

全基因组测序与转录组测序大比拼

过去，人类常见病的研究主要是利用全基因组关联研究（GWAS）来筛查共有的变异体。近年来，测序技术的蓬勃开展以及个性化医疗的目标让科学家们从共性转移到个性，努力发现稀有的变异体。

尽管新一代测序平台的出现让测序费用迅速下降，然而每个碱基的费用仍然阻碍了更多完整测序的基因组出现。除了经济上的限制，人们还普遍意识到很难从如此多的变异体中鉴定出真正起作用的位点。基于这些原因，研究人员通常一开始就关注编码区的变异体。这也就催生了外显子捕获技术。罗氏 NimbleGen 于去年初最先推出了外显子捕获芯片，从基因组 DNA 中抓出外显子部分，然后测序。这一步也是价格不菲，于是人们退一步，以转录组测序（RNA-Seq）作为替代。尽管这种方法无疑将错过一些低表达基因，但它的优势是产生了更多的信息，比如基因表达水平和剪接模式。

这些方法究竟孰优孰劣？美国杜克大学医学院人类基因组变异中心以及国家癌症研究所的研究人员将高覆盖度全基因组测序和 RNA-Seq 所鉴定出的变异体进行了比较，来系统研究 RNA-Seq 在鉴定人类编码变异体方面究竟效果如何。这个研究结果发表在 5 月 28 日的《Genome Biology》上。

研究人员从同一个个体的外周血单核细胞（PBMC）中提取出 DNA 和 RNA，并利用 Illumina 的 [Genome Analyzer II 测序仪](#) 对 cDNA 和 gDNA 分别测序。gDNA 的测序产生了 14.5 亿个读数，每个读长 75 bp。cDNA 的测序产生了 2.8 亿个读数，一半读长 75 bp，一半读长 68 bp。

之后，研究人员利用 SAMtools 来检出两个序列中的单核苷酸变异体（SNV）。插入缺失和大的

结构变异体不包括在内。在 gDNA 中，SAMtools 检出了外显子中的 51,055 个 SNV，在 cDNA 中则检出了 64,128 个。其中，gDNA 的 48,740 个和 cDNA 中的 40,605 个通过了质量控制过滤。

他们还直接评估了 RNA-Seq 在鉴定变异体上的灵敏度和特异性，并评估了覆盖度和基因的表达水平这些关键因素如何相互作用，影响表现。研究人员发现，在全基因组测序中所鉴定出的外显子变异体中，只有 40% 被 RNA-Seq 所捕获，然而，若只集中在 PBMC 表达的基因，这个数字就上升到 81%。研究人员还发现，在处理 RNA-Seq 数据时，假阳性率高可能是个问题，特别是当覆盖度水平高时。

作者认为，只要有高表达基因的组织来源，并执行了适当的质量控制筛选，在发现高水平表达基因的编码变异体时，RNA-Seq 是一种快速廉价的替代方法。（生物通 余亮）

原文检索：

Screening the human exome: a comparison of whole genome and whole transcriptome sequencing

Genome Biology 2010, 11:R57
doi:10.1186/gb-2010-11-5-r57