The human genome is mainly non-protein-coding, which means that ~99% has no ability to produce proteins – complex molecules that are essential for the structure and functions of cells. Interestingly, the non-protein coding DNA is a source of various RNA species that can control the activity of protein coding genes (~19,000) at different levels. Long noncoding RNAs (lncRNAs) – RNA molecules that are longer than 200 nucleotides and do not produce any functional protein – are the most enigmatic and at the same time the most intriguing class among noncoding RNAs. Although many lncRNAs play roles in key physiological and pathological processes, we still do not know much about them. It is not even clear how many lncRNAs in total are encoded by our DNA. We also lack the means to precisely determine *which* lncRNAs are functional. As a result, more than 100,000 genes encoding lncRNAs have been identified in our genome, but <2% of them have been functionally characterized. Therefore, revealing their full, biological potential requires answering the fundamental and longstaning question: *How are lncRNA functions preserved in their primary RNA sequence?*

The most popular definition of „function" in biology refers to the activity of molecules, organs and systems. However, in evolutionary biology „function" is the reason some object or process occurred in the system. Evolution is a process that keeps crucial biological functions by passing them as fragments of DNA sequence from one genome to the other over millions of years. Therefore, evolutionary conservation often refers to the presence of similar genes, their fragments or even large DNA segments in different species. It not only reflects the common origin of species, but also the functional importance of DNA fragments that were preserved by nature over the time. Cross-species comparison has proven to be a powerful method for identifying functional regions in a genome. Over the years, studies based on evolutionary conservation have significantly advanced our understanding of protein coding genes. Therefore, they are also expected to provide some insights into lncRNA genomics and functionality. However, unlike most strongly conserved protein coding genes, lncRNAs evolve rapidly with only <100 lncRNAs showing detectable sequence conservation between human and zebrafish – small tropical fish. At the same time thousands of lncRNAs show positional conservation across those species without detectable sequence conservation, meaning that nature tends to preserve the order of lncRNAs not their sequence. Some positionally conserved lncRNAs share very short (<20 nucleotides) of conserved sequence that are missed by the conventional tools designed for comparative analysis of protein coding genes, containing long stretches of high sequence conservation. Another major difference between lncRNAs and protein coding genes is the way they are processed by the cell. The journey from gene to its protein product is rather similar for all protein coding genes. First, the information stored in the DNA is transcribed to RNA in the nucleus and next, a protein is synthesized in the cytosol. LncRNAs are much more cellular compartment specific, as subcellular localization is a foundation of their functions. In contrast to protein coding genes, the final product of lncRNA gene expression is a mature RNA molecule. Due to its reduced stability, RNA needs to be immediately delivered into a specific subcellular compartment – its place of action. As a result, populations of lncRNA show different cellular localizations, including nucleus, cytoplasm, mitochondria, etc. Interestingly, it has been recently shown that conserved lncRNA molecules undergo distinct processing by cells of different species. The same lncRNA is exported to the cytoplasm by the human, while retained in the nucleus by mouse cells and shows limited functional potential. This brings another fundamental question to the field: „*Are conserved lncRNAs really functional*?"

The goal of this project is to investigate the functional potential of positionally conserved lncRNAs in vertebrate species. First, we will employ our new software ConnectOR (Connecting RNA Orthologues) for detecting positionally conserved lncRNAs in human and mouse genomes. Second, we will develop a new method to accurately annotate them at subcellular resolution in a panel of cells. By comparing lncRNA sequence composition at the subcellular level, we expect to identify stretches of sequence that are the foundation of lncRNA functions. We will not only investigate their contribution to determining specific subcellular localizations, but we will for the first time reveal how the functional potential of lncRNAs changes between closely and distantly related vertebrate species by including zebrafish in those comparisons. Finally, we will experimentally validate the functional potential of positionally conserved lncRNAs in cells and mouse models.

To conclude, this is an innovative and ambitious project that will help to better understand how nature preserves lncRNA functions in their primary sequence. Moreover, it allows us to study lncRNA functions from *why it's there*?' perspective, which is a more accurate justification for the presence of functional elements in the genome than currently employed '*what it does*?' strategy.