Institute of Molecular Biology

IMB TechTalk

Dr. Alexander Seitz

CEO of Lexogen GmbH, Vienna

"Enabling complete transcriptome sequencing through SENSETM strand specific mRNA-Seq and SQUARETM full length transcript amplification technologies."

07 May 2013, 11:00 (*s.t.*)

Venue: 2nd Floor Seminar Room

Institute of Molecular Biology (IMB)

Johannes Gutenberg University Campus Mainz

All are welcome to attend

Host: Dr. Tina Han, Head of CF Genomics, IMB

For further information, please contact: Bianca Steinbach, E-mail: b.steinbach@imb-mainz.de, tel.: +49-(6131)-39-21471

Institute of Molecular Biology

IMB TechTalk

Abstract:

Enabling complete transcriptome sequencing through SENSE™ strand specific mRNA-Seq and SQUARE™ full length transcript amplification technologies.

Antisense Transcripts and Overlaping Genes. The introduction and widespread adoption of strand specific RNA-seq library preparation protocols has made the detection and quantification of antisense transcripts commonplace. In spite of this, current library preparation protocols are limited by incomplete strand specificity. The generation of spurious antisense reads by undesired second strand synthesis during reverse transcription is well-documented and produces extensive background with some protocols, obscuring the detection of antisense transcripts.
Using SENSE protocol, highly complex libraries were consistently generated with >99.99% strand specificity, calculated with ERCC spike-in transcripts that lack antisense transcripts. The high levels of strand specificity reduced antisense background and allowed greater sensitivity when detecting antisense transcripts. Various other quality measures were also examined, and the protocol described provides a fast and high quality library preparation method for applications relying on strand specificity as well as general mRNA-Seq.

Transcript Variants. While the vast majority of genes are expressed as multiple transcript variants, current Next Generation Sequencing (NGS) technologies struggle to accurately describe these variants, and usually generate instead catch-all per-gene expression values. Differences in transcript splicing and promoter usage produce transcripts which can be regulated differently and often encode different proteins, with phenotypic effects in a number of known human diseases.

The assembly and quantification of transcript variants from NGS data presents several unique challenges. A transcript can be defined by several features including its exon structure and splice sites, the promoter used and exact starting nucleotides, and its polyadenylation site. Currently, per-transcript expression values can be estimated based on the coverage distribution and the abundance of reads spanning each splice junction. While these tools work well when defining major transcript variants, they perform poorly with low read depth and for genes with complex splicing patterns. Lexogen has developed a SQUARE technology for the assembly and quantification of full-length transcript variants, enabling the discovery of isoforms which would normally not be hypothesized, even with high sequencing depth.