

Biography

Project

European Research Council Advanced Grant

Project acronym: RNA-MAPS

Project full title: Uncovering and understanding RNA through Massively Parallel Sequencing

Overview

It is increasingly evident that RNA plays a complex, multi-faceted role in the cell, rather than serving as merely a messenger carrying information from DNA to protein sequences. Understanding RNA, and the mechanisms of its synthesis and processing is thus essential to understand cell's Biology and the phenotypic consequences of genetic variation and intervention. Recent technical developments in massively parallel sequencing provide the throughput required for accurate transcriptome characterization. The accurate profiling of transcriptomes across multiple conditions (cell types, species, individuals, cellular compartments, RNA classes, perturbations etc.), together with the profiling of other players involved in RNA synthesis and processing (DNA methylation, chromatin structure and modifications, transcription factors, RNA polymerase, splicing regulators, etc.) makes possible for the first time a global, holistic view of the transcriptional activity of genome, allowing for a systemic approach to the understanding and modeling of the pathways involved in RNA synthesis and processing. The overarching goal of the proposal presented here is the use of massively parallel sequencing for the identification of novel genetic and epigenetic factors playing role in RNA processing. Specifically, we plan to focus on the role of long non-coding RNAs and histone modifications in splicing regulation during cell differentiation. Towards that end, we plan to monitor RNA and chromatin changes across many time points through RNASeq and ChIPSeq in two cell differentiation models: 1) the induced transdifferentiation of human pre-B cells into macrophages, and 2) organ morphogenesis during fly development. By investigating, through these time course experiments, patterns of molecular co-variation—for instance, between inclusion levels of alternative exons and histone modifications—we can learn how these factors participate, and eventually cooperate, to modulate the condition specific abundance of the alternative splice forms. We plan therefore to develop statistical methods to infer significant patterns of molecular co-variation through the analysis of datasets profiling RNA and chromatin structure across many data points. Based on these findings, we plan to develop a mathematical model of splicing in which the relative abundance of alternative splice forms is predicted from the relevant genetic and epigenetic factors. To facilitate the storage, organization, access, and analysis of the data produced through the project we propose to develop a robust, efficient, and scalable software system. The system will be open source, so that it can be used and enhanced by other researchers. Realizing that the current model for the storage of publicly available biomolecular data, based on centralized repositories, may not be longer sustainable, we will design our software system assuming a peer-to-peer like network of distributed RNASeq data and resources, across which bioinformatics analyses will be transparently performed.