

王康宇,王 义,孙春玉,等. RNA 测序的研究进展[J]. 江苏农业科学,2014,42(10):12-16.

RNA 测序的研究进展

王康宇,王 义,孙春玉,蒋世翠,张美萍

(吉林农业大学,吉林长春 130118)

摘要:RNA 测序研究是基因功能及结构研究的基础,能够从整体水平研究基因功能及其结构。随着高通量测序和定量检测技术的不断发展,能够通过 RNA 测序对转录组进行更深度更完整的研究。该研究进展包括改善转录起始位点的预测、链特异性测序、融合基因的检测、microRNA 定量的分析以及 RNA 可变剪切的识别。目前利用单分子测序技术可以实现 RNA 的直接测序,通过二代测序技术与单分子测序技术相结合的方式,能更深层次、更全面地获得转录组信息。

关键词:转录组信息;测序;编码 RNA;非编码 RNA;高通量测序技术;单分子测序技术

中图分类号: Q75 **文献标志码:** A **文章编号:** 1002-1302(2014)10-0012-05

1995 年 Velculescu 等首次提出了关于转录组的概念^[1],转录组广义上是指某一特定功能状态下,细胞内所有转录表达的基因总和,其中包括编码 RNA(mRNA)和非编码 RNA 如 tRNA、rRNA、snRNA、miRNA 等,而非编码 RNA 不能被转录识别,不能翻译成蛋白质,但是能参与某些蛋白质翻译过程;狭义上是指所有 mRNA 的总和^[2]。1995 年第 1 个转录组是由 Velculescu 等在酿酒酵母细胞中获得的,当时的技术共获得了 60 633 个转录本,揭示了 4 665 个基因,其中有 1 981 个基因是具有已知功能的,其他 2 684 个基因尚未被鉴定过^[1]。从人类基因组计划^[3]的实施开始,截至 2013 年 10 月已有 68 种植物和 119 种动物的基因组文章相继发表。高通量测序在过去十几年中快速发展,促使关于生物的功能基因组研究日

益兴起,人们利用测序技术研究了从简单模式生物(如酵母、拟南芥、水稻等)到人等一些高等物种的基因组中 DNA 修饰和 RNA 的定性定量变化等动态的基因组位点的特性。在对基因组测序和分析研究的同时,关于复杂的转录组研究也广泛发展起来。利用高通量测序技术平台分析转录组的结构和表达水平,更能挖掘未知转录本和稀有转录本,精确地识别 RNA 的可变剪切以及编码序列的单核苷酸多态性(SPN),更进一步解析复杂的转录组信息^[4]。

最初的转录组研究主要以基因芯片微阵列技术为基础,由于基因芯片技术的检测范围取决于芯上的探针信息,所以只能检测已知序列的特征,缺少发现新基因的能力,而高通量测序技术可以很好地弥补基因芯片技术在这方面的不足。因此,现阶段转录组的研究是借助于高通量的二代 DNA 测序技术(NGS)^[5]来完成的,通过构建 cDNA 文库并对 cDNA 进行高通量测序,分析测序结果进而解析转录组学中的复杂变化,这使 RNA 测序技术对基因芯片微阵列技术是极大的挑战。目前,以基因测序技术为核心的新技术平台支撑体系已经相对成熟和完善,例如:illumina 公司的 Solexa 测序技术、罗氏公

收稿日期:2013-12-31

基金项目:国家科技计划农村领域项目(编号:2013AA102604-3)。

作者简介:王康宇(1983—),吉林通化人,博士研究生,从事植物功能基因组学研究。E-mail: wky427@sina.com。

通信作者:张美萍,教授,博士生导师,从事植物基因组学与系统生物学研究。E-mail: wanglaoshi0606@163.com。

[16]冯立娟,苑兆和,尹燕雷,等. 槭属 2 品种叶变色期花青苷含量与相关酶活性的变化[J]. 林业科学,2009,45(8):56-60.

[17]姜卫兵,庄 猛,韩浩章,等. 彩叶植物呈色机理及光合特性研究进展[J]. 园艺学报,2005,32(2):352-358.

[18]晁月文,李竞芸,张广辉. 彩叶植物呈色机理及其育种研究进展[J]. 江苏林业科技,2008,35(4):46-48,52.

[19]Oren-Shamir M,Levi-Nissim A. UV-light effect on the leaf pigmentation of *Cotinus coggygia* 'Royal Purple'[J]. Scientia Horticulturae,1997,71(1/2):59-66.

[20]Deal D L,Raulston J C,Hinesley L E. Leaf color retention, dark respiration, and growth of red-leaved Japanese maples under high night temperatures[J]. Amer Soc Hort Sci,1990,115(1):135-140.

[21]张 琰,卓丽环,赵亚洲. 遮荫处理对“血红鸡爪槭”叶片色素及碳水化合物含量的影响[J]. 上海农业学报,2006,22(3):21-24.

[22]李玉娟,张 健,李 敏,等. 蔗糖和不同外源激素处理对美国红枫色叶的影响[J]. 广西农学报,2009,24(6):27-28,42.

[23]杨铁华,梁 鸣,佟 斌,等. 外源物质对元宝槭生长及生理特性的影响[J]. 国土与自然资源研究,2010(2):80-80.

[24]陈 睿,徐书霞,吕建洲. 外源蔗糖对红花槭叶色参数和色素含量的影响[J]. 天津农业科学,2012,18(2):14-16.

[25]孙 波,刘晓东,郑德丞. 假色槭叶色变化对土壤施 FeSO₄ 酸化处理的响应[J]. 东北林业大学学报,2008,36(9):51-52,58.

[26]韩 辉,宫 伟. 不同土壤酸碱度对紫花槭秋季叶色变化的影响[J]. 吉林农业,2010(6):76,80.

[27]唐 玲,李倩中,李淑顺,等. 秋季模拟酸雨对鸡爪槭叶片呈色相关生理的影响[J]. 江苏农业学报,2010,26(6):1357-1361.

[28]马 晓,陈 刚,张冬梅. 生态因子及矿质元素对红花槭叶片色素含量的影响[J]. 北方园艺,2012(13):86-88.

[29]安龙杰. 鸡爪槭叶片花色苷合成关键基因的克隆和序列分析[D]. 保定:河北农业大学,2012.

[30]洪 丽,王金刚,龚束芳. 彩叶植物叶色变化及相关影响因素研究进展[J]. 东北林业大学学报,2010,41(6):152-156.

司的 454 测序技术、ABI 公司的 SOLID 测序技术以及美国螺旋生物科学公司的新型纳米孔测序技术等^[5]。RNA 测序技术平台随着 NGS 技术的不断更新和提高而日益成熟完善,如测序通量、测序长度、错配率、碱基配对读取能力等测序性能方面技术的提高均有利于转录组的研究。

新 RNA 测序技术的不断更新和创新,为人类逐步全面了解真核生物和原核生物的转录组信息提供了新的定性和定量的生物信息学方法。本文阐述了改善转录起始位点的映射、RNA 特异链的测定、融合基因的检测、小 RNA 定性的分析以及 RNA 可变剪切的识别的发展,综述了利用单分子测序技术实现 RNA 的直接测序,通过二代测序技术与单分子测序技术相结合的方式更深层次、更全面的获得转录组信息,并展望了 RNA 测序在研究转录组的潜能。

1 RNA 测序的研究内容

1.1 转录起始位点的预测

转录起始位点(TSSs)是指 RNA 聚合酶识别和结合的位点,并且能够识别和调控每个转录本表达的启动子。第 1 个高通量获得转录起始位点预测方法是帽分析基因表达法(CAGE),该方法是由最早的 Sanger 测序法^[6]发展而来的,并能通过 cDNA 克隆获得完整的 RNA 帽子结构。该方法虽然对转录起始位点的预测有效,但需要大量高质量的 RNA 并且获得的转录起始位点很短,仅是 20~21 个核苷酸长度。

通过 NGS 技术发展,CAGE 方法得到了改进,经研究发现,通过该方法可以获得整个基因组范围内复杂的转录起始位点分布和独特的启动子。因此,CAGE 与 RNA 测序技术结合后衍生出了以 CAGE 策略为基础的 DeepCAGE 法^[7]、nano-CAGE 法^[8]、CAGEscan 法^[8]以及 PEAT 法^[9],同时以 Sanger 测序为基础的分析转录起始位点预测的方法受到了 RNA 测序技术与 CAGE 结合的挑战。例如,nanoCAGE 法^[8]解决了 GAGE 法需要大量 RNA 的缺点,可以通过放大技术从 10 ng 的总 RNA 量获得转录起始位点的映射;PEAT 法和 CAGEscan 法^[8-9]双末端测序可以获得转录起始位点映射以及转录起始位点下游区间,具有良好的连通性并能促进识别特殊的转录本。此外,双末端测序缓解了对单个短读取重复区的校对,通过 RNA 测序可以获得序列的重复特性。虽然这些方法结合了 NGS 技术,克服了一些 CAGE 方法的不足,但是也存在一定的弊端。例如,在检测扩增结果过程中的操作步骤,可能影响了转录起始位点出现的频率^[8];此外,在 cDNA 合成和测序的过程中会产生引物二聚体,这就减少了测序的有效结果^[8]。因此,这些与 RNA 测序技术结合的方法,虽然在定性检测方面很有效,但是在定量检测方面还需要进一步的提高和优化。

以 RNA 测序为基础的转录起始位点预测研究具有依赖 cDNA 合成和测序技术的局限性,这个局限性主要是由 RNA 的结构和序列特点造成的。此外,以 RNA 测序为基础的转录起始位点映射很难捕获那些转录水平高且自身能够快速降解的转录本,如 microRNA。解决这些限制需要 RNA 测序技术与其他方法相结合,如以染色体为基础的转录起始位点的预测,依赖于对组蛋白修饰后对转录本进行有效的检测^[10-11]。转录后加工出现 5'帽子结构的 RNA 片段有利于被检测^[12]。

因此,单单依赖于 GAGE 法获得的转录起始位点具有在转录后加工中难分离的难题。

1.2 链特异性测序

在关于转录组学的研究中发现,物种中普遍存在反义转录现象。反义转录的生物学功能明晰,它能在生物体正常生理状态和病理状态下发挥各种作用^[13]。因此,在更深层次研究转录组时,对正义链和反义链的测序和分析研究成为了一个重点。标准的 RNA 测序方法一般需要合成双链的 cDNA,这样会丢失 RNA 链的部分信息。此外,在 cDNA 第 1 条链合成后需要依赖 DNA 聚合酶(DDDP)反转录产生 cDNA 的第 2 条链,此过程会引入虚假的信息^[14-15],这能混淆转录检测的分辨率。作为抑制反转录酶 DDDP 活性的物质,放线菌素 D 的有效抑制作用尚未被报道^[16]。为克服这些难题,已经开发了链特异性 RNA 测序的分析策略。

对特异链信息的获得依赖于 3 种方式:第 1 种方法是在 RNA 尾部或 cDNA 第 1 条链预定方向连上接头,已知方向上的接头被作为获取 RNA 链信息的参考点;第 2 种方法是直接对 cDNA 的第 1 条链进行测序;第 3 种方法是在合成的 cDNA 第 2 条链或 RNA 上进行选择性标记。这些方法可以让大家更了解反义转录过程,包括反义 RNA 转录位点图谱的建立如核糖体转录位点和识别新的反义转录的启动子。以酿酒酵母的基因组作为参考,比较这 3 种方法^[17],结果发现关于链特异性水平、平均覆盖率、注释信息、建库复杂性以及定量表达谱分析等具有差异。关于外界添加酶的添加对这些数据造成的特异性偏差依然缺乏更深入的研究,这些数据合理处理的问题也成为日后研究的重点。

首先,逆转录酶在合成 cDNA 的第 1 条链并转录形成 cDNA 第 2 条链时具有倾向性,目前尚不明晰这个依赖于 cDNA 第 1 条链测序的方法是否完全具有单链的特异性^[15-16]。这些方法通过对单链特异性进行比较分析获得已知反义链的方向、基因注释以及相对读取位置。有研究表明,小部分的读取方向是反义方向,所以这些链可能不完全具有单链特异性。此外,cDNA 第 1 条、第 2 条链不能恰当地对参考序列进行定位^[17]。对基因组正义和反义的转录给出了不完全的注释,即使在酿酒酵母这样模式物种中,这些方法也不能完全确定链的特异性。

其次,添加接头的方法具有序列的偏好。依赖于接头的方法会存在不同表现性的偏差,这种偏差存在于转录组分析和核糖体分析中^[18-19]。与使用 3'端多聚腺嘌呤酶的获得的文库相比使用接头获得的文库存在覆盖率不均匀的现象^[20]。

最后,其中一些方法包含使用溶液或者添加步骤都会添加外来物如 DNA 聚合酶的使用,例如,对 GC 形式的偏差和重复形式的读取。RNA 模板对 GC 形式的存在有一定的偏差,所以应该拥有中度的 GC 含量^[21]。而对重复序列的读取是特异链 RNA 测序主要解决的问题,这些影响因素能通过链特异性 RNA 测序技术的发展或已有测序技术的改变和提高得到解决。

1.3 可变剪接的识别

目前,已知 15%~60%的突变来源于 RNA 可变剪接,完整的 RNA 可变剪接事件是分析和了解细胞分化和疾病发生的关键点^[22]。可变剪接有 6 种基本形式,即内含子保留、可

变的 5'端、可变的 3'端、外显子盒、互斥外显子、可变的起始或末端外显子(这 2 种形式更有可能是可变启动子、可变 polyA 位点造成的)。因此,可变剪切事件的识别在 RNA 测序中具有重要性。最早利用 RNA 测序方法对可变剪切位点识别的研究受测序读取长度的限制,因此最初的 RNA 测序对可变剪接的研究通过使用计算机来弥补这一限制。在人的基因组中超过 95% 的外显子基因具有可变剪切的发生,每个组织中含有 110 000 个新型剪接位点,可变剪切事件改变了对人类基因组的组装,从而获得了人的基因,数量为 35 000^[23]。通过计算每个外显子读取基因的数量和每个剪接点的生成,决定每个接点剪接效率和不同种类亚型的水平^[24]。

通过改进目前的 RNA 测序技术来增加读取的长度,能更好地映射到具有可变剪切的外显子上。通过改进测序技术能提高读取区的分块^[25],调整定位基因组每个独立的区块。此外,通过改进双末端测序方法能从转录本的预计读取间距离 2 点上获得更多的测序信息。现在在不需要先前已知转录信息的基础上就能识别可变剪接事件的发生,识别可变剪接、转录本连通性以及基因组组装方式需要获得全长转录本序列,这些在未来可能产生新兴技术。

1.4 基因融合检测

基因融合技术是将不同的基因连接起来,从而表达具有复合功能的融合蛋白,融合蛋白除了具有衍生因子的双重活性外,还具有融合蛋白的活性高于衍生因子相加的活性。RNA 测序技术结合计算机进行分析不仅可以对转录起始位点进行识别,还可以用来检测有组织细胞中的基因融合现象,其中用于生物学方面的研究尤为重要。利用单端测序和双末端测序相结合的策略更易检测到基因组 DNA 的易位和基因组重排现象^[26]。然而,RNA 测序技术能够更好地识别物种中产生异常变化的 RNA 种类,使检测因功能性或互作关系引起物种基因融合导致病变的研究具有可能性。此外,以基因组 DNA 为基础方法不能识别基因融合是由于非基因组因素影响,如转移剪接或者相邻转录本间通读所引起的融合现象^[26]。双末端配对的 RNA 测序能够提高基因覆盖率,因此对检测基因融合具有特别的优势,现在该方法被用于病理学的研究,并为调控与治疗提供潜在的可能性。

对基因融合检测所面临的最大难题是与之并行存在的 RNA 可变剪接。此外,RNA 测序分析不能检测包括拥有编码序列的其他基因的启动子融合现象,同时在 RNA 测序时含有嵌合有外界添加制品的 cDNA 测序模板也可能导致基因融合识别中假阳性的出现^[27]。但是,通过增加足够 RNA 测序的测序通量和测序读取片段长度等技术的改进,可能会降低检测基因融合过程中假阳性的产生^[5]。

1.5 microRNA 定量分析

NGS 技术对 sRNA 的发现和鉴定的影响特别明显。对 miRNA 的研究已经成为目前世界上 RNA 测序技术研究的热点,miRNA 最初的发现和鉴定是通过焦磷酸测序进行的^[28],后来随着高通量 NGS 平台的使用,大量的 miRNA 被发现并引起了研究的重视。但 NGS 样品的准备需要更长的 RNA (>200 个核苷酸),因此二代测序对 miRNA 研究不合适^[29]。可见反转录和随机引物测序的策略,为其提供了解决问题的新途径。

目前,研究 miRNA 的 RNA 测序方法的一个重要问题是 miRNA 测序数据的标准化。随着测序技术的进步,关于 miRNA 的研究思路逐渐变得清晰,通过以 NGS 为基础的 miRNA 定量分析可以通过差异表达分析完成,但是还存在如何获得每个 miRNA 的读取数量不是找出其实际表达丰度的有效办法^[30],这些差异更有可能是由在样品准备和测序时的偏差所造成的。所以,新兴技术能否提高 sRNA 定量测定与分析让人拭目以待。

综上所述,高通量测序技术的不断发展使得测序量和成本逐渐降低,但是还存在如何获得足够通量的测序覆盖度和其他一些潜在的问题。在转录过程中,等位基因在表达上存在差异,导致低丰度转录组本与基因型的测定和分析很难完成,这就要求获得转录过程越丰富越好,优选优质的转录组测序可以尽量减少测序的总成本并使分析样品数最大化。

2 单分子测序技术对 RNA 的直接测序

2.1 RNA 测序的影响因素

目前,cDNA 的合成和其他 RNA 操作极大地限制了 RNA 测序技术的应用。现今许多 RNA 测序方法依赖于 cDNA 的合成和以及一系列的合成后操作步骤,这使得 RNA 测序的应用具有一定的局限性。例如,人工条件下合成的 cDNA 第 2 条链在特异链 RNA 测序时相当困难。为了在特异链建库时避免这一问题,使用 RNA-RNA 连接和建库的方法费力且难构建,因此可以通过合成 cDNA 作为模板来突破这个限制^[31]。在反转录时,合成初期的 cDNA 有时会从 RNA 中分离出来,通过再退火与 RNA 序列特异性结合到初始模板上进行延伸,会人工产生嵌合的 cDNA。通过模板切换能解决在外显子-内含子界限的识别和嵌合转录上所产生的问题。同时,反转录酶在无引物条件下由于 RNA 的二级结构也能自由合成 cDNA,这将导致随机的 cDNA 被合成^[32],这是由于反转录酶与其他酶相比保真度低,缺乏校正机制^[32]。所以,在 RNA 反转录成 cDNA 时转化效率受试验条件影响。

RNA 测序技术除了 cDNA 影响因素外,还有其他影响因素。第一,RNA 测序信号具有不均匀覆盖性,这就可能在转录过程中因一些因素的影响产生偏差^[33],如随机引物选择、cDNA 合成以及连接等。第二,使用统一的 RNA 测序策略,导致在 RNA 或 cDNA 读取长度选择的不合理,以致转录本长度存在偏差,给下游分析造成了严重后果。第三,RNA 测序定量分析需要考虑到不确定的读段长度和每个转录本读段正常化^[34]。尽管随着转录方法的提高和生物信息学的发展允许转录组从头合成测序,但存在的方法通常难以检测确定的转录以及其覆盖全部长度。因 RNA 存在可变剪切,启动子、转录位点以及转录边界和长度存在不确定性,正常长度是 RNA 测序定量分析的潜在错误源。第四,RNA 测序策略往往涉及到 mRNA 的富集,因此利用 RNA 转录聚合酶 I 对 mRNA 的 polyA 进行富集可以获得理想的 RNA 产物。

2.2 单分子测序的研究进展

目前,RNA 测序和分析技术发展受到诸多条件的限制,单分子测序技术的出现解决了现阶段二代测序技术解决不了的问题,因为单分子测序技术可以不经反转录成 cDNA,不改变 RNA 的原本特性,直接对 RNA 分子进行测序。但是单

分子测序也有其不足之处,使得 RNA 直接测序技术发展前景充满了未知。

二代高通量测序的读取序列片段太短,使得后续的组装工作困难且繁琐。而三代高通量测序采用的是单分子测序原理,不仅可以降低测序每个碱基的费用,简化测序样品的制备流程,加快测序速度,缩短分析数据的过程,而且还具有读取数百个碱基甚至更长片段的优点,运用在从头测序时使数据分析过程简化了^[35]。这种长片段的测序技术与短片段测序相比在后续组装中得到了简化,因而省去了不必要的麻烦,同时还增加了转录组数据分析的新内容,如拷贝数变异情况、可变剪切识别、转位情况、嵌合转录本识别以及基因型分期等。

RNA 直接技术是随着 Helicos 公司的 Heliscope 并行单分子测序技术发展起来的。现阶段主要的单分子测序技术分为 Helicos 公司的 Heliscope 并行单分子合成测序技术、Pacific 公司的 SMRT 单分子实时合成测序技术以及 Oxford Nanopore 公司的纳米孔单分子测序技术。单分子测序技术具有二代测序技术不具备的优点,如不在需要制备测序所用的 DNA 文库可以对样品片段直接进行单分子扩增、测序的速度更快、测序过程具有并行度以及读取长度更长等。

Heliscope 并行单分子合成测序技术的原理是该技术基于边合成边测序的思想,将待测的 RNA 样品连接上已知序列的接头[通常为 poly(A)]并于单通道与含有 poly(T)的共价引物进行杂交,再用过精确定位的成像系统鉴定具有标记的测序片段,其具有以下特点:第一,测序可以不是同步的,可以同时读取不同长度的片段,这就改善了二代测序读取长度均一的策略,读取长度不均一的测序模板可造成测序速度快慢不等。第二,在荧光标记的核苷酸上没有终止基团。在二代测序时,测序模板存在同聚物时带有荧光标记的碱基之间与同聚物结合会发生猝灭的现象,这样就无法识别读取的碱基数,而 Heliscope 单分子测序技术能记录所读取的每个碱基数据,这样就避免了猝灭现象的发生。第三,可以通过 2 步测序(正向测序和反向测序)提高测序的准确性。第四,测序过程中会有未标记的碱基、不发荧光的碱基以及污染碱基的混入,会使测定缺失突变时的错误率较高,2 步测序 2 次误差率分别为 2%~7%、0.2%~1.0%,但是在测定碱基替换突变时的错误率非常低,2 步测序 2 次错误率分别在 0.01%~1.00%、0.001%,这是目前测序技术中准确率最高的^[36]。

SMRT 单分子实时合成测序技术的原理同样也是基于边合成边测序的思路,该技术以 SMRT 芯片作为测序载体进行测序反应。SMRT 芯片是一种带有需要零模波导孔(zero-mode waveguides,ZMW)的且孔的厚度为 100 nm 的金属片。将 DNA 聚合酶、测序模板以及带有不同荧光标记的 dNTP 加入到 ZMW 孔中,进行边合成边测序反应^[37]。dNTP 的磷酸基团被荧光标记,当 dNTP 被添加到合成链上时,进入 ZMW 孔的激光束会激发荧光,根据不同的荧光成像就能获得测序结果;而添加到合成链上的 dNTP 的磷酸基团会被剪切并释放,这样不再具有荧光标记,便不会再被识别。因此,SMRT 单分子实时测序技术的测序速度很快,其测定碱基的速度可以达到 10 个/s^[38],测序过程中对每条碱基链的数据都会进行评估,更易发现稀有变异;具有超长的读长度,即平均读长度超过 1 000 bp 甚至更长等。但是,也存在缺陷,如不能高效利用

DNA 聚合酶且容易在 ZMW 中降解失活、总体上每个碱基测序成本高、机器相当的昂贵等。

纳米孔单分子测序技术的原理与前两者不同,采样的是电信号测序技术,通过借助电泳驱动单个分子逐一通过纳米孔来实现测序^[39]。纳米孔的直径非常小,仅允许单个核苷酸通过,因此可用于高通量测序。此外,纳米级别的孔径保证了测序过程中良好的检测过程,使测序结果的准确性非常高。RNA 测序时省略的扩增阶段和标记手段可直接测序,使简便快捷的测序过程更可能成为现实。虽然纳米孔单分子测序技术具有测序成本很低和测序长度很长的优点,但是在如何控制纳米孔径、如何通过纳米孔径控制速度以及纳米孔材料和制造等问题是该技术面临的重大问题。纳米孔径过大会造成一次性通过的核苷酸过多,过小会造成单个核苷酸无法通过;通过纳米孔径的速度会影响测序速度,通过速度过慢不能实现快速高通量测序,而通过过快也不能确保识别信号的稳定性;纳米孔制作的材料要求高,制造起来极其费时且其价格很贵。这些因素影响了该技术发展的脚步,解决这些问题对直接测序技术具有里程碑的意义。

总之,单分子测序技术是 RNA 直接测序的主要技术支持。单分子测序技术在测序成本、错配率、读取长度以及测序通量等技术指标上比二代测序技术有了更大的改进和提高。用于 RNA 测序时也解决了反转录成 cDNA 以及 RNA 含量低和长短不均一等造成无法测序的问题,通过简化测序步骤和对测序样品不需要预处理就使得 RNA 直接测序得以实现。但是单分子测序技术还存在着不足,所以利用二代测序技术和单分子测序技术相结合的方法可以解决目前研究中存在的不少问题,如 SNP 位点识别、基因缺失研究以及基因拷贝数的变异分析等。

3 展望

测序技术从 Sanger 测序技术发展到目前阶段的单分子测序技术,在这个过程中 RNA 测序技术的发展还在继续,并且朝着更快速、更灵敏、更精准、更廉价的方向发展。RNA 测序研究是基因功能研究及结构研究的基础,能够从整体水平上研究基因功能以及基因结构。单分子测序技术推动 RNA 测序技术更进一步发展,RNA 直接测序技术研究转录组已经逐步替代基因芯片微阵列技术成为现在功能基因组学研究基因表达的主流方式。

参考文献:

- [1] Velculescu V E, Zhang L, Vogelstein B, et al. Serial analysis of gene expression[J]. Science, 1995, 207: 484-487.
- [2] Berretta J, Morillon A. Pervasive transcription constitutes a new level of eukaryotic genome regulation[J]. EMBO Reports, 2009, 10(9): 973-982.
- [3] Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome[J]. Nature, 2001, 409(6822): 860-921.
- [4] Denoeud F, Aury J M, da Silva C, et al. Annotating genomes with massive-scale RNA sequencing[J]. Genome Biology, 2008, 9(12): R175.
- [5] Metzker M L. Sequencing technologies the next generation[J]. Nature Rev Genet, 2010, 11: 31-46.

- [6] Carninci P, Sandelin A, Lenhard B, et al. Genome – wide analysis of mammalian promoter architecture and evolution[J]. *Nature Genetics*, 2006, 38(6): 626 – 635.
- [7] Valen E, Pascarella G, Chalk A, et al. Genome – wide detection and analysis of hippocampus core promoters using DeepCAGE [J]. *Genome Research*, 2009, 19(2): 255 – 265.
- [8] Ni T, Corcoran D L, Rach E A, et al. A paired – end sequencing strategy to map the complex landscape of transcription initiation[J]. *Nature Methods*, 2010, 7(7): 521 – 527.
- [9] Plessy C, Bertin N, Takahashi H, et al. Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEScan [J]. *Nature Methods*, 2010, 7(7): 528 – 534.
- [10] Marson A, Levine S S, Cole M F, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells [J]. *Cell*, 2008, 134(3): 521 – 533.
- [11] Ozsolak F, Poling L L, Wang Z X, et al. Chromatin structure analyses identify miRNA promoters[J]. *Genes & Development*, 2008, 22(22): 3172 – 3183.
- [12] Affymetrix ENCODE Transcriptome Project, Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post – transcriptional processing generates a diversity of 5′ – modified long and short RNAs [J]. *Nature*, 2009, 457: 1028 – 1032.
- [13] Faghihi M A, Wahlestedt C. Regulatory roles of natural antisense transcripts[J]. *Nature Reviews Molecular Cell Biology*, 2009, 10(9): 637 – 643.
- [14] Gubler U. Second – strand cDNA synthesis: mRNA fragments as primers[J]. *Meth Enzymol*, 1987, 152: 330 – 335.
- [15] Spiegelman S, Burny A, Das M R, et al. DNA – directed DNA polymerase activity in oncogenic RNA viruses [J]. *Nature*, 1970, 227(5262): 1029 – 1031.
- [16] Perocchi F, Xu Z Y, Clauder – Münster S, et al. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D[J]. *Nucleic Acids Research*, 2007, 35(19): e128.
- [17] Levin J Z, Yassour M, Adiconis X, et al. Comprehensive comparative analysis of strand – specific RNA sequencing methods[J]. *Nature Methods*, 2010, 7(9): 709 – 715.
- [18] Faulhammer D, Lipton R J, Landweber L F. Fidelity of enzymatic ligation for DNA computing[J]. *Journal of Computational Biology*, 2000, 7(6): 839 – 848.
- [19] Housby J N, Southern E M. Fidelity of DNA ligation: a novel experimental approach based on the polymerisation of libraries of oligonucleotides [J]. *Nucleic Acids Research*, 1998, 26(18): 4259 – 4266.
- [20] Ingolia N T, Ghaemmaghami S, Newman J R, et al. Genome – wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling[J]. *Science*, 2009, 324(5924): 218 – 223.
- [21] Kozarewa I, Ning Zemin, Quail M A, et al. Amplification – free illumina sequencing – library preparation facilitates improved mapping and assembly of GC – biased genomes[J]. *Nature Methods*, 2009, 6(4): 291 – 295.
- [22] Nilsen T W, Graveley B R. Expansion of the eukaryotic proteome by alternative splicing[J]. *Nature*, 2010, 463(7280): 457 – 463.
- [23] Wang E T, Sandberg R, Luo S J, et al. Alternative isoform regulation in human tissue transcriptomes [J]. *Nature*, 2008, 456(7221): 470 – 476.
- [24] Jiang H, Wong W H. Statistical inferences for isoform expression in RNA – Seq[J]. *Bioinformatics*, 2009, 25(8): 1026 – 1032.
- [25] Trapnell C, Pachter L, Salzberg S L. TopHat: discovering splice junctions with RNA – Seq [J]. *Bioinformatics*, 2009, 25(9): 1105 – 1111.
- [26] Korbel J O, Urban A E, Affourtit J P, et al. Paired – end mapping reveals extensive structural variation in the human genome [J]. *Science*, 2007, 318(5849): 420 – 426.
- [27] McManus C J, Duff M O, Eipper – Mains J, et al. Global analysis of trans – splicing in drosophila[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(29): 12975 – 12979.
- [28] Rajagopalan R, Vaucheret H, Trejo J, et al. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana* [J]. *Genes & Development*, 2006, 20(24): 3407 – 3425.
- [29] Taft R J, Glazov E A, Cloonan N, et al. Tiny RNAs associated with transcription start sites in animals [J]. *Nature Genetics*, 2009, 41(5): 572 – 578.
- [30] Linsen S E, de Wit E, Janssens G, et al. Limitations and possibilities of small RNA digital gene expression profiling[J]. *Nature Methods*, 2009, 6(7): 474 – 476.
- [31] Cocquet J, Chong A, Zhang G L, et al. Reverse transcriptase template switching and false alternative transcripts [J]. *Genomics*, 2006, 88(1): 127 – 131.
- [32] Mader R M, Schmidt W M, Sedivy R, et al. Reverse transcriptase template switching during reverse transcriptase – polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA[J]. *The Journal of Laboratory and Clinical Medicine*, 2001, 137(6): 422 – 428.
- [33] Hansen K D, Brenner S E, Dudoit S. Biases in illumina transcriptome sequencing caused by random hexamer priming [J]. *Nucleic Acids Research*, 2010, 38(12): e131.
- [34] Li B, Ruotti V, Stewart R M, et al. RNA – Seq gene expression estimation with read mapping uncertainty[J]. *Bioinformatics*, 2010, 26(4): 493 – 500.
- [35] 邱超, 孙含丽, 宋超. DNA 测序技术发展历程及国际最新动态[J]. *硅谷*, 2008(17): 127, 129.
- [36] 汪正范, 刘娜. Helicos 公司单分子基因测序仪[J]. *现代仪器*, 2010(1): 95.
- [37] Levene M J, Korlach J, Turner S W, et al. Zero – mode waveguides for single – molecule analysis at high concentrations [J]. *Science*, 2003, 299(567): 682 – 686.
- [38] Clarke J, Wu H C, Jayasinghe L, et al. Continuous base identification for single – molecule nanopore DNA sequencing [J]. *Nature Nanotechnology*, 2009, 4(4): 265 – 270.
- [39] Rusk N. Cheap third – generation sequencing[J]. *Nature Methods*, 2009, 6(4): 244 – 245.