Genome 'dark matter'

Computing the missing components in inflammatory diseases and diabetes…

Imagine a library with billions of books but the only books you read are the ones by Shakespeare. This is comparable to the way we currently look at the genome.

The central dogma is that a gene 'stored' on the DNA encodes for a protein through an intermediate step where it takes a so-called RNA form. This is written in numerous textbooks and is what many students still learn today, even though this is changing.

However, a number of genes remain as non-coding RNA (ncRNA) and carry out their function in this form, for example by regulating protein coding genes. Over the years ncRNAs have been shown to be key players in a large number of scenarios and they attract more and more attention. For example, with the release of the human genome in 2001, it turned out that only ~1.2% of the three billion DNA building blocks in the human genome encode for proteins. The potential to host ncRNA genes is therefore enormous. The number of discovered ncRNA genes is increasing rapidly – now in the order of thousands in the human genome. Given the large space in the genome for ncRNA genes, this provides a large potential to discover ncRNA that may have many diverse roles. A key question is what role does this large space 'dark matter' of genome play for our understanding of disease?

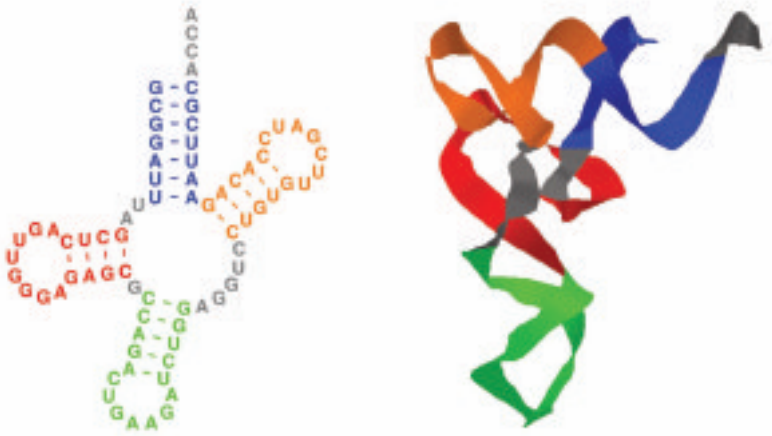

Fig. 1: Non-coding RNA – an unexploited field with a huge potential. Left is a style of the RNA (secondary) structure, generated by computer. On the right is the full three-dimensional structure, which often is unknown. We use computational methods to generate RNA structure candidates in genomic DNA sequence

© Image courtesy of Naturens Verden

A novel strategic research centre

Attention to ncRNAs is also increasing in the areas of medical and health science research. In 2010, the Center for non-coding RNA in Technology and Health (RTH) was established through funding from the Danish Strategic Research Council with the aim of searching for ncRNAs involved in inflammatory diseases and diabetes.

The centre comprises national as well as international partners both from academia (University of Copenhagen, the Technical University of Denmark, Glostrup Hospital, University of Leipzig, University of Vienna, University of Freiburg, University of Washington, Seattle and University of Lausanne) and industry (Exiqon A/S and Novo Nordisk A/S). Its base is located at University of Copenhagen. The partners are experts within different scientific disciplines, including bioinformatics, molecular biology, veterinary and medical sciences. Thus, while a main tool is bioinformatics, experimental methods such as high-throughput sequencing, animal models, human tissue engineering and expression studies are closely integrated within the project.

Computing gene candidates in the dark matter

An important component in the project is computational predictions of ncRNA candidates.

Since RNA structure is a main characteristic for many ncRNAs, it is of interest to predict RNA structure in genomic (DNA) sequence. However, predictions in a single organism have turned out to be unreliable. Instead the sequences of multiple organisms related to humans can be exploited to search (while aligning them by their sequence) for common RNA structure, which often is more conserved than the sequence itself.

Computational methods are developed and applied in combination with RNA biology experiments and then exploited in the further search to look for presence (expression) in a disease related tissue sample. Due to the nature of the RNA structure, the computational analysis is far more extensive than a regular type of comparison of genomic sequence; it requires computational facilities far exceeding what can be accomplished on a desktop computer. Our multidisciplinary approach willl enable us to study and explore the dark matter of the genome by combining the

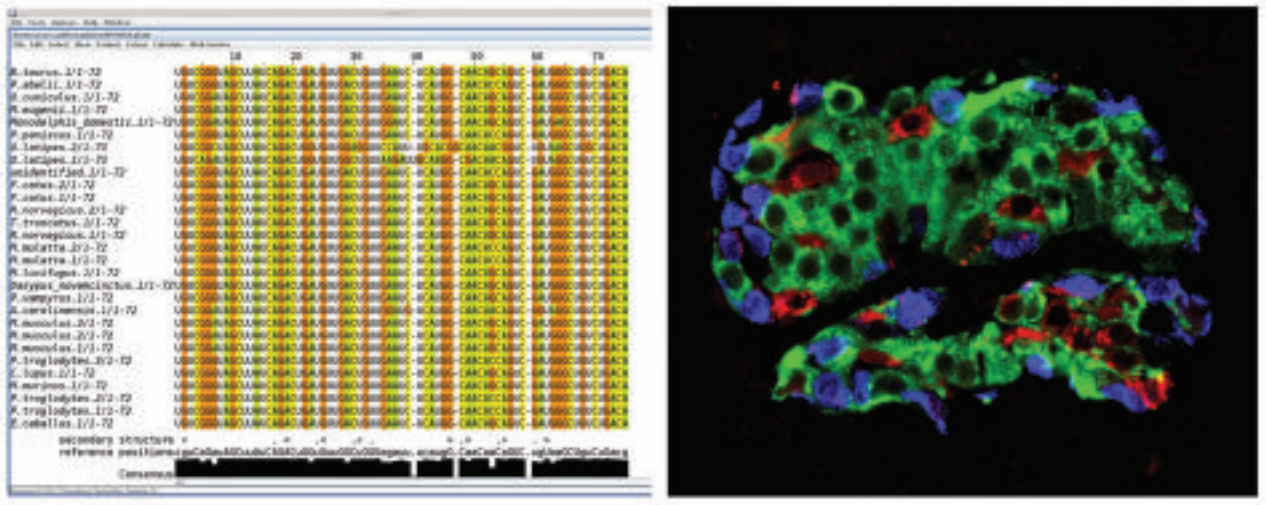

Fig. 2: Computational approaches are important tools in the hunt for novel ncRNAs. We use traditional methods like immunological staining for insulin and glucagon in the pancreas to see if the found ncRNAs play a role in regulation of diabetes

predicted RNA candidates in large scale experiments (such as RNAseq) – searching eg for overlap to recently reported so-called long ncRNA, which are totally uncharacterised.

A new perspective on diabetes and inflammatory diseases

Our focus is to explore the role of ncRNAs in the very complex processes involved in inflammatory diseases, including diabetes. Diabetes is a huge healthcare problem both for the individuals affected and for society. Today, 346 million people worldwide have diabetes.

Without urgent action, diabetes related deaths are likely to increase by more than 50% in the next 10 years. Despite considerable knowledge about Type 1 diabetes pathogenesis, monitoring of disease progression, optimisation of treatment regimes and introduction of new treatment modalities, the mortality of patients with Type 1 diabetes is still four to seven times that of the matched background population, leaving much room for improvement in translational science within this field.

We will test whether some of our computer generated ncRNA candidates are so far overlooked players, which are present together with known diabetes genes in tissue samples from diabetic as well as healthy individuals. Conversely, we will also integrate the (sequence) data and use them to compute whether we see patterns of processing resembling a putative ncRNA gene.

Our hope is that we will identify and describe a number of molecular targets that in turn can be used for drug development, as biomarkers in disease prediction and prevention, and for improved therapeutic responsiveness. We use pedigree pigs as a model animal system. Challenging their insulin response by feeding them with a high concentration of glucose, some of the animals reach high glucose levels in the blood, comparable to Type 2 diabetes patients. We aim to reveal that ncRNAs are novel molecular players with regulatory capacity and will therefore constitute a highly valued resource in systems biology, which currently (with the exception of the emerging studies of the single class of ncRNAs, called microRNA) essentially ignores the potential of ncRNAs.

Educating new scientists

Educating new scientists and providing a fruitful research environment are important issues for us. Since the establishment of the centre we have employed 12 full-time PhD students and postdoc fellows, and several more do work on related topics with interaction to the core centre staff.

We co-organise a summer school for PhD students in RNA bioinformatics in Copenhagen in August. In addition, we arrange seminars with highly recognised international and national speakers and local scientific workshops for the research groups. We wish to attract the best qualified students and researchers. Thus, the centre consists of researchers from many different countries, giving an inspiring international working environment.

RTH
CENTER FOR NON-CODING RNA IN TECHNOLOGY AND HEALTH

Professor Jan Gorodkin
Center Director

Center for non-coding RNA in Technology and Health
Faculty of Health and Medical Sciences
University of Copenhagen
Grønnegårdsvej 3
1870 Frederiksberg C
Denmark

Tel: +45 353 34704

contact@rth.dk
rth.dk