

**Comprehensive Identification and Annotation of Non-  
protein-coding Transcriptomes from Vertebrates  
Indicates Most ncRNAs are Regulatory**

Zhipeng Qu

A thesis submitted for the degree of Doctor of Philosophy

Discipline of Genetics

School of Molecular and Biomedical Science

The University of Adelaide

October 2012

# Table of Contents

<b>Contents .....</b>	<b>I</b>
<b>Abstract.....</b>	<b>II</b>
<b>Declaration.....</b>	<b>III</b>
<b>Acknowledgements .....</b>	<b>V</b>
<b>Chapter 1 Introduction.....</b>	<b>1</b>
<b>Chapter 2 Bovine ncRNAs Are Abundant, Primarily Intergenic, Conserved and Associated with Regulatory Genes .....</b>	<b>3</b>
<b>Chapter 3 Identification And Comparative Analysis Of ncRNAs In Human, Mouse And Zebrafish Indicate A Conserved Role In Regulation Of Genes Expressed In Brain .....</b>	<b>5</b>
<b>Chapter 4 Re-construction and Annotation of Human Protein-coding and Non-coding RNA Co-expression Networks .....</b>	<b>7</b>
Introduction .....	7
Results .....	8
Discussion.....	12
Materials and methods.....	14
References .....	16
<b>Chapter 5 Conclusions and Future Directions .....</b>	<b>26</b>
<b>Supplemental Materials.....</b>	<b>29</b>

## **Abstract**

Non-coding RNAs (ncRNAs), in particular long ncRNAs, represent a significant proportion of the vertebrate transcriptome and probably regulate many biological processes. Initially, I developed a robust pipeline for the genome wide identification and annotation of ncRNAs and used publically available bovine ESTs (Expressed Sequence Tags) from many developmental stages and tissues as input. The pipeline yielded 23,060 annotated bovine ncRNAs, the majority of which (57%) were intergenic, and were only moderately correlated with protein coding genes. I then used this pipeline to annotate ncRNAs from human, mouse and zebrafish ESTs. Comparative analysis confirmed some previously described findings about intergenic ncRNAs, such as a positionally biased distribution with respect to regulatory or development related protein-coding genes, and weak but clear sequence conservation across species. Furthermore, comparative analysis of developmental and regulatory genes proximate to long intergenic ncRNAs indicated that the relationship of these genes to neighbor long ncRNAs was not conserved, providing evidence for the rapid evolution of species-specific gene associated long ncRNA. In addition, I built protein-coding and non-protein-coding gene co-expression networks based on available human transcriptome data. More than 30,000 human protein-coding and non-coding transcripts were annotated into tissue-specific co-expression sub-networks, indicating the possible regulatory connections between ncRNAs and protein-coding genes. In conclusion, I have reconstructed and annotated over 130,000 long ncRNAs, most of which are un-annotated, in human, mouse and zebrafish. Together with the annotated bovine ncRNAs, we provide a significantly expanded number of candidates for functional testing by the research community.

## **Declaration**

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Zhipeng Qu and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

*Signed*..... *Date*.....

## **List of Publications**

Qu Z and Adelson DL (2012) Evolutionary conservation and functional roles of ncRNA. *Front. Gene.* 3:205. doi: 10.3389/fgene.2012.00205

Copyright: © 2012 Qu, Adelson. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Qu Z, Adelson DL (2012) Bovine ncRNAs Are Abundant, Primarily Intergenic, Conserved and Associated with Regulatory Genes. *PLoS ONE* 7(8): e42638. doi:10.1371/journal.pone.0042638

Copyright: © 2012 Qu and Adelson. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

Qu Z, Adelson DL (2012) Identification and Comparative Analysis of ncRNAs in Human, Mouse and Zebrafish Indicate a Conserved Role in Regulation of Genes Expressed in Brain. PLoS ONE 7(12): e52275. doi:10.1371/journal.pone.0052275

Copyright: © 2012 Qu, Adelson. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## **Acknowledgements**

I would like to express my sincere gratitude to the following people:

Dave Adelson, for supervision and guidance, and for giving me the opportunity and support to do the PhD. I am so lucky to have such a nice supervisor. Dave, the knowledge that I learned from you is not just valuable for my PhD, and will support me for my whole academic career.

Jack Da Silva, my co-supervisor, for providing me the machine and office for the first several months of my PhD.

Joy Raison, for helping me to generate nice figures; Dan Kortschak, who helped me to read manuscripts and always provided valuable suggestions, and all other members of the Adelson lab, past and present, for making it such a supportive and enjoyable place to work.

Dong Wang in the Timmis lab, and everyone else in the Genetics Discipline and MLS building who has helped me along this journey.

Sandy McConachy, Richard Russell and Iris Liu, for always giving me support and advices in both research and life in Adelaide through this long journey.

My parents, for always encouraging and supporting me throughout my PhD, and for their endless love throughout my life. My sister and brother who have also always given me support and encouragement.

I would like to also specially thank the China Scholarship Council (CSC) and The University of Adelaide, who provided me this scholarship to support my PhD.