Function and dysfunction of repeated tracts in protein-coding and non-coding transcripts

The aim of this project is to unravel the functioning of a broad spectrum of RNAs containing multiple repeats of three-base sequences, and to show how the length of the repeat tract affects the activity of the RNA or encoded protein. We will focus on CAG (cytosine-adenine-guanine) repeat tracts in the context of nervous system. For a better understanding of how our brain functions and how we can effectively treat neurodegenerative diseases, it is important to understand the network of molecules that interact there, including RNA and proteins.

Repeated sequences of units up to 6 nucleotides (known as microsatellites or short tandem repeats, STRs) constitute about 3% of the human genome. Some microsatellites are transcribed and RNA molecules containing repeated tracts are produced, and some of the transcribed STRs (almost exclusively trinucleotide tracts) are translated, resulting in different homo-amino acid tracts in proteins. A specific type of microsatellites, CAG repeat tracts, are present in many protein-coding RNA (messenger RNAs, mRNAs) and non-coding (ncRNAs). Interestingly, repeat tracts are characterized by polymorphism of tract length what contributes to population phenotypic diversity. In the last two decades, we have only just begun to understand how these sequences influence the molecules they are in.

Almost three decades years ago several genes with CAG repeats, encoding proteins with polyglutamine (polyQ) tracts, were identified to be responsible for rare neurological diseases, which are still incurable. The shared gene mutation is repeated tract expansion that results in 3-10 times longer tract than usually occurring also in both, encoded RNA and protein. Importantly, genes responsible for individual polyQ diseases are not related, but the main pathogenic pathways are consistently occurring in the brain. This suggest that in nervous system there is an unique and crucial function of CAG repeat tracts, as well as a special vulnerability for polyQ tracts dysfunction. Moreover, a phenomenon of somatic expansions in neurons is an important contributor to polyQ diseases pathogenesis and leads to occurrence of much longer tracts than initial mutation

What is worth emphasizing, in the last decade transcriptome-wide techniques allowed for identification and precise description of numerous ncRNAs, including those containing repeat tracts. Nevertheless, we do not understand the role and functioning of majority of them. There is a variety of ncRNA molecules, for which the sequence and tissue expression level is already known but much less is known about their interactome and exact role in cellular processes. Concerning this aspect and project assumptions, it is important that ncRNAs are abundant in nervous system and flagship examples of their functioning were described in brain.

Detailed scientific questions to be addressed in frame of this project include:

- What are the functions important for nervous system of RNA molecules containing CAG repeat tracts?
- Is there a specific network of interacting RNAs containing triplet repeat tracts?
- What is the network of protein-protein interactions mediated by polyQ tract?
- Are some of the RNAs containing CAG repeats functioning both, as ncRNAs and encoding proteins/peptides?
- How the CAG/Q repeat tract length (in normal range) influences functioning of specific RNA/protein?
- How dysfunctional are somatically expanded CAG/Q repeat tracts?

To answer these questions, we will use precisely designed cellular models characterized with neuronal phenotype, in order to find clear relation to nervous system. This includes induced pluripotent stem cells (iPSCs)-derived models, among others. The methodological concept is to use sets of cell lines with the same genetic backgrounds that should enable to expose CAG/Q tract length-specific effects. Moreover, observations will be verified in mouse brain tissue.

Long-term aim of this area of research is also addressing crucial questions referred to efficient therapy for polyQ diseases. Precise description of the hallmarks of CAG repeats in RNAs and Q repeats in proteins will enable design of new strategies and provide important hints for therapy for polyQ diseases.