



"...the SOLiD™ System holds the potential of sequencing the transcriptome in its entirety using short reads..."
– Dr. Sean Grimmond

Next Generation Sequencing Unravels Transcriptome

SOLiD™ sequencing enables deeper examination of transcriptional complexity in mouse embryonic stem cells

Assessing Transcriptional Complexity

Sean Grimmond and his Expression Genomics Laboratory (Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia) perform global surveys of the transcriptome. His laboratory uses computational mining and genomic screening techniques to uncover the genes and transcriptional programs controlling important cellular processes such as cell differentiation, organogenesis, renal damage and repair, and tumor initiation and progression.

The recent FANTOM transcriptome annotation project (see sidebar, *Definitions*) sought to assess transcriptional complexity in humans and mice and found that, on average, every gene transcribes 3–4 different mRNAs and these mRNAs are transcribed from 2–3 distinct promoters [1]. The potential functional consequences for this expanded transcriptional complexity are considerable. Firstly, variant transcripts often encode novel proteins, expanding the organism's protein repertoire. Secondly, multiple transcripts from each locus often encode proteins which

localize to different cellular compartments, expanding and specializing the spatial activity of gene products. Thirdly, alternate 5' exon and promoter usage allows for expanded control of a locus. Finally, many examples of transcriptional complexity involve the inclusion/exclusion of sequences that modulate translational efficiency.

Determining Which Transcripts are Relevant

So how are the biological consequences of such complexity determined? The Grimmond lab has catalogs of

transcriptional complexity maps derived from microarrays used to survey hundreds of libraries from an entire spectrum of cell lines and tissue types. According to Dr. Grimmond, "While this shows the diversity in the entire mammalian transcriptome, it doesn't tell us which transcripts are relevant in specific biological systems." He adds, "To do this, you need to accurately measure which variants are expressed in your model system. The current gold standard of using microarrays for doing this is falling way short due to inherent limitations of hybridization based technologies. Sensitivity is insufficient to measure low expressers—the quality of hybridization signal is dependent entirely on probe quality, and the depth of coverage cannot be achieved by comparison to the latest next generation sequencers about to launch onto the market."

New Technology for Detecting Rare Transcripts

As a solution to this Dr. Grimmond saw the unsurpassed high throughput (up to 2GB of data per run) and high accuracy data quality of Applied Biosystems SOLiD™ Sequencing System as a potential technology to detect rare transcripts and rare variants of known transcripts. This is done using sequence tags generated from the RNA molecules isolated from mouse embryonic stem cells. "Due to its ability to generate millions of sequence tags, the SOLiD System holds the potential of sequencing the transcriptome in its entirety using short reads. These sequence tags could then be used to identify and quantify exon sequences associated with alternate splicing and alternate promoter usage. This can then be directly translated to biological context."

A preliminary analysis of the data generated from a single SOLiD sequencing run suggests the presence of tens of thousands more transcripts than found when the same RNA pool was analyzed using the Illumina® Bead Station

DEFINITIONS

Embryonic Stem Cells

A unique subset of stem cells, derived from the inner cell mass of an early stage embryo (blastocyst), that are totipotent. Under the appropriate conditions, these cells can differentiate into all derivatives of the three primary germ layers, including the >220 cell types found in the adult body. When given no signals for differentiation, embryonic stem cells retain their totipotency through multiple generations, and thus, can be maintained in vitro in the laboratory.

FANTOM

Functional Annotation of the Mouse cDNA; characterization of the mouse transcriptome. More information about the history and goals of the FANTOM project can be found at: <http://bib.oxfordjournals.org/cgi/reprint/5/3/249.pdf>.

Sequence Tag

A short stretch of DNA sequence that unambiguously maps to a unique region of the genome. As an example, it has been shown that specific mRNA molecules contain unique sequence tags that are not found in any other mRNA molecule. The sequence of this unique tag therefore serves as a surrogate for the sequence of the entire mRNA. This principle is being used for a wide range of genetic analysis applications.

Transcriptome

The full complement of RNA transcripts expressed in a cell or tissue in a particular time and space.



Grimmond Laboratory, Institute for Molecular Bioscience, University of Queensland, Australia.

(whole genome genotyping system). "We are seeing all those transcription factors and developmental signaling molecules that you know are there by real-time PCR but you don't see them on arrays," says Dr. Grimmond. The vast majority of the previously unidentified transcripts are present at levels below the limit of detection of the array. His group is now validating this finding and analyzing their data for novel splice variants, transcription start sites, and RNA polymorphisms.

Applied Biosystems looks forward to a continuing collaboration with Sean Grimmond's laboratory to further develop and utilize the full potential of SOLiD to assist in translating the complexity of the transcriptome into biological context.

See solid.appliedbiosystems.com for more details about the SOLiD technology.

References

1. Carninci *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science* **309**:1559–1563.
2. Erratum in: (2006) *Science* **311**:1713.