

DOI: 10.3724/SP.J.1005.2011.00017

单细胞基因组学分析的技术前沿

潘星华, 朱海英, MARJANI Sadie L.

耶鲁大学医学院遗传系, 美国康耐狄克州纽黑文市 CT 06520-8005

摘要: 基因组学已经深刻地改变了生命科学的诸多领域的面貌。目前它的主要内容是新的全基因组碱基序列的测定和在全基因组范围内鉴定那些在不同水平上影响生命活动的基因群的功能和相互作用。为达此要求, 近年出现的第二代测序(深度测序)技术和基因芯片技术发挥了关键作用, 但是两者都需要足够的高质量的核酸样品。所以, 在只有或只能用单细胞或极少量细胞的情况下, 如果没有特殊手段, 上述分析往往不能常规、方便地进行。文章以 DNA 扩增为主线, 综合阐述了目前在单细胞(特别是微生物)全基因组测序和大基因组的靶向重测序, 以及对单细胞或微量细胞进行的基于深度测序或芯片杂交的功能基因组分析, 如转录组、ChIP 和 DNA 的 CpG 甲基化分析等的最新策略和技术, 评价了单细胞基因组测序和功能基因组学各技术的特点并对发展前景进行了展望。

关键词: 单细胞基因组测序; 第二代测序技术; 功能基因组学; 转录组分析; 染色质免疫共沉淀; CpG 甲基化图谱

Technological advances in single-cell genomic analyses

PAN Xing-Hua, ZHU Hai-Ying, MARJANI Sadie L.

Department of Genetics, Yale University School of Medicine, New Haven CT 06520-8005, USA

Abstract: The technological progress of the genomics has transformed life science research. The main objectives of genomics are sequencing of new genomes and genome-wide identification of the function and the interaction of genes and their products. The recently developed second generation or next generation sequencing platforms and DNA microarray technology are immensely important and powerful tools for functional genomic analyses. However, their application is limited by the requirement of sufficient amounts of high quality nucleic acid samples. Therefore, when only a single cell or a very small number of cells are available or are preferred, the whole genomic sequencing or functional genomic objectives cannot be achieved conventionally and require a robust amplification method. This review highlights DNA amplification technologies and summarizes the strategies currently utilized for whole genome sequencing of a single cell, with specific focus on studies investigating microorganisms; An outline for targeted re-sequencing enabling the analysis of larger genomes is also provided. Furthermore, the review presents the emerging functional genomic applications using next-generation sequencing or microarray analysis to examine genome-wide transcriptional profile, chromatin modification and other types of protein-DNA binding profile, and CpG methylation mapping in a single cell or a very low quantity of cells. The nature of these technologies and their prospects are also addressed.

Keywords: single-cell sequencing; next generation sequencing; functional genomics; transcriptome profiling; chromatin immunoprecipitation; CpG methylation mapping

收稿日期: 2010-08-23; 修回日期: 2010-10-31

基金项目: 单细胞转录组和甲基化组分析, NIH 项目(编号: 1R21HD066457-01)资助

通讯作者: 潘星华(1963-), 男, 博士, 副研究员, 研究方向: 功能基因组学。E-mail: xinghua.pan@yale.edu; xinghua.pan@gmail.com

致谢: 感谢耶鲁大学 Sherman Weissman 院士的长期指导及 Robert LaMotte、Flora Vaccarino、InHyun Park 和斯坦福大学 Michael Snyder 等教授的支持。

单细胞基因组学包括单细胞全基因组测序和以单细胞和微量细胞为材料的全基因组范围内的基因功能研究,体现了中华智慧中的“见微知著”和“质重于量”的思想精髓,近年来取得了巨大的进展,涓涓溪水正在汇成浩浩巨流,进入基因组学的海洋,成为目前基因组学研究甚至生命科学研究的一个重要发展方向。

1 基因组学当前研究的主要内容和最新技术

迄今,从结构基因组学来讲,包括人类在内的数千种生物的以全基因组测序为目标的基因组结构初步分析已经完成;但是,尚有成千上万与人类生活和健康息息相关的动植物及微生物等低等生物有待研究。其中在医学和经济上有重要意义的一大类重要生物——微生物基因组测序是目前本领域的研究热点之一。另外,DNA序列的多态性(主要是SNP和CNV)和突变(如癌症)不仅构成了个体生理差异的内在原因,还是决定多数人类疾病发生的根本因素;既是对基因组结构研究的深化,又有重要的功能意义。

功能基因组学以结构基因组学所取得的静态碱基序列信息为基础,在全基因组水平上,高通量大规模的动态分析多种基因的表达及其调节,基因及其表达产物之间的相互作用,以及在不同层次、静态和动态相结合及主动干预下基因作用对生物功能的影响^[1,2]。研究发现蛋白质有相互修饰(乙酰化、磷酸化、甲基化等)、结合并可调节相关基因表达和蛋白质的功能^[3,4];生物体内占基因组主要信息的大量种类的RNA(如microRNA等非编码RNA)也参加复杂的基因调控^[5,6];DNA的表观修饰(如DNA甲基化)不改变DNA序列但在分化、发育和疾病中也起着重要作用^[7,8]。这些发现和研究大大丰富了功能基因组学的内容。其中,以RNA和DNA为目标的研究(包括一些DNA、RNA与蛋白质相互作用的研究)是目前功能基因组学的主要内容,具体包括mRNA表达谱、非编码RNA表达谱、DNA甲基化谱、组蛋白甲基化修饰谱、染色质免疫共沉淀(ChIP,分析与蛋白因子结合的DNA序列)等^[1,9,10]。其中,DNA甲基化谱、组蛋白甲基化谱(反映染色质结构)和DNA酶高敏位点谱等是表观基因组学(epigenomics)的任务,因为这里修饰属于表观修饰,具有可调节性,并不改

变DNA序列本身,但可在细胞复制后忠实保持^[11-14]。生命过程常涉及多种基因和多种调控机制的参与,而且受环境因素的影响。这些机制之间往往互相关联,共同调节特定生命过程,综合分析各种机制可构成一个整合的全基因组水平的分子图像。通过实验技术如大规模人工突变、基因打靶、RNA干扰(RNAi)等来主动干预生物功能,在系统地鉴定相关基因和其功能方面也发挥着积极作用^[15,16]。

在基因组学领域,技术革命往往是科学进步的先声。基因芯片风靡一时,其最根本的优点在于同时检测大量基因的相对量,高通量地在全基因组水平上系统研究参与有关作用的几乎全部基因,而不是之前的一次只研究一个或少数几个基因。当然,这类研究已从早期的基因表达谱发展到SNP谱、CNV谱,并应用到功能基因组分析的各个方面。近年来新发展起来的集群平行焦磷酸测序(Massively parallel pyrosequencing),又称深度测序(Deep-sequencing)或第二代测序技术(相对于经典的桑格(Sanger)测序),主要有Illumina/Solexa、Roche/454 FLX和ABI/SOLiD平台^[9,17,18]。这3个平台各有优点,Solexa测序性价比最高,454的测序片段比较长,而SOLiD据称测序的准确度高(但不少人认为其并无特殊)。它们共同的原理是:(1)每个位点以单分子原DNA模板合成新DNA同时测序(sequencing by synthesis),在此过程中添加不同种dNTP会释放出不同的荧光;(2)巨大数量位点上各DNA短片段测序以阵列方式同时平行进行。这样,测定的每种分子序列的频率定量地反应其在原DNA库中的频率,同时测出的序列准确、定性地反应其在原DNA库中的序列,如果有背景信号造成序列不符很容易在结果中滤除。深度测序不仅在数量上更上一层楼,并有质的飞跃,因为它可以提供更大量的数据(目前每个样品一次测序可以获得一千万个以上独立的序列数据),提供精确可靠的具体序列的拷贝数(而基因芯片只提供相对信号量),能检测更大的动态范围,能更全面覆盖样品中的DNA信息,而且不依赖与已知DNA序列从而便于发现新序列,但也对分析样品提出了更高的质量要求,特别是分析样品中不能有过多非特异DNA^[19]。深度测序技术不仅在新基因组测序,而且在功能基因组学研究领域也发挥着重要作用,

正成为功能基因组学研究的主流技术^[20]。

一般深度测序(或基因芯片杂交)需要的DNA往往较多,事实上在测序之前有一个扩增建库过程,斯坦福大学Quake研究组利用其发明的Helicos 单分子测序技术用较少的DNA快速测定了人类个体的全基因组序列,无需克隆或扩增,获取到了高质量的数据和许多重要发现^[21];该技术也显示mRNA库良好的直接测序结果^[22];该技术的的一个最新结果是,结合流程池捕获(flow-cell capture)和原位反转录技术,只用少至 250 至 16 000 个小鼠或人体细胞,成功深度测定了mRNA库的序列^[23]。总之这项技术开创了一个新的方向,但与其说是“单分子测序”不如说是“直接测序”更能准确反应其特点,应用于单细胞并不容易,到目前为止我们尚未见到相关报道。不久前,原 454 公司的创始人、美国工程院院士 Jonathan Rothberg 又创立了一个基于全新概念的公司——Ion Torrent,其关键技术已经基本建立,还有数个其他公司同时在作各自独特的创新工作,预期不久将可提供更快捷、更方便的新一代测序系统。

2 基因组学分析所面临的细胞材料的量和质的问题

在结构基因组方面,从海洋、泥土、温泉及其他恶劣自然环境下甚至人体外通腔道(如口腔、呼吸道、食道、肠道和生殖道等)可能分离出大量种类的微生物(泛指细菌、真菌、病毒等各种微生物)。其中一些微生物对人类疾病或医药制造甚至工业经济有重要的作用。例如,嗜油菌可用于清理石油污染;从特定微生物分离出来的某些生物酶由于功能独特,在生物医学上意义重大,常常被专利保护,一(微)“克”千金;人体内的正常和病理微生物谱可为保健和医疗提供依据。新的和更全面的基因组序列的完整信息,为后续更深入的研究提供坚实的基础^[24,25]。自然界大约有 5×10^{30} 的微生物,而寄生在人体的微生物数是人体自身细胞总数的十倍以上(构成“人体微生物组”),与人类健康息息相关^[26]。这是一个非常广阔的领域,有极大的生物、医学和经济价值。但微生物种类千差万别,对这些混合材料的深度测序将会得到大量纷繁芜杂的序列,称为宏

基因组(metagenome),只能获得最优势生长的微生物的信息,但无法获得一个完整的单一微生物基因组序列,不能提供足够的有用信息^[27,28]。但是绝大多数或 99% 的微生物目前并不能在实验室培养生长,达不到全基因组测序所需要的DNA质量^[29]。单个微生物细胞肯定无疑是绝对“纯”的单一种类的微生物,所以,单细胞测序大有作为^[30]。

功能基因组学分析也如此^[31],由于需要同时系统地分析一个完整基因组内的遗传全景图和表观遗传全景图(profile),因而往往需要大量的样品才能完成。不仅细胞数量要求超多成为很多实验的一个经济上的瓶颈,而且在一些复杂的情况下,我们只能获得极其有限数量的细胞,或者只有用单个细胞或极其有限数目的细胞才能获得理想的结果,因为每个组织都是由多种细胞组成,而每种细胞的数量都非常有限。例如,我们可能需要了解某种大脑某一核团的或外周神经节内的特定神经元的RNA表达或DNA甲基化调节,但是神经组织是非常异质的,有多种细胞混杂,很难找到两个完全相同的神经元,更不可能得到大量(数以百万计)的绝对“纯种”的细胞,分别分析一系列单个神经元可以找出他们的共性和个性,而混合分析可能相互抵消了他们各自的特质^[2]。目前有最新建立的独特技术正好可以在模式生物上逐个活体鉴定和分离完整单个特定的神经元胞体^[32]。又如,众所周知胚胎或干细胞或iPS细胞的早期分化、癌症组织和多种免疫细胞也是非常异质性的。单细胞(包括微量细胞)应用于此类研究将更好地揭示它们的发生发展过程和规律,从而最终有助于人类疾病的认识、预防、诊断和治疗^[33,34]。总之,单细胞功能基因组学分析技术以少见多、独辟蹊径,正在开创一片新的天地。

3 微生物的单细胞全基因组测序和大基因组的靶序列捕获重测序

临渊羡鱼不如退而结网。事实上,科学家早就开始探索在单个细胞水平进行全基因组测序分析的可行技术并成就斐然^[24,30,35-38]。前提在于,每个人体细胞大约含有 6 到 7 皮克(1 皮克= 10^{-12} 克)基因组DNA, 10 到 30 皮克总RNA或大约 1 皮克的信使RNA(mRNA);一个细菌基因组只有飞克数量级

(1 飞克 = 10^{-15} 克)的基因组DNA。而基因芯片和深度测序一般需要微克级(1 微克 = 10^{-6} 克)的DNA。目标和现实之间有至少百万数量级(10^6)倍数的差距,故核酸扩增是一个必不可少的步骤。PCR在现代分子生物学的发展中功勋卓著,自然成为人们的首选,并改进建立了一批以PCR为基础的全基因组DNA扩增(WGA)方法,如简并引物(DOP)-PCR、引物延伸预扩增(PPE)-PCR、接头连接-PCR等,特别是Sigma公司的GenomePlex PCR试剂盒被认为较为可靠。但是,进一步研究发现,对不同的序列来说,PCR扩增的效率存在相当大的偏差:如富含CG的DNA序列和相应基因座位的非随机丢失,等位基因的非随机丢失,以及与DNA片段大小相关的偏差(倾向于扩增更多短片段),尤其是在应用于单细胞时,大量短片段导致片段间序列丢失,从而使其应用受限。笔者曾参与 Roger Lasken(现 J. Craig Venter Institute)领导的研究组,优化建立了MDA(多重置换扩增)第一代试剂盒^[39]。MDA是目前公认的最好的单细胞基因组扩增技术,该技术应用耶鲁大学专利化的Phi29 DNA聚合酶。该酶具有多重置换的特性,在反应中,后一引物的延伸能超越其前面已经结合的DNA而不受其阻挡;具有超强的模板DNA结合能力,能连续合成10 kb到50 kb长的产物,最大可达100 kb;同时具有3'-5'外切酶活性和自我修复错误的能力,从而具有高保真性(其错误率仅为一般Taq DNA聚合酶的百分之一)。该技术还能在恒温条件下进行扩增,其扩增倍数可达百万倍,单个扩增反应即可满足目前大多常规分子生物学需要,包括全基因组测序。

由于MDA的优点,该技术得到了广泛的应用。例如,在DNA样品量极有限的条件下用于SNP分析、STS分析、CNV分析,古代人类样品(如尼安德特人H. Neanderthalensis)DNA测序等,直接从口腔脱落细胞、法医血斑、固定的组织玻片扩增DNA等和大量单个微生物的全基因组测序等都获得了成功^[36,40]。虽然单个微生物DNA一次扩增并测序会有遗漏(一般覆盖面只有60%~70%),超深度的测序及数次重复不同细胞的扩增和测序往往能达到比较全面的覆盖。但是,MDA技术的原始试剂盒在应用中暴露出一些明显缺点,特别是显著的非特异扩增,往往空白对照样品也总是“无中生有”地产生大量的

DNA,另外就是仍然存在序列偏差。为克服这些缺陷,哈佛大学斯、坦福大学等科学家相继提出多种改进版本^[41],如:采用不同特异修饰的引物来抑制引物自源性产物,采用极微量的反应体积(60 纳升)或用较短的反应时间,虽有改进但由于操作不便或不易重复而未解决全部问题。针对这样的挑战, Pan等^[41]近来在MDA理论上首次在反应中引入一种特殊的海藻糖,并结合实时监测建立的全库扩增(Whole pool amplification, WPA)技术和其他优化策略,不仅消除了非特异扩增,而且显著改善了序列偏差,成为一种高特异、高灵敏度的WGA技术。该方法能有效扩增人类单细胞基因组DNA;扩增产物能检出小于1 kb的CNV;即使低到十分之一飞克(10^{-16} gram)的DNA也能显著扩增而无明显非特异产物出现。本技术公布后,反响热烈。

迄今,只有1 000种左右的微生物基因组获得了测序,主要包括大多数能在实验室培养的微生物。与微生物种类的总数 5×10^{30} 相比,无疑单细胞基因组测序具有广泛的应用前景。但是,科学可以鼓励我们一斑窥豹、一叶知秋,但并不允许我们拘泥于一孔之见、以偏概全。虽然结合细胞分选(FACS)可能部分地解决样品微生物的分离和同种微生物的收集问题^[42],但由于测序样品的异质性,即使是同一种微生物可能也有个体之间的变异,所以若干的不同微生物个体的测序和结果的综合分析往往是不可缺少的。不经培养直接分离单一微生物并扩增测序的策略不仅可以获得某一个微生物的基因组序列的全貌,研究大量单个细胞也可获得特定微生物种群的全貌。而人工培养会造成具生长优势的少数微生物种类的人为正挑选,同时遗漏了其他更大量的劣势生长的微生物种类。事实上,大多数微生物在目前条件下无法在实验室生长。近年来,宏基因组学方法^[27,28]能直接解析不经培养或分离的天然(野生)微生物群的基因组,避免了人工培养的缺点。但是,宏基因组学方法很少机会得到数量上天然弱势菌种的完整基因组信息,大量单细胞深度测序可弥补这一缺点。最近,宏基因组学方法的鸟枪法旨在深度测序野生菌群复杂样品的特定靶基因,往往从16S rRNA基因入手(生物细胞内含量最丰富,基因拷贝数最多,进化稳定但又具有物种特异性的RNA

基因), 可同步鉴定多种微生物的相对数量^[26, 43], 虽然有只测定个别基因及优先检出少量优势菌种的缺点, 但相对来讲, 其菌种覆盖面广, 与单细胞测序在一定程度上相互补充。

对于人类或其他高等真核生物的超大型基因组, 目前尽管有少数顶尖实验室进行较多的个体基因组的全测序, 但大多数实验室尚不能常规进行, 在单细胞水平的全测序尚未见报道。在研究领域, 通常应用靶序列捕获(Target capture)并测序的策略, 把测序局限在一个很小的易于操作的序列范围, 如外显子组(Exome)捕获测序则分离和测定占大约 5% 基因组的全部外显子(编码)序列。因为这种测序并不是该物种基因组的首次测序, 故称为靶向重测序(Targeted resequencing)^[9], 为此采用芯片捕获、分子倒置探针(Molecular inversion probes)或暴雨(Rainstorm)PCR 等方案分离富集预先要得到的靶序列^[44]。但对于连锁分析或关联分析所需的大样本来说分析效率仍十分低下, 所以在技术上非常有改进的空间。另外, 在 DNA 研究或诊断等方面需要分析极少量的细胞甚至单个细胞来获得全基因组范围内的 SNP 和 CNV 等多态或突变信息, MDA 结合芯片或测序分析也大有用武之地^[45, 46]。单个细胞基因组扩增和靶序列捕获技术的结合应该有独特的价值。

4 单细胞和微量细胞的功能基因组学研究技术

功能基因组学研究在样品处理的过程中由于涉及许多个连续的步骤, 需要酶或/和抗体的参与, DNA 或 RNA 多有损伤或丢失, 最后往往需要高度特异地富集特定相关的 DNA, 而且 DNA 的分析无论以基因芯片还是深度测序都需要较大量的高质量 DNA, 所以这类试验往往需要非常大量的细胞, 成为该类研究的一个瓶颈。另外一方面, 如上所述, 如果允许以单个细胞或微量细胞进行整个全基因组范围的分析, 不仅取材更经济, 更重要的是细胞可能达到真正的“纯种”从而结果更精确也更可靠。

由于涉及不同性质和特点的核酸元件, 结构基因组学和功能基因组学分析所用的扩增技术并不完全相同^[9, 47]。转录谱的分析在现代生物学中应用非常

广泛, 分析平台可根据需要选择定量 PCR(qPCR)、基因芯片和最新的深度测序技术等, 可作为功能基因组学分析的一个范例。芯片分析和深度测序一般从 20 微克总 RNA 或 1~2 微克 mRNA 开始, 也有好的操作程序优化到其十分之一^[19, 48]。由于样品往往不足或者有部分降解而 RNA 研究又非常重要, 激起几代科学家为转录组的扩增而殚精竭虑, 迄今成就斐然^[49~51], 仅仅方法学方面的论文就有很多篇已发表在 *Nature*、*Science* 等杂志上, 迄今从事有关试剂生产的国际知名生物公司有十数家。尽管试剂盒众多, 但归纳起来, 现今转录组扩增方法主要有两大类: 以 PCR 技术为基础的指数扩增方法和以体外转录为基础的线性扩增方法, 但前者具有 PCR 本身的缺点但效率高, 后者(称为 Eberwine 方法)扩增效率较低, 操作较复杂但忠实性高。虽然有几种试剂盒都宣称可以扩增多到单个细胞的 mRNA 转录组, 但实际应用中并不多见单个细胞的研究报道。而且由于深度测序对样品的要求高于基因芯片分析, 后者又高于 qPCR, 即使相当数量细胞(而不是单个细胞)的 mRNA 扩增的质量也不易在深度测序平台上分析。目前, 虽然不断有新方法提出(如 Nugen 公司建立的以 Ribo-SPIA 技术为基础的等温线性扩增最近较引人注目), 上述两大类方法仍然是 mRNA 扩增的主力军。目前, 最新比较成功的一个尝试是由剑桥大学和 Applied Biosystems 公司建立的一种转录组扩增暨后续深度测序(mRNA-seq)方案, 它以 PCR 为基础扩增单个细胞的 mRNA 转录组^[52]。但是, 正如论文作者承认, 该方法不能扩增较长 mRNA 全长(一般不超过 3 kb)故不能完全检出这类基因(占 36%)不同剪切异构体的变化; 同时该方法扩增产物也不能保留原转录子的方向信息。microRNA 由于极短(只有 17 到 25 个碱基), 不便操作, 也不能用常规 mRNA 转录方法转化为 DNA。好在 microRNA 数量有限, 只有几百个, 有不同研究者建立的多种方法对单个 microRNA 分子或微量细胞多个 microRNA 作扩增和分析。Applied Biosystems 公司研究小组也建立了同时进行 220 个微 RNA 扩增的方法, 成功对单个胚胎干细胞的微 RNA 进行扩增, 并用 qPCR 检测最后结果^[53]。

染色质免疫共沉淀(ChIP)主要是通过研究转录因子的结合序列进而研究基因调节谱;组蛋白甲基化谱(组蛋白修饰)研究主要也是用ChIP方法,两者都是相关蛋白质和DNA的相互作用,一般都需要大量细胞作材料,单细胞水平的研究方法尚不能实现或接近实现。事实上,通常ChIP芯片或测序至少需要数千万个(10^7 至 10^8)细胞才能进行,减少ChIP所需细胞数量可以大大减少工作量,同时由于可以分析少量但更“纯”的细胞材料也使研究更深入,方法上也取得了重要进展。早期曾尝试的接头连接-PCR方案被发现序列偏差较大。加州大学(UC Davis)的Farnham研究组较早从事这方面研究^[54],达到仅用10 000个或更少细胞进行ChIP启动子芯片分析和有限基因的qPCR分析。奥斯陆大学Collas研究组用Sigma公司的WGA4全基因组PCR扩增方法,结合相应的操作程序优化,能够从1 000个细胞进行ChIP处理和芯片分析(ChIP-chip),也可进行100个细胞的ChIP及其有限相关基因的qPCR分析^[55]。基于深度测序的优点,最近,哈佛大学Bernstein研究小组采用与上述类似但改进的PCR扩增技术,从10 000个细胞开始经过ChIP获得10到50皮克DNA,扩增后成功进行了测序分析(ChIP-seq)并以此研究了血系细胞前体的发育程序^[56],这是目前的最好水平。

CpG岛是指基因组序列中广泛分布的富含CpG双碱基的DNA序列,一般长度在几百到几千碱基对,主要分布于基因的近5'末端调控区。全基因组范围内的DNA CpG甲基化状态构成甲基化组(DNA methylome)^[57]。多数基因的启动子被包含在CpG岛中,其超(高)甲基化或去(低)甲基化调节着基因的表达而又不改变基因序列结构。最近甚至有人利用那些甲基化程度变化显著的DNA甲基化位点作多态标记进行GWAS分析。DNA CpG甲基化检测方法主要有3类^[8, 58, 59]:用甲基化“CpG”结合蛋白或抗体(DNA免疫沉淀)富集超甲基化(或去甲基化)序列的方法;用甲基化敏感的限制内切酶特异剪切去甲基化位点(如HpaII)或超甲基化位点(如McrBC)的方法;用重亚酸盐转化DNA非甲基化“C”的方法。重亚酸盐转化DNA法与全库深度测序相结合,如果能实现全基因组单碱基水平分辨率的检测,将代表CpG甲基化作图的终极方向,但该方法在加工DNA时,DNA遭到

极大破坏,所需起始细胞数目很大。总之,作为重要的表观遗传学指标CpG甲基化检测问题,一般需要微克数量级甚至100微克DNA。虽然有CpG甲基化分析试剂盒声称可以用操作纳克级的DNA,但未见基因组水平的实际研究报道。当细胞数量较多时,PCR方法和MDA方法用来扩增DNA以达到可以在基因组规模上检测的水平。单细胞全基因组水平的DNA CpG甲基化分析一直是人们梦寐以求的目标。

核酸的扩增需要高效率、高灵敏、高特异、忠实而无偏差的扩增而且要求易于规范操作。我们研究小组在建立WPA技术的基础上,根据RNA和CpG甲基化分析规程的特点,进一步建立了在微量细胞(一般数个至数百个)甚至单细胞水平上进行全局(全基因组或转录组)功能基因组分析的若干技术,特别是单细胞全库全长mRNA扩增(例如,长至15 kb的mRNA可以完整扩增)技术和单细胞或微量细胞CpG甲基化作图技术,以解决上述有关问题,可选用基因芯片或深度测序平台来分析,获得了初步成功,正在神经元、干细胞等样品上测试并完善,期望在后基因组学的时代浪潮中推波助澜。

类似于单细胞基因组测序,在单细胞和微量数目细胞的功能基因组学研究上,我们也不能一叶障目而不见泰山。由于不同细胞个体,即使是由克隆培养的细胞群体,可能具有细胞之间的内在功能多态性,加上实验误差,采用平行样品的重复实验自然也是必不可少的。

5 展 望

从最小量的样品(终极目标是单细胞)获得全基因组的最大信息,和更高效率的经济实用的深度测序是目前基因组学迫切需要解决的两大技术问题^[20, 60]。基因组学的发展随着基因重组和克隆、桑格测序、PCR、基因芯片、深度测序等技术而后浪催前浪,波澜壮阔。我们相信,单细胞基因组学研究技术不仅将在生命科学研究的大海风生水起,一展身手,使研究和检测更深入更细致,从而带动基础科学新的发现,也将给人类对抗疾病、保障健康和提高生命寿命和质量带来很多新的机会。

参考文献(References):

- [1] Liu ET. Functional genomics of cancer. *Curr Opin Genet Dev*, 2008, 18(3): 251–256. [\[DOI\]](#)
- [2] Geschwind DH, Konopka G. Neuroscience in the era of functional genomics and systems biology. *Nature*, 2009, 461(7266): 908–915. [\[DOI\]](#)
- [3] Pieroni E, de la Fuente van Bentem S, Mancosu G, Capobianco E, Hirt H, de la Fuente A. Protein networking: insights into global functional organization of proteomes. *Proteomics*, 2008, 8(4): 799–816. [\[DOI\]](#)
- [4] Abu-Farha M, Elisma F, Zhou H, Tian R, Zhou H, Asmer MS, Figeys D. Proteomics: from technology developments to biological applications. *Anal Chem*, 2009, 81(12): 4585–4599. [\[DOI\]](#)
- [5] Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNAworld. *Genes Dev*, 2009, 23(13): 1494–1504. [\[DOI\]](#)
- [6] Chen PY, Meister G. MicroRNA-guided posttranscriptional gene regulation. *Biol Chem*, 2005, 386(12): 1205–1218. [\[DOI\]](#)
- [7] Butcher LM, Beck S. Future impact of integrated high-throughput methylome analyses on human health and disease. *J Genet Genomics*, 2008, 35(7): 391–401. [\[DOI\]](#)
- [8] Beck S, Rakyen VK. The methylome: approaches for global DNA methylation profiling. *Trends Genet*, 2008, 24(5): 231–237. [\[DOI\]](#)
- [9] Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 2008, 92(5): 255–264. [\[DOI\]](#)
- [10] Wold B, Myers RM. Sequence census methods for functional genomics. *Nat Methods*, 2008, 5(1): 19–21. [\[DOI\]](#)
- [11] 谭建新, 孙玉洁. 表观基因组学研究方法进展与评价. *遗传*, 2009, 31(1): 3–12. [\[DOI\]](#)
- [12] Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, 2008, 9(6): 465–476. [\[DOI\]](#)
- [13] Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat Rev Genet*, 2007, 8(4): 253–262. [\[DOI\]](#)
- [14] Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*, 2007, 8(4): 286–298. [\[DOI\]](#)
- [15] Gondo Y. Trends in large-scale mouse mutagenesis: from genetics to functional genomics. *Nat Rev Genet*, 2008, 9(10): 803–810. [\[DOI\]](#)
- [16] Gondo Y, Fukumura R, Murata T, Makino S. Next-generation gene targeting in the mouse for functional genomics. *BMB Rep*, 2009, 42(6): 315–323. [\[DOI\]](#)
- [17] Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*, 2008, 26(10): 1135–1145. [\[DOI\]](#)
- [18] Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet*, 2008, 24(3): 133–141. [\[DOI\]](#)
- [19] Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10(1): 57–63. [\[DOI\]](#)
- [20] Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*, 2010, 11(1): 31–46. [\[DOI\]](#)
- [21] Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. *Nat Biotechnol*, 2009, 27(9): 847–850. [\[DOI\]](#)
- [22] Ozsolak F, Platt AR, Jones DR, Reifengerger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. *Nature*, 2009, 461(7265): 814–818. [\[DOI\]](#)
- [23] Ozsolak F, Ting DT, Wittner BS, Brannigan BW, Paul S, Bardeesy N, Ramaswamy S, Milos PM, Haber DA. Amplification-free digital gene expression profiling from minute cell quantities. *Nat Methods*, 2010, 7(8): 619–621. [\[DOI\]](#)
- [24] Walker A, Parkhill J. Single-cell genomics. *Nat Rev Microbiol*, 2008, 6(3): 176–177. [\[DOI\]](#)
- [25] Ochman H. Single-cell genomics. *Environ Microbiol*, 2007, 9(1): 7. [\[DOI\]](#)
- [26] Hamady M, Knight R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res*, 2009, 19(7): 1141–1152. [\[DOI\]](#)
- [27] Weckx S, De Vuyst L. Metagenome and metatranscriptome analysis: does the flag always cover the cargo? *Int J Food Microbiol*, 2009, 133(3): 292–293. [\[DOI\]](#)
- [28] Baveye PC. To sequence or not to sequence the whole-soil metagenome? *Nat Rev Microbiol*, 2009, 7(10): 756; author reply 756–757. [\[DOI\]](#)
- [29] Rajendhran J, Gunasekaran P. Strategies for accessing soil metagenome for desired applications. *Biotechnol Adv*, 2008, 26(6): 576–590. [\[DOI\]](#)
- [30] Hutchison CA, 3rd, Venter JC. Single-cell genomics. *Nat Biotechnol*, 2006, 24(6): 657–658. [\[DOI\]](#)
- [31] Polyak K. Going small is the new big. *Nat Methods*, 2010, 7(8): 597, 599–600. [\[DOI\]](#)
- [32] Ma C, Donnelly DF, LaMotte RH. *In vivo* visualization and functional characterization of primary somatic

- neurons. *J Neurosci Methods*, 2010, 191(1): 60–65. [\[DOI\]](#)
- [33] Zhao R. From single cell gene-based diagnostics to diagnostic genomics: Current applications and future perspectives. *Clin Lab Sci*, 2005, 18(4): 254–262. [\[DOI\]](#)
- [34] Klein CA, Hölzel D. Systemic cancer progression and tumor dormancy: mathematical models meet single cell genomics. *Cell Cycle*, 2006, 5(16): 1788–1798. [\[DOI\]](#)
- [35] Lange BM. Single-cell genomics. *Curr Opin Plant Biol*, 2005, 8(3): 236–241. [\[DOI\]](#)
- [36] Lasken RS. Single-cell genomic sequencing using Multiple Displacement Amplification. *Curr Opin Microbiol*, 2007, 10(5): 510–516. [\[DOI\]](#)
- [37] Musmann M, Hu FZ, Richter M, de Beer D, Preisler A, Jorgensen BB, Huntemann M, Glockner FO, Amann R, Koopman WJ, Lasken RS, Janto B, Hogg J, Stoodley P, Boissy R, Ehrlich GD. Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol*, 2007, 5(9): e230. [\[DOI\]](#)
- [38] Ishoev T, Woyke T, Stepanauskas R, Novotny M, Lasken RS. Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol*, 2008, 11(3): 198–204. [\[DOI\]](#)
- [39] Dean FB, Hosono S, Fang LH, Wu XH, Faruqi AF, Bray-Ward P, Sun ZY, Zong QL, Du YF, Du J, Driscoll M, Song WM, Kingsmore SF, Egholm M, Lasken RS. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA*, 2002, 99(8): 5261–526. [\[DOI\]](#)
- [40] Lasken RS. Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochem Soc Trans*, 2009, 37(Pt 2): 450–453. [\[DOI\]](#)
- [41] Pan XH, Urban AE, Palejev D, Schulz V, Grubert F, Hu YP, Snyder M, Weissman SM. A procedure for highly specific, sensitive, and unbiased whole-genome amplification. *Proc Natl Acad Sci USA*, 2008, 105(40): 15499–15504. [\[DOI\]](#)
- [42] Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng JF, Eisen JA, Sieracki ME, Stepanauskas R. Assembling the marine metagenome, one cell at a time. *PLoS One*, 2009, 4(4): e5299. [\[DOI\]](#)
- [43] Janda JM, Abbott SL. 16s rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol*, 2007, 45(9): 2761–2764. [\[DOI\]](#)
- [44] Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. Target-enrichment strategies for next-generation sequencing. *Nat Methods*, 2010, 7(2): 111–118. [\[DOI\]](#)
- [45] Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, Melotte C, Debrock S, Amyere M, Vikkula M, Schuit F, Fryns JP, Verbeke G, D'Hooghe T, Moreau Y, Vermeesch JR. Chromosome instability is common in human cleavage-stage embryos. *Nat Med*, 2009, 15(5): 577–583. [\[DOI\]](#)
- [46] Ling JW, Zhuang GL, Tazon-Vega B, Zhang CH, Cao BQ, Rosenwaks Z, Xu KP. Evaluation of genome coverage and fidelity of multiple displacement amplification from single cells by SNP array. *Mol Hum Reprod*, 2009, 15(11): 739–747. [\[DOI\]](#)
- [47] Werner T. Next generation sequencing in functional genomics. *Brief Bioinform*, 2010, 11 (5): 499–511. [\[DOI\]](#)
- [48] Marguerat S, Bähler J. RNA-seq: from technology to biology. *Cell Mol Life Sci*, 2010, 67(4): 569–579. [\[DOI\]](#)
- [49] Wang E. RNA amplification for successful gene profiling analysis. *J Transl Med*, 2005, 3(1): 28. [\[DOI\]](#)
- [50] 李轶女, 胡英考, 张志芳, 沈桂芳. RNA扩增的研究进展. *遗传*, 2007, 29(8): 907–914. [\[DOI\]](#)
- [51] Ginsberg SD. RNA amplification strategies for small sample populations. *Methods*, 2005, 37(3): 229–237. [\[DOI\]](#)
- [52] Tang FC, Barbacioru C, Wang YZ, Nordman E, Lee C, Xu NL, Wang XH, Bodeau J, Tuch BB, Siddiqui A, Lao KQ, Surani MA. mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods*, 2009, 6(5): 377–382. [\[DOI\]](#)
- [53] Tang FC, Hajkova P, Barton SC, O'Carroll D, Lee C, Lao KQ, Surani MA. 220-plex microRNA expression profile of a single cell. *Nat Protoc*, 2006, 1(3): 1154–1159. [\[DOI\]](#)
- [54] Acevedo LG, Iniguez AL, Holster HL, Zhang XM, Green R, Farnham PJ. Genome-scale chip-chip analysis using 10,000 human cells. *Biotechniques*, 2007, 43(6): 791–797. [\[DOI\]](#)
- [55] Dahl JA, Reiner AH, Collas P. Fast genomic μ CHIP-chip from 1,000 cells. *Genome Biol*, 2009, 10(2): R13. [\[DOI\]](#)
- [56] Adli M, Zhu J, Bernstein BE. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nat Methods*, 2010, 7(8): 615–618. [\[DOI\]](#)
- [57] Fouse SD, Nagarajan RP, Costello JF. Genome-scale DNA methylation analysis. *Epigenomics*, 2010, 2(1): 105–117. [\[DOI\]](#)
- [58] Zilberman D, Henikoff S. Genome-wide analysis of DNA methylation patterns. *Development*, 2007, 134(22): 3959–3965. [\[DOI\]](#)
- [59] Shames DS, Minna JD, Gazdar AF. Methods for detecting DNA methylation in tumors: from bench to bedside. *Cancer Lett*, 2007, 251(2): 187–198. [\[DOI\]](#)
- [60] Wang DJ, Bodovitz S. Single cell analysis: The new frontier in 'omics'. *Trends Biotechnol*, 2010, 28(6): 281–290. [\[DOI\]](#)