Long non-coding RNAs in human pluripotent stem cells

Ribonucleic acid (RNA) is a varying in length single strand made up of combinations of four building blocks called nucleotides (guanine, cytosine, adenine and uracil). In contrast to the simplicity of its four building blocks, the RNA is the most functionally versatile molecule in our cells. It can function as carrier of information by transferring the information encoded in genomic DNA to the cellular machinery that makes proteins. Additionally, RNAs can fold into complex structures that catalyze chemical reactions, bind other molecules and alter their activity, as well as store genetic information (the latter, is case of some viruses).

Recent studies using massive parallel sequencing technologies revealed that most of our genome is transcribed into RNA. Protein-coding genes account for only a small percentage of these transcripts – 1-2%. The remaining pool of RNAs consists of non-coding species with various cellular functions. Among them, a large group falls into the category of long non-coding RNAs (lncRNAs) defined as transcripts longer than 200 nucleotides without protein-coding potential. Importantly, only a part of those RNAs have been attributed a biological function based on specific experimental studies reported in literature. Detailed mechanistic studies explaining their mode of action are still sparse. lncRNAs are essential regulators of numerous cellular processes and are being related to a variety of disease states. They are emerging as crucial players in stem cell biology. They both maintain pluripotency and drive differentiation to specialized cells, like neurons, cardiomyocytes, hepatocytes or endothelial cells. Genome-wide screens in pluripotent stem cells identified many lncRNA that are required for pluripotency, among them, lncRNA-ES1 and lncRNA-ES3. Another recently reported lncRNA, Heart Brake lncRNA (HBL1) regulates the differentiation of human stem cells towards cardiomyocytes. These lncRNAs are expressed in pluripotent cells, but their level is downregulated in differentiated cells. Importantly, while binding partners are reported for these lncRNAs, the molecular details of those interactions are unknown.

In the current project, we plan a combination of biochemical and cell biology approaches in order to fully understand the mode of action of the selected stem cell-specific lncRNAs: HBL1, lncRNA-ES1 and lncRNA-ES3. Within the first project aim, we plan to focus on the molecular bases of selected lncRNA cellular functions. We will visualize their subcellular localization and define how it is regulated. Importantly, we will identify with which proteins the lncRNAs are interacting in cells, and to which genomic areas they bind. The second project aim is to determine lncRNA structure alone and together with protein binding partners. We will start with analyses that will reveal the two-dimensional topology of the molecules and define probable structural modules. Next, we will confirm their interaction with protein binding partners reported in literature and identified in our study and indicate the localization of interaction sites on the lncRNAs. Finally, we plan to uncover three-dimensional structures of lncRNAs with and without protein-binding partners by cryo-EM and crystallography.

Through our combination of biochemical and cell biology methodology, the successful completion of the proposed project will increase our knowledge of the mode of action of selected lncRNAs linked to pluripotency and differentiation. It will contribute to enlarging the so far limited collection of lncRNAs with defined secondary structure. Moreover, the study of the interaction mode between lncRNAs and protein binding partners at atomic-level will provide solid bases for the understanding of their cellular functions. As two of the studied lncRNAs - lncRNA-ES1 and lncRNA-ES3 - have elevated expression levels in tumor cells, our results could also be used in the future as a basis to elaborate new approaches in cancer treatment. Finally, our analysis will be conducted on human lncRNAs and proteins and in human cells, making the results directly applicable to human biology. This is especially important when considering the low conservation of lncRNAs between species. Altogether, the importance of the project consists in its high potential to provide the stem cell-, RNA-, and structural biology-interested scientific communities with interesting, novel and potentially therapeutically-applicable data.