Systematic experimental and computational strategies to identify and functionally characterize lncRNAs and RNA modifications in development and disease

ShT-01.5-3

P. Amaral[I,II], T. Leonardi[III], A. Leger[IV], N. Han[V], T. Kouzarides[V,VI], H. Nakaya[II]

[I]Insper Institute of Education and Research, Rua Quatá 300, Vila Olímpia, São Paulo, Brazil, [II]Hospital Israelita Albert Einstein, Av. Albert Einstein 627, Morumbi, São Paulo, Brazil, [III]Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia (IIT), Milan, Italy, [IV]Oxford Nanopore Technologies, Gosling Building, Oxford Science Park, Oxford, United Kingdom, [V]The Milner Therapeutics Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Puddicombe Way, Cambridge, United Kingdom, [VI]The Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge, United Kingdom

The mammalian genome is transcribed into several thousand long noncoding RNAs (lncRNAs) whose properties and functions remain largely elusive. Our group has developed a combination of genome-wide *in silico* and functional strategies to explore their regulation and biological roles. Using a computational pipeline[1] to analyze RNA-seq datasets from human and mouse tissues, tumors and cell lines, we identified thousands of lncRNAs in syntenic locations, whose promoters overlap chromatin loop anchor points containing conserved CTCF sites, dubbed tapRNAs (topological anchor point RNAs). These RNAs are associated with developmental genes with which they: (a) are co-expressed and able to regulate each others' expression, (b) are similarly misregulated in cancers, and (c) influence differentiation of embryonic stem cells and metastatic characteristics of cancer cells *in vitro*. We find novel associations of ncRNAs with specific chromatin compartments and evidence for the role of tapRNAs in chromatin organization. To also explore the potential impact of RNA post-transcriptional modifications (PTMs), we are using Nanopore RNA-seq and a robust statistical framework to determine the distribution of PTMs in lncRNAs and mRNAs. We do so by comparing native RNAs isolated from biological samples of interest (including mammalian disease models) against modification-depleted RNA controls, also mapping known and novel modification sites in conserved polyA-noncoding RNAs. We are extending similar experimental and guilt-by-association expression network analyses to study lncRNAs in different systems, identifying lncRNAs functionally associated with key biological pathways, such as with immune protection in EBOLA vaccine cohorts. Altogether, we find support for widespread roles of RNAs and their modifications in the control of gene expression in developmental and disease models, as well as in other biological systems.

1. Previously published in: Amaral PP et al. (2018) Genome Biol 19, 32.