DOI]
- 19 Ma X H, Hu S J, Ni H, et al. Serial analysis of gene expression in mouse uterus at the implantation site. *J Biol Chem*, 2006, 281(14): 9351—9360[DOI]
- 20 Huang J, Sheng H H, Shen T, et al. Correlation between genomic DNA copy number alterations and transcriptional expression in hepatitis B virus-associated hepatocellular carcinoma. *FEBS Lett*, 2006, 580(15): 3571—3581[DOI]
- 21 Xu C, Zheng P, Shen S, et al. NMR structure and regulated expression in APL cell of human SH3BGRL3. *FEBS Lett*, 2005, 579(13): 2788—2794[DOI]
- 22 Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*, 2001, 29(4): 365—371[DOI]
- 23 Edgar R, Barrett T. NCBI GEO standards and services for microarray data. *Nat Biotechnol*, 2006, 24(12): 1471—1472[DOI]
- 24 Ji H, Davis R W. Data quality in genomics and microarrays. *Nat Biotechnol*, 2006, 24(9): 1112—1113[DOI]
- 25 Tong W, Lucas A B, Shippy R, et al. Evaluation of external RNA controls for the assessment of microarray performance. *Nat Biotechnol*, 2006, 24(9): 1132—1139[DOI]

- 26 Patterson T A, Lobenhofer E K, Fulmer-Smentek S B, et al. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol*, 2006, 24(9): 1140—1150 [\[DOI\]](#)
- 27 Canales R D, Luo Y, Willey J C, et al. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol*, 2006, 24(9): 1115—1122 [\[DOI\]](#)
- 28 Shippy R, Fulmer-Smentek S, Jensen R V, et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat Biotechnol*, 2006, 24(9): 1123—1131 [\[DOI\]](#)
- 29 Shi L, Reid L H, Jones W D, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*, 2006, 24(9): 1151—1161 [\[DOI\]](#)
- 30 Guo L, Lobenhofer E K, Wang C, et al. Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol*, 2006, 24(9): 1162—1169 [\[DOI\]](#)
- 31 Grant G R, Manduchi E, Stoeckert C J Jr. Analysis and management of microarray gene expression data. *Curr Protoc Mol Biol*, 2007, Chapter 19: Unit 19.6
- 32 Stoeckert C J Jr, Causton H C, Ball C A. Microarray databases: standards and ontologies. *Nat Genet*, 2002, 32 (Suppl): 469—473 [\[DOI\]](#)
- 33 Cordero F, Botta M, Calogero R A. Microarray data analysis and mining approaches. *Brief Funct Genomic Proteomic*, 2007, 6(4): 265—281 [\[DOI\]](#)
- 34 Shiu S H, Borevitz J O. The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity*, 2008, 100(2): 141—149 [\[DOI\]](#)
- 35 Xiao H S, Huang Q H, Zhang F X, et al. Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain. *Proc Natl Acad Sci USA*, 2002, 99(12): 8360—8365 [\[DOI\]](#)
- 36 Van 't Veer L J, Dai H, van de Vijver M J, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 2002, 415(6871): 530—536 [\[DOI\]](#)
- 37 The International HapMap Project. *Nature*, 2003, 426(6968): 789—796
- 38 Gunderson K L, Steemers F J, Lee G, et al. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet*, 2005, 37(5): 549—554 [\[DOI\]](#)
- 39 Rebsamen M C, Desmeules J, Daali Y, et al. The AmpliChip CYP450 test: cytochrome P450 2D6 genotype assessment and phenotype prediction. *Pharmacogenomics J*, 2008
- 40 Burton P R, Clayton D G, Cardon L R, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet*, 2007, 39(11): 1329—1337 [\[DOI\]](#)
- 41 Carter N P. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet*, 2007, 39(7 Suppl): S16—21 [\[DOI\]](#)
- 42 Lee C, Iafrate A J, Brothman A R. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet*, 2007, 39(7 Suppl): S48—54 [\[DOI\]](#)
- 43 Manikandan J, Aarthi J J, Kumar S D, et al. Oncomirs: The potential role of non-coding microRNAs in understanding cancer. *Bioinformation*, 2008, 2(8): 330—334
- 44 Grosshans H, Filipowicz W. Molecular biology: the expanding world of small RNAs. *Nature*, 2008, 451(7177): 414—416 [\[DOI\]](#)
- 45 Yin J Q, Zhao R C, Morris K V. Profiling microRNA expression with microarrays. *Trends Biotechnol*, 2008, 26(2): 70—76 [\[DOI\]](#)
- 46 Blenkinson C, Goldstein L D, Thorne N P, et al. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol*, 2007, 8(10): R214 [\[DOI\]](#)
- 47 Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, 2006, 9(3): 189—198 [\[DOI\]](#)
- 48 Mattie M D, Benz C C, Bowers J, et al. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Mol Cancer*, 2006, 5: 24 [\[DOI\]](#)
- 49 Meier A, Fiegler H, Munoz P, et al. Spreading of mammalian DNA-damage response factors studied by ChIP-chip at damaged telomeres. *EMBO J*, 2007, 26(11): 2707—2718 [\[DOI\]](#)
- 50 Guenther M G, Levine S S, Boyer L A, et al. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 2007, 130(1): 77—88 [\[DOI\]](#)
- 51 Lupien M, Eeckhoutte J, Meyer C A, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription.

- Cell, 2008, 132(6): 958—970[DOI]
- 52 Schuster S C. Next-generation sequencing transforms today's biology. *Nat Methods*, 2008, 5(1): 16—18[DOI]
 - 53 Sultan M, Schulz M H, Richard H, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 2008, 321(5891): 956—960[DOI]
 - 54 Harris T D, Buzby P R, Babcock H, et al. Single-molecule DNA sequencing of a viral genome. *Science*, 2008, 320(5872): 106—109[DOI]
 - 55 Mardis E R. The impact of next-generation sequencing technology on genetics. *Trends Genet*, 2008, 24(3): 133—141
 - 56 van Orsouw N J, Hogers R C, Janssen A, et al. Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE*, 2007, 2(11): e1172[DOI]
 - 57 Hillier L W, Marth G T, Quinlan A R, et al. Whole-genome sequencing and variant discovery in *C. elegans*. *Nat Methods*, 2008, 5(2): 183—188[DOI]
 - 58 Mortazavi A, Williams B A, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 2008, 5(7): 621—628[DOI]
 - 59 Sugarbaker D J, Richards W G, Gordon G J, et al. Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proc Natl Acad Sci USA*, 2008, 105(9): 3521—3526[DOI]
 - 60 Zhao T, Li G, Mi S, et al. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev*, 2007, 21(10): 1190—1203[DOI]
 - 61 Houwing S, Kamminga L M, Berezikov E, et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, 2007, 129(1): 69—82[DOI]
 - 62 Stark A, Kheradpour P, Parts L, et al. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res*, 2007, 17(12): 1865—1879[DOI]
 - 63 Ruby J G, Jan C, Player C, et al. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 2006, 127(6): 1193—1207[DOI]
 - 64 Morin R D, O'Connor M D, Griffith M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res*, 2008, 18(4): 610—621[DOI]
 - 65 Berezikov E, Thuemmler F, Van Laake L W, et al. Diversity of microRNAs in human and chimpanzee brain. *Nat Genet*, 2006, 38(12): 1375—1377[DOI]
 - 66 Johnson D S, Mortazavi A, Myers R M, et al. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, 2007, 316(5830): 1497—1502[DOI]
 - 67 Robertson G, Hirst M, Bainbridge M, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*, 2007, 4(8): 651—657[DOI]
 - 68 Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 2008, 133(6): 1106—1117[DOI]
 - 69 Euskirchen G M, Rozowsky J S, Wei C L, et al. Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res*, 2007, 17(6): 898—909[DOI]
 - 70 Oudes A J, Roach J C, Walashek L S, et al. Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression. *BMC Cancer*, 2005, 5: 86[DOI]
 - 71 Zhang G H, Shi D R, Liang X M, et al. Comparison of HER2/neu oncogene detected by chromogenic in-situ hybridization and immunohistochemistry in breast cancer. *Zhonghua Bing Li Xue Za Zhi*, 2006, 35(10): 580—583
 - 72 Dahl F, Stenberg J, Fredriksson S, et al. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc Natl Acad Sci USA*, 2007, 104(22): 9387—9392[DOI]
 - 73 Albert T J, Molla M N, Muzny D M, et al. Direct selection of human genomic loci by microarray hybridization. *Nat Methods*, 2007, 4(11): 903—905[DOI]