



UNIVERSITAT DE
BARCELONA

FACULTAT DE BIOLOGIA
DEPARTAMENT DE GENÈTICA
Programa de Doctorat en Genètica

Unravelling the Role of Long Noncoding RNAs in the Context of Cell-growth and Regeneration

Memòria presentada per
Raziel Amador Rios
per a optar al grau de Doctor per la
Universitat de Barcelona

Treball realitzat al Departament de Genètica, Microbiologia i Estadística de la
Facultat de Biologia de la Universitat de Barcelona i al Centre de Regulació
Genòmica (CRG)

Doctorand
Raziel Amador Rios

Co-Director
Montserrat Corominas
Universitat de Barcelona

Co-Director
Roderic Guigó
Centre de Regulació Genòmica

Barcelona, Novembre de 2021

Acknowledgments

I would like to express my profound gratitude to my thesis co-directors, Roderic Guigó and Montserrat Corominas, for their feedback, guidance, and for providing a research environment that was intellectually stimulating and in which I was able to improve my research skills. My work was also co-supervised by Assaf Bester, to whom I am also grateful for his guidance, and for the feedback in the second part of this study.

Besides Roderic, Montse and Assaf, I also thank Carlos Camilleri, for his remarkable scientific contributions to the improvement of this thesis. I am grateful to all members of Roderic's lab. The administrative support provided by Montse Ruano, and Romina Garrido have been exceptional. This work also benefited from the computational help of Emilio Palumbo.

Many thanks to the following people for useful feedback on various parts of the Thesis manuscript: Montserrat Corominas, Manuel Muñoz, David Brena, Reza Soudaei, Marc Elosua, Iman Sadeghi, and Roderic Guigó.

Additionally, the following non-exhaustive list of people, which positively impacted my PhD experience and this work:

Cecilia C. Klein, for your mentorship during the first year of PhD and share your insights in *Drosophila*, regeneration and bioinformatic. From you I learned how to combine biology and computational analysis to obtain novel results.

Manuel Muñoz, for making my adaptation process in the lab easier; help in statistics, plotting, programming, coding best practices; for introducing me into the Emacs world (the most efficient text editor/IDE) and, last but not least, the fun hackathon times we shared, we learned a lot. Your examples helped me realize the importance of hard working, curiosity-driven projects and never stop learning.

Reza Sodaee and Valentin Wucher, for giving the opportunity to collaborate with you in the circadian-seasonal manuscript. I learned why the science-core should be based on collaboration and use different expertise to solve a problem, and achieve a more complete conclusion.

Iman Sadeghi, for your friendship, all the adventures we had through our PhD years, personal advice, and time to time scientific counseling in this research. Julien Lagarde, it was very helpful in sharing his LaTeX code to everybody, using open-source software, providing thorough instructions of how his code works and helping when I needed it. Vasilis Ntasis, for the geek talks, sharing your linux/R tools and philosophy of mostly using the keyboard made me more productive and efficient.

Thanks to my love Vanessa Vega for her constant support, in graphic design, revisions, polishing plots, and otherwise.

Por último, agradezco a mis padres y hermanos por todo el apoyo, amor y proporcionarme las herramientas necesarias para ser quien soy. Sin ustedes esto no sería posible.

Abstract

Long noncoding RNAs (lncRNAs) have proven biological roles in plethora cellular contexts. Nonetheless, only a handful have been clearly characterized, leaving thousands of newly discovered lncRNAs without an associated function, and sometimes considered as transcriptional by-products. To this end, this thesis work had focused on exploring lncRNA functionality in two scenarios. First, in order to discern between lncRNAs affecting cell-growth rate (lncRNA-hits) and lncRNA-not-hits, we built a tree-based classifier based on high-throughput CRISPRi functional screen data in seven human cell lines, as well as, cell-specific ENCODE transcription factor ChIP-seq data; finding that the genomic features used in our study showed small effects and tend to be transcript-specific. Our classifier outperformed previous algorithms, displayed balanced sensitivity and specificity values, and uncovered a lncRNA (*LINC00879*) involved in cell-growth. Additionally, we unveiled a list of 40 lncRNAs as candidates for experimental validation. Second, we characterized the lncRNA profile during regeneration, using *Drosophila* wing imaginal disc as a regeneration-model. We selected a candidate lncRNA (*CR40469*) and evaluated its role in regeneration at the early stage of cell-damage. Subsequently, using RNA-seq data, we observed significant transcriptomic alterations in consequence of the *CR40469* genetic deletion, suggesting its role in regeneration. In this study we have generated a list of lncRNAs whose possible biological role in cell-growth and in regeneration can be further studied.

Table of Contents

Acknowledgments	iii
Abstract	v
Table of Contents	vii
List of Figures	ix
List of Tables	xi

List of Figures

List of Tables

List of Abbreviations

bp - basepair
cDNA - complementary DNA
CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi - CRISPR interference
Ctrl - Control
DE - Differentially expressed
DEG - Differentially expressed genes
DNA - Deoxyribonucleic Acid
Down - Downregulated
ENCODE - ENCyclopedia Of DNA Elements
FC - Fold change
FPKM - Fragments per kilobase million
GEO - Gene expression omnibus
GTEx - Genotype-Tissue Expression
Gtf - Gene transfer format
H3K4me1 - monomethylation of histone H3 at lysine 4
H3K4me2 - dimethylation of histone H3 at lysine 4
H3K4me3 - trimethylation of histone H3 at lysine 4
H3K9ac - acetylation of histone H3 at lysine 9
H3K9me3 - trimethylation of histone H3 at lysine 9
H3K27ac - acetylation of histone H3 at lysine 27
H3K27me3 - trimethylation of histone H3 at lysine 27
H3K36me3 - trimethylation of histone H3 at lysine 36
H3K56ac - acetylation of histone H3 at lysine 56
KO - KnockOut
lincRNA - long intervening (sometimes intergenic) noncoding RNA
lncRNA - long noncoding RNA
ML - Machine Learning
modENCODE - model organisms ENCODE
mRNA - messenger RNA (protein-coding)
ncRNA - noncoding RNA
NDE - Not differentially expressed

nt - nucleotide
PCG - Protein coding gene
PCR - Polymerase Chain Reaction
Reg - Regeneration
RFE - Recursive Feature Elimination
RNA - Ribonucleic Acid
TPM - Transcripts per kilobase million
TSS - Transcription Start Site
Up - Upregulated
UTR - Untranslated Region
Wt - Wild-type