

Novel insights in the regulation of mRNA transcription, processing and translation through integration of mRNA sequencing data

Eleonora de Klerk^{1,*}, Seyed Yahya Anvar^{1,2,*}, Ivo F.A.C. Fokkema¹, Martijn Vermaat^{1,2}, Klaske A.M.H. Thiadens³, Marieke von Lindern³, Stephen W. Turner⁴, Johan T den Dunnen^{1,2} and Peter A.C. 't Hoen^{1,#}

¹*Department of Human Genetics and* ²*Leiden Genome Technology Center, Leiden University Medical Center, Leiden, 2300 RC, The Netherlands;* ³*Department of Hematopoiesis, Sanquin Research and Landsteiner Laboratory, AMC/UvA, Amsterdam, The Netherlands;* ⁴*Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025, USA*

**Equal contributions*

#Presenting author; E-mail: p.a.c.hoen@lumc.nl

To date, the human transcriptome is known to contain around 80,000 protein-coding transcripts, and the estimated number of proteins synthesized range from 250,000 to 1 million. All these transcripts and proteins are coded by less than 20,000 genes, suggesting extensive regulation at transcriptional, post-transcriptional and translational level. I will discuss how integration of data obtained from diverse RNA sequencing technologies (RNA-seq, deepCAGE, ribosome footprinting) improves our understanding of these regulatory mechanisms and I will illustrate how these mechanisms jointly orchestrate the changes in protein demands during muscle differentiation. The individual regulatory layers appear to be tightly linked, with extensive cross-talk and feedback between them. To decipher the cross-talk between transcriptional and posttranscriptional regulation, we analysed PacBio® single-molecule long sequencing reads capturing full-length mRNA molecules. These data show that the vast number of potential combinations between alternative transcription start sites, alternatively spliced exons and alternative polyadenylation sites result in a relatively limited number of mRNA species, supporting the tight coupling between these processes. Further integration of RNA sequencing data will elucidate the true complexity of the transcriptome and its multi-layered regulation.

Resume: Peter-Bram 't Hoen is Associate Professor in Bioinformatics at Leiden University Medical Center. Since 2010, he has been responsible for all bioinformatics activities within the department of Human Genetics. He is leading a multidisciplinary team of researchers (molecular biologists and bioinformaticians) working on transcriptomics (RNA-seq) and proteomics data analysis, modeling of transcriptional networks, (cross-species) data integration, analysis of biological networks, and discovery of molecular biomarkers. He is an expert in RNA sequencing and coordinator of the yearly BioSB “Advanced RNA sequencing data analysis” course. He is also a member of the management team of BBMRI's Biobank-based integrative omics study (BIOS). His main research interest is the regulation of gene expression and the mechanisms controlling alternative transcription, splicing, polyadenylation, and translation.