RNA structure in health and disease - investigation of lncRNA structure function relationship in imprinting control regions

Dr Russell S. Hamilton

Department of Genetics, rsh46@cam.ac.uk

Project Description

Genomic imprinting is an epigenetic phenomenon leading to specific genes being expressed in a parent-specific manner. Imprinted genes have key roles in placenta and embryo development, as well as an emerging role in brain development and function. Clusters of imprinted genes with shared regulatory elements are termed Imprinting Control Regions (ICRs) and are conserved in human and mouse. Within these ICRs, long non-coding RNAs (lncRNAs) are transcribed and are fundamental to the function of the ICRs. LncRNAs are defined by their length (>200nt) and lack of translation into functional proteins (McDonald et al, 2020, PLoS Genet). The lncRNAs encoded within these ICRs, including H19, Airn, Gtl2, Nespas, Snhg14 & Kcnq1ot1, are themselves imprinted, and therefore also expressed in a parent-specific manner.

Disruptions in imprinted genes, and specifically within lncRNAs, have been linked to human disease. For example, patients with disruptions in the Gtl2 ICR display growth retardation and facial dysmorphism (Enfield *et al*, 2016, Oncotarget) and disruption in the lncRNA of the Snhg14 ICR have been shown to be a direct cause of Prader-Willi syndrome and postnatal growth retardation (Jauregi *et al*, 2013, J. Neurodev; Skryabin *et al*, 2007, PLoS Genet). Structural motifs within the lncRNAs have also been demonstrated to be important to their function. Motifs with in Gtl2 lncRNA have been shown to be key to the stimulation of the p53 pathway (Uroda *et al*, 2019, Mol. Cell) and a secondary structure altering genetic variant in H19 lncRNA has been linked to cardiomyopathy (Martens *et al*, 2021, bioRxiv).

Without a functional protein product, lncRNAs function via their specific sequences and structures. The key to understanding the functional role of these imprinted lncRNAs is to study their structures (2D and 3D) and determine if they can adopt non-canonical structural motifs such as the calcium sensitive kink-turn motif. Conservation of lncRNA sequences between species, human and mouse, is likely be under evolutionary constraint to form specific structures required for their function. We will take advantage of this conservation to model 2D and 3D structures and therefore determine key structural features required for the lncRNA functions. Accurate 3D models will also inform potential protein binding surfaces which can be explored with structural bioinformatics approaches. Using open source computational tools we will model and characterise the local, canonical and non-canonical structural motifs within the imprinted lncRNAs (Taylor & Hamilton, 2017, Sci. Reports). We will then create an interactive web resource and database to perform comparisons across the imprinted lncRNA structures. Characterisation and, crucially, comparison of the structures within these enigmatic lncRNAs, and across species, will provide valuable insight into their functions and tissue specific roles.

Short Summary (80 words)

Long non-coding RNAs (lncRNAs) are fundamental to the function of clusters of imprinted genes, however little is known about their structural features, likely to be key to understanding their functional roles and links to human disease. Here we propose to structurally characterising 5 key imprinted lncRNAs alongside an extensive assessment of their key structural features. These lncRNAs have been shown to have key roles in placenta and embryo development, as well as an emerging role in brain development and function

Key aims and tasks:

1. Generate accurate 3D models of imprinted lncRNAs
2. Perform sequence (1D), 2D and 3D comparison across the lncRNAs
3. Assess 1D/2D/3D conservation between human and mouse
4. Present results in a publicly available web resource

Specific details for your project

- Duration 8 weeks
- The project is purely computational so can be conducted remotely. If restrictions allow the project can be conducted in within the Department
- Essential skills: basic competency in R and using bioinformatics tools, although training will also be provided