Poster Abstract – GCB

Regulatory action of lincRNAs on alternative transcript events

Nadine Albrecht, Hans-Werner Mewes, Thorsten Schmidt

Eukaryotic transcriptomes harbor a large fraction of non-coding transcripts (ncRNAs) of which a significant proportion exceeds lengths of more than 200 nucleotides (1). Such ncRNAs are commonly referred to as long ncRNAs (1) or lincRNAs (large intergenic ncRNAs) (2). There are increasing indications that long ncRNAs are conserved and functional (3; 4; 5). The first long ncRNAs were reported to participate in the regulation of splicing (6; 7; 8) besides other functions. For instance, the long ncRNA Saf is sequence complementary to an intron of the human protein-coding gene Fas and fosters alternative splicing (8). Furthermore, alternative splicing is a key function in eukaryotic gene expression (9). Gaining more knowledge about the regulation of splicing is still an important issue. We suggest, that a part of alternative isoforms might be explained by the influence of RNA:RNA duplex structures of lincRNAs. In this work, we investigate the impact of lincRNAs on alternative transcript events as described in the work of Wang (10) and Katz et al. (11) for the first time systematically at large scale.

We used recently published strand-specific RNA-seq data obtained from three mouse cell lines: embryonic stem cells (ESC), neural progenitor cells (NPC) and mouse lung fibroblasts (MLF) from the work of Guttman and colleagues (12) for our analysis. RNA-sequencing allows the reconstruction and quantification of the expression levels of (i) coding and non-coding transcripts (13) and (ii) of mixtures of transcript isoforms (11). Based on the experimental RNA-seq data, we analyzed the influence of novel antisense lincRNAs on alternative transcript events. First, we identified all protein-coding transcripts (targets) that are in principle capable to form a duplex RNA structure with a lincRNA. Furthermore, functional characteristics of target sites including for example GC content, SNPs, conservation and entropy were analyzed. In the main analysis step, we determined different alternative transcript events such as exon skipping for affected coding exons or transcripts using the tool MISO (11).

In our work, we found that 18% (344 of 1931) of novel lincRNAs in ESC, 8% (98 of 1266) in NPC and 7% (52 of 786) in MLF are antisense to protein-coding genes, especially introns. We further observed that the expression of lincRNAs is correlated with the expression and frequency of their target transcripts (Pearson Correlation Coefficient, $R^2 = 0.8$ in ESC). Intronic target sites have significant functional characteristics, such as increased GC content and entropy (Wilcoxon rank sum test $p < 2.2e\text{-}16$ for target sites and background). As one of the most important results, we could show that lincRNAs regulate 'alternative transcript events' of their target protein-coding genes. For example in ESC 48 % (166 of 344), 31%

(106 of 344) antisense lincRNAs regulate notably the events ‘Skipped exon (SE)’ and ‘Alternative First/Last exons (AFE/ALE)’ of their affected coding exons. The findings of our work are confirmed in three different cell lines. This implies a strong and widespread influence of lincRNAs on alternative transcript events.

1. *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* **Kapranov, Philipp, et al.** 2007, Science, Vol. 316, pp. 1484-1488.

2. *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.* **Guttman, Mitchell, et al.** 2009, Nature, Vol. 458, pp. 223-227.

3. *Long non-coding RNAs: insights into functions.* **Mercer, Tim R, Dinger, Marcel E and Mattick, John S.** 2009, Nat Rev Genet, Vol. 10, pp. 155-159.

4. *The Long Arm of Long Noncoding RNAs: Roles as Sensors Regulating Gene Transcriptional Programs.* **Wang, Xiangting, et al.** 2010, Cold Spring Harb Perspect Biol.

5. *Long noncoding RNAs: functional surprises from the RNA world.* **Wilusz, Jeremy E, Sunwoo, Hongjae and Spector, David L.** 2009, Genes Dev, Vol. 23, pp. 1494-1504.

6. *A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition.* **Beltran, Manuel, et al.** 2008, Genes Dev, Vol. 22, pp. 756-769.

7. *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.* **Tripathi, Vidisha, et al.** 2010, Mol Cell, Vol. 39, pp. 925-938.

8. *Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas.* **Yan, Ming-De, et al.** 2005, Hum Mol Genet, Vol. 14, pp. 1465-1474.

9. *Epigenetics in alternative pre-mRNA splicing.* **Luco, Reini F, et al.** 2011, Cell, Vol. 144, pp. 16-26.

10. *Alternative isoform regulation in human tissue transcriptomes.* **Wang, Eric T, et al.** 2008, Nature, Vol. 456, pp. 470-476.

11. *Analysis and design of RNA sequencing experiments for identifying isoform regulation.* **Katz, Yarden, et al.** 2010, Nat Methods, Vol. 7, pp. 1009-1015.

12. *Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs.* **Guttman, Mitchell, et al.** 2010, Nat Biotechnol, Vol. 28, pp. 503-510.

13. *Advancing RNA-Seq analysis.* **Haas, Brian J and Zody, Michael C.** 2010, Nat Biotechnol, Vol. 28, pp. 421-423.

14. *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* **Pan, Qun, et al.** 2008, Nat Genet, Vol. 40, pp. 1413-1415.

15. *ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome.* **Croft, L., et al.** 2000, Nat Genet, Vol. 24, pp. 340-341.