Single cell transcriptional landscape of long non-coding RNAs orchestrating mouse heart development and its potential association with the fetal gene program

Thaís A. R. Ramos[1,2,3], Soo Young Kim[4], Thomas G. Gillette[4], Joseph A. Hill[4,5], Sergio Lavandero[1,6,*], Thaís G. do Rêgo[2,3,*], Vinicius Maracaja-Coutinho[1,2,7,*]

[1] *Advanced Center for Chronic Diseases (ACCDiS), Faculty of Chemical & Pharmaceutical Sciences & Faculty of Medicine, Universidad de Chile, Santiago, Chile.*
[2] *Programa de Pós-Graduação em Bioinformática, Bioinformatics Multidisciplinary Environment (BioME), Instituto Metrópole Digital, Universidade Federal do Rio Grande do Norte, Brazil.*
[3] *Departamento de Informática, Centro de Informática, Universidade Federal da Paraíba, João Pessoa, Brazil.*
[4] *Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center Dallas, Texas, USA.*
[5] *Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas, USA.*
[6] *Corporación Centro de Estudios Científicos de las Enfermedades Crónicas (CECEC), Santiago, Chile.*
[7] *Instituto Vandique, João Pessoa, Brazil.*
* *Corresponding authors:* slavander@uchile.cl, gaudenciothais@gmail.com or vinicius.maracaja@uchile.cl

Long non-coding RNAs (lncRNAs) comprise the most representative transcriptional units of the mammalian genome, and are known to possess a tissue-specific expression pattern and a key association with organ development. In a variety of pathophysiologic conditions the postnatal heart returns to the "fetal gene program", in which expression changes reported in the failing hearts are consistent with the re-expression of genes related to the fetal heart development. Here, we explored public available single-cell RNA-seq (scRNA-seq) data from 4 embryonic (E9.5, E11.5, E14.5, E18.5) and 4 postnatal (P0, P3, P7, P21) stages across 5 different compartments of the mice heart during development. After the mapping of scRNA-seq raw reads to the mice genome and its filtering to avoid sparsity issues, we performed feature selection and differential expression analysis to identify potential cell-specific gene markers. Next, we performed hierarchical clustering to group the cells, using silhouette method to estimate the optimal number of clusters, PCA to reduce the dimensionality, and t-SNE for cluster visualization. We used literature and database searchers to assign marker genes to cell-types and performed chi-squared and adherence tests to evaluate significance (p-value < 0.05). We identified eight cell-types, including two sub-populations of cardiomyocytes and macrophages; novel markers for the identified cell-types, including coding genes and lncRNAs; and specific signatures for coding genes and lncRNAs for the same cell-type at different heart compartments. Differential expression and functional enrichment analysis reveal cardiomyocyte subpopulations associated with metabolic diseases, mitochondria, and cardiac function; meanwhile modular co-expression analysis reveals cell-specific functional insights for lncRNAs during myocardial development, including a potential association with key genes related to disease and the "fetal gene program". Our results evidence the role of particular lncRNAs in heart development, and highlights the usage of co-expression modular approaches in the cell-type functional definition.

Funding: FONDECYT-ANID #1211731.