基于第二代测序技术的基因资源挖掘

陈昊[1,2], 谭晓风[1,2,*]

中南林业科技大学[1]经济林培育与保护省部共建教育部重点实验室, [2]经济林育种与栽培国家林业局重点实验室, 长沙 410004

摘要: 转录组研究一直是生命科学研究的一个重要方向, 在第二代测序技术问世以前, 已经产生了一些行之有效的转录组研究方法, 但这些方法存在一定的局限性。第二代测序技术的出现不仅使转录组研究很快进入了高速发展期, 同时也为遗传资源的挖掘提供了一套全新的技术平台。本文简要介绍了第二代测序技术的化学原理和特性, 重点阐述了利用第二代测序技术进行转录组测序, 从而在此基础上挖掘遗传资源的研究。

关键词: 第二代测序技术; 化学原理; 转录组测序; 遗传资源

Excavation of Genic Resources Based on Next Generation Sequencing Technologies

CHEN Hao[1,2], TAN Xiao-Feng[1,2,*]

[1]The Key Laboratory of Cultivation and Protection for Non-wood Forest Trees, Ministry of Education, [2]The Key Laboratory of Non-wood Forest Products of State Forestry Administration, Central South University of Forestry and Technology, Changsha 410004, China

Abstract: Transcriptome research has been an important part of life science research. Several effective methods used in transcriptome research have been developed before the advent of next-generation sequencing technologies. However, these methods have some limitations in transcriptome researches. The advent of next-generation sequencing technologies not only prompt transcriptome researches into a high-speed development period but also introduce a whole new technology platform into studies of excavating genetic resources. This review briefly introduced the chemical principles and features of next-generation sequencing technologies and focused on summarizing the excavation of genetic resources based on the transcriptome sequencing using next-generation sequencing technologies.

Key words: next-generation sequencing technologies; chemical principles; transcriptome sequencing; genetic resources

研究基因的表达模式和细胞功能之间的关系一直是生物学家们首要关注的课题, 但其前提是找到表型相关基因。与基因组DNA相比, 由于排除了转录水平基因表达调控等因素的干扰, mRNA与基因在机体中行使的功能的关系更加密切。因此, 在转录组水平进行新基因的挖掘和功能研究越来越受到生物学家们的重视。最早研究细胞转录组的尝试是检测来源于不同器官、组织的总RNA以鉴定感兴趣的转录本存在与否或确定其数量。第一个基于候选基因的转录组研究应用了Northern杂交技术, 但该技术操作复杂, 需要的起始RNA量大, 且使用对人体有害的放射性同位素标记探针, 因此只能在一次实验中检测能够大量获得RNA样本的已知转录本(Alwine等1977)。反转录定量PCR (RT-qPCR)技术的发展使得转录本的检测更为容易和方便, 并增加了实验的通量, 显著减少了RNA的用量(Becker-André和Hahlbrock 1989; Noonan等1990)。尽管如此, 从第一次应用RT-PCR到现在已经过去了几十年, 一次RT-PCR实验还是只能检测几百个转录本, 远远未达到转录组研究的实验通量要求(VanGuilder等2008)。Liang和Pardee (1992)发明的差异显示技术基于使用3′端oligo dT引物和5′端随机引物进行的cDNA合

收稿 2014-03-18 修定 2014-06-19

资助 国家林业公益性行业科研专项重大项目(201204403)。

* 通讯作者(E-mail: tanxiaofengcn@126.com; Tel: 0731-85623416)。

成。这一特定条件使得差异显示技术很难利用已有的数据库进行分析(Bachem等1996)。

微阵列技术能够同时检测上万个已知或推测的转录本表达水平, 它的出现极大的促进了转录组研究的发展(Schena等1995)。通过一定的技术手段, 在1.28平方厘米的面积上能装载409 000个点, 理论上拟南芥(*Arabidopsis thaliana*)所有基因的表达模式可以通过一块这样的微阵列进行分析(Kehoe等1999)。微阵列的灵敏度很高, 能在十万分之一或五十万分之一的水平上检测到特定的mRNA (Gerhold等1999)。此外, 随着微阵列技术的发展, 该技术还被用于noncoding RNA、SNP (single nucleotide polymorphism)和mRNA可变剪接的研究(Mockler等2005)。尽管微阵列能够同时分析上万个基因的表达, 但很难检测未知的转录本, 也不能获得检测到的转录本的编码序列。另外, 由于微阵列是通过杂交信号强度来间接的推断而不是直接测定转录本的表达丰度, 这样获得的数据存在很多不确定的因素, 从而影响了实验的可重复性和各样本数据之间的比较。进行微阵列实验所需费用高, 需要开发合适的软件和标准化分析方法来比较不同批次实验和不同实验室的实验结果, 且在一些实验体系中一块微阵列只能使用一次, 这些因素制约了微阵列技术的发展和应用(van Hal等2000)。

利用DNA测序的手段来研究转录组已表现出取代微阵列研究方法的趋势。随着人类和一些主要模式生物基因组测序工作的完成, 重测序的研究被推向了前沿。这些研究的发展得益于一些被称为“第二代测序技术”的测序方法(Roche/454、Illumina、Applied Biosystems SOLiD)的出现, 这些测序方法比Sanger测序方便快捷且成本便宜得多, 测序通量也远远高于最先进的Sanger自动测序仪, 由此增大了对转录组的覆盖率, 从而大大加快了新基因发现与鉴定的进程。本文重点介绍利用第二代测序技术进行转录组测序, 从而在此基础上挖掘遗传资源的研究。

1 第二代测序技术的化学原理

焦磷酸测序(pyrosequencing)技术成为除了Sanger法之外第一种可选择的商业化测序方法, 其在Roche/454测序平台中得到了应用(Margulies等2005)。焦磷酸测序技术基于4种酶(DNA polymerase、sulfurylase、luciferase、apyrase)在同一PCR反应体系中催化的酶级联化学发光反应, 无需电泳, 使用dNTP而不是ddNTP且无需荧光标记, 因此操作简单易行。另一种第二代测序平台Illumina Genome Analyzer与454/Roche相同, 也是应用了边合成边测序(sequencing-by-synthesis)的原理, 但两者的信号检测方法不同(Bentley 2006)。Illumina测序反应所用的核苷酸类似物经四种不同的荧光标记, 且其具有DNA链延长终止的作用, 但这种终止反应通过一定的化学手段处理后是可逆的, 经过改造的DNA聚合酶可将这种核苷酸类似物整合进合成的DNA链中。与上述基于DNA聚合酶合成反应的边合成边测序方法不同, SOLiD (Supported Oligonucleotide Ligation and Detection System)系统是使用连续的分子杂交和连接来间接推测序列的测序方案(Shendure等2005)。用于测序的DNA样本随机打断后两端加上接头构建测序文库, 文库片段与磁珠上的接头引物配对。磁珠沉积在特殊的玻片表面后加入接头配对引物和8碱基单链荧光探针混合物开始第一轮测序, 通过荧光信号可以知道1、2位的碱基序列。探针和引物通过连接酶连接后经化学处理使探针在3′端的5、6位碱基之间断开并除去6至8位的碱基和荧光标记, 再加入探针进行连接, 这时由于DNA链已经延长了5个碱基, 通过荧光信号检测到的则是模板链6、7位的碱基序列, 如此重复几次。重置后进行第二轮测序, 第二轮测序所用的接头配对引物比第一轮的引物少了一个碱基, 因此第二轮测序经第一次连接探针后所得的模板序列为0、1位, 第二次为5、6位, 依次类推。经过几轮测序后, 每轮测序获得的错开序列就可以拼接成完整的序列。SOLiD测序的最大特点是测序过程中DNA链的延伸是通过连接酶催化的DNA片段的连接而不是通过DNA聚合酶加上单个碱基。

2 基于第二代测序技术的基因资源挖掘

2.1 蛋白质编码基因的注释

尽管包括人类在内的一些生物已经利用Sanger测序法进行了全基因组测序, 但目前对这些基因组数据的了解并不深入, 还有大量有用的信息等待挖掘(Brent 2008)。第二代测序技术能够轻

易的填补Sanger法测序留下的数据空缺, 例如454测序仪的一轮反应可以产生400 000个EST (expressed sequence tags), 而以往的Sanger测序研究一次只能产生720个EST (Bainbridge等2006)。在基因组注释研究中, EST被用来与参考基因组进行比对从而区分基因的外显子与内含子, 进而找到基因转录的边界。转录组测序数据可以与同一生物的基因组进行比对(*cis* alignment), 在同类生物没有可用的参考基因组时则可以与相关物种的基因组进行比对(*trans* alignment)。到目前为止, 第二代测序技术已被用来构建了许多模式生物的EST文库(Morozova和Marra 2008)。通过将测序结果与相似物种的基因组序列进行比对, EST测序能够高效的获取无参考序列物种的基因组序列并对其进行注释, 例如Novaes等(2008)利用454测序技术获得了14 Mb无任何基因组信息的大桉(*Eucalyptus grandis*) EST数据。同样, 学者们利用454测序技术对玉米(*Zea mays*) (Emrich等2007)基因组进行了注释。由于每个454测序反应产生的序列在第二代测序技术中是最长的, 因此由其测序产生的EST序列能够高效的进行转录组重头组装等序列分析(Vera等2008)。与传统测序技术相比, 利用454 测序系统对拟南芥转录组进行重测序时检测出更多基因座位上的转录本, 使整体序列覆盖度增加了50%, 数量增多10% (Wall等2009)。与454技术相比, 尽管Illumina和SOLiD每个反应产生的序列更短, 从而更有利于进行序列的重头组装, 但需要开发出更高级的进行短序列组装的算法(Zerbino和Birney 2008)。在一些物种中, 这些短序列数据已经成功的被应用于检测新的外显子和未知的mRNA剪接位点(Cloonan等2008)。

2.2 基因表达谱分析

从上世纪九十年代开始, 基因芯片逐步被应用于检测大规模基因表达水平变化的研究。然而基于核酸分子杂交的芯片技术最大的局限性在于其只能检测已知基因的mRNA而不能检测未知的mRNA, 且杂交信号的灵敏度有限, 很难检测到低丰度表达量基因在特定环境下表达量的变化。第二代测序技术无需知道物种的核酸序列, 无需标记探针, 灵敏度高, 理论上可以检测某种核酸分子在细胞中的绝对数量, 因而在基因表达谱研究中表现出逐步取代基因芯片的趋势。

Marioni等(2008)比较了Illumina测序和基因芯片在发现差异表达基因研究时的灵敏度, 结果表明, 在相同错误发现率(false discovery rate, FDR)下, Illumina比基因芯片多检测到了30%的差异表达基因, 且其测序数据具有高度的可重复性。疾病影响、逆境胁迫或生物体的不同发育时期等特定环境下生物基因表达谱研究已成为第二代测序技术的重点应用领域。Sajnani等(2012)利用454测序技术发现了一些在口腔癌变组织中表达而在正常组织中不表达的基因。Kakumanu等(2012)利用Illumina技术研究了干旱对受精后子房以及叶片分生组织表达谱的影响, 结果表明受精后子房对干旱胁迫的敏感度要远高于叶片。Eveland等(2010)利用Illumina技术对玉米发育不同时期的转录组进行了研究, 发现了一些对花序发育进行调控的基因。近年来, 基于第二代测序技术的表达谱分析在疾病研究、药物开发、生物体发育、生物抗逆性等研究领域扮演越来越重要的角色。

2.3 非编码RNA (noncoding RNA)的检测

第二代测序技术的出现极大的促进了有关ncRNA的研究。虽然基因芯片被广泛的应用于ncRNA表达谱的研究, 但不同物种的ncRNA序列保守性不高从而导致通过生物信息学的方法来发现未知的ncRNA基因存在一定的局限性(Xu等2008)。相反, 产生大量短序列测序数据的第二代测序技术能够有效的在基因组范围内发现新的miRNA和siRNA基因。除此之外, 第二代测序技术也能够有效的检测到已知miRNA的突变体、RNA的剪接加工以及miRNA的目标RNA配对(German等2008; Reid等2008)。到目前为止, 已经利用454技术在多种植物中鉴定了ncRNA (Axtell等2006; Barakat等2007; Yao等2007; Zhao等2007)。更高通量的Illumina和SOLiD技术同样能够用来建立深度miRNA文库。Glazov等(2008)在鸡胚胎中检测到了449个未知和所有已知的miRNA。利用Illumina技术进行ncRNA的研究在一些植物中得到了很好的开展。玉米*MOP1*基因是拟南芥*RDR2*基因的同源基因, Nobuta等(2008)通过Illumina测序发现, 与拟南芥*rdr2*突变体不同的是, 在玉米*mop1-1*突变体中, 尽管24 nt的siRNA显著减少, 但维持了较高水

平的22 nt siRNA, 由此推测存在另一种异染色质siRNA的产生机制。Vidal等(2013)结合RNA-Seq和sRNA-Seq的数据进行分析, 在拟南芥根中发现了一些新的硝酸盐应答基因。第二代测序技术在小RNA研究方面的另一突出成绩是发现了一类以前未知的不同于miRNA和siRNA的小RNA。这类小RNA被命名为piRNA (Piwi-interacting RNA), 它们与特定的蛋白质作用形成RNA-蛋白质复合体后在许多物种生殖细胞系的基因转录沉默过程中起作用(Lau 等2006; Houwing等2007)。

2.4 eQTL (expression quantitative trait loci)分析

Jansen和Nap (2001)提出了eQTL作图的概念, 即将来自分离群体的各基因型的表达数据作为一个数量性状, 利用传统的QTL分析方法进行分析。直到第二代测序技术问世之初, 基因芯片仍是进行eQTL分析的主流技术并在拟南芥、玉米等植物中得到成功应用(West等2007; Shi等2007)。第二代测序技术表现出了基因芯片所不具备的技术优势, 尤其是通过第二代测序技术能够获取表达谱芯片无法提供的等位基因特异性表达(allele-specific expression, ASE)信息(Skelly等2011; Sun和Hu 2013), 因此近年来, 越来越多的eQTL研究基于第二代测序技术展开。Li等(2013)利用Illumina技术对玉米重组自交系茎尖组织的基因转录丰度进行了eQTL分析, 结果表明一些基因在后代中的表达水平和变异程度无法通过其在父母本之间的差异进行预测, 这表明植物体内存在一类基因表达影响因子, 这些因子使得基因表达数据偏离了孟德尔遗传定律的预测。Lowry等(2013)对拟南芥基因组Illumina重测序数据和干旱胁迫下的基因芯片表达谱数据进行关联分析后发现, 仅有极少数的eQTLs与干旱胁迫相关, 并且这些eQTLs定位于基因组的高重组率区域, 它们的启动子序列具有较高的多态性。随着统计学算法的不断完善, 通过第二代测序数据进行eQTL作图, 将大大加快基因定位研究的进程(Sun 2012; Sun和Hu 2013)。

2.5 单核苷酸多态性的检测

基因组测序表明, 单核苷酸的改变在同一物种的不同个体之间是大量存在的(Wheeler等2008)。发生在编码区的单核苷酸突变很可能会引起基因编码蛋白质功能的改变。尽管所有类型的遗传多样性都能够通过对基因组的重测序来鉴定, 但这一方法的成本高昂, 不利于其推广。由于研究对象仅限于基因组的编码区, 因而转录组测序显著降低了测序成本。相比而言, 利用相同的方法对基因组测序则需要获得数据量大得多的测序数据, 以达到相同的外显子覆盖率, 从而可靠检测编码区多态性(Wang等2008)。

利用第二代测序技术检测SNP时一个值得注意的问题是测序过程中的碱基读取错误率。Illumina测序发生单碱基错误的概率在0.3%至3.8%之间(Dohm等2008), 454/Roche测序的单碱基错误率为4% (Huse等2007)。SOLiD测序的正确率高达99.94%, 但这一数值并不能直接与Illumina和454/Roche的数值进行比较, 因为这三种测序平台产生的数据格式不同(Ondov等2008)。第二代测序数据的错误率可以通过深度冗余测序来纠正, 从而准确的检测到SNP (Harismendy等2009)。冗余测序不可避免的会增加测序的成本且结果最后仍需用Sanger测序法加以验证。尽管如此, 一些计算机程序被用来剔除低质量数据以降低错误率(Huse等2007)。近年来, 已有越来越多的植物SNP研究用到了第二代测序技术。Huang等(2010)利用Illumina测序技术对517个水稻地方品种进行重测序, 获得了3 600万个SNP, 随后在此基础上对这些地方品种群体的14个重要农艺性状进行了全基因组关联分析(genome wide association studies, GWAS), 这一研究成果标志着水稻全基因组关联分析时代的到来(Clark 2010)。Trick等(2012)利用Illumina测序技术对小麦(*Triticum turgidum*)的近等基因系(near isogenic lines)进行了转录组测序, 开发了一些差异染色体片段特异性的SNP标记, 研究结果证实了通过RNA-Seq在多倍体物种中发现SNP的可行性。Hendre等(2012)利用Illumina测序技术对赤桉(*Eucalyptus camaldulensis*)群体41个生长相关基因的SNP位点进行了分析。

3 高通量转录组研究方法存在的问题和展望

高通量测序技术在转录组研究中相对于传统方法来说具有很大的优势, 但其局限性也不容忽视。首先, 如何更好的分析越来越大量的数据对研究者来说是一大挑战(van Vliet 2010)。由于第二代测序获得的数据量巨大, 比对和拼接速度是

衡量某一分析方法好坏的第一要素, 这导致基于伯罗斯-惠勒变换(Burrows-Wheeler transform)算法的工具如Bowtie (Langmead等2009)和SOAP (Li等2009)的出现, 从而大大加快了数据分析的进程。将reads错误比对到序列相似基因则会导致“影子转录本(transcript shadowing)”的出现(Pepke等2009; Trapnell等2010)。另外, 剪接变异现象也对数据比对和拼装造成了干扰, 但双末端测序产生的更长的链特异性reads具有更大的几率跨越剪接位点, 从而降低序列拼装的难度(Levin等2010)。如何经济有效的存储第二代测序产生的大量数据也是一个重大的挑战(Baker 2012)。其次, 高通量测序的成本仍然偏高。虽然在大规模测序方面, 高通量测序技术相比于Sanger测序节省了大量的人力物力, 但在质粒测序、PCR产物测序等小规模测序方面, 高通量测序没有成本优势。第二代测序1次反应动辄数千到数万的费用使一般的研究者难以接受, 这时Sanger测序仍然是首选。第三, 高通量测序所需的起始样本量大, 这使得来源有限的样本的分析受到了限制。Tang等(2009)通过对小鼠四细胞胚胎期单个卵裂球的mRNA进行PCR扩增, 成功分析了单个细胞的转录组。但这一方法仍有缺陷, 例如不能检测到没有多聚A尾的mRNA, 不能保留转录本的原始方向信息等。尽管第二代测序技术有上述缺陷, 但作为新兴的高通量测序手段, 其在转录组研究方面已表现出其它传统转录组研究方法无可比拟的优势。相信随着测序成本的逐步降低以及数据处理和分析方法的不断完善, 第二代测序技术必将在转录组学的研究中占据主导地位并得到更加广泛的应用。

参考文献

Alwine JC, Kemp DJ, Stark GR (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci USA, 74 (12): 5350~5354

Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006). A two-hit trigger for siRNA biogenesis in plants. Cell, 127 (3): 565~577

Bachem CW, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG (1996). Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J, 9 (5): 745~753

Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V et al (2006). Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genomics, 7: 246

Baker M (2012). Gene data to hit milestone. Nature, 487 (7407): 282~283

Barakat A, Wall K, Leebens-Mack J, Wang YJ, Carlson JE, Depamphilis CW (2007). Large-scale identification of microRNAs from a basal eudicot (*Eschscholzia californica*) and conservation in flowering plants. Plant J, 51 (6): 991~1003

Becker-André M, Hahlbrock K (1989). Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). Nucleic Acids Res, 17 (22): 9437~9446

Bentley DR (2006). Whole-genome re-sequencing. Curr Opin Genet Dev, 16 (6): 545~552

Brent MR (2008). Steady progress and recent breakthroughs in the accuracy of automated genome annotation. Nat Rev Genet, 9 (1): 62~73

Clark RM (2010). Genome-wide association studies coming of age in rice. Nat Genet, 42 (11): 926~927

Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G et al (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods, 5 (7): 613~619

Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res, 36 (16): 105

Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res, 17 (1): 69~73

Eveland AL, Satoh-Nagasawa N, Goldshmidt A, Meyer S, Beatty M, Sakai H, Ware D, Jackson D (2010). Digital gene expression signatures for maize development. Plant Physiol, 154 (3): 1024~1039

Gerhold D, Rushmore T, Caskey CT (1999). DNA chips: promising toys have become powerful tools. Trends Biochem Sci, 24 (5): 168~173

German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R et al (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol, 26 (8): 941~946

Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML (2008). A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res, 18 (6): 957~964

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S et al (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol, 10 (3): 32

Hendre PS, Kamalakannan R, Varghese M (2012). High-throughput and parallel SNP discovery in selected candidate genes in *Eucalyptus camaldulensis* using Illumina NGS platform. Plant Biotechnol J, 10 (6): 646~656

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB et al (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell, 129 (1): 69~82

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z et al (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet, 42 (11): 961~967

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol, 8 (7): 143

Jansen RC, Nap JP (2001). Genetical genomics: the added value from segregation. Trends Genet, 17 (7): 388~391

Kakumanu A, Ambavaram MM, Klumas C, Krishnan A, Batlang U, Myers E, Grene R, Pereira A (2012). Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant Physiol, 160 (2): 846~867

Kehoe DM, Villand P, Somerville S (1999). DNA microarrays for studies of higher plants and other photosynthetic organisms. Trends Plant Sci, 4 (1): 38~41

Langmead B, Trapnell C, Pop M, Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10 (3): R25

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE (2006). Characterization of the piRNA complex from rat testes. Science, 313 (5785): 363~367

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods, 7 (9): 709~715

Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MC et al (2013). Mendelian and non-Mendelian regulation of gene expression in maize. PLoS Genet, 9 (1): 1003202

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics, 25 (15): 1966~1967

Liang P, Pardee AB (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science, 257 (5072): 967~971

Lowry DB, Logan TL, Santuari L, Hardtke CS, Richards JH, DeRose-Wilson LJ, McKay JK, Sen S, Juenger TE (2013). Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in *Arabidopsis*. Plant Cell, 25 (9): 3266~3279

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z et al (2005). Genome sequencing in open microfabricated high-density picolitre reactors. Nature, 437 (7057): 376~380

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res, 18 (9): 1509~1517

Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR (2005). Applications of DNA tiling arrays for whole-genome analysis. Genomics, 85 (1): 1~15

Morozova O, Marra MA (2008). Applications of next-generation sequencing technologies in functional genomics. Genomics, 92 (5): 255~264

Nobuta K, Lu C, Shrivastava R, Pillay M, De Paoli E, Accerbi M, Arteaga-Vazquez M, Sidorenko L, Jeong DH, Yen Y et al (2008). Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the *mop1-1* mutant. Proc Natl Acad Sci USA, 105 (39): 14958~14963

Noonan KE, Beck C, Holzmayer TA, Chin JE, Wunder JS, Andrulis IL, Gazdar AF, Willman CL, Griffith B, Von Hoff DD (1990). Quantitative analysis of MDR1 (multidrug resistance) gene expression in human tumors by polymerase chain reaction. Proc Natl Acad Sci USA, 87 (18): 7160~7164

Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D, Sederoff RR, Kirst M (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics, 9: 312

Ondov BD, Varadarajan A, Passalacqua KD, Bergman NH (2008). Efficient mapping of Applied Biosystems SOLiD sequence data to a reference genome for functional genomic applications. Bioinformatics, 24 (23): 2776~2777

Pepke S, Wold B, Mortazavi A (2009). Computation for ChIP-seq and RNA-seq studies. Nat Methods, 6 (11): 22~32

Reid JG, Nagaraja AK, Lynn FC, Drabek RB, Muzny DM, Shaw CA, Weiss MK, Naghavi AO, Khan M, Zhu H et al (2008). Mouse let-7 miRNA populations exhibit RNA editing that is constrained in the 5'-seed/cleavage/anchor regions and stabilize predicted mmu-let-7a:mRNA duplexes. Genome Res, 18 (10): 1571~1581

Sajnani MR, Patel AK, Bhatt VD, Tripathi AK, Ahir VB, Shankar V, Shah S, Shah TM, Koringa PG, Jakhesara SJ et al (2012). Identification of novel transcripts deregulated in buccal cancer by RNA-seq. Gene, 507 (2): 152~158

Schena M, Shalon D, Davis RW, Brown PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270 (5235): 467~470

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science, 309 (5741): 1728~1732

Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Lübberstedt T (2007). Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population. BMC Genomics, 8: 22

Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome Res, 21 (10): 1728~1737

Sun W (2012). A statistical framework for eQTL mapping using RNA-seq data. Biometrics, 68 (1): 1~11

Sun W, Hu Y (2013). eQTL mapping using RNA-seq data. Stat Biosci, 5 (1): 198~219

Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X,

Bodeau J, Tuch BB, Siddiqui A (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods, 6 (5): 377~382

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol, 28 (5): 511~515

Trick M, Adamski NM, Mugford SG, Jiang CC, Febrer M, Uauy C (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biol, 12: 14

VanGuilder HD, Vrana KE, Freeman WM (2008). Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques, 44 (5): 619~626

van Hal NL, Vorst O, van Houwelingen AM, Kok EJ, Peijnenburg A, Aharoni A, van Tunen AJ, Keijer J (2000). The application of DNA microarrays in gene expression analysis. J Biotechnol, 78 (3): 271~280

van Vliet AH (2010). Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol Lett, 302 (1): 1~7

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol, 17 (7): 1636~1647

Vidal EA, Moyano TC, Krouk G, Katari MS, Tanurdzic M, McCombie WR, Coruzzi GM, Gutiérrez RA (2013). Integrated RNA-seq and sRNA-seq analysis identifies novel nitrate-responsive genes in *Arabidopsis thaliana* roots. BMC Genomics, 14: 701

Wall PK, Leebens-Mack J, Chanderbali AS, Barakat A, Wolcott E, Liang H, Landherr L, Tomsho LP, Hu Y, Carlson JE et al (2009). Comparison of next generation sequencing technologies for transcriptome characterization. BMC Genomics, 10: 347

Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J et al (2008). The diploid genome sequence of an Asian individual. Nature, 456 (7218): 60~65

West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. Genetics, 175 (3): 1441~1450

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT et al (2008). The complete genome of an individual by massively parallel DNA sequencing. Nature, 452 (7189): 872~876

Xu Y, Zhou X, Zhang W (2008). MicroRNA prediction with a novel ranking algorithm based on random walks. Bioinformatics, 24 (13): 50~58

Yao Y, Guo G, Ni Z, Sunkar R, Du J, Zhu JK, Sun Q (2007). Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). Genome Biol, 8 (6): 96

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res, 18 (5): 821~829

Zhao T, Li G, Mi S, Li S, Hannon GJ, Wang XJ, Qi Y (2007). A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. Genes Dev, 21 (10): 1190~1203