Transcriptomics

- 产生背景
- 概念、特点及意义
- 研究方法
- 经典案例
- 应用

产生背景

- 基因组研究可以通过序列预测潜在的基因，从而进行研究。但遗传信息从DNA传递到蛋白质的过程中，存在着RNA水平和蛋白质水平上遗传信息的加工，为研究带来困难。
- 越来越多的基因测序工作已完成，接下来的问题是这些基因的功能是什么、不同的基因参与了哪些细胞内不同的生命过程、基因表达的调控、基因与基因产物之间的相互作用、以及相同的基因在不同的细胞内或者疾病和治疗状态下表达水平等等。因此，在人类基因组项目后，转录组的研究迅速受到科学家的青睐。

概念

- 转录组（Transcriptome）的概念由Velculescu VE等首先提出。**转录组是指在一个细胞、组织或生物体中出现的全套RNA，包括：信使RNA（mRNA）、rRNA、tRNA及其他非编码RNA。**转录组指特定细胞在某一功能状态下全部表达的基因总和，代表了每一个基因的身份和表达水平。同一细胞在不同的生长时期及生长环境下，基因表达情况是不完全相同的，具有特定的空间性和时间性特征。

Cell, Vol. 88, 243–251, January 24, 1997,

Characterization of the Yeast Transcriptome

Victor E. Velculescu,*† Lin Zhang,‡ Wei Zhou,‡
Jacob Vogelstein,† Munira A. Basrai,§
Douglas E. Bassett Jr.,*§‖ Phil Hieter,*§
Bert Vogelstein,*†‡ and Kenneth W. Kinzler†
*Program in Human Genetics and Molecular Biology
†Oncology Center
‡Howard Hughes Medical Institute
§Department of Molecular Biology and Genetics
The Johns Hopkins University School of Medicine
Baltimore, Maryland 21231
‖National Center for Biotechnology Information
National Library of Medicine
Bethesda, Maryland 20894

Summary

We have analyzed the set of genes expressed from the yeast genome, herein called the transcriptome, using serial analysis of gene expression. Analysis of 60,633 transcripts revealed 4,665 genes, with expression levels ranging from 0.3 to over 200 transcripts per cell. Of these genes, 1981 had known functions, while 2684 were previously uncharacterized. The integration of positional information with gene expression data allowed for the generation of chromosomal expression maps identifying physical regions of transcriptional activity and identified genes that had not been predicted by sequence information alone. These studies provide insight into global patterns of gene expression in yeast and demonstrate the feasibility of genome-wide expression studies in eukaryotes.

特点

- 与基因组具有静态实体的特点不同，转录组是受外源和内源因子调控的。因此，它是物种基因组和外部物理特征的动态联系，是反映生物个体在特定器官、组织或某一特定发育、生理阶段细胞中所有基因表达水平的数据。可用来比较不同组织或生理状况下基因表达水平差异，发现与特定生理功能相关的基因，推测未知基因。

意义

- 转录组学是功能基因组学的重要组成部分，是一门在整体水平上研究某一阶段特定组织或细胞中全部转录本的种类、结构和功能，以及转录调控规律的科学。转录组学从一个细胞或组织基因组的全部mRNA 水平研究基因表达情况，它能够提供全部基因的表达调节系统和蛋白质的功能、相互作用的信息。转录组学研究作为一种整体的方法，改变了单个基因的研究模式，将基因组学研究带入了高速发展的时代。

从转录组概念提出至今，转录组的研究内容已经逐步丰富，包括：

- **1. 非编码区域功能研究：Non-coding RNA研究、microRNA前体研究等**
- **2 .转录本结构研究：UTR鉴定、Intron边界鉴定、可变剪切研究、Start codon鉴定等**
- **3. 基因转录水平研究**
  4. 全新转录区域研究

研究方法

- 随着研究的深入，人们建立了一系列的方法和技术用于转录组学的研究，目前研究转录组主要有以下几种方法：表达序列标签（Expressed Sequence Tag，EST）、基因表达序列分析（Serrial Analysis of Gene Expression，SAGE）、基因芯片、cDNA-AFLP、大规模平行信号测序系统MPSS (massively parallel signature sequencing， MPSS) 、RNA-Seq、生物信息学（Bioinformatics）等等。

RNA-Seq

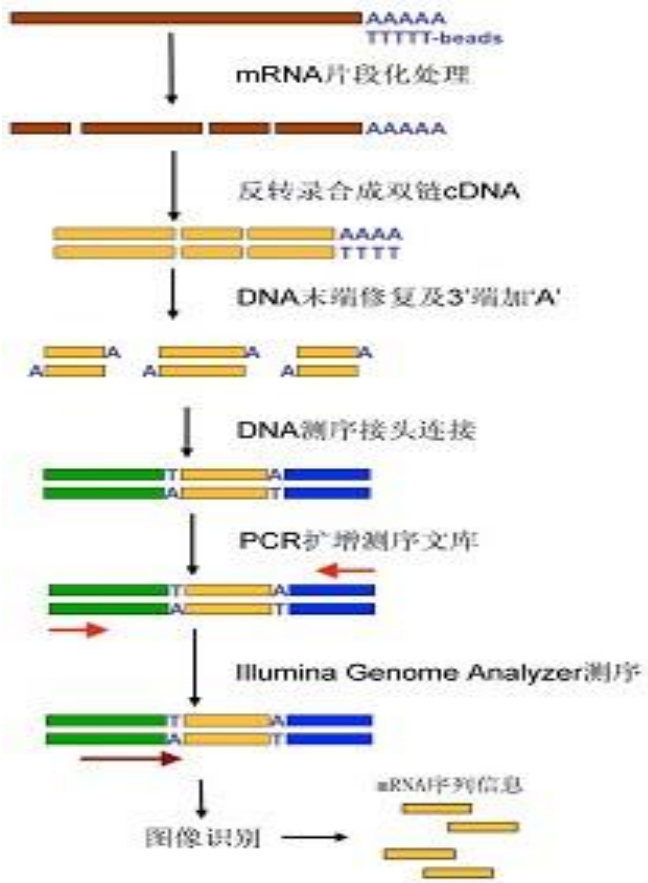
RNA-Seq即全转录组测序，通过高通量测序技术对cDNA测序，通过统计相关reads数计算出不同mRNA的表达量，发现转录水平的SNP，新的mRNA。如果有基因组参考序列，可以把转录本映射回基因组，确定转录位、剪切情况等遗传信息。

实验流程：

1.样品RNA测序文库构建
(1)使用oligo dT微珠纯化mRNA及mRNA片段化处理
(2)反转录反应合成合成双链cDNA
(3)双链DNA末端修复及3’末端加‘A’
(4)使用特定的测序接头连接DNA片段两端
(5)高保真聚合酶扩增构建成功的测序文库
2.DNA成簇（Cluster）扩增
3.高通量测序（Illumina Genome Analyzer IIx）

- 4.数据分析：原始数据读取，与数据库比对并进行注释，深层次数据分析。如基因覆盖率和测序深度分析，基因表达差异分析，基因结构分析，鉴定选择性剪切现象，发现新基因，鉴定基因融合现象。

测序实验流程：


AAAAA
TTTTT-beads
mRNA片段化处理
AAAAA
反转录合成双链cDNA
AAAA
TTTT
DNA末端修复及3'端加'A'
DNA测序接头连接
PCR扩增测序文库
Illumina Genome Analyzer测序
mRNA序列信息
图像识别

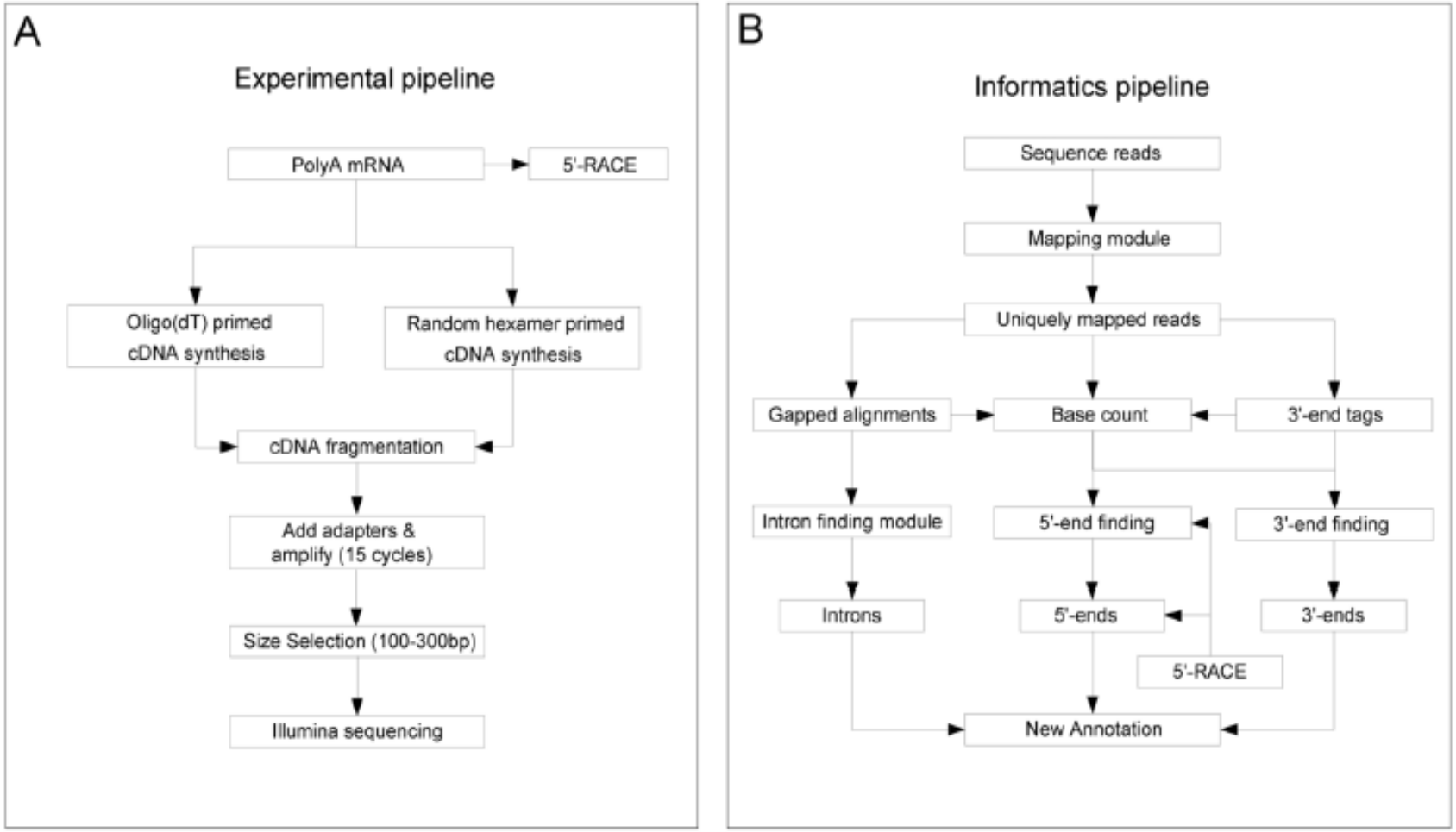
Published in final edited form as:
Science. 2008 June 6; 320(5881): 1344–1349. doi:10.1126/science.1158441.

The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing

Ugrappa Nagalakshmi, **Zhong Wang**, **Karl Waern**, **Chong Shou**, **Debasish Raha**, **Mark Gerstein**, and **Michael Snyder**
Department of Molecular, Cellular, and Developmental Biology, Program in Computer Science, Department of Molecular, Biophysics and Biochemistry, Yale University, New Haven, CT 06520

Abstract

Although many genome sequences have been determined, identification of genes and their elements such as untranslated regions (UTRs), introns, and coding regions is still a significant challenge. We have developed a novel sequencing-based method called RNA-Seq in which cDNA fragments are subjected to high throughput sequencing using the Illumina platform, and short reads are computationally mapped to the genome to identify the transcribed regions. We have successfully applied RNA-Seq to generate a high-resolution transcriptome map of the yeast genome. We demonstrate that most (74.5%) of the unique sequence of the yeast genome is transcribed. We used this method to globally map 5′ UTR and 3′ UTR boundaries and confirmed many known and predicted introns and demonstrated that others are not actively used. Our results suggest an alternative initiation codon from that annotated for a number of known genes and demonstrate that many yeast genes contain upstream open reading frames (uORFs). We also found unexpected 3′ end heterogeneity and the presence of many overlapping genes. We also found many novel transcribed regions not identified by other methods. These results indicate that the yeast transcriptome is more complex than previously appreciated. Furthermore, RNA-Seq is demonstrated to be at least as accurate as DNA microarrays for quantifying RNA expression levels and has a much larger dynamic range. We expect that RNA-Seq will be a valuable general approach for high resolution mapping of transcriptomes in many organisms.


A
Experimental pipeline
PolyA mRNA
5'-RACE
Oligo(dT) primed cDNA synthesis
Random hexamer primed cDNA synthesis
cDNA fragmentation
Add adapters & amplify (15 cycles)
Size Selection (100-300bp)
Illumina sequencing
B
Informatics pipeline
Sequence reads
Mapping module
Uniquely mapped reads
Gapped alignments
Base count
3'-end tags
Intron finding module
5'-end finding
3'-end finding
Introns
5'-ends
3'-ends
5'-RACE
New Annotation

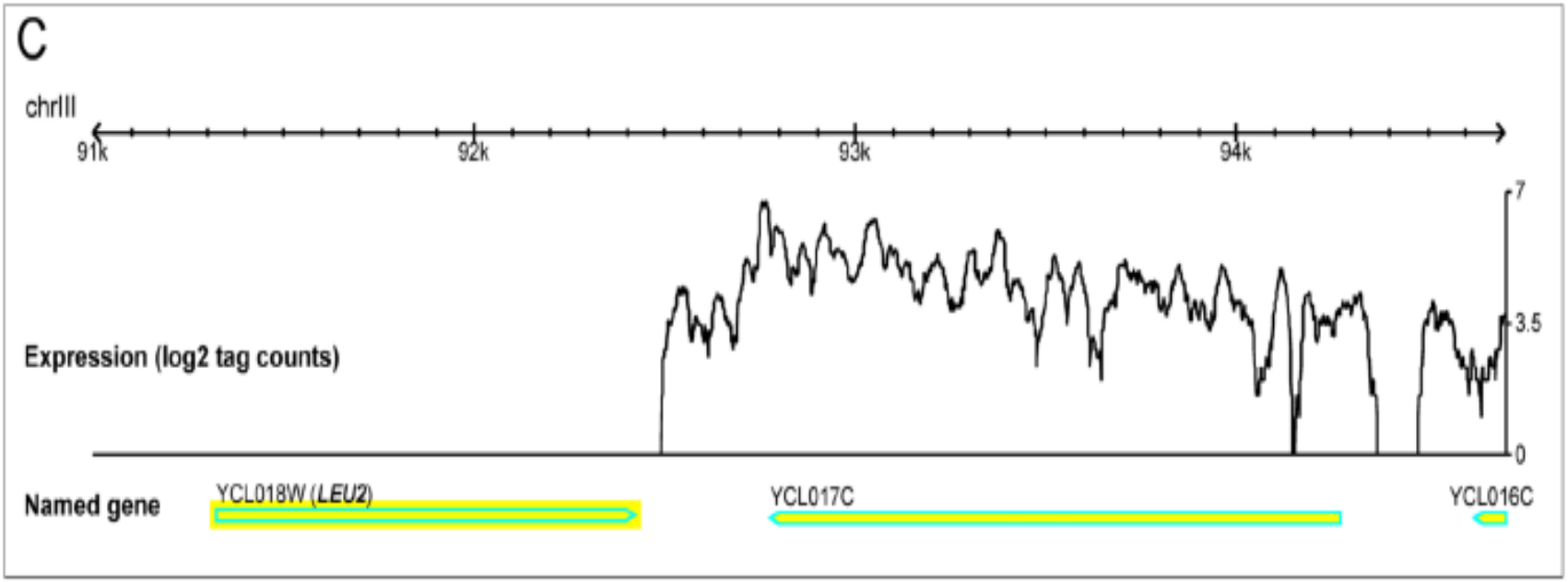


Figure 1. Flowchart of experimental and informatics of RNA-Seq method

A) RNA Seq experimental pipeline. B) Informatics pipeline. C) A snapshot of the mapped RNA-Seq reads showing no expression in a deleted gene (*LEU2*) and an expressed neighboring gene (YCL017C).

- **Mapping Transcribed Regions in the Yeast Genome Using RNA Sequencing**
- **Extensive Expression of the Yeast Genome**
- **Mapping Gene Boundaries and UTRs Using RNA-Seq**
- **Reannotation of the Yeast Genome**

- **Upstream ORFs are Present in Many 5′ UTRs of Yeast Genes**
- **Detection of Novel Transcribed Regions**
- **Quantitative Monitoring of Gene Expression Levels Using RNA-Seq**

Extensive Expression of the Yeast Genome

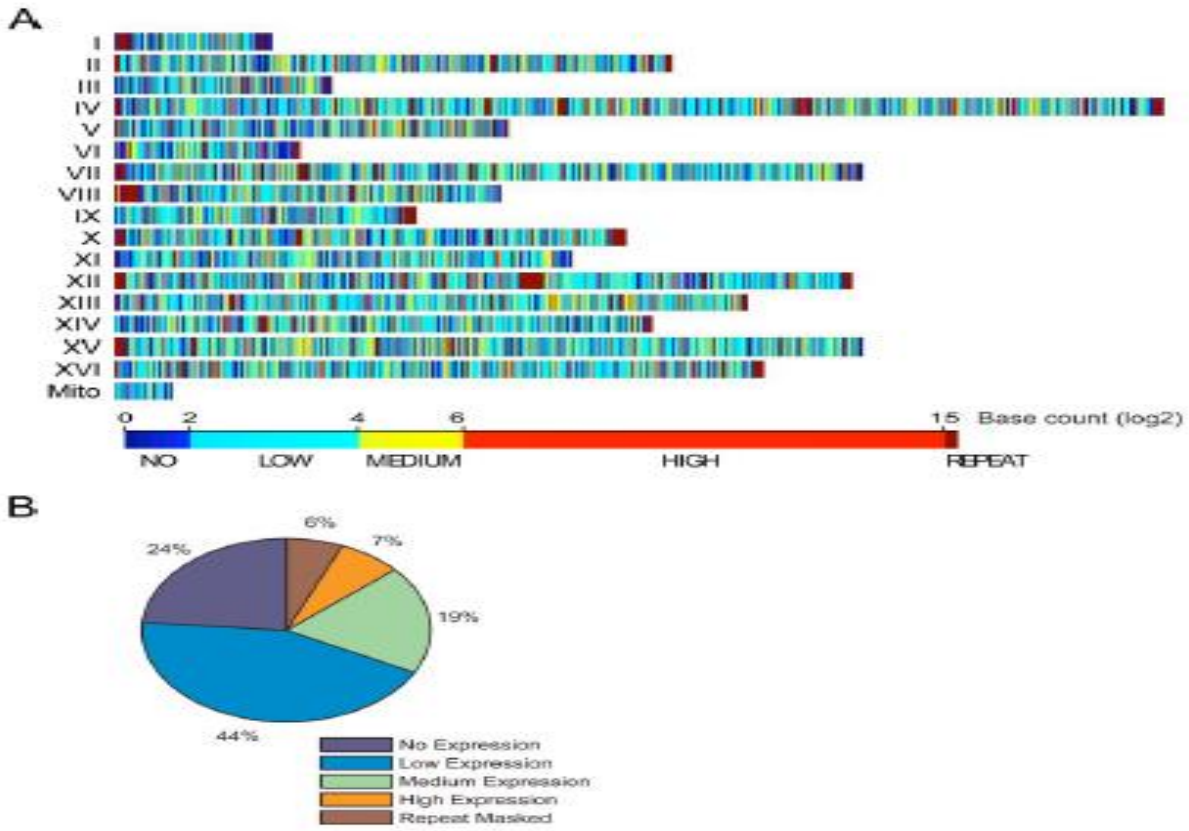


Figure 2. Extensive expression of the yeast genome revealed by RNA-Seq
A) The genome distribution of transcribed regions. Colors represent different transcription levels for each base (log2 tag count). B) Distribution of transcribed regions on chromosome VI. C) Histogram of transcribed bases. D) A summary of the transcription level of the transcriptome.

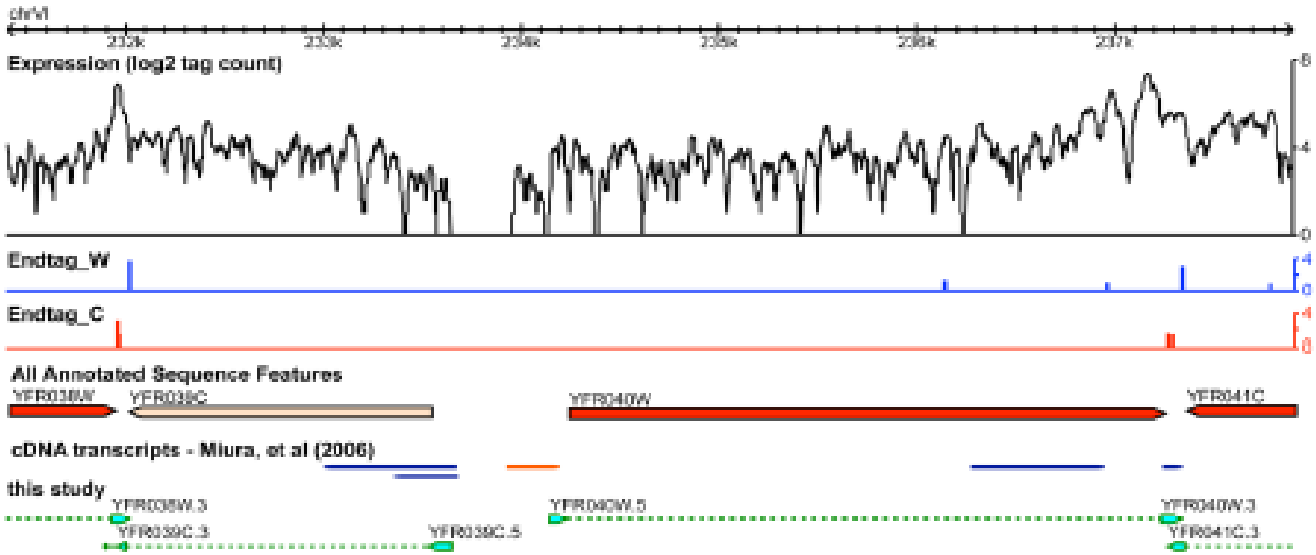


Figure 4. Precise annotation of UTRs using RNA-Seq

New annotations of the UTRs in a previously well annotated region on chrVI (A) and a relatively poor annotated region on the same chromosome (B). In the new annotation, ORFs are denoted by dotted lines, and arrows denote transcription direction. UTRs are denoted by green shaded boxes flanking the ORFs. cDNA transcripts in red are high confident ones and those in blue are low confident ones (7)

应用

预防医学

- 转录组谱可以提供什么条件下什么基因表达的信息，并据此推断相应未知基因的功能，揭示特定调节基因的作用机制。通过这种基于基因表达谱的分子标签，不仅可以辨别细胞的表型归属，还可以用于疾病的诊断。
- 例如：阿尔茨海默病(Alzheimer's diseases, AD)中，出现神经原纤维缠结的大脑神经细胞基因表达谱就有别于正常神经元，当病理形态学尚未出现纤维缠结时，这种表达谱的差异即可以作为分子标志直接对该病进行诊断。

应用

分子诊断

- 同样对那些临床表现不明显或者缺乏诊断金标准的疾病也具有诊断意义，如自闭症。
- 目前对自闭症的诊断要靠长达十多个小时的临床评估才能做出判断。基础研究证实自闭症不是由单一基因引起，而很可能是由一组不稳定的基因造成的一种多基因病变，通过比对正常人群和患者的转录组差异，筛选出与疾病相关的具有诊断意义的特异性表达差异，一旦这种特异的差异表达谱被建立，就可以用于自闭症的诊断，以便能更早地，甚至可以在出现自闭症临床表现之前就对疾病进行诊断，并及早开始干预治疗。