

MOLECULAR AND SYSTEMIC FUNCTIONS OF THE VERTEBRATE-SPECIFIC  
TATA-BINDING PROTEIN N TERMINUS

by

Olivier Lucas

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

Doctor of Philosophy

in

Veterinary Molecular Biology

MONTANA STATE UNIVERSITY  
Bozeman, Montana

March 2009

©COPYRIGHT

by

Olivier Lucas

2009

All Rights Reserved

APPROVAL

of a dissertation submitted by

Olivier Lucas

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Edward E. Schmidt

Approved for the Department of Veterinary Molecular Biology

Dr. Mark Quinn

Approved for the Division of Graduate Education

Dr. Carl A. Fox

## STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Olivier Lucas

March 2009

## ACKNOWLEDGEMENTS

First, I would like to thank my parents for providing me moral and financial supports during my studies. They have always encouraged me in my endeavors. In addition, I am lucky to have such a fantastic wife, encouraging me and supporting me through rough seas. I don't know if I would have made it without her.

I came to Bozeman determined to learn about the molecular biology toolbox. For this Ph.D. thesis, I felt that I had the most exciting possible project. I am thankful to my advisor, Dr. Ed Schmidt, to have handed me out such a fantastic study. Dr. Schmidt's love and dedication for science have been contagious.

That study could not have been performed without my labmates and collaborators. I am especially thankful to Dr. Justin Prigge and Dr. Elena Suvorova. Alla Bondareva was and will be a reference to me. Tedious genotyping required the help of Sonya Iverson, Marie-Clare Rollins, Ashley Siders, Stephanie Wallin and Carla Weisend. Gayle Callis has been of great help for histological procedures. My gratitude goes toward Jean Kundert for managing our mice colonies and helping me on whole animal procedures anytime I needed it. I feel honored to have generated mutant mice in collaboration with Dr. Mario Capecchi, Nobel Prize of Physiology or Medicine 2007.

I am very thankful to Dr. Valérie Copié for her help, encouragements and availability. I am thankful to all my committee members: Dr. Valérie Copié, Dr. Mark Jutila, Dr. Adam Richman, Dr. Michael White and Dr. John Paxton, for their critical comments over the years.

## TABLE OF CONTENTS

|                                                                |    |
|----------------------------------------------------------------|----|
| 1. MOLECULAR ORIGINS OF VERTEBRATE INNOVATIONS .....           | 1  |
| Introduction .....                                             | 1  |
| Canonical Markers .....                                        | 4  |
| Overview .....                                                 | 4  |
| Neural Crest Cells .....                                       | 4  |
| Placodes.....                                                  | 7  |
| Brain.....                                                     | 10 |
| Vertebrate Specializations.....                                | 14 |
| Cartilage, Bone, and Teeth in Relation to Predation.....       | 14 |
| Ear .....                                                      | 15 |
| Pharyngeal Arches .....                                        | 18 |
| Pharynx Derived Organs .....                                   | 19 |
| Thyroid .....                                                  | 19 |
| Pituitary .....                                                | 20 |
| Pancreas.....                                                  | 21 |
| Eye.....                                                       | 22 |
| Multichambered Heart .....                                     | 23 |
| Circulatory System and Endothelial Cells .....                 | 26 |
| The Adaptive Immune System.....                                | 28 |
| Morphological Asymmetry .....                                  | 30 |
| Liver .....                                                    | 32 |
| Novel Molecules .....                                          | 33 |
| Steroid Hormones.....                                          | 33 |
| Fibrillar Collagen .....                                       | 34 |
| Genome Complexity .....                                        | 35 |
| Evidence of New Vertebrate-Specific Domains and Proteins ..... | 35 |
| Role of Duplication: .....                                     | 36 |
| Exonization: .....                                             | 38 |
| Alternative and Atypical Splicing:.....                        | 39 |
| Transposition Events:.....                                     | 39 |
| Influence of Duplication on PPI Network.....                   | 39 |
| CRE Duplication and Subfunctionalization .....                 | 40 |
| Copy Number Variation .....                                    | 41 |
| MicroRNA .....                                                 | 41 |
| Conclusion .....                                               | 42 |
| 2. INTRODUCTION .....                                          | 45 |
| Evolutionary Perspectives on Transcriptional Regulation.....   | 45 |
| Intrinsic Protein Disorder and the TBP N terminus .....        | 48 |

## TABLE OF CONTENTS - CONTINUED

|                                                                                                               |           |
|---------------------------------------------------------------------------------------------------------------|-----------|
| Function of the TBP N Terminus.....                                                                           | 52        |
| Knowledge From the <i>tbp</i> <sup>ΔN/ΔN</sup> Mice.....                                                      | 53        |
| <b>3. CO-EVOLUTION OF PITX TRANSCRIPTION FACTORS<br/>WITH TBP IN METAZOANS .....</b>                          | <b>56</b> |
| Introduction.....                                                                                             | 56        |
| Materials and Methods.....                                                                                    | 61        |
| Yeast Two-Hybrid.....                                                                                         | 61        |
| Isolation of Orthologous cDNAs and Interaction Mapping Tests .....                                            | 64        |
| Mouse Embryonic Fibroblast Isolation and Transfection.....                                                    | 65        |
| RT-PCR and RNase Protection Assays .....                                                                      | 65        |
| Whole Mount In Situ Hybridization .....                                                                       | 65        |
| Results.....                                                                                                  | 66        |
| Identification of TBP-Interacting Proteins.....                                                               | 66        |
| mPitx2 is a TBP-Binding Protein.....                                                                          | 72        |
| The TBP-Pitx Interaction Involves the Metazoan-Specific Domain of TBP...                                      | 74        |
| TBP-N Attenuates Pitx2-Dependent Transcription Activation.....                                                | 76        |
| The TBP Phyla-Specific Region is involved in Nppa Regulation <i>in vivo</i> .....                             | 79        |
| Discussion .....                                                                                              | 82        |
| Summary .....                                                                                                 | 82        |
| Implications.....                                                                                             | 83        |
| The Unstructured TBP-N is a Functional Protein-Protein Interaction<br>Domain:.....                            | 83        |
| TBP-Pitx Co-evolution in Relation to Heart Development: .....                                                 | 83        |
| Consequences for the <i>tbp</i> <sup>ΔN/ΔN</sup> Mice:.....                                                   | 85        |
| Changes in the Basal Transcription Machinery:<br>a Novel Mechanisms for Evolution?.....                       | 85        |
| <b>4. GENETIC MODIFICATION OF THE TBP N TERMINUS SUGGESTS IT HAS A<br/>VERTEBRATE-SPECIFIC FUNCTION .....</b> | <b>86</b> |
| Introduction.....                                                                                             | 86        |
| Materials and Methods.....                                                                                    | 88        |
| Hagfish Targeting Vector.....                                                                                 | 88        |
| Splicing Analysis .....                                                                                       | 92        |
| Genotyping.....                                                                                               | 93        |
| Results.....                                                                                                  | 94        |
| Discussion .....                                                                                              | 96        |

## TABLE OF CONTENTS - CONTINUED

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| 5. GENE EXPRESSION AND PHYSIOLOGY OF THE “ΔN” AND “HFN” ALLELES .....   | 99  |
| Background and Summary .....                                            | 99  |
| Materials and Methods .....                                             | 100 |
| Microarrays and RT-PCR .....                                            | 100 |
| Glycemia, Insulin Tolerance Tests, and Blood Analytical Chemistry ..... | 101 |
| Histology .....                                                         | 101 |
| Bioinformatics .....                                                    | 102 |
| Liver Transcriptome Analysis .....                                      | 102 |
| Introduction .....                                                      | 102 |
| Results .....                                                           | 103 |
| <i>cd36</i> and Putative Regulatory Mechanisms .....                    | 107 |
| Introduction .....                                                      | 107 |
| Results .....                                                           | 108 |
| Lipid Accumulation and Insulin Response .....                           | 111 |
| Introduction .....                                                      | 111 |
| Results .....                                                           | 114 |
| Pre-weaning and Adult Phenotype Similarities .....                      | 117 |
| Introduction .....                                                      | 117 |
| Results .....                                                           | 118 |
| Discussion .....                                                        | 127 |
| Steatosis, <i>cd36</i> , and Putative Regulatory Elements .....         | 128 |
| Insulin Signaling .....                                                 | 129 |
| Growth Retardation and Cholesterol Metabolism in P11 Pups .....         | 130 |
| Vertebrate-Specific Genes and TBP-N .....                               | 132 |
| Adipose and Hepatic Tissue Crosstalk .....                              | 132 |
| 6. SUMMARY AND CONCLUSION .....                                         | 134 |
| REFERENCES .....                                                        | 137 |
| APPENDICES                                                              |     |
| APPENDIX A: Luciferase Assay Protocol .....                             | 185 |
| APPENDIX B: Generation of Mouse Embryonic Fibroblast Culture .....      | 192 |
| APPENDIX C: Whole Mount In Situ Hybridization Protocol .....            | 194 |
| APPENDIX D: Adult Livers Microarray Data .....                          | 199 |
| APPENDIX E: P11 Microarray Data .....                                   | 206 |

## LIST OF TABLES

| Table                                                                                                                                                                                          | Page |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 3.1. Oligonucleotides sequences .....                                                                                                                                                          | 62   |
| 3.2. Yeast two-hybrid cDNA libraries characteristics .....                                                                                                                                     | 63   |
| 3.3. Yeast two-hybrid screens .....                                                                                                                                                            | 68   |
| 3.4. TBP-interacting proteins identified by yeast two-hybrid screens .....                                                                                                                     | 69   |
| 3.5. Amino acid sequences used for phylogenetic analyses .....                                                                                                                                 | 71   |
| 4.1. Oligonucleotides primers used in the generation and<br>characterization of the <i>tbp<sup>hfN</sup></i> allele .....                                                                      | 94   |
| 4.2. Effect of the <i>tbp<sup>hfN</sup></i> allele on survival at weaning.....                                                                                                                 | 95   |
| 5.1. List of differentially abundant endogenous retroviral mRNAs .....                                                                                                                         | 106  |
| 5.2. Representative events and associated genes in fatty acid<br>metabolism and their abundance in <i>tbp<sup>ΔN/ΔN</sup></i> and <i>tbp<sup>hfN/hfN</sup></i> .....                           | 107  |
| 5.3. List of mRNAs differentially abundant in <i>tbp<sup>ΔN/ΔN</sup></i> versus <i>tbp<sup>+/+</sup></i><br>livers that are transcribed from genes having associated Pitx2 binding sites ..... | 111  |
| 5.4. Functional classification of the differentially abundant mRNA in P11<br><i>tbp<sup>ΔN/ΔN</sup></i> belonging to the lipid metabolism category .....                                       | 121  |

## LIST OF FIGURES

| Figure                                                                                                                                                                  | Page |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1.1. Cladogram of the major metazoan groups discussed in the text.....                                                                                                  | 2    |
| 1.2. Antero-posterior gene expression of CNS marker genes<br>in amphioxus and a model vertebrate .....                                                                  | 12   |
| 1.3. Generic vertebrate embryo.....                                                                                                                                     | 17   |
| 2.1. Basal transcription machinery.....                                                                                                                                 | 46   |
| 2.2. TBP folding prediction .....                                                                                                                                       | 50   |
| 3.1. TBP metazoan phylogeny .....                                                                                                                                       | 58   |
| 3.2. Pitx phylogeny .....                                                                                                                                               | 60   |
| 3.3. Yeast two-hybrid baits and quality controls .....                                                                                                                  | 67   |
| 3.4. Identification of hfPitxA as a TBP-N-interacting protein.....                                                                                                      | 70   |
| 3.5. Identification of mouse orthologs TBP-binding partners .....                                                                                                       | 73   |
| 3.6. The TBP-Pitx interaction involves the MSD .....                                                                                                                    | 75   |
| 3.7. The TBP MSD modulates mPitx2c activity on the <i>nppa</i> promoter .....                                                                                           | 78   |
| 3.8. Cardiac <i>Nppa</i> mRNA expression in wildtype and <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> mice.....                                                 | 80   |
| 4.1. Schematic representation of the <i>tbp</i> alleles.....                                                                                                            | 88   |
| 4.2. Outline of the design strategy that was used to produce the <i>tbp</i> <sup>hfN</sup> allele .....                                                                 | 91   |
| 5.1. Transcriptome analyses of livers from <i>tbp</i> <sup>+/+</sup> , <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> and <i>tbp</i> <sup>hfN/hfN</sup> mice..... | 105  |
| 5.2. <i>cd36</i> is the only gene which regulation is affected by by either<br>removal of TBP-N, or replacement by the hfTBP-N.....                                     | 110  |
| 5.3. Differential fat accumulation in <i>tbp</i> <sup>+/+</sup> , <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> and <i>tbp</i> <sup>hfN/hfN</sup> livers.....    | 115  |
| 5.4. Glycemia, insulin tolerance tests, and blood analytical chemistry .....                                                                                            | 116  |

## LIST OF FIGURES - CONTINUED

| Figure                                                                                                                                                     | Page |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 5.5. Growth retardation and steatosis of 11-day old animals.....                                                                                           | 119  |
| 5.6. Sterol metabolism is compromised in P11 $tbp^{ΔN/ΔN}$ animals.....                                                                                    | 122  |
| 5.7. Identification of the mRNAs affected in young and adult $tbp^{ΔN/ΔN}$<br>and $tbp^{hfN/hfN}$ animals compared to age-matched $tbp^{+/+}$ animals..... | 124  |
| 5.8. Effect of low or high fat diet on CD36 mRNA abundance<br>and hepatic steatosis.....                                                                   | 126  |

## NOMENCLATURE

Genetic loci are represented with lowercase italic. Allelic designation follows as a superscript. The two alleles of a locus are separated by a slash. Thus  $tbp^{\Delta N/\Delta N}$  designates the *tbp* locus where both alleles lack the TBP N terminus protein coding sequence.

Protein names are capitalized. Thus TBP is the protein coding sequence for the TATA Binding Protein encoded in the *tbp* locus.

TBP-N is the abbreviation for the N-terminal protein coding sequence region of the TBP protein.

## ABSTRACT

The invertebrate/vertebrate transition and associated innovations can be regarded as a major event in evolution. Recent molecular progresses invite to an analysis of the events leading to the apparition of vertebrates and the underlying embellishment in gene regulation. In eukaryotes, the TATA-Binding Protein (TBP) has a central role in transcription initiation of most genes. TBP is comprised of a highly conserved DNA binding domain and, in vertebrates, it also contains a novel region: the N-terminal TBP protein coding sequence. The role of the TBP-N is largely unknown, but previous studies suggest that it is important for fetal survival. Most animals lacking the TBP-N (*tbp*<sup>ΔN/ΔN</sup>) die before weaning. The goal of the present work was to establish a deeper knowledge of the vertebrate-specific TBP-N. It was hypothesized that TBP-N could be involved in protein-protein interactions and that the high degree of similarity of TBP-N protein sequences in different species could correlate with similar functions. To test those hypotheses, two independent approaches were taken: (1) Protein-protein interactions involving the TBP-N via unbiased screens were characterized. (2) The mouse TBP-N was replaced by a similar and homologous TBP-N, *in vivo*, through homologous recombination. The TBP-N-replacement mutation was characterized through pathway analyses, bioinformatics, and whole-animal physiology. Screens for proteins interacting with the TBP-N of hagfish (hf), a basal vertebrate, uncovered hfPitxA. The Pitx family of transcription factors are proteins important in vertebrate development. The mouse paralogs of hfPitxA, Pitx1 and Pitx2, were found to interact with the mouse TBP-N. Moreover, the interaction appeared functional as it regulated the expression of *nppa*, a known target gene of Pitx2. *In vivo* replacement of the mouse TBP-N with the similar hfTBP-N did not affect the survival. Gene expression analysis indicated that lipid metabolism pathways were affected in animals lacking the TBP-N or when the hfTBP-N was present. Further analyses pointed toward a potential defect in insulin response and an abnormal hepatic fat storage. The data presented here argues in favor of an important role for TBP-N in vertebrate-specific gene regulation. More specifically, it is likely involved in heart development and in regulation of lipid metabolism.

## MOLECULAR ORIGINS OF VERTEBRATE INNOVATIONS

### Introduction

It has long been thought that evolutionary changes correlate with an increase in complexity<sup>1-3</sup>. Measures of complexity have historically been based on morphological structures (number of organs, cell types, tissues and body segments) while the underlying molecular organization has only recently become accessible.

The Cambrian explosion is the best known period of increased phenotypic complexity<sup>4</sup>; however it does not seem to correlate either with a major genesis of novel genes or duplication and divergence of genomes, suggesting neither mechanism was involved<sup>5</sup>. Nevertheless, it is becoming widely accepted that two whole-genome duplications (WGD), the “2R hypothesis”, occurred at or around the invertebrate/vertebrate transition<sup>6-15</sup> (Fig. 1.1). Studies focused on molecular aspects of extant species<sup>16</sup> suggest that the vertebrate transition involved acquisition of many novel structures. The studies, however, could not include early vertebrate lineages that are now extinct and, for some authors, the notion of a burst of complexity may not be as straightforward when extinct species are considered<sup>17</sup>. Further, the invertebrate/vertebrate transition may have been a more gradual period of character acquisitions rather than a single step.

Here we review the literature and mine data from recent sequencing projects on non-vertebrate metazoans (sea urchin, *Ciona*, and amphioxus) to better understand the origins of “vertebrate novelties”. We show that a great number of vertebrate molecular

innovations had deep roots in non-vertebrate chordates, where partial systems gave rise to discrete vertebrate organs.

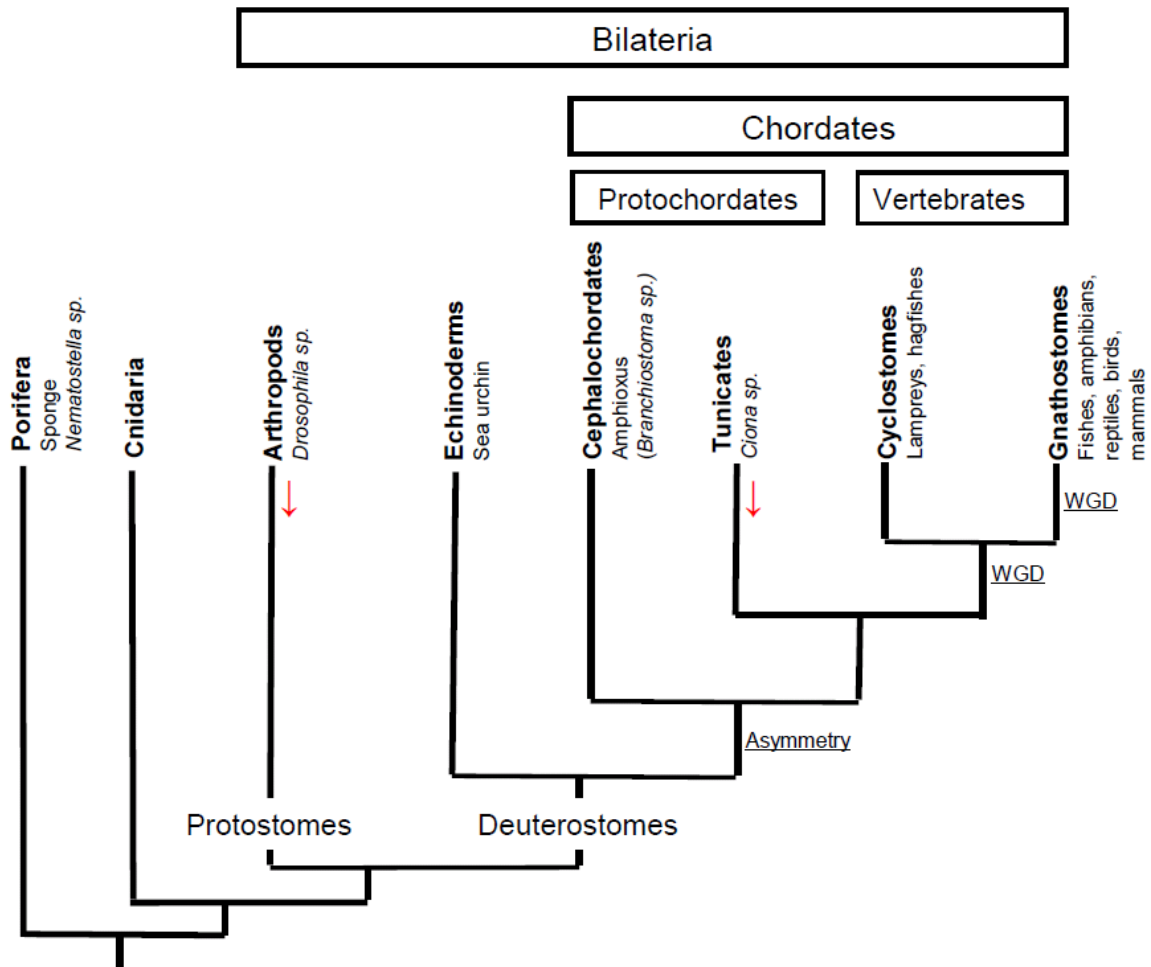


Figure 1.1: Cladogram of the major metazoan groups discussed in the text. Simplified relationship based on molecular and morphological evidence between the organisms discussed in the text is drawn. Relevant species are mentioned in italics along their vernacular name. Red arrows indicate a phyla with small genomes where a gene loss is likely. WGD indicate the position of a putative whole genome duplications. Appearance of asymmetry is indicated. Adapted from <sup>18-20</sup>.

The change from filter-feeding organisms (protochordates) to a more predator-driven ecosystem <sup>21</sup> and two successive WGD could have catalyzed the vertebrate radiation during this period. Most of the new characters unique to vertebrates relate to the

head <sup>21</sup> and enable the predation-driven lifestyle of most vertebrates. Vertebrates have a complex and highly regionalized brain encased in a skull, external sensory organs, a muscularized and specialized pharynx and an endoskeleton. Those characteristics may be seen as additional modules to the basic body plan of protochordates.

A module is a self-contained, independent component of a system which can interface with other components and can be interchanged or replaced to form different structures. Therefore, it can be seen from the organism scale or the molecular scale. Darwin suggested the existence of “core elements” that could be duplicated and adapted to different roles <sup>22</sup>, therefore suggesting modularity of organisms <sup>23</sup> at the morphological level. Analysis of modularity at a molecular level and its role in development and evolution has recently been investigated <sup>24</sup>. A module may therefore refer to a group of proteins and *cis*-regulatory elements (CRE). Modularity is thought to drive evolution through co-option and reduced pleiotropy <sup>24,25</sup> as, for example, an interface of the module may be functional in only one tissue despite its organism-wide distribution.

We will start with an overview of the canonical markers of vertebrates, which are extensively reviewed in the current literature, namely the neural crest (NC) cells, the placodes, and the brain. Some overlooked vertebrate characters will then be considered. Finally, genetic mechanisms potentially responsible for such innovations will be examined.

## Canonical Markers

### Overview

Most of the vertebrate innovations are the result of two cell types: NC cells and placodes. They are both versatile ectodermal cell populations that differentiate into many cell types, including neurons, glia, secretory and sensory cells. They originate near the border of the neural plate (ectoderm) and have migratory abilities. Despite these similarities, some specific characteristics differentiate those two cell types. First, NC cells originate from the head and trunk while placodes are confined to the head. Second, NC cells have a greater repertoire of differentiation than placodes. Third, NC cells are a homogenous delaminating and migrating cell population while placodes are more heterogeneous and originate from thickening of the non-neural ectoderm. Fourth, both have specific gene regulatory networks. In addition to NC cells and placodes, another hallmark of vertebrates is the brain. The brain is a great enrichment of neurons in the anterior part of the neural tube responsible for integrating stimuli and mounting a coordinated response.

### Neural Crest Cells

It is widely thought that the NC cells are central to vertebrate innovations and their evolution <sup>21</sup>. They are so fundamental that some authors even go as far as assimilating them to a fourth germ layer <sup>26</sup> specific to vertebrates. The closure of the neural tube from neural plate border generates a transient migratory and pluripotent cell population, the NC cells, which are progenitor cells of many structures. Multiple clues will specify their differentiation outcomes during and after migration. They can be

classified into two categories: neuroglial or non-ectomesenchymal (neurons, glia, and pigment cells), and skeletogenic or ectomesenchymal (bone, cartilage, dentine) <sup>27</sup>. Outstanding reviews about NC biology have been recently published <sup>27,28</sup>. We will just touch upon some of their relevant aspects for vertebrate innovations.

Recently, NC biology has been described from a gene regulatory network (GRN) perspective, where multiple markers organized in modules are used to define NC cells, leading to a more comprehensive approach <sup>28-31</sup>. Many components of the NC cell GRN have been found in invertebrates <sup>27,30</sup> but the synergies, new genes, and interactions <sup>32</sup> appear to be only found in vertebrates <sup>32-34</sup>. Nevertheless, the concept of GRN for use as a definition for NC cells can be seen as a limitation as it does not consider the evolutionary processes needed to build the GRN. It is conceivable that the pluripotency observed in NC cells gradually arose from use of ancestral signaling pathways. This could have led to multiple temporal and spatial functions for the NC cells.

The recent re-assessment of chordate evolution (Fig. 1.1) <sup>35</sup> suggests that tunicates are closer relative to vertebrates than cephalochordates even though tunicates underwent some drastic specialization leading to loss of some ancestral characteristics still found in cephalochordates <sup>35-37</sup> (Fig. 1.2). *Amphioxus* and *Ciona* express several homologues of NC cells markers, including transcription factors and signaling molecules responsible for vertebrate-like structures (*BMP2/4*, *Pax3/7*, *Msx*, *Snail*) <sup>38-41</sup>. Co-option of novel genes and/or gene duplication, like the *SoxE* family <sup>33,42</sup>, by NC cells can help explain the transcriptional changes required for the novel NC derived specializations <sup>32,43-45</sup>.

Amphioxus has cells representing likely evolutionary source of the vertebrate NC cells. Nevertheless, it is difficult to know if they are homologous. These cells, which originate at the border of the neural plate, exhibit some migratory capabilities and gene expression patterns similar to NC cells (*AmphiDll*<sup>46</sup>, *AmphiPax3/7*<sup>47</sup>, *AmphiSox1/2/3*<sup>48</sup>, *Snail*<sup>49</sup>). The more developed GRN involving those genes found in vertebrates<sup>50,51</sup> are likely the result of gene co-option and novel or modified CRE. As an example of gene co-option, both amphioxus and lamprey express *Id*, which is used as a NC cell marker, but only lamprey exhibits expression at the neural plate border<sup>44</sup>.

Tunicates also have NC-like cells which have migratory and differentiation properties but no evidence of pluripotency<sup>40</sup>. This cell population could also be the result of a convergent evolution restricted to pigmentation<sup>29</sup> and not be homologous to NC cells. This is supported, in part, by the presence of NC-like cells in tunicates. The existence of a NC-like or NC precursor in tunicates suggests that evolution of NC cells was a gradual and continuous process rather than an abrupt innovation at the invertebrate-vertebrate transition. This view is supported by the fact that the extent of NC contribution to the head is lowest in basal vertebrates and gradually increases in more recent vertebrate lineages<sup>52-54</sup>.

To explain the potential gradual appearance of NC cells, it has been proposed that, at the base of the chordates phylum, a population of cells gained a migratory function and a skeletal fate. Some of these cells underwent an epithelial-mesenchymal transition and began to delaminate from the neural plate border and migrate in vertebrates<sup>55</sup>. Therefore, the chordate common ancestor had the basic wiring for NC-like activity.

Refinement of the underlying GRN could have been important for the development of vertebrate NC <sup>56</sup>.

### Placodes

Placodes are regions of the vertebrate head where ectoderm invaginates to form distinct cell types that contribute to sensory organs and cranial sensory ganglia. Extensive reviews on placodes are available for a more throughout analysis <sup>57-59 60,61</sup>. We will focus on their relevance as generators of vertebrate characters. Sensory placodes give rise to the eyes, ears, olfactory organs and lateral lines while the neurogenic placodes differentiate into sensory neurons and cranial ganglia <sup>61</sup>. There are eight placodes: (1) the lens placode for eye formation; (2) the otic placode for the inner ear; (3) the olfactory placode, which produces sensory cells, mucus or neuropeptide-secreting cells and glial cells including Schwann cells responsible for the neuronal sheath <sup>62-65</sup>; (4) the lateral line placode, which gives rise to a peripheral sensory system in fish <sup>66-68</sup>; (5) the adenohypophysal placode, which forms the adenohypophysis (i.e. the Rathke's pouch) through stomodeum invagination; (6) the epibranchial and hypobranchial placodes, which originate in a dorsal region posterior to the pharyngeal pouches and give rise to multiple head and neck sensory neurons; (7,8) the profundal and trigeminal placodes, which form the sensory profundal and trigeminal nerves, respectively.

It is thought that placodes derive from a common panplacodal primordium <sup>59,69</sup>. Development of a mature placode requires the activities of dozens of transcription factors (see below) through a modular and multilayer regulatory process <sup>61</sup>. Several of the key transcription factors for placodes formation will be critically reviewed here.

Many placodes marker genes have been identified in protochordates. Expression of the *Six1/2*, *Six4/5* and *Eya* families is common to the panplacode primordium<sup>61</sup>. The *Six* gene family encodes homeodomain transcription factors that are present in metazoans, not just vertebrates<sup>70</sup>. Also, the *Eya* gene family is found in invertebrates<sup>71</sup>, *Eya* proteins act as coactivators of *Six* expression in eye development of vertebrates<sup>71-75</sup>. Similarly, in *Drosophila*, *Eya* and *Six1/2* homologs are major component of the network responsible for the development of the compound eye<sup>76</sup>. *Amphioxus* has a cluster of sensory epithelial cells located rostrally which has some similar functions to the vertebrate olfactory cells<sup>77,78</sup>. Interestingly, these cells in *amphioxus* express homologs of olfactory placode markers: *AmphiMsx*<sup>78</sup> and *AmphiNeurogenin*<sup>48</sup>. In vertebrates, *Msx* and *Neurogenin* are developmental specific transcription factors<sup>48,78</sup>. Another family of transcription factors required for most placodes is the *Dlx* family. *Dlx* genes are expressed in many vertebrate structures including forebrain, neural crest cells, placodes, pharyngeal arches, limb buds and teeth<sup>79</sup>. Interestingly, *amphioxus* has one *Dlx* gene<sup>46</sup> while lamprey has four<sup>79</sup> and mouse has six<sup>80</sup>. The two *Dlx* genes of *Ciona*<sup>81</sup> likely result from a lineage-specific duplication event<sup>79</sup>. Taken together, the distribution of these transcription factors likely suggests a gradual evolution of placode GRN.

The diversity of placodes and their GRN suggest that such a complex system may have a pre-vertebrate origin rather than an abrupt appearance at the invertebrate/vertebrate junction<sup>58</sup>. Structures with comparable roles and molecular signatures as vertebrate placodes (termed “placode-like”, placode homolog or proto-placode) have been identified in protochordates<sup>19,82-86</sup>. For example, the colonial ascidian

*Botryllus schlosseri* has an olfactory placode-like region of the neurohypophyseal duct that delaminates and migrates<sup>87</sup>. Therefore, even though olfactory placodes are found in vertebrates only, potentially similar functions exist in other chordates. The similarity between the vertebrate olfactory, adenohypophysal, and otic placode with the tunicate ventral organ, ciliary funnel, and Langerhans placodes can be compatible with the unifying placodal primordium evidence<sup>86</sup>. Thus, it is likely that the vertebrate placodes arose through multiple recruiting and increased complexity of “placode-like” structures present in protochordates<sup>30</sup>.

In light of these recent developments, one may ask how relevant the concept of placode as a characteristic of “vertebrateness” is. The concept of placode is therefore probably better seen, at least from an evolutionary and parsimonious perspective, as a character common to all chordates as it has already been suggested elsewhere<sup>86</sup>. A model of accretion and divergence of pre-existing cell types in a chordate common ancestor has been presented to explain the evolutionary origins of placodes in chordates<sup>30,75</sup>.

For 25 years, placodes have been used as a characteristic of vertebrates<sup>21</sup> but the authors were clear on the fact that placodes may derive from an ancestral “epidermal nerve plexus” present in protochordates. Therefore, the recent deeper understanding of the molecular mechanisms blurs the distinction between the vertebrate placode and the “epidermal nerve plexus” and suggests a likely homology with gradual refinement. A recent model was developed where the chordate common ancestor would have had an array of epidermal sensory cluster cells, with comparable genetic markers to the vertebrate placode. Early vertebrates would have recruited them in the anterior part of the

body and embellishments of the GRN would have contributed to generate placodes<sup>30,60</sup>. Alteration of the anterior signaling could have provoked this new patterning<sup>30,88</sup>. From an evolutionary point of view, the term placode regroups ancestral and more recent innovations with a common basal GRN and additional modules in vertebrates<sup>58</sup>. This suggests that placodes did not arise at once but that they gradually appeared in protochordates, and that highly developed sense organs were present in the anterior region of protochordates.

### Brain

In vertebrates, the anterior segment of the neural tube is greatly enlarged and compartmentalized thus forming the brain. In comparison to vertebrates, amphioxus has a more linear and poorly segmented brain and is more reliant on peripheral sensory neurons and plexuses<sup>9</sup>. The vertebrate brain is also more centralized and includes novel NC- and placode-derived structures<sup>89</sup>. Nevertheless, the basal regionalization (forebrain and hindbrain) and the anteroposterior organization, reflected by comparable but less segmented expression of the marker genes *Otx* and *Hox*, is conserved between protochordates and vertebrates (see below). Comparative analysis of expression patterns of brain development genes between amphioxus and vertebrates revealed that the anterior nervous system of amphioxus and vertebrates share some striking similarities, as discussed below. In an anteroposterior order, the vertebrate brain is composed of four regions anterior to the spinal cord: the telencephalon and diencephalon forming the forebrain, the midbrain, and the hindbrain. The midbrain-hindbrain boundary (MHB) is

considered as a major anterioposterior organizer. In addition, the vertebrate hindbrain can be subdivided in segments, or rhombomeres<sup>90</sup>, which are precursors of the cerebellum.

Marker genes are used to define each region (Fig. 1.2). *BF1* is the most anterior marker of the telencephalon in vertebrates<sup>91</sup>. An anterior cell population in the amphioxus neural tube expresses a *BF1* homolog<sup>92</sup> despite the fact that the telencephalon is considered a vertebrate innovation<sup>93,94</sup>. *Otx* is a classical marker of the telencephalon, diencephalon and midbrain in vertebrates<sup>95</sup>. Expression of *AmphiOtx*, the amphioxus homolog of vertebrate *Otx*, is regionalized to the analogous region of the poorly segmented anterior neural tube of amphioxus<sup>84</sup>. Another marker of the vertebrate diencephalon, *Nkx2-1*, has a homolog present in amphioxus (*AmphiNk2-1*) and is similarly expressed in its cerebral vesicle suggesting a likely homology between the two structures<sup>96</sup>.

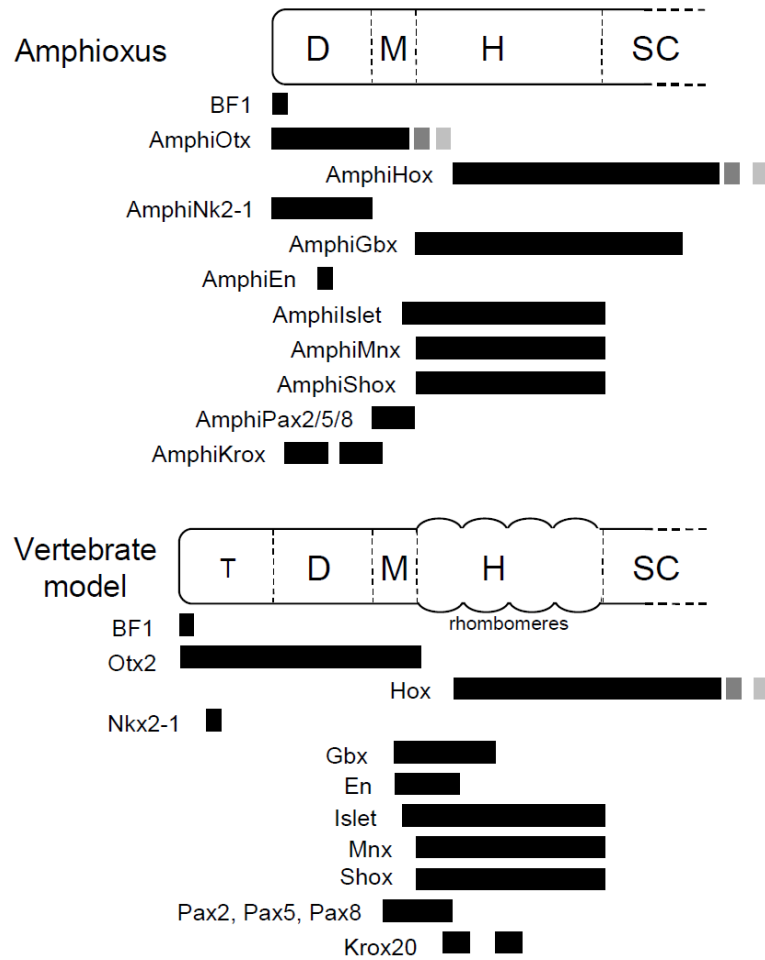


Figure 1.2: Antero-posterior gene expression of CNS marker genes in amphioxus and a model vertebrate (dorsal views). Diagrammed are the anterior regions of the amphioxus and a generic vertebrate CNS. The expression of the genes discussed in the text (left) is qualitatively represented as a black bar. A lower expression level is represented in grey. Multiple *Hox* genes are expressed in fine patterns in the hindbrain and spinal cord of both models, but only a generic expression pattern is shown. CNS: central nervous system; D: diencephalon; M: Midbrain; H: Hindbrain; SC: spinal cord; T: telencephalon. See text for references.

The MHB in vertebrates and amphioxus shares similar expression patterns of the two marker genes *Gbx* and *Otx*, nevertheless, the organizer function of the MHB, conducted through cell fate determination, is not found in amphioxus. The genes responsible for the organizer function are not expressed at the MHB in amphioxus<sup>9</sup>. Studies on *engrailed*, another marker of the MHB in vertebrates<sup>97</sup>, are conflicting in

protochordates. In amphioxus, it is not expressed at the MHB but in the diencephalon<sup>97</sup> while it is in tunicates<sup>98</sup>. Therefore, even though the genetic material needed for a vertebrate midbrain/hindbrain organizer is present, its role, accomplished mostly through co-option in vertebrates, is the real vertebrate innovation.

The hindbrain is marked by comparable expression profiles of *Islet*, *Mnx* and *Shox* in vertebrates and amphioxus<sup>99-101</sup>. The major vertebrate novelty is the great segmentation of the hindbrain in rhombomeres under retinoic acid regulation, regionalized *Hox* genes expression, and ephrin/eph signaling<sup>102-104</sup>. The *Hox* gene cluster is found in all bilateria and dictates segmentation and anteroposterior axis. An increasing copy number and divergence of *Hox* clusters correlates with an increasing body plan complexity. The classical example of increased brain complexity in vertebrates correlates with the evolution of the *Pax* family and the cerebellum. Because there is one *Pax* family member in invertebrates (*Pax2/5/8*)<sup>105</sup> and three in vertebrates (*Pax2*, *Pax5*, and *Pax8*), it is tempting to draw a positive correlation between increased gene family size and brain complexity<sup>16</sup>, in terms of compartments. A marker of hindbrain and rhombomeres in vertebrates is *Krox-20*, which has a very restricted expression pattern<sup>106</sup>. In amphioxus, *Krox* does not exhibit a related expression pattern, but rather is expressed in the anterior neural tube<sup>101,107</sup>. Therefore the existence of a similarly functional MHB and its role as a major anteroposterior organizer like the vertebrate MHB is discussed in protochordates<sup>105,108</sup>.

In conclusion, vertebrates, through segmentation, added complexity to the basic brain organization found in amphioxus. Vertebrates possess a more regionalized brain

with a MHB and multiple rhombomeres involved in the development of cranial nerves. The “vertebrateness” of the telencephalon correlates well with its role in movement, sensory processing, communication, learning, and memory. The MHB and the complex subregionalization of the brain could thus be the true vertebrate innovation.

### Vertebrate Specializations

#### Cartilage, Bone, and Teeth in Relation to Predation

Cartilage, bone, and teeth are the major mineralized parts found only in vertebrates, even though cartilage mineralization is rare<sup>109,110</sup>. In the earliest vertebrates, the mineralized endoskeleton enabled locomotion, the dermal skeleton was used for protection, and teeth enabled predation<sup>21</sup>. The tooth is composed of enamel or enameloid, dentin, and bone<sup>111</sup>. The bone is formed through a deposit of hydroxyapatite while the cartilage is composed of proteoglycans and has been suggested to be an ancestral mechanoreceptor before being co-opted for support of the muscularized pharynx and, later, the skeleton and joints<sup>21</sup>.

Mineralization is regulated by extracellular matrix (ECM) proteins constituting the heterogeneous calcium-binding phosphoprotein (SCPP) family. The vertebrate radiation and the appearance of the mineralized skeleton coincides with the duplication of SCPP<sup>112</sup>. Duplication of the *sparc* gene (also called “osteonectin”), probably after the cyclostomes branched out from the other vertebrates, generated SPARCL1 (secreted protein, acidic, cysteine-rich like 1)<sup>113</sup>. SPARCL1, through duplication and neofunctionalization in lamprey, gave rise to the vast SCPP family<sup>112-114</sup> and SPARCA and SPARCB<sup>113</sup>, suggesting a parallel role for SPARCA/B in cyclostomes and

SPARCL1 in gnathostomes. SCPPs were later co-opted for the eggshell matrix in birds and for milk casein and salivary proteins in mammals <sup>113</sup>. Even though tooth structure in teleosts and tetrapods is similar, different non-homologous SCPPs regulate tooth development <sup>115</sup>. This suggests a flexible role for SCPP in gradual evolution of the skeleton. It is thought that tooth development co-opted the already present GRN found in the dermal skeleton <sup>113</sup>. Limb skeletogenesis derives from the lateral plate mesoderm <sup>116</sup> while the cranium is cephalic mesoderm- and NC-derived <sup>117</sup>. Despite this heterogeneity, molecular pathways in the limb and head share a common regulatory core around the more ancient transcription factors *Sox*, *Msx* and *Dlx* <sup>118</sup>. It has been suggested that subtle regulatory differences could explain some differences between limb and cranial skeletogenesis. Therefore, many of the GRN required for bone formation find homologs in non-vertebrates <sup>119</sup>.

### Ear

During ear formation, the otic placodes give rise to neurons and then to sensory hair cells and support cells derived from a common progenitor <sup>61</sup>. Rhombomeres also contribute cells to the inner ear <sup>120</sup>. In addition to the inner ear, the core element present in all vertebrates, stepwise refinements led to the presence of an outer and middle (ossicle and ear drum) ear in mammals. Hair cells act as sensors and transducers of the external stimuli <sup>121</sup>. The inner ear, through semi circular canals, is responsible, in higher vertebrates, for detection of acceleration and gravity while the cochlea is responsible for sound perception <sup>121</sup>. The inner ear, which is absent in invertebrates, underwent a

stepwise morphological evolution in vertebrates where hagfishes have one semi-circular canal, lamprey have two, and jawed vertebrates have three <sup>121,122</sup>.

The otic placodes (Fig. 1.3) or their comparable structure are thought to be present in protochordates <sup>122</sup> (see section on placodes, above). Expression of the tunicate *Pax2/5/8* and the vertebrate *Pax2*, *Pax5* and *Pax8* in the atrial primordium suggests a homologous process underlying formation of the structure in protochordates and the vertebrate otic placode <sup>84</sup>. Hair cells of the inner ear are functionally comparable to the *Drosophila* external sensory organ, the mechanosensory bristle, <sup>123,124</sup>. Vertebrates and *Drosophila* express homologous genes in otic placodes and bristles respectively <sup>121</sup>. Also, comparative gene expression of the *Drosophila* chordotonal organ (composed of mechanosensory receptors) and the vertebrate inner ear show a remarkably conserved GRN <sup>121</sup>. Interestingly, in the inner ear, the vertebrate novelty arises mainly from recruitment of NC-derived glial cells. Also, mouse *Math1*, a gene essential to hair cell formation in the mammalian ear <sup>125</sup> can functionally replace the *Drosophila* homolog, *atonal*, in the nervous system <sup>126</sup>.

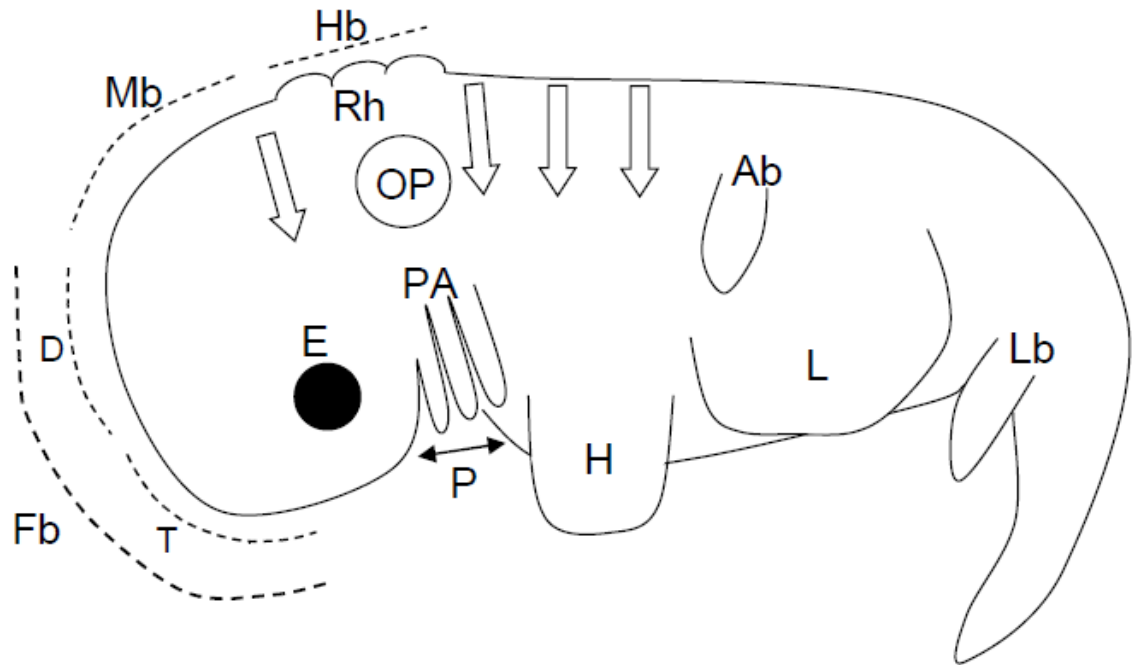


Figure 1.3: Generic vertebrate embryo. Diagrammed is a simplified generic vertebrate embryo showing the structures discussed in the text. Arrows represent migrating neural crest cells. AB: arm bud; D: diencephalon; E: eye; Fb: forebrain; H: heart; Hb: hindbrain; L: liver; LB: leg bud; Mb: midbrain; OP: otic placodes; P: pharynx; PA: pharyngeal arches and pharyngeal pouches; Rh: rhombomeres; T: telencephalon.

Neurotrophins, which are neuron-specific growth factors, and their receptors, the Trk family of receptor tyrosine kinase, have long been thought to be vertebrate-specific and central to the formation of sensory neurons. Nevertheless, recent analysis of non-vertebrate genomes indicates the presence of at least one *Trk* receptor gene in amphioxus, *amphi-Trk*, and presence of neurotrophin-like proteins and Trk proteins in the sea-urchin<sup>127,128</sup>. Interestingly, neurotrophin and Trk have not been found in tunicates<sup>129</sup>. Duplications in the vertebrate lineage greatly diversified the Trk and neurotrophin families<sup>130</sup>, correlating with their diversified roles in vertebrates.

Co-option of pre-existing genes (*Otx*, *Fgf* and *Gata*, reviewed in<sup>121</sup>) is a hallmark of ear evolution despite the vertebrate ear being a seemingly obvious morphological

innovation than can only be partially related to ancestral bilaterian structures. Therefore, a stepwise formation of the vertebrate ear is suggested, maybe through increased co-option, use of new paralogous genes, and modifications of some GRN.

### Pharyngeal Arches

Pharyngeal arches are mesoderm derived structures from the embryonic pharynx that give rise to multiple skeletal, muscular, nervous, and arterial structures in the adult neck and lower head (see below). Pharyngeal pouches are the invagination between two pharyngeal arches. Ventral migration of NC cells interacting with pharyngeal endoderm extensions around the sixth aortic arch artery initiates pharyngeal arch development. The pharyngeal endoderm, along with some NC contribution, further differentiates into cartilage rods adapted to bony, cartilaginous, or ligamentous structures. Each arch can be thought as an endodermal pouch and an ectodermal groove that will derive later structures. Derivative structures of each pouch yield skeleton, viscera, arteries, muscles motor nerves, and sensory nerves<sup>131</sup>. The endoderm component of the pharyngeal system and its remodeling are required for normal morphogenesis<sup>132</sup>. The ventral side of the second pouch will form the thyroid while the thymus is derived from the third pouch. The formation of the tetrapod-specific parathyroid gland from the third and fourth pouches relies on the *Gcm-2* transcription factor, which is a homolog of *Drosophila* glial cells missing 2 (*Gcm-2*). *Gcm-2* has been co-opted for calcium homeostasis, a new function found in tetrapods. Pharyngeal pouches can form without NC cells and the pharyngeal segmentation is not vertebrate-specific<sup>133</sup>. Markers of pharyngeal identity (*Bmp7*, *Fgf8*, *Pax1*, *Tbx1*) are conserved developmental genes that have invertebrate homologs<sup>132</sup>.

Pharyngeal pouches are not vertebrate-specific as they form in amphioxus<sup>46,105</sup>. The addition of the NC component to generate the archeal support is, however, unique to vertebrates.

### Pharynx Derived Organs

The segmentation, muscularisation, sub-specialization, and NC-derived skeletal support of the vertebrate pharynx are vertebrate novelties. As mentioned above, parts of the pharynx gave rise to endocrine organs: thyroid, parathyroid, and pituitary gland.

Thyroid: The thyroid is an endocrine organ involved in energy metabolism. The thyroid is under the control of the pituitary and hypothalamus. The hypothalamus is a region of the brain that links the neural and endocrine systems through the pituitary endocrine gland. The thyroid is also involved in calcium homeostasis through calcitonin secretion. It has long been recognized that there is a common origin to the vertebrate thyroid gland and the protochordates endostyle<sup>134</sup>. In vertebrates, localized budding from the ventral floor of the embryonic pharynx will generate the thyroid gland. Also, some enzymatic machinery and developmental processes overlap between endostyle and thyroid<sup>135-138</sup>. Interestingly, lampreys exhibit an endostyle during their larval stage that will directly convert to a thyroid gland in the adult<sup>137,139</sup>. The chordate endostyle has two functions: (1) protein secreting in the filter feeding process and<sup>140</sup> (2) thyroid-like activity through overlapping expression of the genes responsible for thyroid hormone synthesis<sup>141</sup>. *Nkx2-1* expression in amphioxus suggests that the endostyle may be homologous to the vertebrate thyroid gland<sup>96</sup>. The presence of an endostyle in the larval stage of the lamprey, being replaced by a thyroid in the adult, supports the hypothesis of

gradual modification. Even though the endostyle has no described endocrine function, it is almost certainly homologous to the vertebrate thyroid. Comparative orthologous gene expression patterns are strikingly similar for *Pax8*, *Nkx2-1* compared to *AmphiPax2/5/8* and *AmphiNkx2-1*<sup>96,105</sup>. *FoxE1*, a vertebrate thyroid marker, and *AmphiFoxE4*, its amphioxus homolog, are expressed adjacently to the endostyle<sup>142</sup>. Thyroperoxidase, a marker of thyroid activity, is also conserved in amphioxus<sup>143</sup>. Nevertheless, in vertebrates, the thyroid has both NC cells and endodermal contributions, whereas there is no contribution from NC cells in protochordates.

Pituitary: The pituitary is a bilobal organ composed of the adenohypophysis (i.e. the anterior lobe), originating from the adenohypophyseal placode, and the neurohypophysis (posterior pituitary), composed mainly of neurons originating in the hypothalamus<sup>144,145</sup>. The anterior pituitary is of central importance for the hormonal control of the body<sup>146,147</sup>. The neurohypophysis is responsible for the secretion of oxytocin and vasopressin. The pituitary was first thought to be unique to vertebrates but over the last ten years, organs with similar development and function have been described in tunicates and cephalochordates<sup>86,87,148-150</sup>. Evidence of a separate olfactory and adenohypophysal organ in *Oikopleura* and *Ciona*<sup>75,86</sup> considerably weakens the hypothesis of a double-duty organ in the ancestral chordate<sup>148</sup> followed by subfunctionalization in distinct pituitary and olfactory organs in vertebrates. In cephalochordates, the Hatschek's pit is regarded as an organ with pituitary-like functions<sup>148,151</sup> (see below). Both regions have a comparable expression of *Nkx2-1*<sup>96,143</sup>. In vertebrates, *Pitx1* and *Pitx2* homeodomain transcription factors are viewed as pan-

adenohypophyseal markers <sup>152,153</sup>. Their non-vertebrate chordate homologs have comparable expression patterns <sup>148,149,154,155</sup> and also share the initial patterning under BMP4 and FGF8 signaling <sup>156</sup>. In vertebrates, NC cells will contribute to the cells producing adenocorticotropin (ACTH) and melanocyte-stimulating hormones.

Taken together, the data suggests that, even though a fully functional and organized pituitary does not seem to be present in protochordates, some pituitary functions and underlying molecular determinants are clearly accounted for in protochordates.

### Pancreas

The pancreas arises from a dorsal and ventral budding of the foregut endoderm. It is composed of four endocrine cell types and an exocrine tissue <sup>157,158</sup>. Its endocrine role is central to glycemia homeostasis through insulin secretion from the  $\beta$ -cells and glucagon secretion from the  $\alpha$ -cells. Protostomes have no similar structure but protochordates have functionally comparable endocrine cells. *Amphioxus* expresses an insulin-like peptide <sup>97</sup>. Other insulin superfamily members are found in *Ciona* and *amphioxus* and show a synteny (i.e., the physical linear co-localization of loci on a chromosome) similar to human <sup>8,159</sup>. Glucagon is also found in *amphioxus* and tunicates <sup>157</sup>. *Pdx1*, a member of the ParaHox family of homeobox transcription factors <sup>7</sup>, is involved in the early development of the vertebrate pancreas <sup>160</sup>. Interestingly the *Pdx* family is also present in *amphioxus* <sup>8</sup>, but querying the sea urchin and *Ciona* genomes <sup>10,161</sup> and Ensembl for *Pdx1* sequence similarity did not yield any putative homologs (O. Lucas, August , 2008 unpublished observation). Therefore, this suggests that at least part

of the genetic machinery required for pancreas formation was likely present in the common ancestor to chordates and was possibly lost in tunicates.

## Eye

Photoreceptors or “eyes” are widely found in animals as they are a prime gateway for environmental interaction. Repeated use of common GRN and physical constraints generated eight distinct “eye” types<sup>162</sup>. For example, *Pax6* has long been considered a master regulator of eye development in metazoans<sup>163-165</sup>. The focus, here, will be on the chambered eye with a lens, found in cephalopods and vertebrates. Compound eyes like *Drosophila*’s are of ectodermal origin. The vertebrate lens, a true vertebrate innovation, originates in the lens placode<sup>166</sup>. The vertebrate lens is composed of transparent cells filled with crystallins. Many of these proteins are enzymes that were co-opted for a new function in vision<sup>167</sup>.

Transduction of light to generate a nervous message can be accomplished by two morphologically distinct photoreceptors: ciliary and rhabdomeric. Each differs in its strategy to increase the membrane surface area and the signal transduction pathway. Rhabdomeric photoreceptors use multiple folds in their apical surface while ciliary photoreceptors modify the cilium through folding to resemble a stack of disks. Vertebrates rely on ciliary phototransduction while invertebrates mostly use rhabdomeric phototransduction and both are present in protochordates<sup>168</sup>. The transmembrane opsin proteins of the photoreceptor cell are responsible for the transduction of photons. Opsins are more ancient than the eye and, in vertebrates, along with other phototransduction-related proteins, underwent multiple duplications<sup>169</sup>. For example, six opsins are found in

amphioxus<sup>170</sup> and three in *Ciona*<sup>171</sup>. Also of interest, phylogenetic analysis of the *Ciona* *opsin1* and *opsin2* genes clearly groups them with the vertebrate pigments and not with the invertebrate rhabdomeric *opsins*<sup>171</sup>. Arrestin blocks the downstream G-protein coupled receptor signaling pathway linked to the ciliary or rhabdomeric rhodopsins. Phylogenetic analyses suggest that *Ciona*'s *arrestin* is more closely related to the vertebrate *arrestins* than the invertebrate one<sup>172</sup>. This suggests, in protochordates, an intermediate stage between non-chordates metazoan and vertebrates.

In light of those recent developments, Riyahi and Shimeld<sup>166</sup> present a compelling model where the ancestral chordate would have had a placode-like GRN (*Pax*, *Six*, *Eya*) and an eye GRN ( $\beta\gamma$ -*crystallins*). *Ciona* would have kept those networks separate while their combination in vertebrates would be responsible for the formation of the vertebrate eye. The vertebrate innovation could therefore be the co-option of the already present mechanisms for lens development. Future comparative genomics studies may unravel even more molecular intermediary steps.

### Multichambered Heart

The MesH description of heart is “a hollow, muscular organ that maintains the circulation of the blood” (<http://www.ncbi.nlm.nih.gov/mesh>). The heart must be connected with the body circulatory system and generate an adequate blood flow for the organism's size and changing metabolic needs. Non-vertebrates heart tubes lack chambers, septa, valves, and unidirectional flow. What are the components required for a vertebrate heart? Could a few genes account for such a striking addition to the ancestral barely polarized linear tube?

*Drosophila* heart, also called the dorsal vessel, diffuses blood rather than distributing it through a distinctive vascular system. The tubular heart of tunicates is not lined with endocardiac cells while amphioxus has some endothelium-like lining its vessels<sup>173,174</sup>. In protochordates, both poles of the heart are alternatively used as pacemakers<sup>175</sup>. They have been replaced by synchronous contractions in vertebrates where the heart can be seen as a collection of “modules” (structural or functional) governed by specific GRN added to a primitive GRN found in invertebrates (see below). The complexity of the vertebrate heart-specific GRN was likely enhanced by genome duplication<sup>176</sup>.

The vertebrate heart is composed of chambers (atria and ventricles), valves, endothelium, coronary circulation, and a pacemaker-system. Cardiogenic cells from the lateral mesoderm fuse to form a linear tube. The contractile cells form the myocardium, the future thick-walled ventricle. The ventricle is lined with a single layer of endocardium, a non-contractile endothelial cell type. Different cell fates and expression patterns along the linear tube predict the future chambers. The polarized flow requires a polar pacemaker and a conduction system to synchronize the contraction of the incoming and outgoing chambers in a sequential, coordinated fashion. Another characteristic of the vertebrate heart is its asymmetric morphology. The embryonic linear tube first forms a primary myocardium that latter balloons into differentiated chambers<sup>177,178</sup> and loops to form the four chambered heart of reptiles, birds, and mammals. Chamber formation results in complete separation of blood flows through NC-dependent septation (see<sup>178</sup> for review). Therefore, the vertebrate heart is a highly compartmentalized organ with

multiple innovations. These innovations enable efficient, high pressure delivery of blood throughout a large body following an oxygenation step.

The main genes composing the basal heart GRN can be tracked back to *Drosophila*<sup>179</sup> where the mesoderm, under BMP2/4 (DPP in *Drosophila*) endodermal signals, turns on heart-specific genes downstream of NKX2-5 (Tinman in *Drosophila*), NKX2-3 and GATA-4,-5 (Pannier in *Drosophila*)<sup>180</sup>. The core GRN common to *Drosophila* and human is composed of three genes: *Nkx*, *Gata*, and *Hand*<sup>181</sup>. Vertebrates have an increase number of those genes, providing a greater redundancy than *Drosophila*. For example, there are five *Nkx* and three *Gata* genes in vertebrates compared to only one of each in *Drosophila*. Interestingly, the duplication of ventricles in higher vertebrates coincides with the duplication of the *Hand* gene into *Hand1* and *Hand2*<sup>50</sup> and ventricle development is protected against single gene mutations during cell fate determination. *Amphioxus* was shown to express a *Nkx2-5* homolog<sup>182</sup>. In addition to the *Nkx*-*GATA*-*Hand* core, *Ciona* has a *Mesp* gene<sup>183</sup> that does not appear to be involved in heart formation in protochordates<sup>181</sup>. Vertebrates have three *Mesp* paralogs: *Mesp1*, *Mesp2* and *pMesogenin*<sup>184</sup>. *Mesp1* and *Mesp2* are essential for heart formation<sup>184</sup>. Therefore, *Mesp* may have been co-opted for heart development. Also, the appearance in vertebrates of *connexin* genes and gap junctions enabling rapid cell-cell communication needed for synchronous contractions coincide with the increasing complexity of the vertebrate heart<sup>185</sup>. Duplication and co-option played a role in vertebrate heart evolution despite most of the genes being present in protochordates.

In a nutshell, the peristaltic heart found in primitive chordates was refined to generate a directional, high pressured compartmentalized retrograde-free flow. Understanding of the basal heart GRN modules may help us link molecular mechanisms to congenital heart diseases, which represent ~1% of live births <sup>186</sup>. To this end, *Ciona* is becoming a great model to study heart development from a GRN perspective <sup>51,181,187</sup>. *Ciona* has a small fully sequenced genome (~16,000 genes) <sup>161</sup> and does not have as many paralogs as vertebrates. This enables to functionally test fewer genes and fewer redundancy problems are encountered. A major success was established when Davidson <sup>51</sup> linked *Mesp* to the core *Nkx-Gata-Hand*.

#### Circulatory System and Endothelial Cells

The importance of a closed circulatory system was recognized by Aristotle <sup>188</sup> (see below). Unlike vertebrates, invertebrates do not have a true endothelium, a continuous layer of interconnected epithelial cells <sup>189-192</sup>. Of note, an endothelium-like organ exists in a few non-vertebrates and is regarded as a phyla-specific novelty <sup>193</sup>. The advantage of an endothelium-lined closed regulatory system is threefold <sup>189</sup>: (1) cooperation with the immune system (see below); (2) fine regulation of the blood flow through nitric oxide production and signaling <sup>194</sup>; and (3) cell migration far from the visceral core of the circulatory system, especially needed in the case of limb development, as present in more recent vertebrate lineages.

Multiple hypotheses exist to explain the origin and development of the endothelium. A wide range of blood cells are present in invertebrates and are functionally similar to vertebrates (oxygen transport, blood coagulation, and immunological defense).

In vertebrates, mesoderm-derived angioblasts differentiate into endothelial cells, and form blood vessels. Their growth is controlled by mitogenic factors including basic fibroblast growth factor (bFGF) and vascular endothelial growth factor (VEGF). The VEGF and its receptor (VEGF/VEGFR) are clearly present and, at least partially functional in invertebrates <sup>195,196</sup>, even though true endothelial cells are not present <sup>189</sup>. The apparition of endothelial lining correlates with the vertebrate's adaptive immune system, which relies on bloodstream for movement of its effector cells.

Coincidental with the vertebrate endothelial system, a complex blood coagulation system arose in jawed vertebrates with some basal components already present in cyclostomes <sup>197</sup>. This increased complexity could be due to a WGD, which is thought to have played a great role in the blood coagulation network evolution. A recent review pointed to an astonishing number of genes shared by *Drosophila* and mammals in the vascular system <sup>198</sup>. Interestingly, blood and cardiovascular cells of vertebrates and *Drosophila* share the same origin in hemangioblasts that, in the last common ancestor, gave rise to both hemocytes and vascular cells <sup>198</sup>.

One true vertebrate innovation is the endothelium/vascular smooth muscle interface through the tyrosine kinase receptors with immunoglobulin-like and EGF-like domains TIE1 and TEK (formerly TIE2) and angiopoietins, their ligands <sup>199</sup>. Nevertheless, using BLAST analysis we identified a unique putative TIE1/TEK homolog in the sea urchin genome (XP\_797021). This suggests that the endothelial lining may have some deep roots in invertebrates.

### The Adaptive Immune System

Immune cells function primarily to eliminate pathogens, maintain self/non-self recognition, and police symbiotic microbes. These functions are carried through recognition (MHC, TLR...), regulation/integration (Myd88, GATA), and molecular-based responses (Ig, perforins...). Innate immunity is the first line of defense and can be complemented by adaptive immunity, which involves somatic recombination (see below). The field of cellular immunity originated more than a century ago with studies of the sea urchin innate immunity<sup>200</sup>.

The innate immunity has very ancient roots that can be tracked to basal metazoans (Porifera<sup>201</sup> and Cnidaria<sup>202,203</sup>). A good example is the *Drosophila* Toll receptors<sup>204,205</sup>. Toll-like receptors (TLR) are major components of invertebrates innate immune system (see<sup>206</sup> for review). The recent sequencing of the sea urchin genome revealed an astonishing diversity of innate immune system mediators with an order of magnitude (n=222<sup>207</sup>) more TLRs in comparison to humans<sup>207-209</sup>. An excellent review on sea urchin innate immunity and its unsuspected relationship to vertebrate adaptive immunity has been compiled recently<sup>20</sup>.

In sea urchin, phagocytic cells with varying scavenger receptors, receptors that recognize modified low-density lipoproteins, can be generated through combinatorial assembly<sup>210</sup>. This suggests that diversification mechanisms arose before the clonal expansion characteristic of the vertebrates adaptive immune system lymphocytes, effector cells of the adaptive immune system<sup>211</sup>. Long-lived, V(D)J recombination, clonal expansion, and self-renewal are characteristics of T and B lymphocytes in jawed vertebrates. V(D)J recombination is the vertebrate-specific recombination mechanism of

immune gene segments that generates the molecules involved in recognizing pathogens (T cell receptor, TCR and immunoglobulins, Ig (see below)). Even though no adaptive immune system is found in sea urchin, molecules involved in V(D)J rearrangement and Ig domain proteins have been identified<sup>207,212</sup>. Ig are proteins of the immune system that recognize and neutralize pathogens. The adaptive immune system is based on somatic lymphocytes recombination of molecules recognizing non-self. There is clear evidence of Ig-mediated pathogen recognition in invertebrates<sup>213-217</sup> but no evidence of adaptive immune system, which suggests vertebrates co-opted Ig for the adaptive immune system.

The search for an adaptive immune system outside of jawed vertebrates has been biased by the assumption that it would look like V(D)J recombination. Interestingly, looking for function instead of structure, Pancer and colleagues<sup>206,218,219</sup> elegantly described a new adaptive immune system in cyclostomes based on variable lymphocyte receptors (VLR). VLR are composed of leucine rich repeat (LRR) modules, which genes undergo somatic rearrangement. Interestingly, LRR-containing proteins are found in plants, protostomes, and deuterostomes, where they mediate antimicrobial responses<sup>220</sup>. Also, LRRs form the modular domains of the TLR extracellular domains, which generate a concave binding surface reminiscent of the MHC antigen presentation pocket<sup>221</sup>. It has been suggested that the appearance of the vertebrate adaptive immune system relied on the complex innate immune system ancestral to all deuterostomes<sup>222,223</sup>. The parallel between LRR recombination and V(D)J recombination suggests a case of convergent evolution.

The presence of two distinct adaptive immune systems in vertebrates could coincide with colonization of new ecological niches and increased predation. Moreover, the increased complexity and body size of vertebrates may have been facilitated by an adaptive immune system<sup>206</sup>. As Pancer and colleagues suggested<sup>206</sup>, the high reactivity and autoimmunity of LRR may have been unsuitable for a burst of morphological innovations in basal vertebrates and the maintenance of large families of innate receptors may have been disfavored due to a lesser number of endosymbiotic communities in vertebrate animals<sup>21</sup>. Moreover, vertebrates often show multiple paralogs where sea urchin genome encodes for a single transcriptional regulator<sup>224</sup>. This increased regulatory complexity may have helped generate adaptive immunity. In light of the recent discoveries gleaned from inter-phyla comparison, one may speculate that unusual but analogous adaptive immune system could exist in non-vertebrates.

### Morphological Asymmetry

Despite their usual bilateral symmetry deuterostomes are fundamentally asymmetric. The establishment of the first two body axes (anteroposterior, A/P, and dorsoventral, D/V) is followed by a left-right (L/R) axis, clearly visible in vertebrate organs like heart, stomach, and intestine for example. This L/R axis is also a characteristic of bilaterians. In vertebrates, L/R asymmetry is a four step process: (1) breaking of the symmetry, (2) transfer of signal to the lateral plate mesoderm, (3) asymmetric expression of molecules (the *Nodal-Lefty-Pitx2* pathway, see below), and (4) asymmetric organ morphogenesis (for review, see<sup>225</sup>). Asymmetries in protochordates have been found in amphioxus<sup>226</sup> and tunicates<sup>227</sup>.

Nodal is best known for its role during mesoderm and endoderm formation<sup>228,229</sup>, A/P patterning, and D/V patterning but it is also essential to L/R asymmetry through regulation of *Lefty* and *Pitx2*<sup>230</sup>. Recently, Nodal factors have been found in tunicates<sup>227,231,232</sup>, amphioxus<sup>142</sup> and sea urchins<sup>233,234</sup>. The central *Nodal-Lefty-Pitx* pathway is conserved in chordates<sup>227,235</sup>. *Pitx2* is the downstream transcription factor that regulates asymmetric organogenesis through its expression in the embryonic left lateral plate mesoderm<sup>236</sup>. This asymmetrical *Pitx* expression is also found in tunicates<sup>237</sup>. Recently a L/R asymmetry pathway depending on *Lefty* and *Pitx2* homologs was shown in sea urchin<sup>18,233,238</sup>, suggesting that the Nodal-Lefty-Pitx2 pathway might be quite ancient.

While *Nodal*, *Lefty* and *Pitx2* are expressed in the embryonic left lateral plate mesoderm in vertebrates and amphioxus<sup>235</sup>, they occupy the right side of the sea urchin embryo<sup>233</sup>. This suggests the L/R asymmetry was determined by *Pitx* homologs in the common ancestor of echinoderms and chordates<sup>234</sup>. Most protochordates have a single *pitx* gene<sup>85,227,237</sup> while vertebrates have up to three paralogs<sup>239</sup>. Only one *Pitx* seems to be present in cyclostomes (hagfishes and lampreys)<sup>240</sup>. Asymmetric *Pitx* expression is a chordate-wide determinant of L/R asymmetry<sup>19</sup>. In vertebrates, *Sonic hedgehog* (*Sh*) induces asymmetrical *Nodal* expression before organogenesis<sup>241</sup>, while the amphioxus *Sh* homolog seems to have a very limited role in that process<sup>242</sup> and tunicates do not exhibit L/R *Sh* asymmetry<sup>243</sup>. This suggests that, in vertebrates, *Sh* has been co-opted to add a layer of specificity to the asymmetric development. The uncertainty over the D/V axis inversion at the chordate common ancestor could be correlated with the inverse role of L/R in echinoderms and chordates<sup>233,234</sup>. Based on a parsimony hypothesis and a

conserved molecular pathway, asymmetry appears as a deuterostome and not a vertebrate novelty.

### Liver

The vertebrate liver is a central organ for energy metabolism and detoxification<sup>244</sup>. Liver's interplay with adipocytes and myocytes is central for energy homeostasis. Most invertebrates are constant feeders while vertebrates usually ingest large quantities of food at once, which requires special metabolic mechanisms, partly accomplished by the liver.

Searches for an invertebrate liver from a functional point of view reveal some unexpected recent findings. Even though a true liver with hepatocytes and complex GRN governing liver functions<sup>245</sup> is not mentioned in invertebrates, amphioxus has a hepatic diverticulum or hepatic caecum, an asymmetric extension of the digestive tract that has long been considered a primitive liver<sup>246</sup>. Markers of liver function are numerous. SREBP-1c and PPARs are widely recognized as markers of hepatic fatty acid metabolism and homolog's in GenBank are widely found outside of vertebrates (data not shown). Antithrombin (AT) is an acute phase plasma protein used as a liver marker and is required in the coagulation cascade. Recently, an active AT has been identified in the amphioxus diverticulum<sup>247</sup>. Also, a functional fatty-acid binding protein (FABP) has been described as being expressed in the amphioxus diverticulum (*AmphiFABP*)<sup>248</sup>. There are nine FABP in vertebrates that are mostly involved in fatty acid uptake and transport<sup>249</sup>. Even though no clear bile salts or vitamin D has been described in invertebrates, a vitamin D receptor/pregnane X receptor (VDR/PXR), ancestral to the two

distinct VDR and PXR found in vertebrates, has been identified in *Ciona*<sup>250</sup>. This suggests that protochordates possess some liver-like marker genes.

The family of cytochromes P450 (CYPs) is involved in numerous biotransformation of compounds and is mostly expressed in the liver. Four members of the CYP3 subfamily, widely studied for its vast substrate diversity, have recently been reported outside of the vertebrate lineage in two *Ciona* species<sup>251</sup>.

A hepatocyte-like function for the *Drosophila* oenocytes (a population of secretory cells) has been recently described<sup>252</sup>. Notably, more than 20 fat metabolizing genes, specific to human liver, were found to be expressed in the *Drosophila* oenocytes. The authors also functionally characterized the fly *Cyp4g1* and showed that it is essential to the fatty-acid storage composition. Oenocytes accumulate fats during starvation and body fat-oenocyte communication is comparable to the vertebrate adipose tissue-liver fat flux. *Drosophila* oenocytes seem to be analogous to the vertebrate hepatocytes even though they are not involved in glycogen storage. This suggests a more ancestral origin of liver or liver-like functions thought, until recently, to be confined to vertebrates.

### Novel Molecules

Steroid Hormones: Nuclear receptors are part of a family of transcription factors activated by ligand-binding and found in most metazoan<sup>209,253-255</sup>. Steroid receptors (SR) are one of its subfamilies and are widely studied. SR are thought to be responsible for much of the hormonal-dependent regulatory complexity unique to vertebrates<sup>256-258</sup>. They are responsible for a high degree of organ pleiotropy as one gene can have multiple roles and effects. Cross-talk between their receptors and multiple signaling pathways

provides large regulatory potential<sup>259-263</sup>. Six steroid hormone receptors exist in vertebrates: estrogen receptor (ER $\alpha$  and ER $\beta$ ), progesterone receptor (PR), androgen receptor (AR), glucocorticoid receptor (GR), and mineralocorticoid receptor (MR). They are thought to have evolved by gene duplication followed by divergence at the vertebrate transition through ligand exploitation<sup>257,264,265</sup>. Invertebrates have only two steroid hormone receptors: ER, which appears to be the oldest member<sup>59,257,266</sup>, and a generic SR in amphioxus (bfSR)<sup>267</sup>. Therefore, a major increase in complexity seems to have arisen at the invertebrate/vertebrate transition.

The cytochrome P450 oxidases 11 $\beta$ -HSD, 17 $\beta$ -HSD, and 3 $\beta$ -HSD are essential enzymes in steroid metabolism<sup>59</sup>. The binding of estradiol to the ER receptor was suggested to have evolved later, within the vertebrate lineage<sup>266</sup>. Recently, in amphioxus, bfSR, a putative homolog of the vertebrate ER, was shown to be activated by estrogens<sup>267</sup>. Therefore, steroid signaling, which a few years ago seemed to be a true vertebrate innovation, seems to have some functional roots in amphioxus.

Fibrillar Collagen: Collagens are major scaffolding proteins present in all metazoans<sup>268</sup>. The diversity of collagens present in vertebrates (21 types in mammals<sup>268,269</sup>) strikingly contrasts with the unique or few types present in invertebrates<sup>270</sup>. Vertebrates heavily rely on multiple types of collagens for skeletogenesis and structures proper to vertebrates (bone, teeth, cartilage, skin, tendon, ligament, blood vessel, and cornea). The vertebrate molecular innovation has been elegantly tracked back to a small genomic insertion present only in vertebrates<sup>270,271</sup>. This novel domain, along with gene duplications, was selected for different roles and evolved into the complex vertebrate-

specific collagen family. Vertebrates use a vast number of distinct collagens, not found in invertebrates, which, indirectly, enables predation and locomotion <sup>21</sup>.

### Genome Complexity

Gene duplication has long been thought to be a likely event in genome evolution and, indirectly, generation of new traits and, sometimes, new species <sup>272</sup>. Numerous recent studies exhibit data strengthening this 30-year old hypothesis (see below). In a generic scenario, a copy of a duplicated gene can be co-opted for a novel function while the original copy conserves its original function. Below is an overview of the genetic concepts potentially leading to “gene duplication” and genetic complexity. Their specific contribution to the vertebrate sub-phylum is still very much a work in progress.

Most often, duplication leads to nonfunctionalization and gene loss <sup>273</sup>. Nevertheless, duplication also provides a unique opportunity to evolve novel functional roles with the new genetic material, or provides a source for combination of gene products to generate a novel chimeric protein (accretion or shuffling) <sup>274,275</sup>. The theory goes that duplicated genes are initially equivalent and are under identical constraints <sup>276</sup> but they subsequently diverge asymmetrically <sup>277</sup>.

### Evidence of New Vertebrate-Specific Domains and Proteins

New domains, and even more so new proteins, are rare. One of the most ancient, and likely least complex animal genomes, that of the sea anemone <sup>278</sup>, consists of only 15% protein domains that are specific to animals. Also, more than 98% of the genes in chicken are shared with humans, despite 310 million years (Myr) since divergence from a

common ancestor <sup>279</sup>. Recent comparative genomic studies between vertebrates and invertebrates show that, excluding splicing isoforms, only 22% of the genes found in vertebrates are vertebrate-specific <sup>280</sup>. The faster evolutionary rate could be a reason why orthologs are hard to find in invertebrates. Interestingly, most vertebrate-specific genes do not belong to families but are stand alone genes. Therefore, refinement, or shuffling, of pre-existing exons appears to be widely preferred to *de novo* invention of a “new” gene. Domain shuffling is a likely means for protein architecture evolution <sup>281</sup>, but does not explain the appearance of novel genes. Four distinct mechanisms can be seen as generator of truly new protein sequence: (1) duplication and divergence; (2) exonization of intronic sequence <sup>282,283</sup>; (3) alternate splicing <sup>284</sup>; and (4) transposition events.

Role of Duplication: Duplication of genetic material can be global (i.e., whole genome) or local, also called segmental. In both cases, the novel copy can acquire a novel role <sup>272</sup>. It was shown more recently that duplicated genes can undergo a three-step process: duplication, divergence, and complementation (DDC model), where the functions become carried by the ancestral and the novel gene <sup>285</sup>. Domain duplication is common in eukaryotes, where ~90% of the domains have been duplicated and it is therefore suggested that domains rather than proteins should be considered as the “evolutionary unit” <sup>286</sup>. An increased number of proteins have been correlated with complexity <sup>26,287-289</sup>. Comparing the total gene number of different organisms, one can conclude that expansion of protein families can also be correlated to adaptations to a specific environment and not only complexity <sup>290,291</sup>. These two kinds of protein expansion are called “progressive” (increase complexity) and conservative” (adaptive

complexity)<sup>292</sup>. Interestingly, the authors show that the domain superfamily denomination that correlates best with increased complexity falls under the generic term “regulation”, meaning domains with signal transduction or DNA binding signatures<sup>291</sup>. This indicates an important role for duplication in the increase of regulatory processes observed in vertebrates (see above).

Regardless of the exact mechanism, more large-scale duplications are present in human than protochordates<sup>289,293</sup>. In humans, many loci exhibit evidence of four paralogs, like the *Hox* and *ParaHox* gene clusters<sup>294</sup>, where evidence of a single cluster in protochordate is overwhelming<sup>7,295</sup>. Moreover, in *Ciona*, most gene families do not exhibit more than a single vertebrate ortholog<sup>296</sup>. Analysis of genomic complexity using *Ciona* can be treacherous as *Ciona* is widely recognized as being largely a “regressed” form of the chordate ancestor<sup>297</sup>. Therefore, large-scale duplications likely coincide with the invertebrate/vertebrate transition.

Segmental duplications represent at least 5% of the human genome<sup>289</sup>. Local chromosomal regions can duplicate on the same chromosome or on non-homologous chromosomes<sup>12</sup>. Secondary partial duplications can consequently occur. Truncation of functional protein coding sequence is common through such genetic events but transcription is often retained and can fuse multiple segments in a new transcript<sup>12</sup>. This mechanism potentially generates a novel gene product or may combine different parts from different loci. The resulting combination of existing domains into a novel genetic entity is termed shuffling. Exon shuffling is observed in 20% of eukaryotic exons<sup>298</sup>.

The addition of a new domain to an existing entity can be termed domain accretion. Vertebrates have abundant multidomain proteins<sup>294,299</sup>. Multiple examples of segmental duplications have been reported in humans and are a source of multidomain proteins<sup>300-302</sup>. In eukaryotes, domain accretion is most prominent in proteins involved in transcriptional regulation and chromatin remodeling<sup>303</sup>. In addition, mutation of the novel sequence may result in lack of function, new function, increased polymorphism, fusion of transcripts, (accretion), dynamic mutation, and alteration of gene dosage. Small-scale duplication is a continuous process, not limited to a few occurrences during eukaryotic evolution<sup>224,304</sup> and, therefore, greatly increases the dynamic state of the genome. Segmental duplications may have had a role in vertebrate evolution while their direct correlation to a specific transition might be harder to pinpoint.

Exon shuffling or accretion may result in multidomain proteins thus allowing the coexistence of independent functions in the same physical protein and, therefore, making good bridges between independent GRN through a possible increased propensity for protein-protein interactions (PPI)<sup>305</sup>.

Exonization: Exonization is the process of creating a new exon from non-exonic sequence<sup>283,306</sup>. Thousands of exons are found only in vertebrates<sup>283</sup>. Different mechanisms can lead to exonization<sup>282,283</sup>. Despite the theoretical potential of exonization for the vertebrate radiation, most evidence confine exonization to lineage-specific functions<sup>307</sup> where, as an example, 62% of the novel human exons are restricted to the primate lineage.

Alternative and Atypical Splicing: Human genome-wide analysis revealed that 60-74% of multi-exonic genes are alternatively spliced, which suggests a major functional role for alternative splicing<sup>308,309</sup>. Alternative splicing can be intragenic or intergenic<sup>284</sup>. Intragenic splicing is confined to a single gene and alternate usage of exons through skipping, inclusion, alternate use of promoter, alternate polyA signal, and alternate splice donor or splice acceptor site can generate dozen of functional distinct transcripts from the same genetic locus. In intergenic splicing, the resulting transcript is formed of exons from both genes (reviewed in<sup>284</sup>). An atypical splice form can then replace or complement a predominant form. This is a likely mechanism to rapidly generate genetic diversity, but has not, to our knowledge, been formally involved in the invertebrate/vertebrate transition.

Transposition Events: Transposons or transposable elements are sequences that can move and replicate within the genome. Retrotransposons use an intermediary RNA step before integrating into the genome. The influence of transposons in host evolution has been thoroughly documented<sup>310-314</sup>. It is thought that 25% of human promoters contain transposons-derived sequences<sup>315</sup> and such abundance is postulated to provide raw sequences for CRE evolution<sup>316,317</sup>. In addition, numerous DNA binding proteins and transcription factors are transposons-derived (reviewed in<sup>313</sup>).

#### Influence of Duplication on PPI Network

The effect of duplication must also be considered in terms of connectivity. Highly connected genes/proteins are more likely to increase complexity, in terms of network connection and, therefore, increase the range of network output, than genes/proteins with

few interactions. Moreover, proteins that interface with other modules usually have greater sequence plasticity and often evolve faster than proteins confined to a single module<sup>318,319</sup>. Duplication of a whole network results in a mirror image with, at first, conservation of the previous PPI connections. The nodes are doubled and the connections quadrupled. Subsequent asymmetric divergence and/or inclusion of novel interacting partners results in a new PPI network,<sup>320,321</sup>. In yeast, duplicated PPI increase the likelihood of sharing an interacting partner by 1,000 fold compared to randomly picked protein pairs<sup>320</sup>. This over-simplification does not take into consideration potential indirect interactions, involvement in complexes, homo-oligomerization, and novel binding sites through localized duplication-divergence<sup>321</sup>. Therefore, the effect of genetic duplication at the protein network level can appear fast and may be tremendous.

#### CRE Duplication and Subfunctionalization

A mechanism has been described where partitioning of the duplicated genes leads to subfunctionalization via independent loss or evolution of CRE in the two paralogs. The teleost-specific WGD generated multiple examples of subfunctionalization<sup>285,322,323</sup>. For example, only one *Pax6* gene is present in all vertebrates, except teleosts where two copies are found. *Pax6a* and *Pax6b*, in zebrafish, underwent differential alteration of their CRE resulting in tissue-specific expression of one paralog, in addition to some partial redundancy<sup>322</sup>. In addition to this typical subfunctionalization that comprised both CRE and the linked gene, one can postulate that duplication of the CRE, without the linked coding sequence, could occur and be moved to another genomic region, where it might modulate the expression of an unrelated gene.

### Copy Number Variation

Copy number variants refer to different number of genetic modules in the genome leading to more than the diploid two copies of a gene. Deep analysis of individual human transcriptomes revealed that 12% of the human genome has copy number variants<sup>324</sup>. Nevertheless, only a fraction of the total copy number variants are confined to the germ line and therefore heritable<sup>325</sup>. This unsuspected inter-individual plasticity has a great potential to explain inter-individual penetrance of complex traits and individual disease susceptibility<sup>326</sup>. Even though the role of such individual-based mechanism is undocumented in terms of effect in speciation, its results could theoretically be selected and influence phenotypical traits at a population level, therefore potentially affecting speciation.

Combination of domains through the above-mentioned mechanisms may result in multidomain proteins, which allow the coexistence of independent functions in the same physical protein and therefore make good bridges between independent GRN through an increased propensity for PPI<sup>305</sup>.

### MicroRNA

MicroRNAs (miRNAs) have been linked to the increase of complexity seen in vertebrates<sup>327,328</sup>. Vertebrates exhibit at least 41 miRNA families more than *Ciona* or amphioxus<sup>329,330</sup>. Two WGD alone likely cannot account for such increase<sup>330</sup> and, as miRNAs are continuously added to the genome, the authors suggested a major role for miRNA in the increased complexity of vertebrate's GRN. Also, the duplications of the *Hox* cluster and its associated miRNA that specifically targets *Hox* genes has been

suggested to play a role in the development of vertebrate structural novelties, especially limbs<sup>331,332</sup>. Therefore, it is possible that miRNA played a role in the appearance of vertebrates.

None of the previously listed mechanisms are exclusive and many could be combinatory. The main factors seem to be alternative splicing, novel domain combinations, increased potential of protein-protein interactions, and increased regulational possibilities in GRN. The influence of regulation by ncRNA is possibly vast but hard to quantify to date.

### Conclusion

Here we reviewed the recent molecular evidence characterizing the canonical vertebrate innovations and other often overlooked characteristics. Recent advances in the study of protochordates diminish the idea that vertebrates have unique body plan with many novel structures. Most of what was considered “vertebrate-specific” a decade ago now shows roots in other chordates. Protochordates seem to have many orthologous genes involved in the regulation of “vertebrate-specific” developmental processes. Therefore, a blind dichotomy between vertebrates and protochordates might appear somehow artificial. Even though the basic developmental plan of vertebrates and protochordates are comparable, the versatility of the NC cells and placodes significantly increase the potential for new structures, mostly involved in processes at the animal-environment interface. It seems that modifications of the pharynx through co-option of new genes mostly were a great innovation along with shuffling of protein domains. In light of the extant species studied, molecular evidence strongly supports an

invertebrate/vertebrate transition through functional intermediates where sequence, structure, and function evolved as a whole, as shown for the steroid hormone pathway<sup>264,333,334</sup>.

Combining duplications, domain shuffling, interaction, and differential regulation can generate a staggering complexity without much novel genetic material. A link between gene number and complexity has long been thought as intuitive but does not seem to exist<sup>38,122,179</sup>. In contrast, there is a good correlation between the total number of domains present in a genome and overall complexity<sup>335</sup>. Therefore, the increase of diversity could be thought more in terms of domains and their combinations and interactions than the total number of present proteins.

The robustness of a system is its ability to withstand mutations<sup>336</sup>. Duplication and redundancy are a mechanism to ensure robustness while a highly distributed architecture, through independent modules, is another. Robustness facilitates evolutionary innovations, like the vertebrate heart, where the core GRN has paralogs and potential redundancies<sup>50</sup>. In vertebrates, the greater proportion of gene duplications involved in regulatory processes could be part of the correlated increased morphological complexity<sup>14,280</sup>. In addition, duplication of CRE and their subfunctionalization must be taken into consideration.

Evolution of the genome only accounts for generation of variability. A filter of natural selection must then be applied for an innovation to be retained. For example, teleosts underwent a WGD that did not result in a doubling of “complexity”, even though their plasticity to adapt and specialize to many environmental niche is undeniable<sup>337</sup>.

The living world is a continuum of organisms, some alive and most extinct. Representation of life-forms as discrete sets of entities (e.g., species) defined by binary characteristics is central to Biology. Nevertheless, the underlying molecular processes and their role in speciation and transition can be harder to compartmentalize.

## INTRODUCTION

### Evolutionary Perspectives on Transcriptional Regulation

It has been hypothesized that protein sequence divergence alone may be an insufficient determinant of speciation<sup>338</sup>. For example, protein sequence identity is >98% between man and chimp in spite of many highly different traits<sup>339</sup>. The recent availability of numerous genome sequences and their comparisons support the idea that phenotypic complexity may be more dependent on differences in regulation of gene expression than variation in protein coding sequences.

Gene expression and regulation involve complex mechanisms. They are influenced by multiple events and factors, but generally involve the interaction of regulatory sequences (enhancer, promoter) and transcription factors binding those sequences. Changes in regulatory sequences, or *cis*-regulatory elements (CRE), can have a major impact on evolution of gene expression and may lead to speciation<sup>336,340-344</sup>. Transcription factors are proteins that function through DNA binding and protein-protein interactions aimed at regulating gene expression. Evolution of transcription factors and their binding abilities will be influenced by these interactions. Gene expression is also regulated by additional mechanisms like non-coding RNA, regulatory elements in the untranslated regions, alternative splicing, and epigenetic gene regulation.

At the heart of the transcriptional machinery lays the TATA-Binding Protein (TBP) (Fig. 2.1). TBP is one of the few transcriptional components of multimeric protein complex traditionally required for all three RNA polymerases<sup>345-347</sup>. RNA polymerase I

is responsible for the transcription of ribosomal RNA (rRNA) while RNA polymerase III synthesizes transfer RNA (tRNA). Messenger RNA (mRNA) is transcribed by pol II by integrating signals relayed by the general transcription factors (GTF)<sup>348-350</sup> (Fig. 2.1). The present study focuses on gene expression and only RNA polymerase II (pol II) is discussed hereafter.

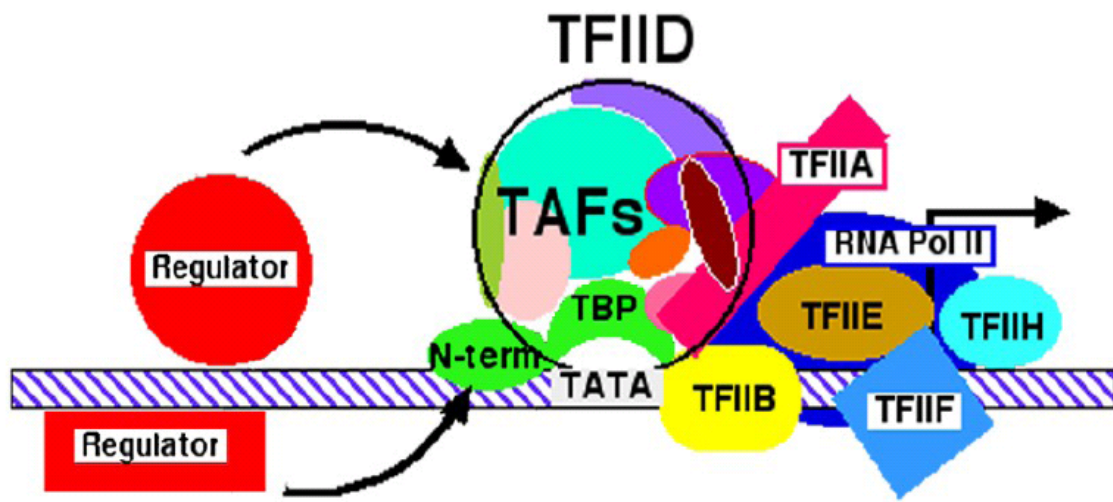


Figure 2.1: Basal transcription machinery (from<sup>351</sup>). Depiction of a model of vertebrate basal transcription machinery, on a mammalian promoter, where signals from regulators (in red) are integrated. TBP, in green, binds the TATA box recognition motif on the DNA promoter. This induces a recruitment of the basal transcription machinery (see text) and RNA polymerase II in blue. The black arrow represents the start of the coding sequence of the gene exemplified here. Mammals harbor many embellishments thought to provide novel signal integration surfaces active in gene regulation. We hypothesize that each of these embellishments co-evolved with specific regulators.

TBP is present in all eukaryotes and archaeabacteria. Like many eukaryotic proteins, and especially transcription factors, TBP is modular, i.e. it is composed of functionally independent motifs. Modular proteins have usually two or more domains separated by linker regions rich in polar, uncharged, and small amino acids (a.a.) (Ala, Asn, Gln, Gly, Pro, Thr, Ser)<sup>352</sup>. A protein domain is defined from experimental

evidence showing that a specific functional region of the protein can fold independently. Modularity of transcription factors in general, and TBP in particular, could be involved in tissue- or developmental stage-specific pathways contributing to morphological changes and possibly speciation<sup>336,353-356</sup>. TBP has a 181 a.a. C-terminal core domain (TBP<sub>CORE</sub>) that is almost identical across phyla<sup>345,357</sup> and, in many species, an amino- (N-) terminal phyla-specific domain of variable size (18 a.a. in *Arabidopsis thaliana* to 172 a.a. in *Drosophila melanogaster*) (see chapter 3, Fig. 3.1B).

TBP recruits the pre-initiation complex (PIC) via protein-protein interactions<sup>350,358</sup>. TBP, in complex with TBP associated factors (TAFs), forms the TFIID complex that recruits TFIIA, thus stabilizing the binding of TBP to DNA<sup>359</sup>. The addition of TFIIB promotes transcription through the recruitment of the other GTFs and pol II, which is sufficient for basal transcription *in vitro*<sup>350</sup>. This assembly is conducted through the TBP<sub>CORE</sub> domain binding to the promoter, often at a “TATA box”<sup>350,357,358</sup>, a common promoter motif<sup>360</sup>. TBP can also be recruited to promoters lacking a TATA box<sup>361,362</sup>. The composition of the core promoter and its ability to recruit a diverse panel of factors helps determine how the signals will be integrated by the PIC, then leading to transcription initiation. An extra layer of complexity in the integration of the signals is channeled through gene specific co-factors and DNA-binding activators and repressors<sup>350,363-367</sup>. Ultimately, all of the protein-protein and protein-DNA interactions result in a finely tuned temporal and spatial pattern of gene expression within the cell. In yeast, TBP is essential for all transcription<sup>345,368-370</sup>. In multicellular organisms TBP paralogs are present, all of which share the conserved TBP<sub>CORE</sub> domain and may also function to

recruit the transcriptional machinery in some situations<sup>350,371</sup>. These could potentially have arisen to increase the complexity of the transcription machinery beyond that found in unicellular organisms. The structured and highly conserved TBP<sub>CORE</sub> domain is therefore central to basal transcription in eukaryotes.

### Intrinsic Protein Disorder and the TBP N terminus

Highly structured and widely conserved core domains characterize most transcription factors<sup>372</sup>. Nevertheless, the additional presence of unstructured domains is fairly common and may be crucial for transcriptional and translational regulation, as well as for signal transduction<sup>373-378</sup>.

The occurrence of intrinsically disordered domains appears to be positively correlated with organism complexity<sup>373,379</sup>. Disordered domains are characterized by low sequence complexity, low hydrophobic a.a. content (Val, Leu, Ile, Met, Phe, Trp and Tyr), and an enriched proportion of polar and charged a.a. (Gln, Ser, Pro, Glu, Lys and Gly, Ala)<sup>380,381</sup>. Therefore, in some cases, it might be useful to picture a transcription factor as a conserved structured core associated with a flexible module poised for diverse functional roles. This is apparently the case with p53<sup>352,382</sup> and TBP, as discussed below.

The highly structured TBP<sub>CORE</sub> domain<sup>383-387</sup> is fused to a phylum-specific N-terminal region that contains very few charged residues<sup>388,389</sup>. In the specific case of TBP-N, it will be referred to as a region based on primary amino-acid sequence inspection, not from experimental evidence indicating an independently folded domain. In vertebrates, the TBP-N evolved through sequence duplication<sup>282,389</sup>. TBP-N is composed of a vertebrate-specific domain (VSD; a.a. 1-118 in mouse) and a metazoan

specific domain (MSD; a.a. 119-130 in mouse). The MSD is composed of 3-7 iterations of the imperfect PXT repeat, where P stands for proline, T for threonine, and X any other amino-acid (see chapter 3; Fig. 3.1A, 3.1B). *Drosophila* has some partial similarity to the chordate PXT region<sup>388</sup> (see chapter 3; Fig. 3.1B). Interestingly, TRF3, one of the TBP paralogs, which is vertebrate-specific, has 2 PXT repeats in its N-terminal region while the other TBP paralog, the pan-metazoan TRF2, has no PXT repeat (O. Lucas, unpublished observation). The differences in N-terminal sequences for TBP and TRF3 have recently been suggested as potentially being involved in recruiting different transcriptional regulatory complexes<sup>390</sup>. The significance of this N-terminal architecture has been revisited recently<sup>391</sup>. This short low complexity (i.e., composed of few different a.a.) region has been considered, so far, as a spacer region or perhaps a hybrid between a short linear motif (SLiMS)<sup>336,392</sup> and a simple sequence repeat (SSR). SSRs are microsatellite-like a.a. repeats and SLiMS are short (3-10 a.a.) sequence patterns known to carry out pivotal functions in protein-protein interaction. They are usually present in structurally undefined regions, suggesting a functional need-driven flexibility<sup>392</sup>. Linear motifs are generally associated with transient, low affinity (low  $\mu\text{M}$  range) protein-protein interactions, like those found in transcriptional regulation<sup>393,394</sup>. A particular motif is rarely conserved across phyla, suggesting that phylum-specific specialization enhances rapid adaptation<sup>392</sup>. SSRs are particularly abundant in regulatory proteins and often have a high rate of indel (insertion, deletion) evolution<sup>395-399</sup> as observed in the TBP PXT repeats<sup>282,391</sup>. In chordates (see chapter 1; Fig. 1.1), a short repeat-based motif bridges the TBP<sub>CORE</sub> to a specific N-terminal region<sup>391</sup>. Each chordate subphyla:

cephalochordates (amphioxus), tunicates (sea squirt, *Ciona sp.*), and vertebrates, has its own non-homologous N-terminal region <sup>391</sup>. Moreover, the amino acid composition of the amphioxus TBP N terminus is quite different (greater proportion of aromatic and basic a.a.) from the vertebrate one and does not seem to be derived from a common ancestor <sup>391</sup>. The structure of TBP-N has not been determined and is thought to be intrinsically disordered <sup>400</sup> (Fig. 2.2). Some intrinsically disordered proteins are known to carry out important biological functions like regulation, signaling, and transcriptional control <sup>401</sup>. Indeed, many non-enzymatic functions of proteins rely on unstructured domains <sup>373,375,402,403</sup>.

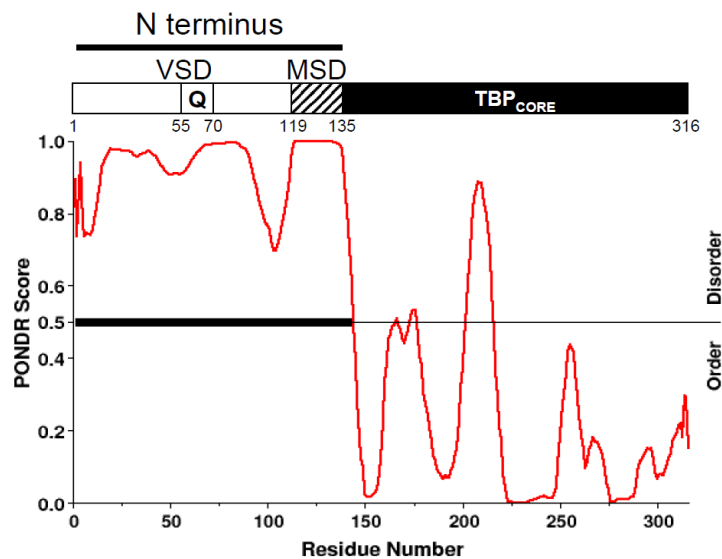


Figure 2.2: TBP folding prediction. Prediction of order (scores <0.5) and disorder (scores >0.5) were determined using the POND VL-XT algorithm <sup>380</sup>. Above are represented the different domains of TBP with the residues number. VSD: vertebrate-specific domain; Q: polyglutamine repeat <sup>404</sup> (see chapter 3; Fig. 3.2B); MSD: metazoan-specific domain. Note the contrast between the TBP<sub>CORE</sub> predicted here, and known <sup>383</sup> to be highly structured, and the predicted highly disordered TBP-N. The predicted disorder is similar across most of the VSD and maximal for the MSD.

One of the current theory is that intrinsic disorder enables folding to be induced through protein-protein interaction <sup>401</sup>. Recent studies suggest that this theory may be

especially relevant to transcription factors binding their targets and is therefore related to the presented study <sup>405</sup>. The induced structures could then be solved <sup>406-409</sup>. For example, more than half the sequence of human transcriptional activator cyclic-AMP-response-element-binding-protein (CBP) is disordered <sup>352</sup> and is nevertheless a major scaffold for protein complex assembly. Therefore, it is thought that the regulatory function can greatly depend on the dynamic flexibility of the protein. An unstable structured region can add flexibility and adaptability to the proteins being bound. Additionally, the distinct bound states can have different functions <sup>378,410,411</sup>. Intrinsically disordered regions can be highly involved in protein-protein interaction <sup>405</sup>. For example, in the tumor suppressor p53, 71% of the known protein-protein interactions involve the disordered N and C termini <sup>382</sup>. This suggests a general role for disordered regions in protein-protein interactions <sup>382,412-415</sup> in general and transcription in particular <sup>405</sup>.

Adaptation to a new environment could favor a highly evolvable protein. However, adaptive mutations leading to pleiotropic effects can be deleterious and be suppressed in the following generations through compensatory mutations of the interacting partner. Intrinsically unstructured domains represent potentially viable evolutionary alternatives to structured domains through the functional advantages (flexibility and evolvability) they offer. Therefore, a limited repertoire of primary sequences could be used for multiple conformational states, facilitating the evolution of new domains and increasing the functions of existing proteins <sup>352,377,416,417</sup>.

### Function of the TBP N Terminus

The low evolutionary rate of the TBP-N sequence suggests some environmental constraints may be responsible for conserving the protein sequence of that region of TBP<sup>391,418</sup>. This correlates well for the TBP<sub>CORE</sub> and its universal function in transcription initiation. TBP<sub>CORE</sub> can be assimilated to a protein “hub” for its low evolutionary rate<sup>391</sup> and its ability to scaffold the PIC through multiple simultaneous protein-protein interaction. TBP<sub>CORE</sub> interacts with more than 100 distinct proteins (www.hprd.org query).

Protein interaction candidates and roles for the TBP-N are more elusive, even though its function has been sought for over fifteen years<sup>419,420</sup>. Multiple biochemical studies have demonstrated different putative roles for the TBP-N. TBP-N was first postulated to be involved in preventing the TBP<sub>CORE</sub> from binding the TATA box and possibly enhancing DNA bending<sup>419-421</sup>. It has also been proposed that TBP-N may regulate pol II and pol III dependent transcription<sup>422</sup>, that it participates in Sp1 coactivator-dependent transcription<sup>388,423</sup>, or that it functions in TATA-less and TATA-containing pol III transcription<sup>424</sup>. TBP-N was also shown to disrupt TBP dimerization<sup>425</sup> and is potentially required for transcriptional repression by NC2<sup>426,427</sup>. NC2 is a TBP-associated transcriptional repressor<sup>426,427</sup>. Moreover, *in vitro*, the removal of TBP-N increases interaction of TBP<sub>CORE</sub> with TBP associated factors<sup>428</sup> and, in RNA polymerase III-dependent transcription, its removal enables the small nuclear RNA activating protein (SNAP) to complex with TBP on the U6 promoter to enhance U6 transcription<sup>429</sup>. The TBP-N contains a polyglutamine repeat, the expansion of which is

associated with neurodegenerative disorders <sup>430-433</sup>. A high mobility group protein (HMG-1) has been shown to bind the TBP-N polyglutamine tract increasing the affinity of TBP for the TATA box by ~20 fold <sup>434</sup>. The polyglutamine repeat has also been shown to influence the intensity of huntingtin-associated protein 1- (HAP1) dependent cytoplasmic TBP sequestration <sup>435</sup> and is associated with spinocerebellar ataxia 17 (SCA17) pathogenesis by influencing TBP dimerization and its interaction with TFIIB <sup>404</sup>. In spite of many proposed roles, mostly through biochemical studies, the function of the N terminus *in vivo* is very much uncertain. Nevertheless, more recent genetic studies have been able to exclude some of these proposed functions (see below).

#### Knowledge From the *tbp*<sup>ΔN/ΔN</sup> Mice

Removal of TBP-N in mouse (*tbp*<sup>ΔN/ΔN</sup>) led to ~89% lethality at midgestation due to a placental defect <sup>351</sup>. Mouse embryonic fibroblasts derived from *tbp*<sup>ΔN/ΔN</sup> were not affected in basal cellular function, suggesting that the absence of TBP-N did not affect the general cellular pathways <sup>357</sup>. This suggested that the TBP-N functions to regulate some temporally or spatially specific genes involved in placental function. *tbp*<sup>ΔN/ΔN</sup> mice surviving after weaning were found to be fertile <sup>351</sup>. Interestingly, breeding of *tbp*<sup>ΔN/ΔN</sup> survivors did not increase the occurrence of *tbp*<sup>ΔN/ΔN</sup> pups, suggesting that a mechanism other than heredity was involved in survival.

Numerous independent occurrences of placenta-like organs exist in different non-mammalian vertebrate lineages, illustrating the benefit of such an organ for internal development <sup>436,437</sup>. For example, some sharks exhibit an analogous yolk sac-based placenta, a result of convergent evolution <sup>438</sup>. The chorioallantoic, or “true” placenta, is a

mammal-specific organ involved in nurturing the developing fetus inside the mother. It is a transient organ derived from embryonic and maternal tissues. The fusion of the fetal chorion, an extra-embryonic membrane, and the fetal allantois, a mesoderm-derived structure, with the uterine epithelium forms the true placenta. The placenta is involved in transport of oxygen, carbon dioxide, nutrients, and wastes; it fosters the maternal tolerance to non-self and is a potent endocrine organ itself<sup>439,440</sup>. Defects in placental development can induce intra-uterine growth retardation of the fetus or even death<sup>441,442</sup>.

TBP-N is conserved in all vertebrates, but most vertebrates do not exhibit a placenta. Therefore it is possible that TBP-N is involved in a more basal function common to all vertebrates. The above-mentioned potential functions for the TBP-N are not satisfactory to explain the phylogenetic distributions of distinct TBP N termini. It has long been suggested that TBP-N could be involved (and co-evolved) with specific transcriptional regulators responsible for species-specific functions<sup>388,390,391,422,423,443,444</sup>.

To test this, two approaches were taken. In chapter 3, we designed an unbiased screening process, using yeast two-hybrid, to find proteins that interact with the hagfish TBP-N. We identified several candidates for such a role and showed that mouse paralogs of one of these proteins also interacted with the mouse TBP-N. Moreover, the mouse TBP-N modulated the activity of one of these paralogs, *in vivo*. In chapter 4, we described and characterized the survival of a line of mice where the endogenous TBP-N was replaced by the homologous, yet diverged, hagfish TBP-N (*tbp<sup>hfN</sup>*). In chapter 5, we further characterized the *tbp<sup>ΔN</sup>* and *tbp<sup>hfN</sup>* alleles through transcriptome and pathway analyses, lipid metabolism analyses, and measurements of physiological parameters. The

results presented herein suggest that the TBP-N may have multiple roles in vertebrate-specific gene regulation, notably during development and in maintaining lipid homeostasis.

## CO-EVOLUTION OF PITX TRANSCRIPTION FACTORS WITH TBP IN METAZOANS

### Introduction

Comparative analyses of genomes from closely related species generally show few differences in protein-coding sequences<sup>222,445</sup>. Instead, most species-specific differences in genomes lie within non-coding sequences<sup>446</sup>, including sequences that participate in regulating genes (e.g., “*cis*-regulatory elements”)<sup>447-450</sup>. The implication, as proposed decades ago<sup>338</sup>, is that speciation and evolution are likely driven more by quantitative changes in the expression of genes than by qualitative changes in the protein products of genes<sup>447,448,451-453</sup>. Considerable recent attention has been given to the role of changes in *cis*-regulatory elements in mediating speciation<sup>343,450,453,454</sup>.

Noteworthy examples of protein-coding changes correlating with evolutionary transitions have been documented in some transcription regulatory proteins. The genes encoding many transcription factors have undergone duplication and divergence to generate paralogous gene families during evolution<sup>47,48,239,455,456</sup>. In domestic dogs, subtle breed-specific differences in repetitive protein-coding sequences within transcription factors have been suggested to cause morphological and perhaps behavioral differences between breeds<sup>398,457</sup>. In another example, TBP, a component of the basal transcription machinery, has a large region that underwent dramatic changes at certain phylogenetic transitions (Fig. 3.1)<sup>282,391,455</sup>. Interestingly, in all of these examples, it is expected that the result of these evolutionary changes in protein-coding capacity, much like evolutionary

changes in *cis*-regulatory elements, will be to quantitatively affect expression of other genes.

The core of the basal transcription machinery has changed little during evolution of eukaryotes<sup>348,350,358,359</sup>. Perhaps most dramatic among the components is the 181-amino acid TBP<sub>CORE</sub> domain, whose primary sequence and crystal structure are almost identical among all eukaryotes and archaeabacteria<sup>345,383,458,459</sup>. However, upstream of the TBP<sub>CORE</sub>, metazoans have a more variable region that can constitute over half of the entire protein. The amino- (N-) terminal portion of this region generally differs between major metazoan clades, with most distinctions being between subphyla. Thus, within the chordate phyla, all vertebrates share a homologous region that is unrelated to those in either tunicates or cephalochordates (Fig. 3.1B). Within insects, all dipterans and hymenopterans share a related N terminus, which differs both from the chordate N termini and from that in lepidopterans (Fig. 3.1A; Table 3.5)<sup>391</sup>. Thus, this region of the protein appears to have been “switched-out” at various major evolutionary transitions, yet to have been subjected to strong purifying selection following these transitions<sup>391</sup>.

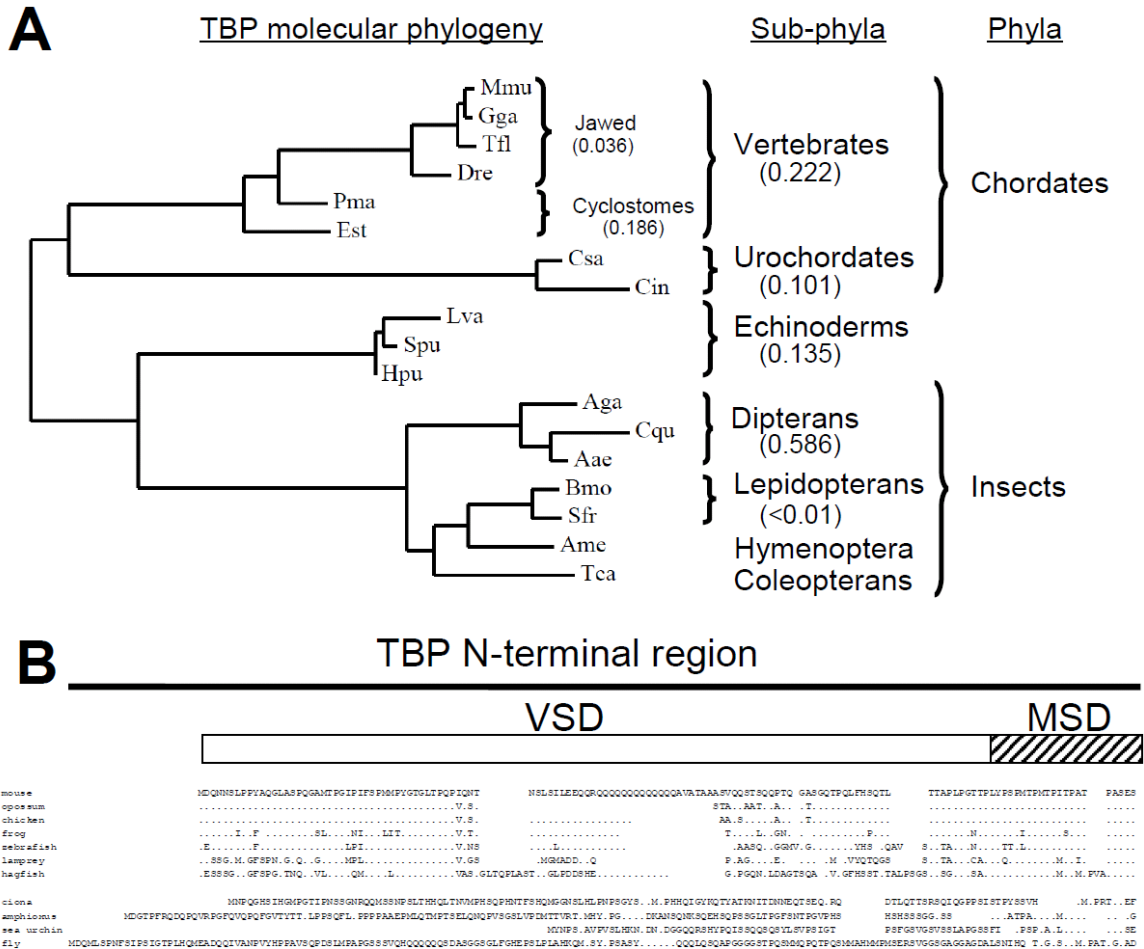


Figure 3.1: TBP metazoan phylogeny. **(A)** Maximum likelihood tree obtained after MUSCLE sequence alignment of full length TBP protein sequences<sup>460</sup>. Species are indicated before gene name (see table 3.5 for accession numbers and full species names). Numbers between brackets are indicative of the ratios of non-synonymous/synonymous mutations (dN/dS), determined using START2<sup>461</sup>. **(B)** Amino-acid (a.a.) sequences of TBP proteins from bilaterian model species. Domains of the vertebrate N terminus are highlighted. VSD: vertebrate-specific domain; MSD: metazoan-specific domain. “.” Represents a.a. identical to mouse, a white space represents a gap.

Between this subphylum-specific region and the TBP<sub>CORE</sub> domain lies a semi-repetitive motif that is common to all metazoans, but not found in archaea, plants, fungi, or protozoans, designated the metazoan-specific domain (MSD; Fig. 3.1B). This region is typically about 15- to 25 amino acids long, including three to seven repeats of a core PXT sequence, where X is typically A, M, I, or S<sup>391</sup>. Based on the ubiquity of this domain, it

has been suggested that MSD is simply a connector that fuses the various subphylum-specific N-termini of TBP to the TBP<sub>CORE</sub><sup>282,391</sup>. Here we show that the TBP MSD has a conserved function that goes beyond serving as a connector between two functional domains. The MSD participates in the interaction between TBP and Pitx family transcription factors.

Pitx transcription factors are found in most or all metazoans, where they typically participate in processes such as patterning and morphogenesis, or regulation of endocrine function<sup>8,149,239,462,463</sup>. Whereas most non-chordate metazoans have a single *pitx* gene, scattered phyla have had independent ancestral gene duplications and divergence, leading to the presence of multiple distinct Pitx proteins (Fig. 3.2). In lamprey, a basal vertebrate, a single *pitx* gene, *pitxa*, has been identified<sup>240</sup>. Higher vertebrates have three *pitx* genes, *pitx1*, *pitx2*, and *pitx3*, with distinct functions. In mammals, *pitx1* participates in pituitary development, pituitary function, and hind-limb morphogenesis<sup>464-468</sup>; *pitx2* is involved in left-right asymmetry, pituitary, cardiac, and craniofacial development, gonad morphogenesis, and cardiac endocrine functions<sup>236,469-473</sup>; *pitx3* has been implicated in eye and midbrain development<sup>474-476</sup>. Added complexity exists in that the *pitx2* gene uses alternative promoters and alternative splicing to issue several N-terminally distinct protein isoforms, including Pitx2a, Pitx2b, and Pitx2c, each of which appears to have a different tissue distribution and pathway specificity<sup>477</sup>.

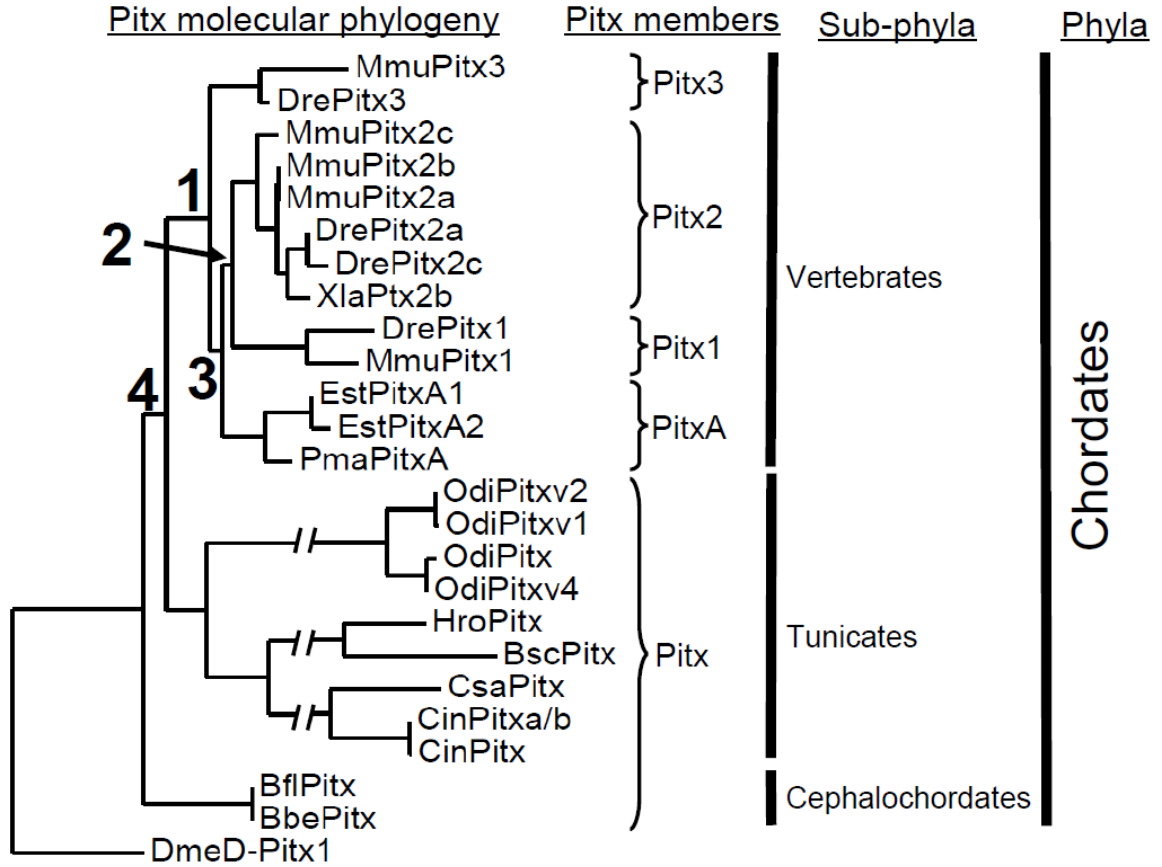


Figure 3.2: Pitx phylogeny. Maximum likelihood tree obtained after MUSCLE sequence alignment of full length protein sequences<sup>460</sup>. Similar topologies were obtained with neighbor joining (BioNJ) and Bayesian inference (MrBayes, v.2.1.2) methods. Bootstrap proportions obtained after 100 (ML) and 1,000 (NJ) replicates and Bayesian posterior probability are indicated, respectively: 1 (34, 88, 1.0), 2 (40, 65, 0.8), 3 (9, 68, 0.6) and 4 (35, 77, 1.0). Species abbreviations are indicated before gene name (see table 3.1 for accession numbers and species names).

In this chapter, we show that TBP and Pitx proteins likely co-evolved. Thus, the TBP-Pitx interaction for different mouse (m) Pitx proteins was differently affected by removal of the TBP vertebrate-specific domain (VSD), MSD, or TBP<sub>CORE</sub> domain. Removal of the TBP MSD disrupted the interaction of TBP with the Pitx2c isoform, and cells or mice lacking the TBP VSD and MSD exhibited increased expression of the Pitx2c target gene *nppa*. Moreover, whereas hagfish (hf) PitxA protein interacted with either hfTBP or mouse (m) TBP, mPitx2c interacted with mTBP but not with hfTBP. These observations

suggest that the TBP-Pitx interaction has undergone intricate refinement and selection during vertebrate evolution. This process might have participated in evolution of various Pitx-regulated functions.

## Materials and Methods

### Yeast Two-Hybrid

The MP34 bait vector (a gift from R. Brazas, Mirus Bio Corp.) and pGADT7<sup>SN</sup> prey vector were used for yeast two-hybrid screens <sup>478</sup>. MP34 places the GAL4 DNA binding domain downstream of the bait (Fig. 3.3A), which may be a favorable arrangement for assessing interactions with bait sequences from found at or near the N-terminus of their endogenous proteins <sup>478</sup>. cDNA fragments encoding the hagfish TBP full-length (hfTBP-FL) and TBP N terminus (hfTBP-N, amino acids 1-143) were generated by PCR from a plasmid containing the hfTBP-FL cDNA <sup>391</sup> using the primers listed in table 3.1. All clones used in this study were verified by sequencing.

Table 3.1. Oligonucleotides sequences.

| Name                            | Sequence <sup>c</sup>                                 |
|---------------------------------|-------------------------------------------------------|
| TBP-N (F)                       | 5'-TATGTCGACATTGACCATGGAATCCTCCA-3'                   |
| TBP-N (R)                       | 5'-ATAGCGGCCGCGAACTCTCAGATGCTGGAGCAAC-3'              |
| TBP-C (R)                       | 5'-TATGCGGCCGCGTGGTCTTCCTGAATCCCTTYAA-3' <sup>d</sup> |
| TBP-N (F)                       | 5'-TATGTCGACAGGTATGGACGGGACAC-3'                      |
| TBP-N (R)                       | 5'-ATAGCGGCCGCGACAATCCCAGACCCCTCTGCT-3'               |
| TBP-C (R)                       | 5'-TATGCGGCCGCTAGATGTCTTCCTGAAGC-3'                   |
| TBP-N (F)                       | 5'-TATGTCGACGATGGACTCGTCGGGCAGCATG-3'                 |
| TBP-N (R)                       | 5'-ATAGCGGCCGCGTGTCTCTCAGAGGCCGGCGTGAT-3'             |
| TBP-C (R)                       | 5'-ATAGCGGCCGCGCTTGTTTTTCGAAAGCCCTTC-3'               |
| TBP-N (F)                       | 5'-ATCGTCGACTATGGACCAGAACAACAGCCCTTCCA-3'             |
| TBP-N (R)                       | 5'-TATGCGGCCGCCCAGAGCTCTCAGAAGCTGGTGT-3'              |
| TBP-C (F)                       | 5'-TATGTCGACCCACCATGGCGAGCTCTGGAATTGTACCGAG-3'        |
| TBP-C (R)                       | 5'-TATGCGGCCGCGTGGTCTTCCTGAATCCCTTTAA-3'              |
| TBP-N <sub>ΔMSD</sub> (R)       | 5'-ATAGTCGACGGGGTGGTGCCTGGCAATGGTGCA-3'               |
| TBP-MSD (F)                     | 5'-TATGTCGACCCCTTCACCAATGACTCCTATGAC-3'               |
| TBP-C <sub>263</sub> (R)        | 5'-TATGCGGCCGCTCTGGCTCATAGCTACTGAACT-3'               |
| Pitx-A-C (R)                    | 5'-ATAGCGGCCGCGAAGCCGTTCTTACACAGTTC-3'                |
| Pitx2c-N (F)                    | 5'-TATGTCGACTCCCTCGTCCATGAACTGCATG-3'                 |
| Pitx2c-C (R)                    | 5'-ATAGCGGCCGCGACCGGCCGGTCGACTGCATACTG-3'             |
| Pitx2a-N (F)                    | 5'-TATGTCGACAATGGAGACCAATTGTGCGAAACTAGT-3'            |
| TCFL5 <sub>70-273</sub> -N (F)  | 5'-TATGTCGACCTTCCAGGAGCTGCGTATGATGC-3'                |
| TCFL5 <sub>152-273</sub> -N (F) | 5'-TATGTCGACTTCAGAGTCCTCACAGTCAAACCTG-3'              |
| TCFL5 <sub>152-273</sub> -C (R) | 5'-ATAGCGGCCGCGCTTGATCTCCATGGCAGGGCTGCTCTGCA-3'       |
| Pitx1 (F)                       | 5'-TATGTCGACCATGGACGCCTTCAAGGGAGGCATGAG-3'            |
| Pitx1 (R)                       | 5'-ATAGCGGCCGCGCTGTTGTACTGGCAAGCGTTGAGC-3'            |
| U2AF2 (F)                       | 5'-TATGTCGACCATGTGCGACTTCGACGAGTTCGAG-3'              |
| U2AF2 (R)                       | 5'-TATGCGGCCGCGCGGTGGTAGGAATCAGGGTCA-3'               |
| Nppa <sub>-638</sub> (F)        | 5'-TATAGATCTGAATTCTTTAGAGCCTGTATCATG-3'               |
| Nppa <sub>+1</sub> (R)          | 5'-TATCCATGGCGTGGGTGCGGGCACGATCTG-3'                  |
| Nppa (F1)                       | 5'-TATGAATTCAGAGACAGCAAACATCAGATCGTGC-3'              |
| Nppa (F2)                       | 5'-TATGAATTCGCTTGAGCAGATCGCAAAAGATC-3'                |
| Nppa (R)                        | 5'-ATAGGATCCGACCACTCATCTACATTTAACATCGATCG-3'          |
| β-actin (F)                     | 5'-GCTGTCTGGTGGTACCACCATGTA-3'                        |
| β-actin (R)                     | 5'-CTCACTGTCCACCTTCCAGCAGAT-3'                        |
| pGADT7 (R)                      | 5'-GAAAGAAATTGAGATGGTGCAC-3'                          |
| MP34 (F)                        | 5'-CTGCACAATATTTCAAGCTATAACC-3'                       |
| MP34 (R)                        | 5'-TGTTCTTCAGACACTTGGCGC-3'                           |

<sup>a</sup>: Species (Sp): hf, hagfish; a, amphioxus; l, lamprey; m, mouse; n/a, not applicable

<sup>b</sup>: Orientation in between brackets, F: forward; R: reverse

<sup>c</sup> Endogenous protein-coding sequences are in bold and engineered restriction sites are underlined

<sup>d</sup> Y = C or T

Oligo-(dT)-primed cDNA prey libraries were constructed and inserted into the pGADT7<sup>SN</sup> <sup>478</sup> as follows. Total RNA was isolated from adult hagfish tissues (live hagfish were generously provided by J. Seaborne and D. Headlee, Oregon Department of Fish and Wildlife) or various stage lamprey embryos (50 embryos each at eleven 2-day intervals from 2- to 22-days after fertilization) as indicated in table 3.2. Poly(A<sup>+</sup>) mRNA was purified using Oligo-(dT)<sub>25</sub> Dynabeads (Invitrogen, Carlsbad, CA) following the manufacturer's recommendations. Each cDNA library was constructed from 2.5 µg of poly(A<sup>+</sup>) mRNA using the Superscript plasmid system for cDNA synthesis and cloning (Invitrogen, Carlsbad, CA), which yields cDNAs containing 5' *Sal* I and 3' *Not* I overhangs. cDNAs were cloned in pGADT7<sup>SN</sup>.

Table 3.2. Yeast two-hybrid cDNA libraries characteristics.

| Screen | Species | Bait     | Prey library | Clones screened       |
|--------|---------|----------|--------------|-----------------------|
| 1      | hagfish | hfTBP-FL | 1            | 3.7 x 10 <sup>6</sup> |
| 2      | hagfish | hfTBP-N  | 1            | 5.3 x 10 <sup>6</sup> |
| 3      | hagfish | hfTBP-FL | 2            | 3.7 x 10 <sup>6</sup> |
| 4      | hagfish | hfTBP-N  | 2            | 3.2 x 10 <sup>6</sup> |
| 5      | lamprey | lTBP-FL  | 3            | 2.2 x 10 <sup>6</sup> |
| 6      | lamprey | lTBP-N   | 3            | 3.8 x 10 <sup>6</sup> |

Two-hybrid screens and assays were performed in *Saccharomyces cerevisiae* (strain AH109), which contains the Ade2, His3, and LacZ reporters under different promoters. Library transformations were performed as described previously <sup>478</sup>. Single and two-component transformations were performed using a standard protocol <sup>479</sup>, modified to include 10% DMSO.

Two-hybrid screens were performed on sterile complete (SC) medium lacking leucine, tryptophan, and histidine (SC-L-W-H). Primary transformants were transferred after 7 to 14 days on a new SC-L-W-H and tested for His3 and Ade2 activity on plates lacking histidine (SC-L-W-H) or adenine (SC-L-W-A). Prey plasmids for the clones that grew were extracted from yeast by the glass beads lysis procedure <sup>480</sup>. Prey plasmids were electroporated into *Escherichia coli* (strain DH10B) and colonies were screened by PCR. Isolated prey plasmids were then tested for autoactivation by retransformation into AH109 with “empty” (no bait insert) MP34 bait plasmid. Non-autoactive clones were retested in AH109 for interaction with hfTBP-FL or hfTBP-N to confirm each interaction. Interacting clones were sequenced and the in-frame DNA-binding domain-fused predicted protein sequence was analyzed with BLASTP through the NCBI server <sup>481</sup> for identification.

#### Isolation of Orthologous cDNAs and Interaction Mapping Tests

A mPitx1 cDNA was isolated by RT-PCR from embryonic day 10.5 (E10.5) pregnant mouse uterus RNA; mPitx2c and mPitx2a cDNAs were isolated by RT-PCR from adult mouse pituitary RNA; an mTCFL5 cDNA was isolated by RT-PCR from adult mouse testis RNA; mTCFL5<sub>152-273</sub> was amplified from the mTCFL5 cDNA using primers indicated in table 3.1. To identify hfPitxA mRNA variants, a 5' RACE (rapid amplification of cDNA ends) was performed using the hagfish libraries described above as templates. PCR reactions used a vector primer (T7) and a reverse primer designed downstream of the homeodomain that would amplify all hagfish PitxA 5'-splice variants (Table 3.1).

### Mouse Embryonic Fibroblast Isolation and Transfection

Mouse embryo primary fibroblast (MEF) cultures were established using E11.5-12.5 fetuses from *tbp*<sup>ΔN/+</sup> X *tbp*<sup>ΔN/+</sup> matings, as described previously<sup>357</sup>. Transfections were performed using 0.5x10<sup>6</sup> cells and plasmid conditions indicated in figure 3.7 under conditions that typically gave greater than 50% transfection efficiency (data not shown). Luciferase assays were performed by standard procedures as described elsewhere<sup>482</sup>. Each condition was performed in triplicate (three independent electroporations). Luciferase activity was normalized to total protein content and to a positive control, and the maximal value was set as 100%.

### RT-PCR and RNase Protection Assays

RT-PCR were performed as previously described<sup>357</sup>. mRNA relative abundance was performed for NPPA and normalized using gene specific primer sets (Table 3.1). RNase protection was performed as described earlier<sup>357</sup>. For *nppa*, a probe hybridizing to sequences between positions 575 and 725 (Table 3.1) of mouse *nppa* cDNA (GenBank accession number NM\_008725) was transcribed from a clone isolated by RT-PCR using mouse adult heart RNA.

### Whole Mount In Situ Hybridization

Whole mount in situ hybridizations were performed as described<sup>483</sup>. A 676-bp mouse *Nppa* cDNA fragment was isolated by RT-PCR using mouse heart RNA and the primers indicated in table 3.1.

## Results

### Identification of TBP-Interacting Proteins

The phyla-specific region of TBP in vertebrates evolved at a rate between those of the common “molecular clocks”, histone H2B and  $\alpha$ -actin<sup>391</sup>, indicating this region has been subjected to relatively stringent purifying selection. Removal of this region from mouse (m) TBP does not disrupt general TBP functions<sup>357</sup>, but results in high fetal- and neonatal-lethality<sup>351</sup>. To gain insights into the functions of the phyla-specific region of TBP, we performed two-hybrid screens for cyclostome proteins that can bind to it. The cyclostome clade, which includes hagfishes (hf) and lampreys (l), are basal vertebrates. They share with higher vertebrates only the most ancient vertebrate ancestor and they are thought to have lower genomic redundancies than higher vertebrates<sup>391,484-486</sup>. Cyclostomes have a TBP-N homologous and similar to mammals<sup>391</sup> (Fig. 3.1B). Interactions between hfTBP and the mouse orthologs of the known TBP<sub>CORE</sub>-interacting proteins TFIIA<sup>359,459,487,488</sup>, Brf1<sup>459,487-490</sup>, PIAS1, and PIAS3<sup>478</sup> verified that the hfTBP-FL bait was functional in yeast (Fig. 3.3B).

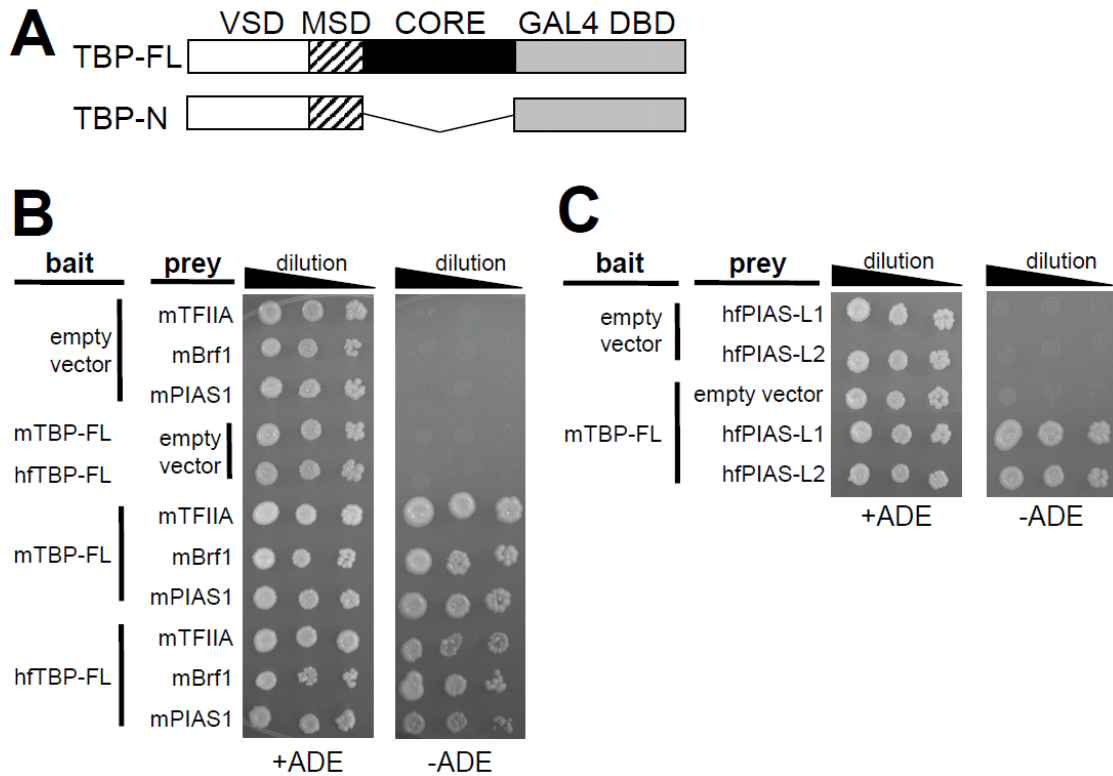


Figure 3.3: Yeast two-hybrid baits and quality controls. **(A)** Yeast two-hybrid bait constructs. TBP-FL and TBP-N were expressed from the MP34 plasmid, which fused the GAL4 DNA binding domain downstream of TBP. N designates the vertebrate-specific N terminus, encompassing VSD and MSD. TBP-FL and TBP-N baits from both hagfish (hf) and lamprey (l) were generated. **(B)** hftBP was functionally expressed in yeast. Yeasts were transformed with the indicated controls. Black wedges represent 10-fold serial dilutions. Growth on +ADE indicates that yeasts contained both prey and bait plasmids. Growth on -ADE indicates interaction between bait and prey constructs. ADE: adenine. **(C)** hftBP-FL interacted with the PIAS homologs hfPIAS-L1 and hfPIAS-L2, identified in the yeast two-hybrid screens (see table 3.4).

cDNA libraries were prepared from mixtures of adult hagfish tissues or from lamprey embryos ranging from 2- to 22-days after fertilization (Table 3.2) and were screened with either the full length (TBP-FL) or the phyla-specific (TBP-N) region of hfTBP or lTBP, respectively (Fig. 3.3A; Table 3.3). Two clones of PIAS-like proteins (termed hfPIAS-L1 and hfPIAS-L2), which are known TBP<sub>CORE</sub> interactors<sup>478</sup>, were isolated in screens using hfTBP-FL as bait (Table 3.3) and were verified as also interacting with mTBP-FL (Fig. 3.3C), thus internally validating the conditions of the screen.

Table 3.3. Yeast two-hybrid screens.

| Screen | Species | Bait     | Prey library | Clones screened   |
|--------|---------|----------|--------------|-------------------|
| 1      | hagfish | hfTBP-FL | 1            | $3.7 \times 10^6$ |
| 2      | hagfish | hfTBP-N  | 1            | $5.3 \times 10^6$ |
| 3      | hagfish | hfTBP-FL | 2            | $3.7 \times 10^6$ |
| 4      | hagfish | hfTBP-N  | 2            | $3.2 \times 10^6$ |
| 5      | lamprey | lTBP-FL  | 3            | $2.2 \times 10^6$ |
| 6      | lamprey | lTBP-N   | 3            | $3.8 \times 10^6$ |

Three proteins were identified that interacted with the hagfish phyla-specific region: a Pitx (paired-like homeodomain transcription factor) family member, a TCFL5 (transcription factor-like 5) homolog, and a clone unrelated to known proteins (Table 3.4). The later protein had several short regions similar to sequences in CPSF1 (cleavage and polyadenylation specificity factor 1), which has been shown to be associated with the TBP-containing TFIID complex at pre-initiation <sup>491</sup>. However, since no orthologous proteins were found in mammals, this protein was not pursued. TCFL5 is a poorly characterized basic helix-loop-helix transcription factor <sup>492</sup>. No orthologues of TCFL5 were found in searches of the sea urchin, *Ciona*, or amphioxus genomes and queries in GenBank did not yield any orthologues outside of the vertebrate subphylum (data not shown). Hence, TCFL5 may be a vertebrate-specific transcription factor. The hfPitx protein isolated in our screen was an N-terminally truncated clone with close similarity to the previously reported lamprey “PitxA” protein <sup>240</sup> and was therefore designated hfPitxA.

Table 3.4. TBP-interacting proteins identified by yeast two-hybrid screens.

| Bait     | Identity  | Clones (n=) | Known TBP interaction | Genbank accession number <sup>a</sup> | Gene ontology <sup>b</sup><br>molecular function |
|----------|-----------|-------------|-----------------------|---------------------------------------|--------------------------------------------------|
| hfTBP-FL | hfPIAS-L1 | 2           | Yes <sup>c</sup>      |                                       | transcription factor                             |
| hfTBP-FL | hfPIAS-L2 | 1           | Yes <sup>c</sup>      |                                       | transcription factor                             |
| hfTBP-N  | hfPitxA   | 1           | No                    |                                       | transcription factor                             |
| hfTBP-N  | hfTCFL5   | 1           | No                    |                                       | transcription factor                             |
| hfTBP-N  | unknown   | 1           | No                    |                                       | n/a                                              |

<sup>a</sup> Sequences have been submitted to Genbank. Final accession numbers will be available before publication.

<sup>b</sup> Gene ontology term associations were compiled from <http://amigo.geneontology.org> and NCBI Gene.

<sup>c</sup> 478

To obtain the missing N-terminal hfPitxA sequences, we performed a 5'-RACE analysis and recovered 16 hfPitx clones. Two alternative splicing-dependent isoforms of hfPitxA protein were identified, designated hfPitxA1 and hfPitxA2. mRNAs for the isoforms exhibited alternative 5' sequences, including the untranslated region and first 20 (hfPitxA1) or 15 (hfPitxA2) predicted amino acids. This suggests the gene contains alternative promoters and 5' exons, and it issues at least two N-terminally distinct protein products. These results are consistent with cyclostome *pitxa* representing the ancestral vertebrate *pitx* gene, which was present at the pre-vertebrate/vertebrate transition, and with expansion of the gene family having occurred later <sup>239</sup>. Cyclostome PitxA is most similar to mammalian Pitx2 (77% identity versus 65% for mPitx1 or mPitx3) (Fig. 3.4 and data not shown).

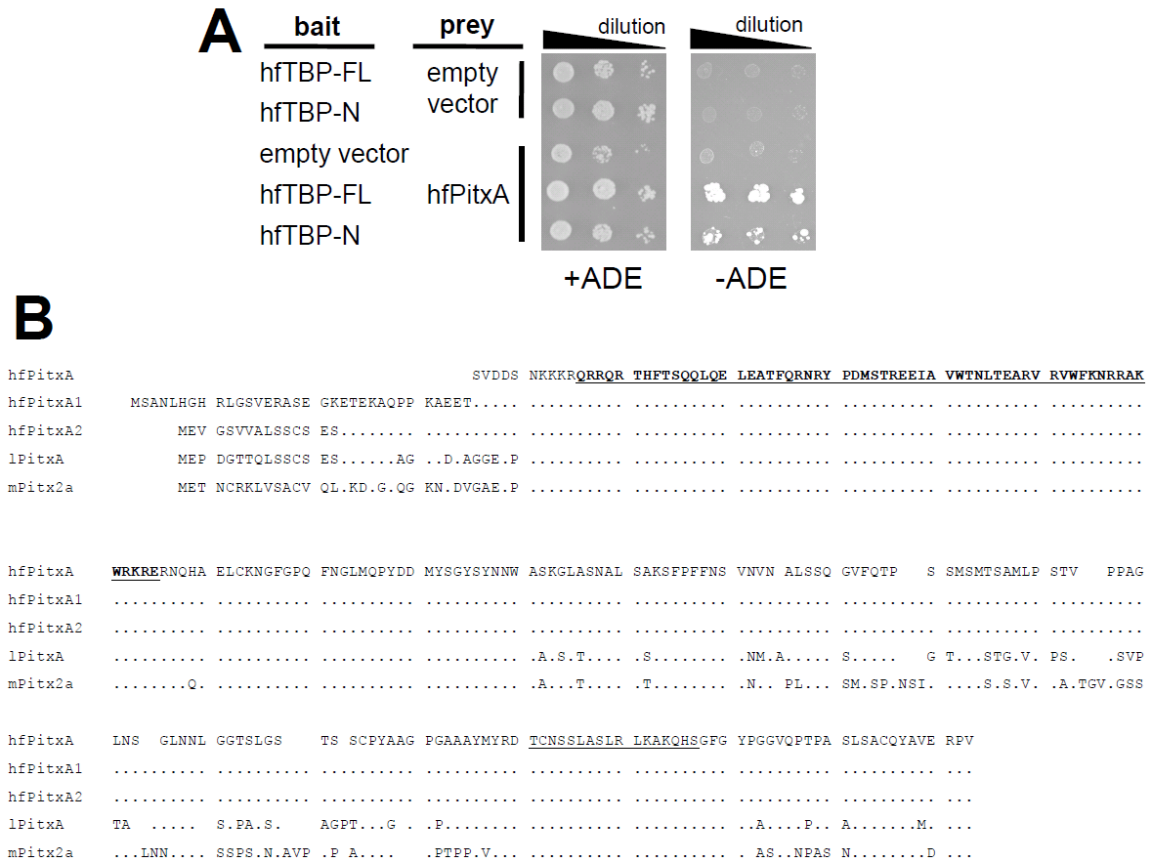


Figure 3.4: Identification of hfPitxA as a TBP-N-interacting protein. **(A)** hfPitxA interaction with hfTBP. Bait and prey identities are tabulated at left. Dilutions were a 10-fold series and were identical for both conditions. Growth on +ADE indicates yeast contained both prey and bait plasmids. Growth on -ADE indicates an interaction between bait and prey constructs. ADE, adenine. **(B)** Nucleotide sequences alignment of the hagfish (hf) Pitx homolog with mouse (m) Pitx2a, lamprey (l) PitxA, and hfPitxA 5' RACE clones. Bold underlined represent the homeodomain, perfectly conserved in all species. Underlined sequence represents the OAR domain, characteristic of Pitx family members. \* indicates the partial clone identified during the yeast two-hybrid screen. Sixteen hfPitxA clones were identified by 5'RACE. MUSCLE alignment<sup>460</sup> revealed two distinct isoforms, hfPitxA1 and hfPitxA2 as indicated. Fourteen clones of hfPitxA1 and two of hfPitxA2 were identified and aligned separately. Translation is indicated for the putative coding sequence.

Table 3.5. Amino acid sequences used for phylogenetic analyses.

| Taxonomy        | Species                      |     | TBP<br>accession<br>number | Name       | Pitx<br>Accession<br>number |
|-----------------|------------------------------|-----|----------------------------|------------|-----------------------------|
| Mammal          | <i>Homo sapiens</i>          | Hsa | CAG33057.1                 | Pitx2a     | NP_700476                   |
| Mammal          | <i>Mus musculus</i>          | Mmu | BAA00840.1                 | Pitx2c     | EAX06264                    |
| Marsupial       | <i>Monodelphis domestica</i> | Mdo | ENSMODP00000               | Pitx2a,b,c | EDL12244,5,6                |
| Bird            | <i>Gallus gallus</i>         | Gga | 0006780                    | n/a        |                             |
| Reptile         | <i>T. flavoviridis</i>       | Tfl | BAA20298.1                 | Pitx2      | NP_990341                   |
| Amphibian       | <i>Xenopus laevis</i>        | Xla | BAA06555.1                 | n/a        |                             |
|                 |                              |     | AAI06645.1                 | Pitx2a,b   | CAA06696,7                  |
| Fish            | <i>Danio rerio</i>           | Dre | NP_956390.1                | Pitx2a     | AAD34391                    |
| Fish            | <i>Oncorhynchus mykiss</i>   | Omy | n/a                        | Pitx2c     | AAD34390                    |
| Fish            | <i>G. cirratum</i>           | Gci | AAO34522.1                 | Pitx2      | AAS92266                    |
| Cyclostome      | <i>L. japonicum</i>          | Lja | n/a                        | Pitx       | BAD12775 <sup>a</sup>       |
| Cyclostome      | <i>Lethenteron reissneri</i> | Lre | BAB78513                   | n/a        |                             |
| Cyclostome      | <i>Petromyzon marinus</i>    | Pma | AAO34515.1                 | PitxA      | CAD30206                    |
| Cyclostome      | <i>Eptatretus stoutii</i>    | Est | AAO34521.1                 | PitxA1     | <sup>b</sup>                |
|                 |                              |     |                            | PitxA2     | <sup>b</sup>                |
| Cephalochordate | <i>B. floridae</i>           | Bfl | AAO34516.1                 | Pitx       | CAD27489 <sup>b</sup>       |
| Cephalochordate | <i>B. belcheri</i>           | Bbe | n/a                        | Pitx       | AAF03901                    |
| Tunicate        | <i>Ciona savignyi</i>        | Csa | ENSCSAVT000                | Pitx       | <sup>c</sup>                |
|                 |                              |     | 00012298                   |            |                             |
| Tunicate        | <i>Ciona intestinalis</i>    | Cin | ENSCINP00000               | Pitxa/b    | AAU93886.1                  |
|                 |                              |     | 005464                     | Pitx       | CAD27490                    |
|                 |                              |     |                            | Pitx       | AAW23072                    |
| Tunicate        | <i>Oikopleura dioica</i>     | Odi | n/a                        | Pitx_v1    | AAZ23132                    |
|                 |                              |     |                            | Pitx_v2    | AAZ23137                    |
|                 |                              |     |                            | Pitx_v4    | AAZ23135                    |
| Echinoderm      | <i>Paracentrotus lividus</i> | Pli | n/a                        | Pitx2      | AAW51825                    |
| Echinoderm      | <i>S. purpuratus</i>         | Spu | AAB47272.1                 | Pitx2      | XP_001176640                |
|                 |                              |     |                            | Pitx2      | XP_001192248                |
| Echinoderm      | <i>Lytechinus variegatus</i> | Lva | AAC35362.1                 | n/a        |                             |
| Echinoderm      | <i>H. pulcherrimus</i>       | Hpu | BAB92075.1                 | Pitx       | BAE66654                    |
| Insect          | <i>D. melanogaster</i>       | Dme | AAA79092.1                 | D-Pitx1    | O18400                      |
| Insect          | <i>Anopheles gambiae</i>     | Aga | EAA45305.2                 | n/a        |                             |
| Insect          | <i>C. quinquefasciatus</i>   | Cqu | EDS30877.1                 | n/a        |                             |
| Insect          | <i>Aedes aegypti</i>         | Aae | AAEL012330-RA.1            | n/a        |                             |
| Insect          | <i>Bombyx mori</i>           | Bmo | CAA11000                   | n/a        |                             |
| Insect          | <i>S. frugiperda</i>         | Sfr | P53361.1                   | n/a        |                             |
| Insect          | <i>Apis mellifera</i>        | Ame | XP_623088.1                | n/a        |                             |
| Insect          | <i>Tribolium castaneum</i>   | Tca | XP_969326.1                | n/a        |                             |

<sup>a</sup> Partial clone<sup>b</sup> Sequences have been submitted and will be available at publication<sup>c</sup>: SINCSAVT00000011282

### mPitx2 is a TBP-Binding Protein

To determine whether the interactions between TBP and these cyclostome proteins were conserved in higher vertebrates, mouse orthologs of Pitx and TCFL5 proteins were tested for interaction with mTBP. cDNA clones of mPitx1, mPitx2a, mPitx2c, and mTCFL5 were isolated. Prey plasmids containing full-length mTCFL5 were strongly auto-active and could not be used in two-hybrid assays (data not shown); however a partial clone containing only the region homologous with hfTCFL5 (mTCFL5<sub>152-273</sub>) was not auto-active. Both mPitx2c and mTCFL5<sub>152-273</sub> interacted with mTBP-FL (Fig. 3.5A, B, and data not shown). mPitx2c also interacted with mTBP-N; however mTCFL5<sub>152-273</sub> did not (Fig. 3.5B). Therefore, attention was focused on Pitx proteins.

During vertebrate evolution, the phyla-specific domain of TBP diverged slowly, such that cyclostomes and mammals share ~60% amino acid identity in this region<sup>391</sup>. Concomitantly, the *pitx* gene family expanded to contain three genes<sup>239</sup>, and the proteins diverged such that, in mice, they share ~ 70% identity with each other, and each shares 65% to 77% identity with hfPitxA (Figs 2.2, 2.4B).

To investigate whether the interaction between Pitx proteins and TBP might have co-evolved in vertebrates, inter-specific protein-protein interaction assays were performed. mTBP-N interacted with hfPitxA (Fig. 3.5C) whereas hfTBP-N did not interact with mPitx2c (Fig. 3.5C). This suggested that amino acid differences between mPitx2c and hfPitxA that prevent the former from binding to hfTBP-N are complemented by amino acid differences in mTBP-N. Besides interacting with hfPitxA and mPitx2c, mTBP-FL interacted with mPitx1 (Fig. 3.5A), suggesting that the Pitx-TBP interaction involved a domain shared between Pitx1 and Pitx2 proteins.

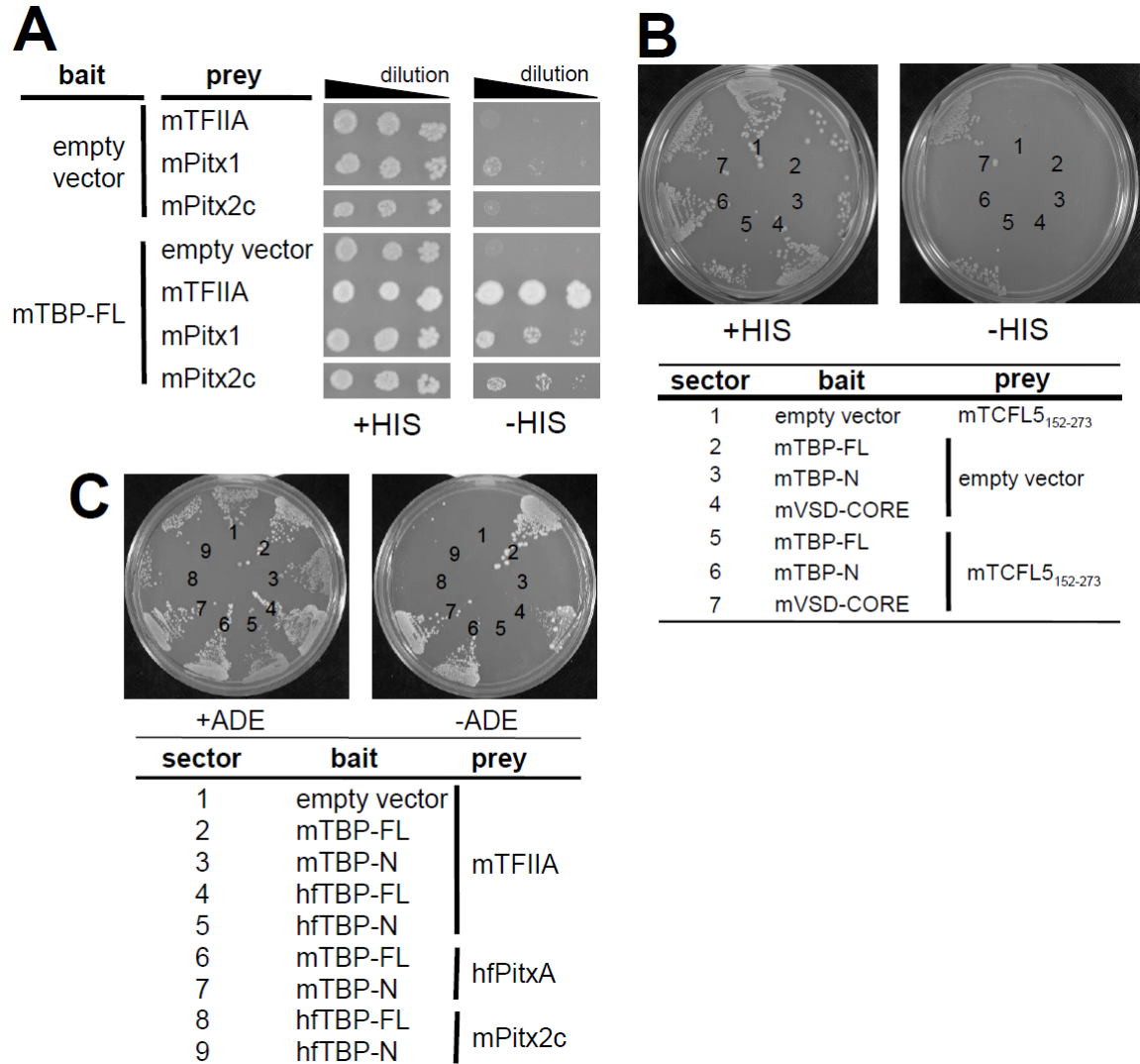


Figure 3.5: Identification of mouse orthologs TBP-binding partners. **(A)** TBP-Pitx1, Pitx2c interaction. mTFIIA, mouse general transcription factor IIA, is a known TBP<sub>CORE</sub>-interactor<sup>358,493</sup> that was used as a positive control. Dilutions were a 10-fold series and were identical for both conditions. Growth on +ADE indicates yeast contained both prey and bait plasmids. Growth on -ADE indicates an interaction between bait and prey constructs. ADE, adenine. **(B)** TBP-TCFL5 interaction. A mTCFL5 clone, encoding for a domain similar to the hfTCFL5 clone identified in the screen, interacted with mTBP-FL but not with the MSD or VSD+MSD. Sector numbers on plate corresponds to sector number in the table. **(C)** Testing interactions of hfTBP with mPitx2c and of mTBP with hfPitxA. mTFIIA specificity control shows TBP<sub>CORE</sub>-dependent species-independent interaction.

### The TBP-Pitx Interaction Involves the Metazoan-Specific Domain of TBP

It was curious that Pitx proteins, which are found throughout metazoan phyla, would interact with the phyla-specific region of vertebrate TBPs. Although a portion of this region, the metazoan-specific domain (MSD), is shared with all metazoans, this proline-rich domain is relatively small and simple (Fig. 3.1B). Indeed, it has been speculated that the MSD likely served only as common “linker” to join different functional phyla-specific N-terminal sequences to the universal TBP<sub>CORE</sub><sup>391</sup>. However, outside of the MSD, the phyla-specific region of TBPs from vertebrates, cephalochordates, tunicates, insects, and nematodes show no evidence of homology with each other (Fig. 3.1A)<sup>345</sup>. As a first test of whether the Pitx-TBP interaction was truly common to other metazoans, and not vertebrate-specific, we tested whether hfPitxA would interact with TBP from amphioxus (a), a cephalochordate. aTBP shares a very similar MSD and TBP<sub>CORE</sub> region with vertebrate TBPs, but it has an extended cephalochordate-specific domain having no sequence identity with vertebrate TBPs (Fig. 3.1B)<sup>391</sup>. Results showed that aTBP interacted with hfPitxA (Fig. 3.6A, B), verifying that the Pitx-TBP interaction was not restricted to vertebrates.

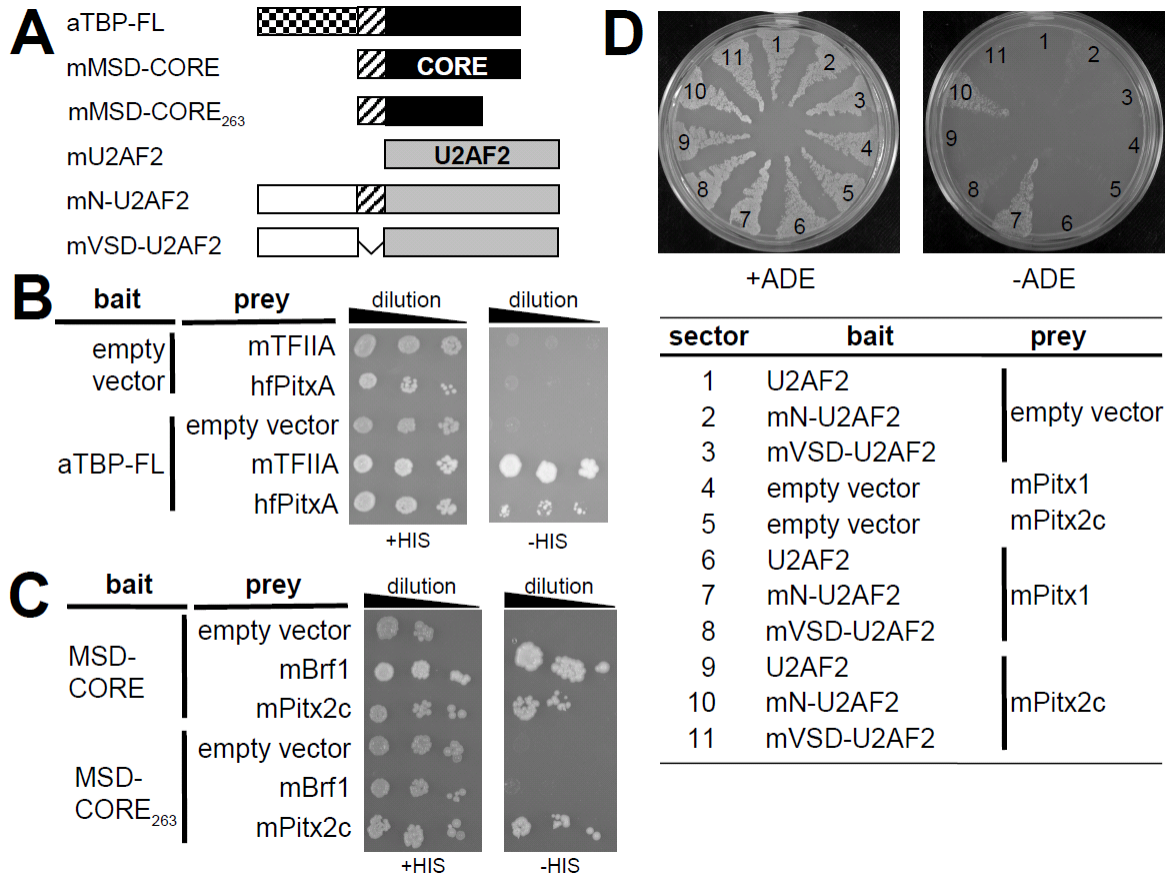


Figure 3.6: The TBP-Pitx interaction involves the MSD. **(A)** Schematic of relevant regions of bait constructs used in this figure. **(B)** Amphioxus (a) TBP-FL interaction with hfPitxA. Designations and dilutions as in figure 3.5. **(C)** Role of VSD in mTBP-mPitx2c interaction. mBRF1, mouse B'-related factor 1, a known TBP<sub>CORE</sub>-interactor<sup>358,478</sup> used as a specificity control for an interaction requiring intact TBP<sub>CORE</sub>. **(D)** Roles of MSD and TBP<sub>CORE</sub> in mTBP-mPitx2c and mTBP-mPitx1 interactions. Numerals on plates correspond to sector designations tabulated below.

The results above suggested that the Pitx-TBP interaction involved the TBP MSD. To test this, mTBP bait constructs containing the MSD but lacking the VSD (mMSD-CORE) or lacking the VSD and the last 45 amino acids of the TBP<sub>CORE</sub> (mMSD-CORE<sub>263</sub>; Fig. 3.6C) were used. Due to removal of part of the TBP<sub>CORE</sub> domain, mMSD-CORE<sub>263</sub> would not interact with Brf1, but both mMSD-CORE and mMSD-CORE<sub>263</sub> retained the interaction with mPitx2c (Fig. 3.6C). We also fused either the mTBP phyla-specific region (VSD + MSD) or only the VSD to an unrelated and irrelevant nuclear protein from

the splicing machinery, U2AF2<sup>494</sup>, and tested whether either of these would interact with mPitx2c (Fig. 3.6D). The bait containing the TBP MSD interacted with mPitx2c whereas the bait lacking the MSD did not interact (Fig 2.6D). This indicated that the MSD was required for the Pitx-TBP interaction, the VSD was not, and the TBP<sub>CORE</sub> was dispensable in the context of either the mN+U2AF2 or the mTBP-N bait construct. Taken together, these data hint that TBP MSD interacts with multiple chordate Pitx proteins. It is therefore likely that this ancestral interaction has been retained and possibly specialized in more recent animals.

#### TBP-N Attenuates Pitx2-Dependent Transcription Activation

Next, we wanted to know if the TBP-Pitx2 interaction was functional *in vivo*. Whether mPitx2c could functionally interact with mTBP was tested in reporter assays. Pitx2 is a DNA-binding transcription factor that has been shown to directly activate genes including *nppa*<sup>472</sup>, *plod1*<sup>495</sup>, *dlx2*<sup>496</sup>, *prl*<sup>497</sup>, and several pituitary-specific genes<sup>464,498</sup>. *nppa* encodes the 35 amino acids atrial natriuretic factor A precursor, from which the 14 amino acids atrial natriuretic factor (ANF), a vertebrate-specific cardiac hormone involved in electrolyte and fluid homeostasis of the cardiovascular system, is cleaved<sup>499-502</sup>. Pitx2 has been shown to bind and activate the promoter of the *nppa* gene<sup>472,503,504</sup>. Sequences from -638 to +1 of the mouse *nppa* gene were cloned upstream of the firefly *luciferase* gene and this reporter was co-transfected with expression cassettes encoding either mTBP-FL or mTBP lacking the central metazoan-specific domain (mTBP<sub>ΔMSD</sub>) into COS7 cells, which lack endogenous Pitx family proteins but contain endogenous full-length TBP (Fig. 3.7A, B). Consistent with published data<sup>504</sup>, expression of Pitx2c

increased the activity of the *nppa* reporter by ~10-fold (white bars). Overexpression of either mTBP-FL or mTBP<sub>ΔMSD</sub> in the absence of Pitx2 expression also increased activity of the *nppa* promoter several-fold (compare white with gray or black bars at left, respectively). However, whereas co-expression of mPitx2c and mTBP-FL had no effect beyond that of mPitx2c alone (gray bar at right), co-expression of mPitx2c and mTBP<sub>ΔMSD</sub> further augmented the activity of the *nppa* promoter (black bar at right). This unexpected result suggested that the MSD may attenuate the ability of mPitx2c to activate the *nppa* promoter. Nevertheless, interpretation of the data was complicated both by uncertainties of the effects of overexpression of TBP on the cells and by the presence of endogenous TBP-FL in the COS7 cells. Therefore, we wanted to test the effects of mPitx2c on the *nppa* promoter in cells in which only endogenous levels of TBP that either possessed or lacked the MSD were expressed.

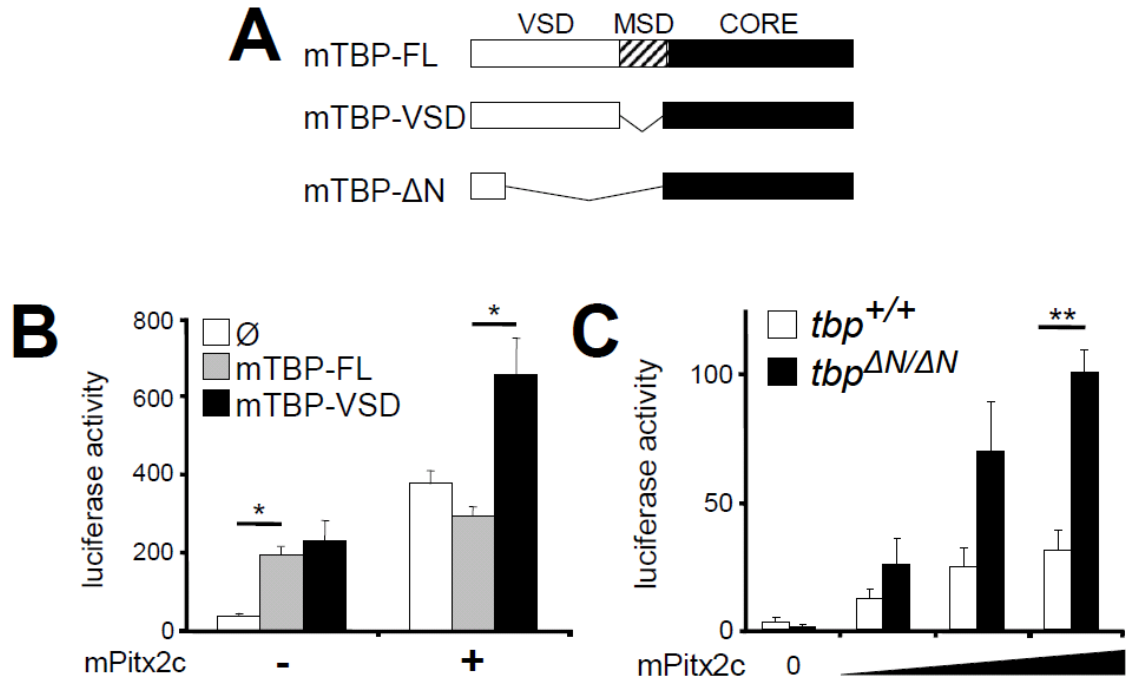


Figure 3.7: The TBP MSD modulates mPitx2c activity on the *nppa* promoter. **(A)** Schematic of TBP proteins expressed in the cells analyzed in this figure. mTBP-ΔN, the protein expressed from the *tbp*<sup>ΔN</sup> allele, lacks the MSD and all except the first 24 amino acids of the VSD<sup>351</sup>. **(B)** Functional interaction of TBP-MSD and Pitx2c on the *nppa* promoter. A mouse *nppa* promoter (-664 to +90 from the cap site<sup>472</sup>) driven *luciferase* reporter gene was co-transfected into COS7 cells with an expression vector for the indicated mTBP protein and with or without an expression vector for mPitx2c protein. Transfection efficiencies were normalized to a transfected positive control plasmid; assays represent equivalent amounts of cell lysate based on protein content. COS7 cells express endogenous TBP-FL in addition to the transfected TBP proteins. **(C)** mPitx2c differentially activates transcription from the *nppa* promoter in cells containing only endogenous TBP-FL versus cells containing only endogenous TBP-ΔN. Individual primary mouse fibroblasts from four wildtype (*tbp*<sup>+/+</sup>) or four mutant (*tbp*<sup>ΔN/ΔN</sup>) fetuses<sup>357</sup> were transfected in triplicate with the *nppa* promoter-driven *luciferase* reporter gene and increasing amounts of plasmid expressing mPitx2c, as indicated. Results represent the average normalized values from three transfections of all four wildtype and all four mutant fibroblast lines. Normalization was as in panel **(B)**, above and *tbp*<sup>ΔN/ΔN</sup> with 10 μg of Pitx2 was set as 100%. Displayed are means plus error bars (SEM); asterisks, significantly different (\*, P < 0.05; \*\*, P < 0.01), Student's unpaired t-test.

Primary fibroblast cultures were explanted from four wild-type fetuses and from four fetuses homozygous for a targeted mutation that removed amino acids 25-135 of the endogenous TBP protein, including the entire MSD (amino acids 114-135; entitled *tbp*<sup>ΔN/ΔN</sup>; mTBP-ΔN in Fig. 3.7A)<sup>351,357</sup>. Using a protocol that gave ~50% transfection

efficiency, the *nppa* promoter-based luciferase reporter was co-transfected into each of these cell lines with varying amounts of the mPitx2c expression plasmid. Results showed that, whereas increased amounts of Pitx2c expression plasmid increased the activity of the *nppa* promoter in a dose-dependent fashion over a ~5-fold range in wild-type fibroblast cultures (Fig. 3.7B, white bars), the same plasmid concentrations activated the *nppa* promoter over a ~20-fold range in the *tbp*<sup>ΔN/ΔN</sup> fibroblast cultures (Fig. 3.7C, black bars). These results suggest Pitx2c activates the *nppa* promoter in a manner that is attenuated or modulated by the TBP phyla-specific region.

#### The TBP Phyla-Specific Region is involved in Nppa Regulation *in vivo*

If the TBP-MSD restricted Pitx2-dependent transcription *in vivo*, one might predict that mPitx2-regulated genes would show expanded levels or domains of expression in *tbp*<sup>ΔN/ΔN</sup> mice. Although *tbp*<sup>ΔN/ΔN</sup> mice exhibit a combined 95-99% lethality in mid-gestation and the early post-natal period, the few surviving adults are overtly normal, including apparently normal heart development<sup>351</sup> (data not shown). Levels of *Nppa* mRNA were compared in whole hearts from wild-type (*tbp*<sup>+/+</sup>) and *tbp*<sup>ΔN/ΔN</sup> adults. Semi-quantitative RT-PCR showed that *Nppa* mRNA was more abundant in *tbp*<sup>ΔN/ΔN</sup> hearts than in wild-type hearts (Fig. 3.8A). Quantitative analyses based on real-time RT-PCR (data not shown) and RNase protection assays (Fig. 3.8B) indicated that this increase was 3- to 4-fold in magnitude.

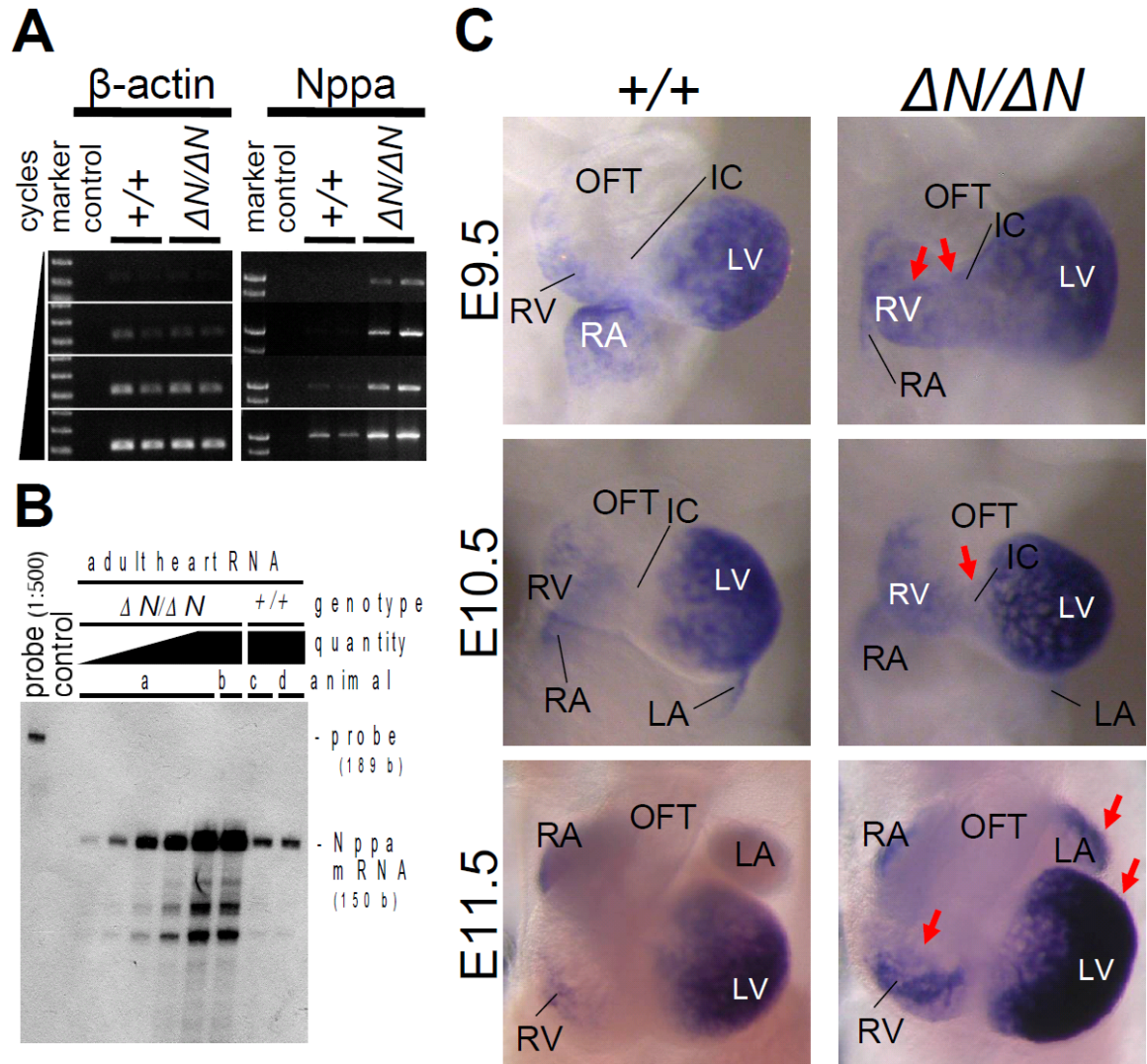


Figure 3.8: Cardiac *Nppa* mRNA expression in wildtype and *tbp*<sup>ΔN/ΔN</sup> mice. **(A)** Relative *Nppa* and  $\beta$ -actin mRNA levels in adult heart determined by semi-quantitative RT-PCR. Oligo (dT)-primed cDNAs were prepared from 2  $\mu$ g of total RNA from adult male hearts of the indicated genotype and 0.2% of the total was used in each PCR reaction. The sets of four vertically stacked panels represent increasing PCR cycles, as indicated by wedge at left (28, 30, 32, and 34 cycles for *Nppa*; 24, 26, 28, and 30 cycles for  $\beta$ -actin). **(B)** Quantitative measure of the difference in cardiac *Nppa* mRNA levels between wildtype and *tbp*<sup>ΔN/ΔN</sup> mice by RNase protection. Non-digested "probe (1:500)" sample (left lane) contained 10 amol of probe and 10  $\mu$ g of yeast RNA, and was not RNase-digested. All other hybridizations contained 5 fmol radiolabeled *nppa* probe and a total of 50  $\mu$ g of RNA, including indicated amounts of adult male heart total RNA and yeast RNA, and were RNase-digested following hybridization. In the four lanes at right, 4.0  $\mu$ g of total heart RNA from adult male mice "A", "B", "C", and "D", respectively, was used. To the left, samples contained two-fold dilutions of mouse "A" heart RNA (from left to right: 0.25, 0.5, 1.0, and 2.0  $\mu$ g, as indicated by wedge above, with the difference being made up with yeast RNA). The negative "control" sample contained 5 fmol of probe and 50  $\mu$ g of yeast RNA (no mouse

Figure 3.8 - continued

heart RNA). Results indicated *Nppa* mRNA levels were 6-fold higher in *tbp*<sup>ΔN/ΔN</sup> hearts than in wildtype hearts. **(C)** Developmental cardiac *Nppa* mRNA expression. Whole-mount *in situ* hybridizations were performed using a complementary digoxigenin-labeled *Nppa* RNA probe on fetuses harvested at embryonic day (E) 9.5, E10.5 and E11.5, as indicated. Genotypes were determined by PCR. Four to six fetuses of each genotype at each stage were included except at E11.5, when, due to large losses of *tbp*<sup>ΔN/ΔN</sup> fetuses by this stage<sup>351</sup>, only three mutant fetuses were recovered. Representative data are shown. *Nppa* mRNA expression is visualized as purple staining. Red arrows point to regions of expanded *Nppa* mRNA abundance in *tbp*<sup>ΔN/ΔN</sup> as compared to wildtype fetuses. OFT, outflow tract; IC, inner curvature; LV, left ventricle; RA, right atrium; RV, right ventricle.

The multi-chambered heart is a vertebrate-specific innovation regulated by a complex gene regulatory network (see chapter 2), where morphological complexity correlates with phylogeny<sup>176,505</sup>. *Nppa* mRNA has been used as a marker of myocardium chamber formation<sup>177,505</sup>, where it preferentially accumulates in a defined pattern in the left side of the developing heart. Heart asymmetry is regulated by Nodal signaling, which activates *Pitx2c*, while other *Pitx* isoforms are not involved in cardiogenesis<sup>506,507</sup>. Since *nppa* is regulated by *Pitx2c*<sup>504</sup>, *Nppa* mRNA expression was assessed in *tbp*<sup>ΔN/ΔN</sup> fetuses by whole-mount *in situ* hybridization. In E9.5 and E10.5 wild-type fetuses, *Nppa* mRNA abundance was primarily in the left ventricle and atrium, with much lower expression in the right ventricle (Fig. 3.8C), as previously reported<sup>177</sup>. By contrast, in *tbp*<sup>ΔN/ΔN</sup> fetuses, there was an increased spatial distribution into the inner curvature of the heart and greater *Nppa* mRNA accumulation in the left atrium and ventricle at E11.5 (Fig. 3.8C). Also, although wild-type hearts showed diminishing *Nppa* level in the right ventricle between E9.5 and E11.5, mutant hearts showed continued or augmented right ventricular abundance by E11.5 (Fig. 3.8C). Taken together, these data suggest that TBP-N likely interacts, *in vivo*, with *Pitx2c* to regulate the expression of *nppa*.

## Discussion

### Summary

To investigate why the phyla-specific region of TBP changed at, or very near to, the pre-vertebrate/vertebrate transition, but has subsequently been conserved, we performed screens for proteins from basal vertebrates that would interact with this domain. We reasoned that basal vertebrates, likely having lower genomic complexity<sup>485</sup>, and thus possibly lower genetic redundancy, might more efficiently yield the “ancestral” interacting proteins than would similar investigations in more recent vertebrates. Despite that yeast two-hybrid screens do not allow exhaustive screening, we identified three proteins that interacted with the phyla-specific region of TBP. Of those three cyclostome proteins, one had no obvious homolog in mammals, one (TCFL5) had a mouse homolog but we could not detect the interaction found in hagfish, in mouse, and one (PitxA) had mouse homologs that interacted with mTBP-N.

Here, we show that TBP-N is a functional domain involved in protein-protein interactions. The interaction between hfTBP-N and hfTCFL5 was not detected between the mTBP-N and mTCFL5. Either we could not detect the interaction due to sensitivity issues or the interaction present in hagfish was not present in mouse. The interaction between hfTBP-N and hfPitxA was conserved between the mouse orthologs: TBP and Pitx2c. Also, TBP interacted with Pitx1, a likely recent paralog of Pitx2. The interaction domain was refined to the TBP MSD, which was until then considered a linker<sup>391</sup> and possibly not thought as a functional domain. *In vivo*, it is likely that TBP-N acts as a repressor of Pitx2c activity on its target gene, *nppa*, as lack of TBP-N (TBP-ΔN)

increased the Pitx2c-dependent gene expression. Therefore, the binding of Pitx2c to TBP-N may “tune down” the activity of Pitx2c.

### Implications

#### The Unstructured TBP-N is a Functional Protein-Protein Interaction Domain:

The TBP variable region (VSD+MSD) has been described as unstructured (see chapter 2). The presence of an unstructured domain is fairly common and is involved in crucial roles for transcriptional and translational regulation as well as signal transduction<sup>373-377,401</sup>. This suggests a role for disordered regions to be involved in protein-protein interactions in general<sup>382,412-415</sup>, and TBP-N in particular. Natively unstructured proteins or polypeptide regions, like TBP-N, have been shown to often be part of protein complexes that present different interfaces for different protein-protein interaction<sup>417,508</sup>. Therefore, the function of the vertebrate specific TBP-N may be to assemble protein complexes or bridge multiple partners in a tissue- or developmental stage-specific fashion and not only to be involved in a binary protein-protein interaction as shown in the present chapter.

#### TBP-Pitx Co-evolution in Relation to Heart Development:

Co-evolution is the genetic change in one entity constraining a compensatory change in the other entity. The TBP<sub>CORE</sub> has been shown to have co-evolved with MBF1, a transcriptional co-activator, through compensatory point mutations<sup>509</sup>. Protein-protein co-evolution has been shown to use interfaces located at functionally important sites like linkers or active sites<sup>510</sup>. Interestingly, the MSD has been thought as such a linker region<sup>391</sup>. It is of interest to

note that the interaction hfTBP-mPitx2c was not detected while mTBP-hfPitxA interaction was present. While there are likely multiple explanations, this could be due to a more general PitxA binding interface. Indeed, only one *pitx* gene is known in cyclostomes<sup>240</sup> while three exist in mammals<sup>239,477</sup>. One may extrapolate that the increasing number of *pitx* genes in more recent vertebrates may be involved in more specialized mechanisms and potentially more selective protein-protein interactions, while the unique cyclostome gene fills in all those roles. Also, it is possible that the mechanisms requiring the other *pitx* genes found only in jawed vertebrates do not exist in cyclostomes. It is thus possible that TBP-N and PitxA/Pitx1/Pitx2c co-evolved and that the interaction became more refined in more recent vertebrates.

The cyclostome heart has only one atrium and one ventricle, while two of each are present in mouse<sup>176,511</sup>. The cyclostome heart also potentially requires a simple gene regulatory network to develop. The identification of an increased expression of *nppa* correlating with a lack of TBP-N, in the left heart, may have some downstream implications as minute alteration of heart development can result in congenital heart defects<sup>176</sup>, which are present in ~1% of live birth in humans<sup>186</sup>. Heart development and evolution is governed by gene regulatory networks and the often observed increase in genes redundancy, respectively<sup>50</sup>. The heart is formed of two distinct organs, the left heart, derived from the first heart field (FHF), and the right heart, derived from the second heart field (SHF), fusing together during development to give one organ<sup>176</sup>. Both heart fields share the same core transcription factors<sup>50,505</sup>. Interestingly, we showed that,

in  $tbp^{AN/AN}$ , Pitx2-dependent transcription of *nppa* is increased four fold in the FHF but little in the SHF.

Consequences for the  $tbp^{AN/AN}$  Mice: The present study focused on potential gene regulatory events that would be TBP-N-dependent. Future studies may measure the level of ANF, the *nppa*-encoded circulating hormone. One may speculate that the ANF circulating level will be increased like the mRNA, and correlate with physiological parameters unique to  $tbp^{AN/AN}$  animals.

Changes in the Basal Transcription Machinery: a Novel Mechanisms for Evolution?: Current evidence shows that alteration of one *cis*-regulatory element can influence gene expression of the one target locus and is, sometimes, involved in speciation<sup>447,448,451-453</sup>. Therefore, one change (i.e., mutation, duplication,...) will affect expression of one locus. Mutations at multiple loci may be required to achieve a multi-trait phenotype leading to an evolutionary transition like speciation. Conversely, change in the basal transcription machinery could affect all genes<sup>391</sup>. Alteration of TBP have been rare and most correlate with major evolutionary transition<sup>345,391</sup> (Fig. 3.1). It is therefore possible that alteration of the basal machinery in general, and TBP in particular, and their potential widespread effects could be another mechanism involved in evolutionary transitions.

## GENETIC MODIFICATION OF THE TBP N TERMINUS SUGGESTS IT HAS A VERTEBRATE-SPECIFIC FUNCTION

### Introduction

TBP plays a central role in transcription. This function is carried through its C terminal domain (TBP<sub>CORE</sub>) and has been extensively characterized<sup>346,348-350</sup>. All vertebrate lineages possess a homologous TBP N-terminal region (TBP-N) fused to TBP<sub>CORE</sub><sup>391</sup>. The vertebrate-specific TBP-N is highly conserved in all species studied to date (see chapter 3; Fig. 3.1B). Unlike the highly structured TBP<sub>CORE</sub>, the TBP-N lacks a defined structure or function<sup>383-387</sup>. Based largely on its conservation, the vertebrate TBP-N has been postulated to play a role in vertebrate-specific transcriptional regulation<sup>391</sup>. Homozygous mice lacking all but the first 24 amino acids of the TBP-N protein coding sequence usually do not survive past mid-gestation (embryonic day 14.5; E14.5) due to a placental defect<sup>351</sup>. In addition to its role in placental regulation, which is restricted to mammals, the sequence conservation suggests that some of the biological functions of TBP-N may be conserved and comparable among all vertebrate species. TBP-N is composed of a vertebrate-specific domain (VSD) and a metazoan-specific domain (MSD) (see chapters 2 and 3; Fig. 3.1). We previously showed (see chapter 3) that TBP-N is involved in protein-protein interactions and that the interaction between the TBP-MSD and Pitx proteins, a multigene family of transcription factors involved in patterning and endocrine functions, is conserved in chordates. This supports the hypothesis that one of the functional roles of TBP-N is to modulate gene transcription.

The following study aims to further characterize the biological role of TBP-N during mouse development.

One of the hypotheses that we have been exploring is that if TBP-N possesses a conserved vertebrate-specific function *in vivo*, expression of the most different vertebrate TBP-N protein-coding sequence may rescue the fetal loss observed in *tbp* <sup>$\Delta N/\Delta N$</sup>  animals<sup>351</sup> (see chapter 2). The vertebrate TBP-N protein-coding sequence that is most different from that in mouse is from hagfish (hf) (see chapter 3; Fig. 3.1B)<sup>391</sup>. Hagfish, a cyclostome, is a living representative of a lineage that has the oldest common ancestor with mammals (see chapter 3; Fig. 3.1). Hagfishes and mice diverged from a common ancestor ~500 Myr ago<sup>512</sup>. We hypothesized that if the placental roles of the mouse TBP-N were, indeed, functions shared by all vertebrates and not specific to placental mammals, the hfTBP-N could allow mouse embryonic development, in contrast to the fetal loss observed with the lack of TBP N terminus, in *tbp* <sup>$\Delta N/\Delta N$</sup>  animals. To test this hypothesis, the endogenous mouse TBP-N was replaced with the hagfish TBP-N through homologous recombination (Fig. 4.1, 4.2).

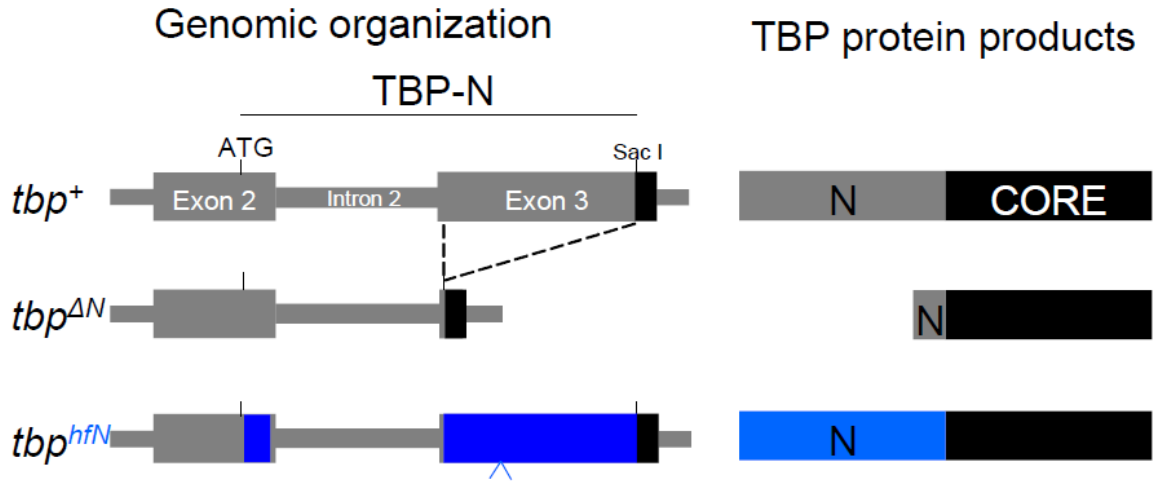


Figure 4.1: Schematic representation of the *tbp* alleles. Representation the *tbp* genomic loci of TBP-N encoded on exon 2 and 3. Grey represents DNA sequences derived from the mouse TBP-N and black represents TBP<sub>CORE</sub> sequence. The wildtype allele is displayed on the top diagram with TBP-N being encoded between the ATG and *Sac* I nucleotide regions, both represented as vertical bars. The second diagram represents the gene structure for the *tbp*<sup>ΔN</sup> mutant allele where the N-terminal coding sequence of exon 3 has been deleted and replaced with two FLAG epitope tags (not shown)<sup>351</sup>. The last schematic represents the constructs engineered in this study where the TBP-N sequence was replaced with the hagfish TBP-N specific sequence (in blue). The arrowheads denote species-specific introns that have been removed. Mouse intron 2 has been inserted in this construct at the nucleotide position where another naturally occurring intron is located in the hagfish *tbp* locus. The mouse splice donor and acceptor sites on each side of intron 2 have been retained to insure correct splicing in mice.

## Materials and Methods

### Hagfish Targeting Vector

The protein coding sequence of mouse TBP-N was replaced with that of the hagfish TBP. Hagfish exon 3 exhibited an additional intron<sup>391</sup> that was removed using a hagfish cDNA template for PCR amplification as follows: Silent mutations were introduced to generate a *Nar* I restriction site at the 5' end of hagfish exon 3. A *Sac* I site was engineered at the N/C junction in hagfish exon 3 (Fig. 4.2D). These restriction sites were used to insert the hagfish N terminus into the mouse *tbp* gene. The protein-coding

sequence of the mouse N terminus was perfectly replaced with sequences encoding the homologous region from hagfish (except a Val to Lys mutation at amino-acid twenty).

Care was taken to conserve the mouse intron 2 and mouse splice donor and acceptor sites to ensure splicing *in vivo* (Fig. 4.2). To this end, the restriction enzyme independent method of Tvrdik et al.<sup>513</sup>, which was generously communicated to us prior publication, was used (Fig. 4.2B and Table 4.1, hfTV2-7). Briefly, overlapping pairs of PCR products are generated, mixed, boiled, and re-annealed. A portion of the strands anneals into hetero-hybrids having unique complementary ends. This strategy was used to precisely insert hagfish TBP N terminus-coding exons into the mouse *tbp* gene while precisely retaining mouse intron 2 in the *tbp*<sup>hfN</sup> allele, respectively (Fig. 4.2). All coding sequences and junctions required for the final assembly of the targeting vector were sequenced.

The targeting vector was linearized using the restriction enzyme *Not* I, and resulting DNA was electroporated into embryonic stem cells. Colonies growing in media containing G418 were analyzed for targeted gene replacement by PCR (Fig. 4.2C).

A Southern blot strategy was designed to assess the correct integration of the construct in the genomic *tbp* locus. The upstream probe hybridized a 2765 bp fragment when wildtype genomic DNA was digested with *Spe* I and *Bam* HI. The downstream probe hybridized to a 4998 bp fragment when wildtype genomic DNA was digested with *Nde* I and *Bam* HI. Southern blot fragment sizes for the replaced allele are shown (Fig. 4.2A); 5023 bp for the upstream probe following digestion with *Spe* I, *Bam* HI, and *Xho* I and 3473 bp for the downstream probe following digestion with *Nde* I and *Bam* HI.

For selection, an expressed neomycin resistance cassette was inserted in the *Bam* HI site. The neomycin resistance was driven by a ubiquitously expressed synthetic promoter composed of the polyoma enhancer fused to the basal promoter of the herpes simplex virus *thymidine kinase* gene (MC1p)<sup>514</sup>. One correctly targeted embryonic stem cell clone was injected into blastocysts (strain C57/Bl6) and these were implanted in pseudo-pregnant surrogate mothers by standard procedures<sup>351</sup>. Weaned pups were analyzed for targeted genomic integration by PCR reaction (Fig. 4.2C and accompanying figure legend).

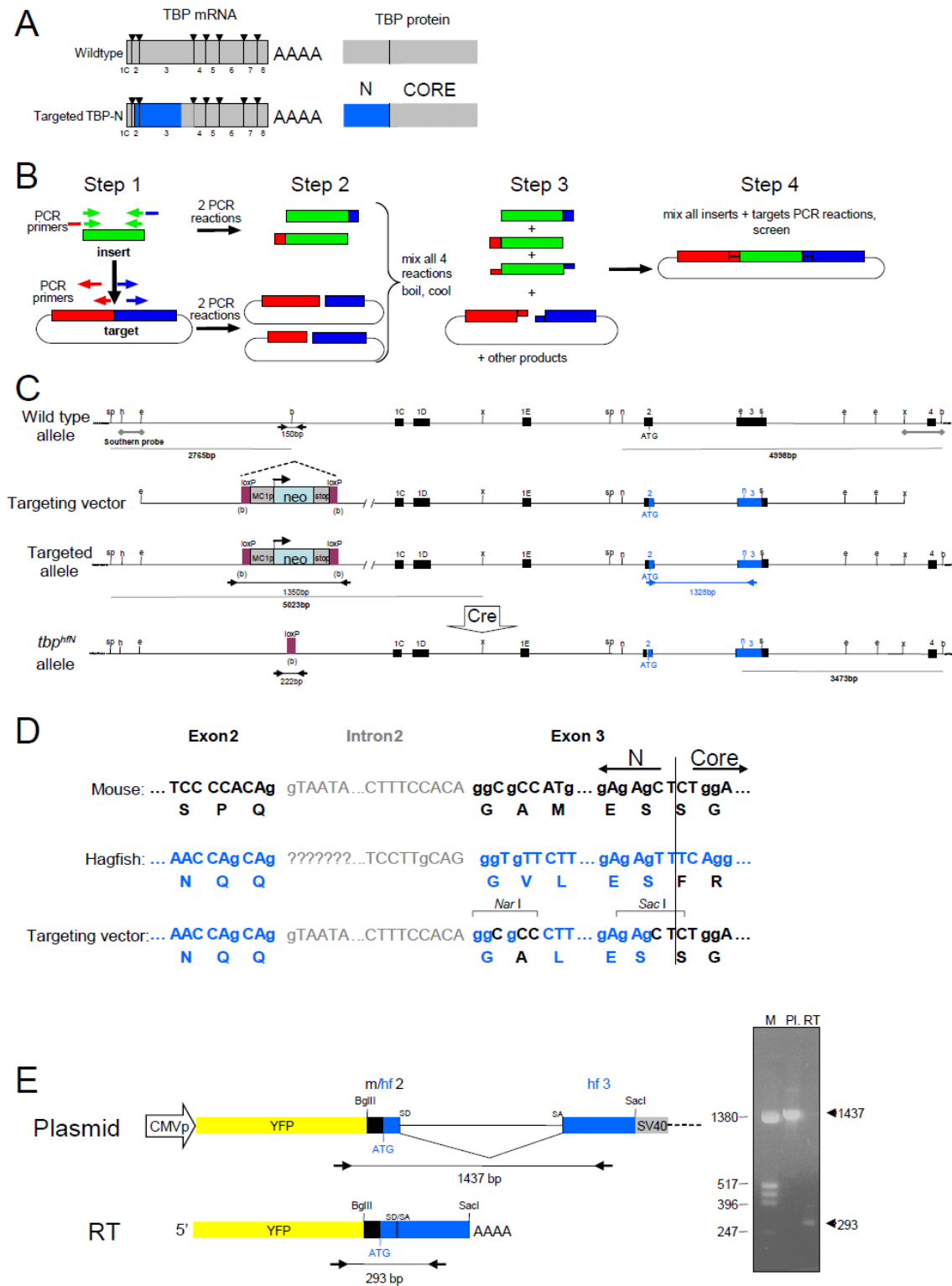


Figure 4.2: Outline of the design strategy that was used to produce the *tbp*<sup>hN</sup> allele. **(A)** Wildtype and mutant TBP mRNAs and TBP proteins. The wildtype TBP mRNA is represented in grey. The

Figure 4.2 - continued

blue in the replaced allele represents the hagfish protein coding sequence. Translation initiates in the second exon (see figure 4.1A, “ATG”) and ends in exon 8 generating a 316 amino-acids protein. The m/hf mutant protein is 324 amino-acids long. **(B)** Strategy for insertion of a desired sequence into a target site lacking restriction enzyme sites. The four steps to insert a desired DNA sequence (in green) in a restriction-independent manner in a target sequence (red and blue) are shown. For clarity, only the desired target is represented in step 3. **(C)** Schematic of the gene targeting strategy and targeting vector design. The 5' end of the mouse wildtype *tbp* gene is diagrammed (fine black line, wildtype allele) with exons represented as black boxes. Selected restriction sites are indicated as: b: *Bam* HI; (b): *Bam* HI/*Bgl* II; e: *Eco* RI; h: *Hind* III; n: *Nde* I; s: *Sac* I; sp: *Spe* I; and x: *Xho* I. Diagnostic PCR primers (arrows) and PCR products are indicated (line between arrows, length indicated). Diagnostic Southern blot probes are drawn in grey and expected fragment and sizes for wildtype allele are shown in grey. Below is shown the targeted vector design. The second schematic from the last shows the targeted allele still containing the *loxP* flanked MC1p neomycin-expressed cassette with the diagnostic PCR from the primers shown above on the wildtype allele. The bottom schematic represents the targeted allele after Cre-dependent removal of the *loxP*-flanked Neo cassette. **(D)** Conservation of the murine splicing sites of intron 2. The hybrid m/hf exon 2 was generated by synthetic oligonucleotides complementary to PCR primer tails specific for the murine sequence upstream of the transcription start site and for the murine intron 2. Mouse-derived exonic nucleotides are shown in black, hagfish-derived exonic nucleotides are in blue, and intronic nucleotides (mouse or hagfish) are grey. The 5' end of hagfish intron 2 has not been sequenced and therefore sequences are denoted as question marks. **(E)** Conservation of the murine splicing sites in the targeted *tbp*<sup>hfN</sup> exon2/exon3 junction by splicing assay in h293 cells. The hybrid region for the *tbp*<sup>hfN</sup> targeting vector was inserted into a yellow fluorescent protein (YFP) expression vector under the control of the cytomegalovirus promoter (CMVp) as depicted. The blue rectangles represent the hagfish exonic sequence. The black squares represent the mouse 5' untranslated region (UTR) exon 2 sequences upstream of the ATG. An oligo(dT)-primed RTase reaction followed by a PCR reaction with the primers indicated on the figure generated a 293 bp fragment after correct splicing of intron 2 from the chimeric minigene (lane RT). As a control, plasmid was used as a template in a PCR reaction yielding a 1437 bp (unspliced) PCR product (lane Pl), which would mimic the expected unspliced pre-mRNA. Molecular size markers are shown in lane M.

### Splicing Analysis

Assessment of the construct splicing pattern was conducted as follows: The targeting vector was digested with *Bgl* II and *Sac* I restriction enzymes to generate a 1437 bp fragment encompassing the coding region of exon 2 and the TBP-N region encoded in exons 2 and 3 (Fig. 4.2E). The resulting fragment was inserted into *Bgl* II/*Sac* I-cut pYFP-C1 plasmids (Clontech; eYFP: enhanced yellow fluorescent protein; CMVp: Cytomegalovirus promoter shown as open arrow in Fig. 4.2E). Human embryonic kidney

carcinoma cells (h293) at 50% confluence were transfected with 3 µg of the resulting plasmid. Total RNA was harvested 21 h later as described previously<sup>357,515</sup>. Random-primed (9-mer) and oligo-(dT)-primed reverse transcriptase (RTase) reactions were performed using 2.5 µg of total RNA. Half a percent of the RT reaction was used in PCR reaction with the primers indicated on Fig. 4.2E. The PCR reactions spanning the splicing region generated a 293 bp fragment following correct splicing of intron 2 from the chimeric minigene. As a control, 1 ng of untranslated plasmid was used as template in a PCR reaction yielding a PCR product of the size expected for unspliced pre-mRNA (1437 bp).

### Genotyping

Fetuses were genotyped as previously described<sup>351</sup>. Tails from weaned pups were collected and DNA was extracted as previously described<sup>351</sup>. Specific primers used for DNA amplification are listed in Table 4.1. For the presence of the *tbp*<sup>hfN</sup> allele, primers G2 (reverse primer; mouse intron 2/exon 3), G3 (forward primer; mouse intron 2), and G4 (reverse primer; hagfish exon 3) were used. A 0.7kb band was indicative of the *tbp*<sup>hfN</sup> allele. A 0.4 kb band was indicative of the wildtype allele. Survival was determined at various developmental stages, as indicated in Table 4.2.

Table 4.1. Oligonucleotides primers used in the generation and characterization of the *tbp*<sup>hfN</sup> allele.

| Name  | Sequence                                        |
|-------|-------------------------------------------------|
| G1    | ATAGAGCTCTCTGCTGGGGTTGCAGGGGTC                  |
| G2    | ATACTCGAGGATCCGTAAGGCATCATTGGACTAAAG            |
| G3    | TATTTGCCTCTGCCTCCAGAGCTAG                       |
| G4    | ATAGAGCTCTCAGATGCTGGAGCAAC                      |
| hfTV1 | TATGGCGCCCTTACACCTGGGATTCAAATG                  |
| hfTV2 | CGGGAGGCCACTGGAGGATTCCATGATTCTCCCTTTCTTC        |
| hfTV3 | TCCATGATTCTCCCTTTCTTCAGTAC                      |
| hfTV4 | CAGGTAATAGAGCAAGAGGGAGGGGCA                     |
| hfTV5 | AGGGGCAGAGAGGGCTCATATC                          |
| hfTV6 | CCCTCTTGCTCTATTACCTGCTGGTTCGTCAATCCAGGAGAGAATCC |
| hfTV7 | ATCCTCCAGTGGCCTCCCGGGATTCTCTCCTGGATTGACGAACCAG  |

## Results

Using the strategies described above, a mouse line that expressed TBP protein containing a TBP-N domain originating from hagfish (hf) coding sequence was generated. No visible defect in intron 2 splicing of the TBP-N genomic locus was detected in the *tbp*<sup>hfN</sup> mutant construct (Fig. 4.2E). The chimeric construct was thus deemed adequate to be inserted into mouse stem cells.

Survival of the mutant animals was assessed as follows: Analysis of survival at weaning (21 days after birth, P21) indicated that *tbp*<sup>hfN/+</sup> and *tbp*<sup>hfN/hfN</sup> mice showed wildtype-like survival (Table 4.2), health, and fertility. The full survival compared advantageously to the ~3% survival observed in mice lacking most of the TBP-N domain (*tbp*<sup>ΔN/ΔN</sup>)<sup>351</sup>. Our data indicate that whereas removal of the mouse TBP-N<sup>351</sup> is severely detrimental to fetal survival, replacement of the mouse TBP-N with that of hagfish is not. Therefore, the functions of the TBP-N, whose disruption causes loss of *tbp*<sup>ΔN/ΔN</sup> fetuses at

mid-gestation, appear to have been conserved during the evolution of these very distant vertebrate lineages. Thus, it is likely that this pre-existing process was co-opted for eutherian-specific functions and not developed specifically for this process.

Table 4.2. Effect of  $tbp^{hfN}$  allele on survival at weaning.

| Genotype             | Age (in days)            |
|----------------------|--------------------------|
|                      | P21                      |
| $tbp^{hfN/+}$        | <b>100%</b><br>n=119     |
| $tbp^{hfN/hfN}$      | <b>94% n.s.</b><br>n=119 |
| $tbp^{\Delta N/hfN}$ | <b>27% ***</b><br>n=243  |

Percentage of animals surviving to weaning based on Mendelian ratio of fertilized eggs and 100% survival for wildtype animals. Under each is indicated the total number of animals (n=) genotyped by PCR reaction (Fig. 4.2A, 4.3E). Chi-square tests were performed to determine if ratios were different from Mendelian ratios (n.s., not significant, \*\*\*  $p < 0.001$ ).

As a more sensitive test to determine if  $tbp^{hfN}$  was truly equivalent to the wildtype allele, crosses of  $tbp^{hfN/+}$  mice with  $tbp^{\Delta N/+}$  animals were performed. Our expectation was that if the  $tbp^{hfN}$  allele is equivalent to wildtype, the survival of  $tbp^{\Delta N/hfN}$  animals should be comparable to that of  $tbp^{\Delta N/+}$  animals. Therefore, we would expect ~70% survival of  $tbp^{\Delta N/hfN}$  at P21<sup>351</sup>. In contrast to this expectation, we observed a ~3 fold lower survival of  $tbp^{\Delta N/hfN}$  animals at P21, which was highly significant (27%; Table 4.2). The survival rate observed by introduction of the  $tbp^{hfN}$  allele in conjunction with a  $tbp^{\Delta N}$  allele was therefore intermediate to the one seen for  $tbp^{+/+}$  and  $tbp^{\Delta N/\Delta N}$  alleles. In  $tbp^{\Delta N/hfN}$  animals, one copy of the  $tbp^{hfN}$  allele did not rescue the effect of the  $tbp^{\Delta N}$  allele when comparing survival rates of  $tbp^{\Delta N/hfN}$  to  $tbp^{\Delta N/\Delta N}$  litters. This observation has allowed us to rule out

the hypothesis that the  $tbp^{hfN}$  allele restored wildtype-like survival. This suggests more subtle specializations have been imparted on TBP-N since hagfish and mouse diverged from their common ancestor ~500 Myr ago<sup>512</sup>.

### Discussion

In this chapter we have presented data that support the notion that the TBP-N participates in a fundamental function likely shared by all vertebrates. Given the homology and similarity between the hagfish and mouse TBP-N<sup>391</sup>, it was hypothesized that TBP proteins from all vertebrates have retained some common functions<sup>391</sup>. To test for such a conserved function, we swapped the mouse TBP-N protein-coding sequence with the hfTBP-N one. The full survival of  $tbp^{hfN/hfN}$  animals (Table 4.2) argues in favor of a conserved role of TBP-N common to all vertebrates. Moreover, this suggests that its functions are present in the primary amino-acids sequence already present in hagfish, a living representative that diverged from the main vertebrate lineage 500 Myrs ago<sup>36</sup>. Contrastingly, lack of TBP-N did not allow full embryonic development<sup>351</sup>. Thus, we can extrapolate that hfTBP-N likely performs a function not found in  $tbp^{AN/AN}$  animals where most (82%) of the TBP-N is absent. In addition, the partial complementation (as assessed by percentage of embryo survival) observed in  $tbp^{AN/hfN}$  (Table 4.2) supported the idea of a common fundamental role of TBP-N in vertebrates. Indeed, the  $tbp^{hfN}$  allele could partially rescue the survival (3% survival in  $tbp^{AN/AN}$ <sup>351</sup> compared to 27% in  $tbp^{AN/hfN}$ ). The partial complementation suggests that most of the defects due to the absence of TBP-N are fulfilled by the hfTBP-N but that more refined functions may have arisen after divergence of cyclostomes and jawed vertebrates. These embellishments were not

accounted for in the lineage leading to modern hagfish and could not be complemented by the *tbp*<sup>hfN</sup> allele. Therefore, this suggests that even though most defects are rescued by *tbp*<sup>hfN</sup>, one would expect some residual *tbp*<sup>hfN</sup>-specific defects.

Our study suggests that vertebrates evolved a highly conserved role for TBP-N. This new function is shared between the most diverged vertebrate lineages, those of cyclostomes and mammals. Our findings suggest that the placental defect observed in *tbp*<sup>4N</sup> animals is possibly the result of the co-option of an earlier system, as defined by use of a gene product predating a novel function<sup>32</sup>. Invertebrate chordates do not possess a TBP-N homologous to the vertebrate TBP-N. The most striking difference between protochordates (see chapter 1; Fig. 1.1) and vertebrates is that protochordates are filter feeders while most vertebrates are predators<sup>21</sup>. Different selection pressures may have resulted in the need for invertebrate chordates TBP-N to function differently than vertebrate TBP-N, perhaps modulating distinct protein-protein interactions facilitating the vertebrate radiation.

TBP-N appears to be switch-out at different evolutionary transition (see chapter 3; Fig. 3.1 and<sup>345</sup>). Our functional data presented here argues in favor of a conserved function that can be tracked back to the base of the vertebrate sub-phylum. Such a system is reminiscent of a “frozen accident”<sup>516</sup>. In the context of the frozen accident theory, a trait (i.e. TBP-N), which might not have any special properties at first, is fixed in a common ancestor, and the subsequent different lineages where any alteration of that trait becomes deleterious, like in the case of the genetic code<sup>516</sup>. Co-evolution of a vertebrate-specific protein interacting with the N-terminal region of TBP involved in embryonic

development is, in the frozen accident context, one possible hypothesis to explain the data presented above. Such interactions could be, perhaps, mediating novel cell signaling pathways through control of gene-specific transcription, for example. We realize that such possibility is one among many that could explain why absence of TBP-N is so deleterious. Therefore, the appearance and retention of TBP-N in vertebrates may suggest it has a key and conserved role in vertebrate's biology.

As the swapping and partial complementation assays showed a likely conserved role of TBP-N in vertebrates, we sought to identify the underlying molecular mechanisms common to *tbp* <sup>$\Delta N/\Delta N$</sup>  and *tbp* <sup>$hfN/hfN$</sup> . For this, we next aimed to identify a common set of misregulated pathways via global gene expression analysis.

## GENE EXPRESSION AND PHYSIOLOGY OF THE “ΔN” AND “HFN” ALLELES

### Background and Summary

In metazoans, TBP is a modular transcription factor composed of a conserved domain (TBP<sub>CORE</sub>) and a more variable N-terminal region<sup>345</sup>. The function of TBP<sub>CORE</sub> in transcription initiation has been well studied<sup>358</sup> and is conserved among all eukaryotes<sup>345</sup>. Within the N-terminal region and adjacent to the TBP<sub>CORE</sub>, all metazoans share a related metazoan-specific domain (MSD), which is not found in other eukaryotes<sup>391</sup>. Non-homologous N-terminal regions (TBP-N) are found in other metazoan lineages (see chapter 3; Fig. 3.2) with no evidence of homology between subphyla<sup>391</sup>. The function of these TBP-N remains largely uncertain<sup>391</sup>. Within the vertebrate subphyla, all species share a homologous subphyla-specific domain, hereafter termed the vertebrate specific domain (VSD, see chapter 3; Fig. 3.2).

In order to gain insights into the function of the TBP-N, transcriptome analyses were performed on livers from wildtype (*tbp*<sup>+/+</sup>), *tbp*<sup>ΔN/ΔN</sup>, and *tbp*<sup>hfN/hfN</sup> mice (see<sup>351</sup> and chapter 4) as well as for mice bearing combinations of the three different alleles.

Liver was chosen as a target tissue for its cell type homogeneity (hepatocytes make up more than 90% of the liver mass<sup>517</sup>), its active metabolic role, and its large size. The liver is an organ central to energy homeostasis through carbohydrates, proteins, and lipids metabolism. Animals tend to preferentially burn carbohydrates to generate energy when requiring fast replenishment of energy supplies, and surplus is frequently stored as triglycerides (a combination of three fatty acids esterified to a molecule of glycerol)

<sup>245,518</sup>. Most dietary lipids are long-chain triglycerides that are absorbed by the gut and transported, via the blood stream, to different tissues, including the liver, where they are metabolized <sup>245,518</sup>. The liver is also the major organ producing and clearing lipoproteins, the soluble vehicle for plasma lipid transport. This briefly illustrates the central role of the liver in energy metabolism.

The results presented in this chapter showed that liver mRNA abundance induced by the different *tbp* alleles revealed alterations in energy metabolism-related pathways. The main gene expression differences, based on functional grouping of differential mRNA abundance, were as follows: (1) *tbp* <sup>$\Delta N/\Delta N$</sup>  had an altered abundance of endogenous retroviruses mRNAs; (2) *tbp* <sup>$h^{fN}/h^{fN}$</sup>  was not, overall, an intermediate phenotype between *tbp* <sup>$+/+$</sup>  and *tbp* <sup>$\Delta N/\Delta N$</sup> , but rather revealed specific alterations in mRNA abundance; (3) *tbp* <sup>$\Delta N/\Delta N$</sup>  and *tbp* <sup>$h^{fN}/h^{fN}$</sup>  exhibited a large set of differentially abundant mRNAs in common that were involved in lipid-metabolism. Attention was further focused on *tbp* <sup>$\Delta N/\Delta N$</sup> , where histological analysis and serum analytical chemistry indicated a defect in lipid metabolism. Also, *tbp* <sup>$\Delta N/\Delta N$</sup>  has an impaired insulin response. Finally, juveniles and adults *tbp* <sup>$\Delta N/\Delta N$</sup>  mice fed a high fat diet, presented an increased accumulation of lipids in the liver.

## Materials and Methods

### Microarrays and RT-PCR

Male mice aged of 2 to 8 months were singly-housed at the Montana State University Animal Resources Center for 10 days before liver RNA harvest, at 2 pm. Unless otherwise noted, animals were fed *ad libitum* with a diet containing 9% of fat by

weight. For pre-weaning animals, pups of each genotype aged 11 days were sacrificed. Three animals of each genotype were harvested and RNA was extracted and purified. Biotinylated complementary RNA (cRNA) probe synthesis and hybridization to Affymetrix chips (MOEA\_2.0) were performed, as previously described<sup>391</sup>. Reverse transcription was performed as previously described<sup>351,483</sup>.

#### Glycemia, Insulin Tolerance Tests, and Blood Analytical Chemistry

For glycemia determination, following overnight (13h) fasting, blood was collected from the saphenous vein and glucose levels were determined using a TrueTrack glucometer (CVS pharmacy). Insulin tolerance tests were performed following overnight fasting by administration of bovine insulin (0.75U/kg, intraperitoneally; Sigma) and blood glucose was measured at specific time points following insulin injection. For blood chemistry, animals (n=3-4 per genotype) were fasted overnight before blood collection. Resulting serum samples were sent for blood chemistry analysis at the Marshfield clinic (Marshfield, WI).

#### Histology

For preparation of paraffin sections, tissues were fixed in neutral buffered formalin (10%), dehydrated through an alcohol gradient, cleared with xylene, and embedded with paraffin. Microtome sections (5 µm) were cut, dried, and stained with hematoxylin and eosin (H&E)<sup>519</sup>.

For preparation of oil red-O stained samples, fresh biopsies were embedded in OCT (Optimal Cutting Temperature; Sakura Finetek, CA) embedding medium and frozen. A cryotome was used to cut 5 µm-thick sections that were fixed overnight in

neutral buffered formalin (10%). Slides were stained with oil red-O and counterstained with hematoxylin (E. Suvorova, personal communication).

### Bioinformatics

Computational analyses of microarray data were performed using the GeneSpring 7.2 software package (Agilent) and the DAVID webtool<sup>520</sup>. A 1.5 fold differential signal cut-off and a maximal t-test value ( $p < 0.05$ ) were defined as threshold for datasets that reliably represented differentially abundant mRNAs. Benjamini and Hochberg post-hoc was used to identify patterns not expected prior to analysis. After identifying redundant probe sets, genes encoding differentially abundant mRNAs were further characterized. A probe set represents the collection of 11 different probes used to unequivocally identify a specific mRNA. (Some genes are represented by multiple probe sets.) Genes were classified into functional categories based on the gene ontology classification ([www.geneontology.org](http://www.geneontology.org)). Identification of potential transcription factor binding sites on genomic sequence was performed using the rVISTA<sup>521</sup> and DiRe<sup>522</sup> algorithms.

## Liver Transcriptome Analysis

### Introduction

Transcriptome analyses performed in this chapter revealed defects in fatty acids metabolism and in regulation of endogenous retroviruses. Fatty acids provide a source of energy that needs to be absorbed from the diet, transported, and oxidized before use<sup>518,523,524</sup>. Fatty acids can also be synthesized de novo<sup>518</sup>. In the liver, fatty acids are metabolized mostly through  $\beta$ -oxidation in mitochondria. Partial  $\beta$ -oxidation leads to

production of ketone bodies (a condensation of acetyl-CoA molecules), which are exported as energy source for other tissues. Full  $\beta$ -oxidation occurs when acetyl-CoA molecules enter the tricarboxylic acid cycle for in site energy production, releasing CO<sub>2</sub> and H<sub>2</sub>O as final products<sup>518</sup>.

About 10% of the mouse genome is made of endogenous retroviruses (ERV) sequences<sup>445</sup>. An increased number of ERV correlates with the appearance of placentation<sup>525</sup>. In their normal infectious cycle, some ERV can mediate the formation of syncytium<sup>526,527</sup>. HERV-W, a protein family expressed in human placenta includes the highly fusogenic glycoprotein, syncitin<sup>526,528</sup>. HERV-W has been shown to induce syncytium formation in syncytiotrophoblasts through interaction with the sodium-dependent amino acid transporter (SLC1A5) in human cell lines<sup>526-528</sup>. In addition, many ERV protein-coding sequences were reported as being under continued selective pressure, suggesting they maintain important functional roles<sup>527</sup>. Recently, it was shown that knocking down expression of a sheep-specific ERV through the use of morpholino anti-sense oligonucleotides could inhibit placental growth and differentiation<sup>529</sup>. ERVs have the capacity to evade immune surveillance and it has therefore been postulated that the non-rejection of the embryo could be induced by activation of ERV<sup>525</sup>.

## Results

In order to gain a better understanding of the underlying molecular perturbations imparted by the lack of the native TBP-N in *tbp*<sup>ΔN/ΔN</sup> mice or its replacement by the homologous hfTBP-N in *tbp*<sup>hfN/hfN</sup> mice, whole transcriptome analyses were performed on liver RNA samples from three adult homozygous male mice of each genotype. A total of

209 probe sets, representing 184 different mRNAs exhibited differential hybridization signals (appendix D). Of these 184 mRNAs, 127 (69%) genes were more abundant in *tbp* <sup>$\Delta N/\Delta N$</sup>  (Fig. 5.1A, B, and appendix D). Similar analysis from *tbp*<sup>*hfN/hfN*</sup> animals revealed that 71 mRNAs were differentially abundant, including 52 (73%) that were more abundant. As a qualitative verification of the data obtained from the microarrays experiment, RT-PCR analyses were performed on 19 mRNAs (Fig. 5.1D and data not shown). Genes encoding those mRNAs were classified into categories using the publicly available gene ontology classifications<sup>530</sup> (Fig. 5.1B; see Materials and Methods).

In *tbp* <sup>$\Delta N/\Delta N$</sup>  animals, 11 mRNAs were classified under the “immunity” category as defined by gene ontology classification, including 8 (72%) mRNAs that were more abundant in mutants. The list included mRNAs from two potentially interesting genes: *ly6a*, which is involved in CD4+ T cell proliferation and cytokine production<sup>531</sup>; and *pstpip2*, which encodes a signaling protein that regulates macrophage motility<sup>532</sup>. Histological examination of *tbp*<sup>+/+</sup>, *tbp* <sup>$\Delta N/\Delta N$</sup> , and *tbp*<sup>*hfN/hfN*</sup> livers (n=3) revealed infiltration of lymphocytes and some rare neutrophils in the liver of *tbp* <sup>$\Delta N/\Delta N$</sup>  (Fig. 5.1C; A. Harmsen, personal communication). Infiltration of immune cells in liver tissue of *tbp* <sup>$\Delta N/\Delta N$</sup>  animals could be one explanation for the increased abundance of “immunity”-related mRNAs in *tbp* <sup>$\Delta N/\Delta N$</sup> . Therefore, the differential mRNA abundance may be due to the presence of a different cell population, specific to *tbp* <sup>$\Delta N/\Delta N$</sup> , and not of an increased expression of “immunity” genes from the hepatocytes.

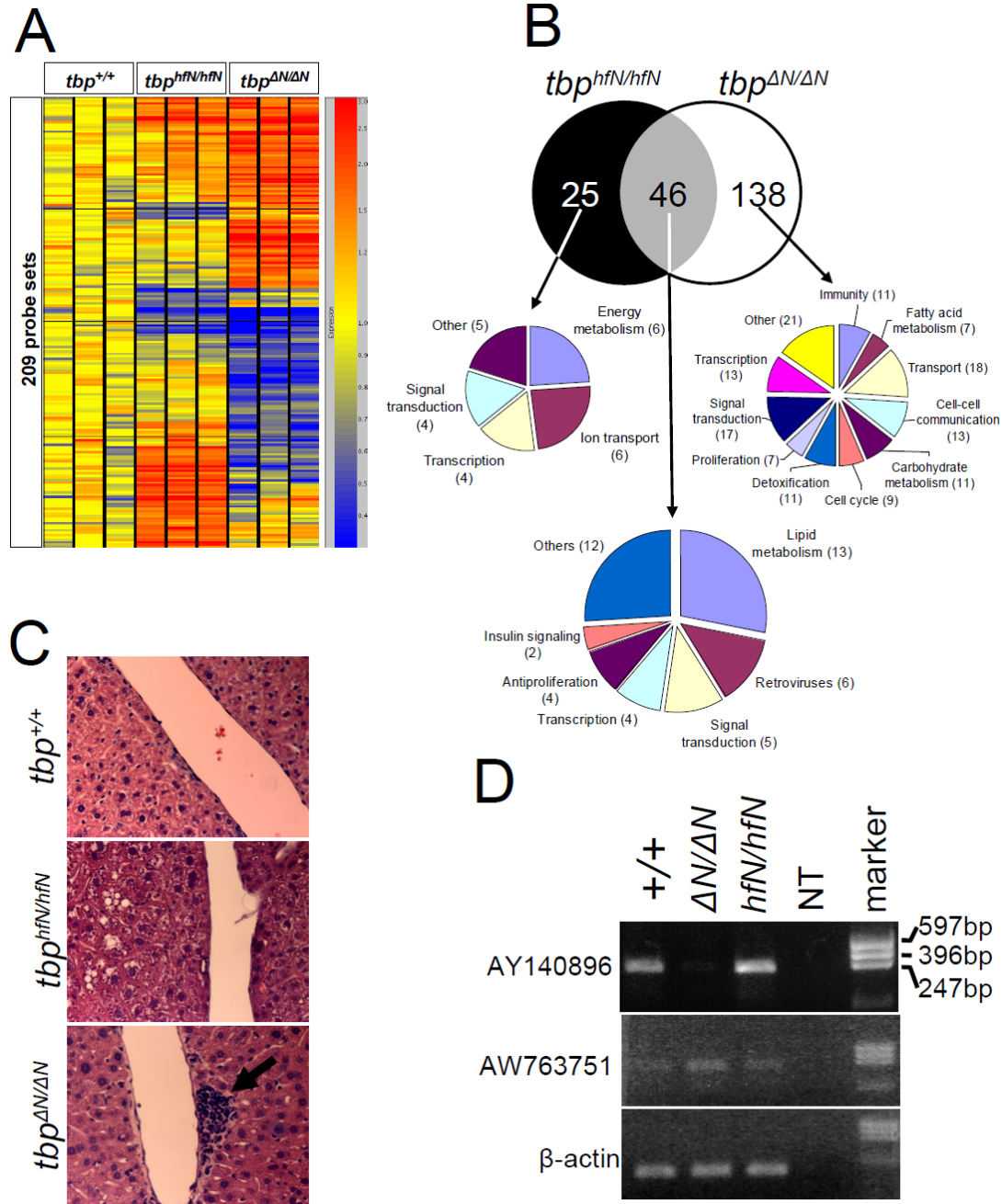


Figure 5.1: Transcriptome analyses of livers from  $tbp^{+/+}$ ,  $tbp^{\Delta N/\Delta N}$ , and  $tbp^{hfN/hfN}$  mice. **(A)** Liver RNA was harvested from the indicated genotypes and submitted to microarray hybridization followed by expression-based clustering. **(B)** mRNAs deemed significant, as defined in Materials and Methods, were classified based on their functions. Analysis of mRNAs shared in both lists (in grey) revealed a common set of 46 mRNAs that were differentially abundant in  $tbp^{\Delta N/\Delta N}$  and  $tbp^{hfN/hfN}$  livers. **(C)** Differential cell composition of  $tbp^{\Delta N/\Delta N}$  livers. Lymphocyte and neutrophil infiltration was observed in  $tbp^{\Delta N/\Delta N}$  animals. **(D)** Representative RT-PCR of two endogenous retroviruses qualitatively confirmed data obtained in the microarrays. A matched  $\beta$ -actin control was included. N.T.: no template.

To gain insights into the functions of TBP-N, the biological importance of mRNAs that were differentially abundant in both *tbp* <sup>$\Delta N/\Delta N$</sup>  and *tbp*<sup>*hfN/hfN*</sup>, animals was further examined. Interestingly, many of these mRNAs changed in abundance in response to one or the other mutation, but not to both (Fig. 5.1B). Analysis of both lists indicated that only 46 out of 106 probe sets were differentially abundant in both *tbp* <sup>$\Delta N/\Delta N$</sup>  and *tbp*<sup>*hfN/hfN*</sup> livers (Fig. 5.1B). Interestingly, 6 probe sets for mRNAs arising from ERV revealed differential mRNAs abundance between *tbp*<sup>*+/+*</sup>, *tbp* <sup>$\Delta N/\Delta N$</sup> , and *tbp*<sup>*hfN/hfN*</sup> livers (Table 5.1). Further analyses indicated that most differentially abundant ERV mRNAs were related to mouse leukemia viruses genes, which are vertebrate-specific ERV<sup>525</sup>. RT-PCR analysis of some of these representatives ERV confirmed the different mRNA levels observed in the microarray data (Fig. 5.1D). These observations led us to conclude that TBP-N may be involved in modulating expression of some endogenous retroviral genes.

Table 5.1. List of differentially abundant endogenous retroviral mRNAs.

| Genbank  | Relative expression level        |                                                      |                                      | Number of genomic hits            | Encoded protein | Annotation                      |
|----------|----------------------------------|------------------------------------------------------|--------------------------------------|-----------------------------------|-----------------|---------------------------------|
|          | <i>tbp</i> <sup><i>+/+</i></sup> | <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | <i>tbp</i> <sup><i>hfN/hfN</i></sup> |                                   |                 |                                 |
| AF070719 | 1.0                              | -26.3                                                | 1.8                                  | 53 <sup>a</sup> , 36 <sup>b</sup> | LTR             | Mxv2 endogenous retrovirus, LTR |
| AY140896 | 1.0                              | -3.7                                                 | 2.5                                  | 16 <sup>a</sup> , 69 <sup>b</sup> | Pol./env.       | Defective leukemia virus        |
| BB794742 | 1.0                              | -1.8                                                 | 2.3                                  | >10                               | Integrase       | Leukemia virus-related          |
| AW763751 | 1.0                              | 2.7                                                  | -2.0                                 | 745                               | Polymerase      | none                            |
| BI438039 | 1.0                              | 3.7                                                  | 1.0                                  | 9                                 | Envelope        | Leukemia virus-related          |
| BC002257 | 1.0                              | 4.3                                                  | 2.4                                  | 923                               | Envelope        | Leukemia virus-related          |

<sup>a</sup>: highly conserved (BLASTN>80)

<sup>b</sup>: somewhat conserved (BLASTN<80)

In the group of mRNAs that were differentially abundant and common to *tbp* <sup>$\Delta N/\Delta N$</sup>  and *tbp*<sup>*hfN/hfN*</sup>, many encoded lipid metabolism enzymes (n=13, 28% of the total). Most of

these genes (85% or 11/13) were related to fatty acids metabolism. To further explore the roles of TBP-N in fatty acids metabolism, the abundance of diagnostic marker mRNAs for the major relevant pathways were extracted from the above-mentioned microarray data. Results indicated that uptake and transport of fatty acids may be one of the most affected processes in *tbp*<sup>ΔN/ΔN</sup> and *tbp*<sup>hfN/hfN</sup> animals (Table 5.2).

Table 5.2. Representative events and associated genes in fatty acid metabolism and their abundance in *tbp*<sup>ΔN/ΔN</sup> and *tbp*<sup>hfN/hfN</sup>.

| Event                                   | mRNA    | Genbank                 | Fold change compared to <i>tbp</i> <sup>+/+</sup> |                               |
|-----------------------------------------|---------|-------------------------|---------------------------------------------------|-------------------------------|
|                                         |         |                         | <i>tbp</i> <sup>ΔN/ΔN</sup>                       | <i>tbp</i> <sup>hfN/hfN</sup> |
| <b>Uptake</b>                           | CD36    | BB534670                | 3.1                                               | -1.9                          |
| <b>Transport</b>                        | ApoA-IV | AV027367                | 4.1                                               | 3.8                           |
|                                         | FABP2   | NM_007980               | 2.5                                               | n.s.                          |
| <b>Beta-oxidation</b>                   |         |                         |                                                   |                               |
| Oxidation (FAD-dependent)               | ACAD9   | AV235115                | 2.1                                               | n.s.                          |
| Hydration                               | EHHADH  | NM_023737               | -1.8                                              | -1.8                          |
| Oxidation (NAD <sup>+</sup> -dependent) | ACADM   | NP_031408               | n.s.                                              | n.s.                          |
| Thiolysis                               | ACAA2   | AI255831                | n.s.                                              | n.s.                          |
| <b>Synthesis/elongation</b>             | ELOVL5  | NM_134255               | 1.7                                               | 2.1                           |
| Triglycerides                           | DGAT1,2 | EDL29576,<br>Q9DCV3     | n.s.                                              | n.s.                          |
| <b>Storage</b>                          | FIT1,2  | NP_081084,<br>NP_775573 | n.s.                                              | n.s.                          |

n.s: fold change was less than |1.5|

### cd36 and Putative Regulatory Mechanisms

#### Introduction

The CD36 cell surface receptor is encoded by the *cd36* locus<sup>524</sup>. CD36 is a major lipid receptor involved in uptake of long-chain fatty acids (C>14)<sup>524,533-535</sup> and lipoproteins, following their hydrolysis by lipoprotein lipase (LPL)<sup>534-536</sup>, which is

encoded by the *lpl* locus. *cd36* is a vertebrate-specific gene, but a similar gene that lacks most of the N terminus region is present in *Ciona intestinalis* ([www.ensembl.org](http://www.ensembl.org) accession number: ENSCINP00000010908), a tunicate (see chapter 1, Fig. 1.1). The CD36 protein has been shown to have a complex role, influencing both glucose and lipid metabolism. Thus, both a CD36 deficiency and a defect in cellular lipid uptake correlate with insulin resistance<sup>524</sup> (see below).

## Results

CD36 mRNA displayed an interesting pattern in our analyses being more abundant in *tbp* <sup>$\Delta N/\Delta N$</sup> , but less abundant in *tbp*<sup>*hfN/hfN*</sup> livers, as compared to wildtype livers (Fig. 5.2B). Interestingly, *lpl* was not differentially abundant, suggesting that the lipid metabolism defect is potentially specific to the uptake of long chain fatty acids. Because of its connection to carbohydrate and lipid metabolisms, *cd36* could be a prime candidate gene whose expression may be dependent on the presence of native TBP-N.

In order to determine whether the differential abundance of CD36 mRNA observed in homozygous animals was a dominant effect of each mutation, microarray analyses were performed on livers from *tbp* <sup>$\Delta N/+$</sup>  and *tbp*<sup>*hfN/+*</sup> animals (Fig. 5.2). Of the 60 mRNAs that were differentially abundant between *tbp* <sup>$+/+$</sup>  and *tbp* <sup>$\Delta N/+$</sup>  livers, only 11 mRNAs were also different between *tbp* <sup>$\Delta N/\Delta N$</sup>  and *tbp* <sup>$+/+$</sup> , including *cd36*. Also, the only overlap between *tbp*<sup>*hfN/hfN*</sup> and *tbp*<sup>*hfN/+*</sup> was *cd36* (Fig. 5.2A, B). Microarrays and RT-PCR verifications indicated that *cd36* liver mRNA levels in *tbp* <sup>$\Delta N/+$</sup>  were intermediate between those in *tbp* <sup>$+/+$</sup>  and *tbp* <sup>$\Delta N/\Delta N$</sup>  livers (Fig. 5.2B). The mRNA abundance pattern observed between *tbp*<sup>*hfN/hfN*</sup> and *tbp*<sup>*hfN/+*</sup> was more complex (Fig. 5.2B). CD36 mRNA was

strikingly less abundant in  $tbp^{hfN/hfN}$  when compared to  $tbp^{+/+}$ ,  $tbp^{\Delta N/\Delta N}$ ,  $tbp^{\Delta N/+}$ , or  $tbp^{hfN/+}$ . Further RT-PCR analysis of  $tbp^{\Delta N/hfN}$  heterozygotes indicated that CD36 mRNA abundance was increased when compared to wildtype (Fig. 5.2B). Decreased abundance of CD36 mRNA observed in  $tbp^{hfN/hfN}$  was not observed in  $tbp^{\Delta N/hfN}$  suggesting a likely distinct role of the  $tbp^{hfN}$  allele compared to the  $tbp^{\Delta N}$  allele in *cd36* gene expression.

Taken together, these observations and possible correlations suggested that the  $tbp^{hfN}$  and the  $tbp^{\Delta N}$  alleles each have distinct effects on *cd36* gene expression. This might result from an indirect effect of the mutation. Indeed, one possibility would be that some uncharacterized endogenous sensor can detect abnormal blood levels of lipids and, in response, alter expression of the *cd36* gene. This response could be context and/or environment dependent, providing a rationale for why distinct mutations generate different *cd36* gene expression responses (Fig. 5.1A, B).

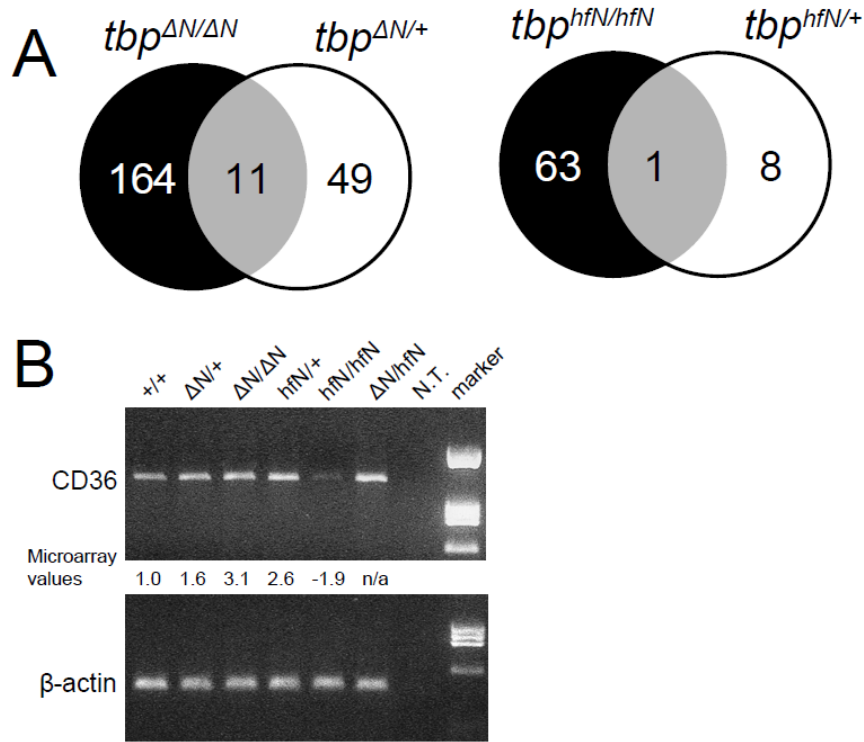


Figure 5.2: CD36 is the only mRNA whose abundance is affected by either removal of TBP-N or replacement by hfTBP-N. **(A)** Venn diagrams of the significant differentially abundant mRNAs in the indicated genotypes. The common mRNA to both intersections (i.e., to both lists depicted in grey) is CD36. **(B)** RT-PCR analysis of *cd36* mRNA abundance in the indicated *tbp* allelic combinations compared to the observed relative abundance values from the microarray analysis. A RT-PCR for  $\beta$ -actin was performed as a control. N.T.: no template.

We hypothesized that alterations of the TBP-N primary sequence could be involved in the regulation of the genes encoding these mRNAs through a common set of DNA-binding transcription factors. Bioinformatics-based examination of regulatory elements for the 184 above-mentioned mRNAs was performed with DiRE<sup>522</sup>. DiRE is an algorithm that identifies enrichment of putative proximal and distant regulatory elements over-represented in a user-defined set of genes when compared to 5,000 randomly picked genes (“background”). Unexpectedly, among the ten most represented binding sites found in genes encoding the mRNAs showing differential abundance in  $tbp^{\Delta N/\Delta N}$  was the binding site for Pitx2 (found in 10.57%, or 12/184, of the genes). No Pitx2 binding site

was identified in the genes exhibiting differential mRNA abundance in  $tbp^{hfN/hfN}$  or the intersection or the gene lists from  $tbp^{AN/AN}$  and  $tbp^{hfN/hfN}$ . Of note, all 12 genes showed an increased mRNA abundance in  $tbp^{AN/AN}$ .

Table 5.3. List of mRNAs differentially abundant in  $tbp^{AN/AN}$  versus  $tbp^{+/+}$  livers that are transcribed from genes having associated Pitx2 binding sites.

| Genbank accession number | Gene name                               | Fold change in $tbp^{AN/AN}$ | Location of Pitx2 binding sites |
|--------------------------|-----------------------------------------|------------------------------|---------------------------------|
| AF461145                 | Flavin-containing monooxygenase 4       | 1.9                          | Intergenic                      |
| NM_026405                | RAB32 (Ras family)                      | 1.9                          | Intergenic                      |
| AF071068                 | DOPA decarboxylase                      | 2.1                          | Intron                          |
| NM_138654                | Endogenous retrovirus, envelope         | 2.3                          | Intron                          |
| NM_134069                | Solute carrier 17, 3 (sodium phosphate) | 1.7                          | Intergenic                      |
| NM_010702                | Leukocyte cell-derived chemotaxin 2     | 3.0                          | Intergenic                      |
| BC021619                 | Methionine sulfoxide reductase B2       | 1.8                          | Intergenic                      |
| NM_020255                | Sca domain-containing 1                 | 2.2                          | Intergenic                      |
| AA717142                 | Makorin 1                               | 3.2                          | Promoter                        |
| AF061972                 | Homolog to HIV-1 TIP2                   | 2.3                          | 3'UTR                           |
| BB307136                 | IP phosphatase A                        | 2.1                          | Intron                          |
| NM_011734                | Sialic acid acetyltransferase           | 1.8                          | Intron                          |

### Lipid Accumulation and Insulin Response

#### Introduction

Lipids as an Energy Source: Fatty acids catabolism is a major source of energy for the heart and adipose tissue<sup>523,537</sup>. Adipose and hepatic tissues are intertwined in the regulation of this highly complex mechanism<sup>244</sup>. Most dietary lipids are long-chain triglycerides that are absorbed by gut enterocytes. Hydrolysis of triglycerides is mediated by liver-produced bile and the resulting fatty acids are packaged and released in the blood stream as chylomicrons. Chylomicrons are molecular complexes of triglycerides, cholesterol, phospholipids, and apolipoproteins A-IV, B, and E<sup>244,538</sup>. Chylomicrons are

then taken up by the liver, heart, muscle, and adipose tissue, where they are dissociated by lipoprotein lipases (LPL, see above)<sup>538</sup>.

The structural unit of the liver is the hepatic acinus, which is a collection of hepatocytes centered around a portal tract composed of a portal vein, an hepatic artery, and a bile ductile<sup>539</sup>. Blood flows from the portal tract to the hepatic veinule at the center of the acinus and a gradient of oxygen and metabolites forms three zones. Zone 1 is closest to the portal tract and receives the most oxygen, while zone 3, nearest to the veinule, is poorly oxygenated<sup>518</sup>. The liver is also the major organ producing and clearing lipoproteins, the soluble vehicle for plasma lipid transport. Energy homeostasis is a complex, highly intertwined process.

Fatty Liver Disease and Lipid Transport: Pathological accumulation of fat in the liver results in intracellular formation of lipid droplets, which may induce hepatic steatosis, mostly observed in cases of obesity<sup>540</sup>. Steatosis can be minor, a hallmark of fatty liver disease (FLD), but often leads to cirrhosis<sup>244,540,541</sup>. FLD is not a well understood disease. It can be manifested by microvesicular (multiple small lipid droplets in an hepatocyte) or macrovesicular (a unique lipid droplet displacing the nucleus to the periphery of the hepatocyte) forms<sup>540</sup>. It is most frequently associated with excessive alcohol or caloric consumption, imbalance of lipid uptake, oxidation, export, or synthesis, or altered signaling pathways regulating these events<sup>518</sup>. Increased CD36 mRNA expression has been correlated with FLD in response to long chain fatty acids levels<sup>524,542</sup>. FLD is the most common chronic liver condition, affecting almost a third of adults in the USA<sup>541</sup>. In addition, FLD is often associated with insulin resistance<sup>244,540</sup>.

Insulin Resistance: Another hallmark of FLD is insulin resistance <sup>244</sup>. Insulin resistance is the condition where a normal level of serum insulin fails to clear an increased glycemia. Often, this, in turn, induces an impaired glycogen synthesis in the liver and an increased hydrolysis of triglycerides in the adipocytes resulting in elevated blood glucose. Insulin resistance is defined by increased circulating insulin and a decreased systemic action of insulin <sup>543</sup>. The increased level of circulating insulin may lead to metabolic syndrome and type 2 diabetes <sup>244,543</sup>. Metabolic syndrome, a combination of medical symptoms usually associated with obesity, may affect 25% of American people <sup>544</sup>. Type 2 diabetes is now considered as an epidemic by the Center for Disease Control (<http://cdc.gov/diabetes/>). Symptoms of insulin resistance include low blood glucose, weight gain, elevated triglycerides levels, and increased blood pressure <sup>543</sup>. Injection of insulin and examination of the glycemia over a few hours is often performed to detect insulin tolerance (insulin tolerance test). A reduced intensity of the hypoglycemia and a shorter time needed to recover to baseline levels are preliminary indicators of insulin resistance <sup>524,545,546</sup>. Insulin modulates transcription of the *glucokinase* gene through sterol response element-binding protein 1c (SREBP1c) signaling. The glucokinase protein (GCK) then phosphorylates glucose to glucose-6-phosphate and is therefore a major step in the cellular commitment to glycolysis. Another factor involved in glycemic regulation is glucagon. The hormone glucagon is synthesized in the pancreas and induces hepatic glycogenolysis and glucose secretion, thereby increasing the serum glucose level <sup>547</sup>.

## Results

Histological assessment for lipid accumulation in livers was performed because differential regulation of genes involved in lipid transport, like *cd36* (Fig.5.1B), could correlate with accumulation of lipids in the liver (see below). Staining with oil-red-O (PubChem 5841742), a fat soluble dye specific for neutral triglycerides, lipids, and some lipoproteins, was performed. This revealed an increased hepatic fat accumulation in *tbp* <sup>$\Delta N/\Delta N$</sup>  compared to *tbp*<sup>+/+</sup> and, to a lower extent, *tbp*<sup>*hfN/hfN*</sup> mice were also affected (Fig. 5.3). This further suggested alteration of hepatic lipid metabolism and potentially that *tbp* <sup>$\Delta N/\Delta N$</sup>  may develop FLD.

To examine whether there was a defect in blood lipid homeostasis in *tbp* <sup>$\Delta N/\Delta N$</sup>  mice, serum lipid levels and composition were assessed through serum metabolomic analysis and blood analytical chemistry. Preliminary metabolome analysis did not indicate major differences in serum composition between *tbp*<sup>+/+</sup> and *tbp* <sup>$\Delta N/\Delta N$</sup>  animals. No major lipid enrichment was detected, indicating that circulating lipid levels may not be dramatically altered (data not shown). Analytical chemistry of serum from *tbp*<sup>+/+</sup> and *tbp* <sup>$\Delta N/\Delta N$</sup>  indicated comparable profiles on 15 out of 17 common blood markers (Fig. 5.4A, C). Only the fasting glycemia and the total cholesterol level of *tbp* <sup>$\Delta N/\Delta N$</sup>  animals were statistically different from *tbp*<sup>+/+</sup> animals ( $p < 0.01$ , Wilcoxon ranked test,  $n = 9-12$  and  $p < 0.02$ , unpaired t-test with Welch correction,  $n = 3-4$ , respectively).

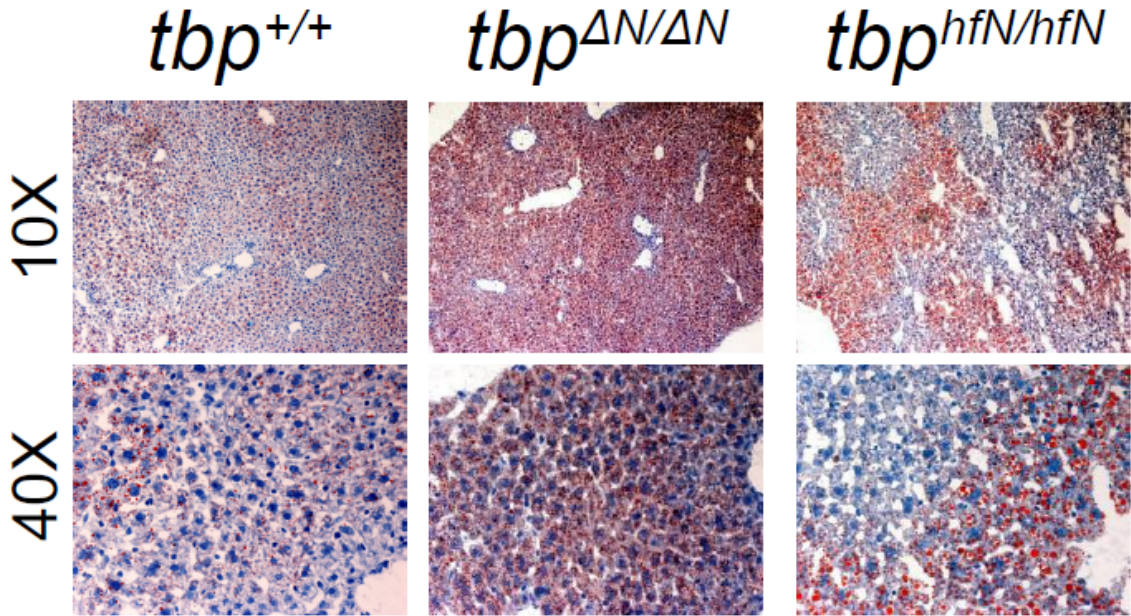


Figure 5.3: Differential fat accumulation in  $tbp^{+/+}$ ,  $tbp^{\Delta N/\Delta N}$ , and  $tbp^{hfN/hfN}$  livers. Liver sections of age matched adult males under standard care conditions were stained with oil red-O (red) and counterstained with hematoxylin (blue) for the indicated genotypes.

Blood chemistry analyses showed no difference in serum triglyceride levels between  $tbp^{+/+}$  and  $tbp^{\Delta N/\Delta N}$  animals (Fig. 5.4C). Also, glycemia of overnight-fasted animals was lower in  $tbp^{\Delta N/\Delta N}$  compared to  $tbp^{+/+}$  mice (Fig. 5.4A). Given the number of symptoms pointing towards an insulin associated defect, we examined glycemic regulation. Following overnight fasting, injection of insulin to  $tbp^{+/+}$ , and  $tbp^{\Delta N/\Delta N}$  indicated a normal immediate insulin response at 15 min and 30 min, but a highly significant difference between  $tbp^{+/+}$  and  $tbp^{\Delta N/\Delta N}$  starting at 90 min ( $p < 0.01$  and  $p < 0.001$ ; two-way ANOVA and Bonferroni posttest). The  $tbp^{\Delta N/\Delta N}$  animals recovered to near-baseline levels within an hour, while  $tbp^{+/+}$  animals had roughly 50% of their starting glycemia at the same time, as expected<sup>545</sup>. The comparable glucose levels between  $tbp^{+/+}$  and  $tbp^{\Delta N/\Delta N}$  at 15 min and 30 min indicated a likely normal immediate insulin-induced response in the mutants. However, the rapid increased of glycemia

observed at 60 min and the above-baseline levels observed at later time points was unexpected and, to our knowledge, does not compare to any of the defined patterns of insulin response defects: namely insulin resistance<sup>244,543</sup> or insulin sensitivity<sup>545,546</sup>. This suggested that the defect observed in  $tbp^{\Delta N/\Delta N}$  might be the result of a combination of different pathways (see discussion). GCK mRNA, a target gene of insulin, was 2.6-fold less abundant in  $tbp^{\Delta N/\Delta N}$  livers, suggesting a potential compensatory mechanism for the lower glycemia baseline (Fig. 5.4A).

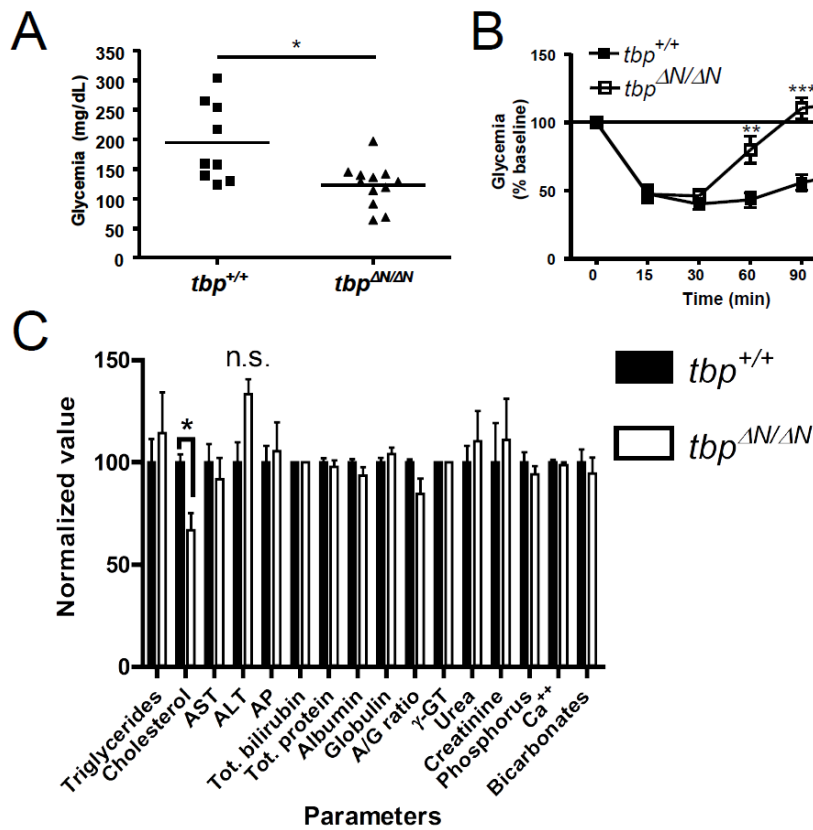


Figure 5.4: Glycemia, insulin tolerance tests, and blood analytical chemistry. **(A)** Fasting glycemia of  $tbp^{+/+}$  and  $tbp^{\Delta N/\Delta N}$  animals (n=9-12, \*p<0.01, Wilcoxon signed ranked test). Results are mean  $\pm$  SEM. **(B)** Insulin tolerance tests. Fasted animals were subjected to insulin intraperitoneal injection (0.75U/kg) and glycemia was measured at the indicated time (n=8-12, \*\*p<0.01, \*\*\*p<0.001, two-way ANOVA). Results displayed represent mean  $\pm$  SEM. **(C)** Blood chemistry. Blood was extracted from fasted animals and analyzed for the indicated parameters (n=3-4, \*p<0.05, unpaired t-test with Welsch correction, n.s., not significant). Results are mean  $\pm$  SEM.

Taken together, these data point to a possible role of TBP-N in regulation of some lipid metabolism-specific genes. Removal or replacement of TBP-N resulted in hepatic steatosis. Therefore, we hypothesized that a diet rich in fat should exacerbate those defects. To test this, we examined 1) the pups feeding on milk, a high fat diet (see below) and 2) the influence of high-fat and low-fat fed adults on *cd36* gene expression.

### Pre-weaning and Adult Phenotype Similarities

#### Introduction

As *tbp*<sup>ΔN/ΔN</sup> animals seem to show accumulation of lipids due to a defect in lipid transport and/or uptake, we decided to look at pre-weaning animals that rely on their mother's milk, which is rich in lipids (26-28% in mouse<sup>548</sup>). We reasoned that it was possible that pre-weaning *tbp*<sup>ΔN/ΔN</sup> animals would exhibit some lipid-related defect given the high level of lipids in their diet.

Wasting syndrome, or acute malnutrition, refers to a disease state where muscle and fats are consumed or “wasted”, often resulting in a runt phenotype<sup>538</sup>. At the gene expression level, hallmarks of the wasting syndrome are an up-regulation of *asparagine synthetase*, a classical starvation response factor used to detect unbalance of amino acids or glucose levels<sup>549</sup> and a down-regulation of *insulin-like growth factor 1*.

Cholesterol is a major intermediate compound in steroid biogenesis<sup>550</sup>. Cholesterol is the main sterol synthesized in mammals<sup>551</sup>. Cholesterol is essential for cell membranes and therefore growth, but is also required for bile acids synthesis. Bile acids are cholesterol derivatives involved in emulsifying dietary fat, thus helping lipid

absorption. The main sensor of sterol synthesis and uptake is the transcription factor sterol regulatory element binding protein (SREBP)<sup>551</sup>. Lack of sterol in the cell results in translocation of SREBP to the nucleus, where it binds co-activators, and the sterol-response element DNA recognition motif<sup>551</sup>. SREBP then turns on its target genes, especially *hmg-CoA reductase* and *ldl receptor*<sup>551</sup>. HMG-CoA reductase (hydroxymethylglutaryl coenzyme A reductase) catalyzes the rate-limiting step of cholesterol synthesis<sup>550</sup>.

## Results

First, comparison of pre-weaning body weight between *tbp*<sup>+/+</sup> and *tbp*<sup>ΔN/ΔN</sup> indicated a significant growth retardation around 11 days after birth (P11) (p<0.01; ANOVA and Bonferroni posttest, n=8-10, Fig. 5.5A). Interestingly, neither *tbp*<sup>hfN/+</sup> nor *tbp*<sup>hfN/hfN</sup> exhibited growth retardation at P11 (data not shown, n=3-7). Livers of *tbp*<sup>ΔN/ΔN</sup>, *tbp*<sup>ΔN/+</sup>, *tbp*<sup>hfN/hfN</sup>, and *tbp*<sup>hfN/+</sup> animals were histologically analyzed for lipid accumulation. All genotypes, except *tbp*<sup>+/+</sup>, exhibited a marked accumulation of hepatic lipids with very intriguing patterns visible in heterozygous animals (Fig. 5.5B).

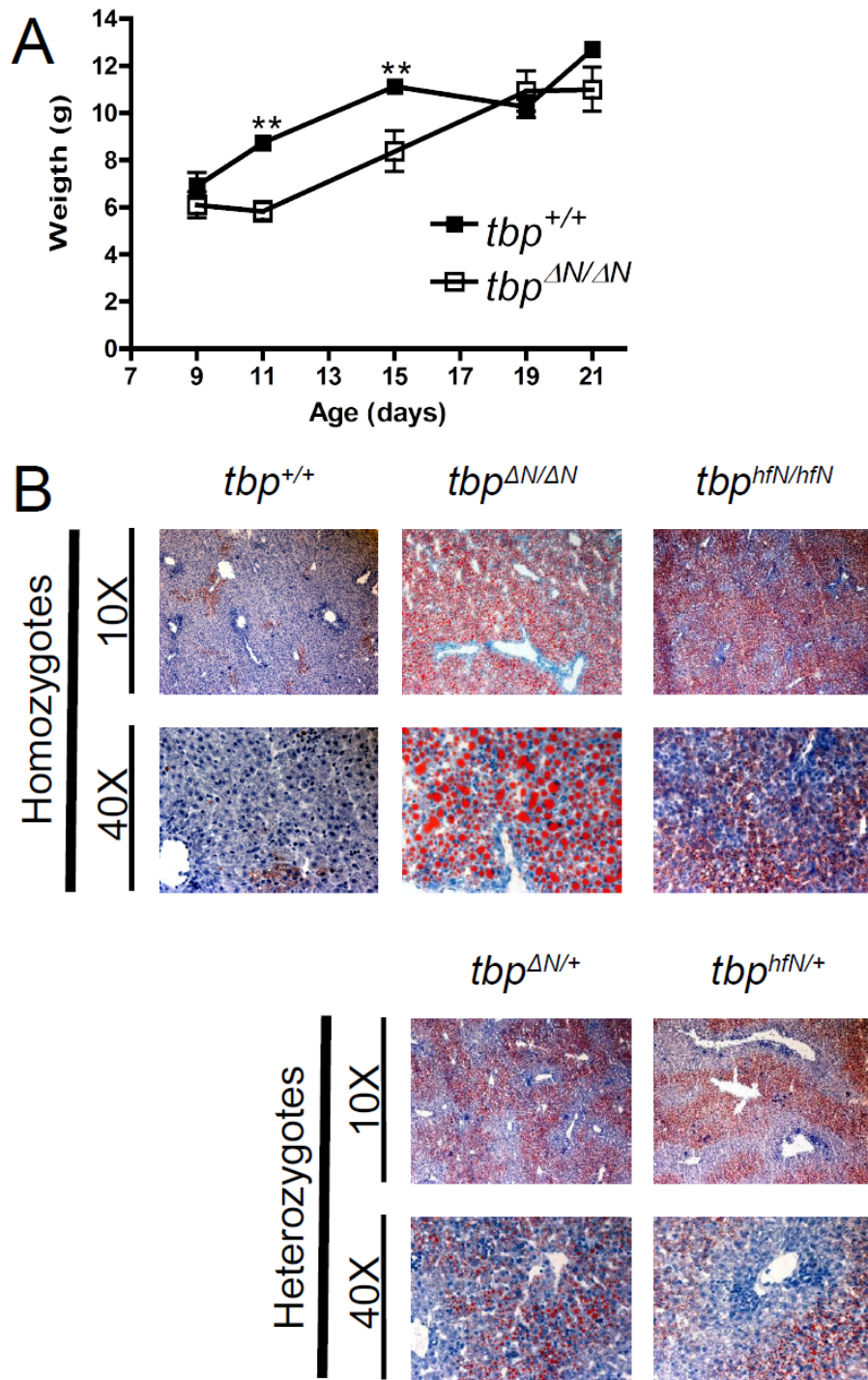


Figure 5.5. Growth retardation and steatosis of 11-day old animals. **(A)** Body weight of pre-weaning animals. The body weight of pups from multiple litters at the indicated age was recorded ( $n=8-10$  at 11 and 15 days old,  $**p<0.01$ , ANOVA and Bonferroni posttest). Results are mean  $\pm$  SEM. **(B)** Liver sections were stained with oil red-O (red) and counterstained with hematoxylin (blue) for the indicated genotypes.

As there are many candidate genes that may explain such a dramatic steatosis to be tested, microarrays analyses from P11 liver and kidney were performed. 237 probe sets, representing 228 unique genes, yielded significantly different hybridization levels in between *tbp* <sup>$\Delta N/\Delta N$</sup>  and wildtype, in both liver and kidney of P11 animals (see appendix E). None of the two mRNA markers of wasting syndrome (asparagine synthetase and insulin-like growth factor 1) were found to be differentially abundant in *tbp* <sup>$\Delta N/\Delta N$</sup>  pups when compared to wildtype animals. Moreover, stomachs of *tbp* <sup>$\Delta N/\Delta N$</sup>  pups were full of milk, which suggested that the growth defect was likely not due to malnutrition but maybe malabsorption of dietary lipids. Thus it is unlikely that P11 pups succumb to wasting syndrome.

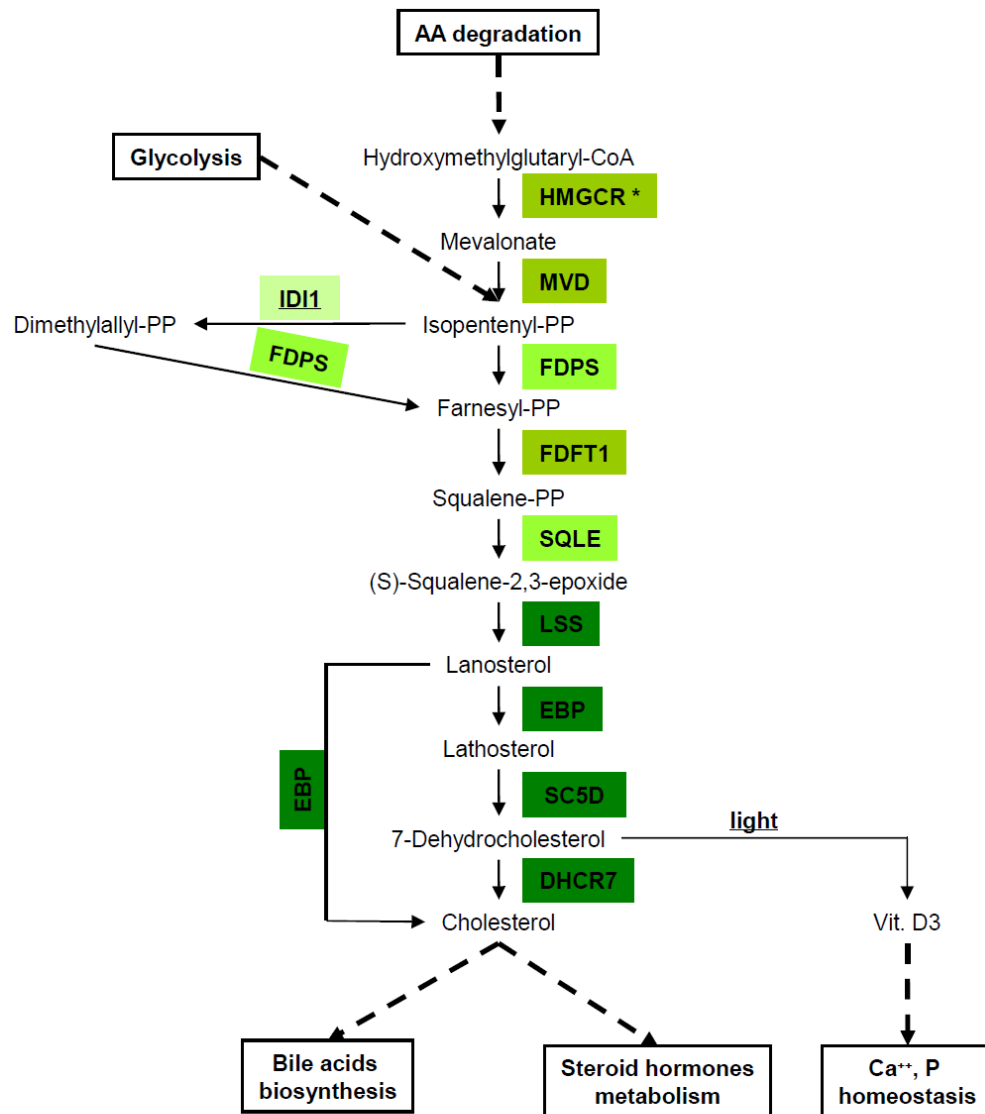
Attention was focused towards the 228 mRNAs identified as differentially abundant in *tbp* <sup>$\Delta N/\Delta N$</sup>  pups. Classification of the genes into functional categories, as previously described <sup>552</sup>, revealed a predominant group: lipid metabolism mRNAs. Markers of adipogenesis (*c/ebp $\alpha$* ), lipogenesis (*fatty acid synthase*), peroxisome  $\beta$ -oxidation (*cyp4a*, *ppar*), and insulin-dependent lipolysis (*perilipin*) were not affected, while uptake of lipoproteins through LPL might be altered as LPL mRNA was 2 fold less abundant in *tbp* <sup>$\Delta N/\Delta N$</sup> . Moreover, wildtype levels of apolipoprotein B mRNA and microsomal triglyceride transfer protein (MTTP) mRNA were observed. These two genes are used as markers of very low density lipoproteins (VLDL) production <sup>538</sup>. This indicated that some VLDL lipoproteins, involved in serum lipid transport, may be synthesized and exported from the liver. Therefore, based on gene expression analysis, it

seemed that transport and uptake of lipids were likely not the cause of a smaller body weight.

mRNAs encoding functions involved in steroid metabolism, a sub-category of lipid metabolism, showed a great reduction in abundance in P11 *tbp*<sup>ΔN/ΔN</sup> animals (Table 5.4). The pathway leading to cholesterol synthesis from acetyl-coA was retrieved from the Kyoto Encyclopedia of Genes and Genomes database (KEGG)<sup>553</sup>. For *tbp*<sup>ΔN/ΔN</sup> relative abundance of mRNA encoding for the corresponding enzymes are indicated in Fig. 5.6. Based on mRNA abundance, this pathway appeared strongly impaired in *tbp*<sup>ΔN/ΔN</sup>. Cholesterol can also be taken up from the blood. Uptake from the blood is performed through low-density lipoprotein (LDL) endocytosis by the LDL receptor, which showed a decreased mRNA abundance of 1.6 fold in *tbp*<sup>ΔN/ΔN</sup> animals. This suggested that a major defect in sterol regulation was associated with the lack of TBP-N. Neither hmg-CoA reductase nor LDL receptor mRNAs were more abundant and hmg-CoA reductase mRNA was actually less abundant (see appendix E and Fig. 5.6).

Table 5.4. Functional classification of the differentially abundant mRNA in P11 *tbp*<sup>ΔN/ΔN</sup> belonging to the lipid metabolism category.

| Ontology                    | Number of genes |
|-----------------------------|-----------------|
| Steroid metabolism          | 14              |
| Transport/storage           | 8               |
| Miscellaneous               | 6               |
| Arachidonic acid metabolism | 5               |
| Fatty acids elongation      | 3               |
| Fatty acids degradation     | 5               |



Legend:

HMGCR: enzyme name  
\*: rate limiting step

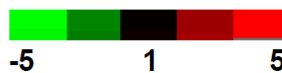


Figure 5.6: Sterol metabolism is compromised in P11 *tbp*<sup>ΔN/ΔN</sup> animals. The sterol metabolism pathway obtained from KEGG ([www.genome.ad.jp/kegg](http://www.genome.ad.jp/kegg)) was manually curated and microarrays values obtained for the enzymes listed in P11 animals are indicated. HMGCR: 3-hydroxy-3-methylglutaryl-Coenzyme A reductase; MVD: mevalonate (diphospho) decarboxylase; IDI1: isopentenyl-diphosphate delta isomerase 1; FDPS: farnesyl diphosphate synthase; FDFT1: farnesyl-diphosphate farnesyltransferase 1; SQLE: squalene epoxidase; LSS: lanosterol synthase; EBP: emopamil binding protein or sterol isomerase; SC5D: sterol C5-desaturase; and DHCR7: 7-dehydrocholesterol reductase.

The data indicated that the failure to thrive of the pre-weaning pups correlated with a likely defect in cholesterol biogenesis gene expression. It is thus possible that cholesterol levels were lower in P11 *tbp* <sup>$\Delta N/\Delta N$</sup> . Interestingly, lower circulating cholesterol levels were observed in *tbp* <sup>$\Delta N/\Delta N$</sup>  adults (Fig. 5.4C). Nevertheless, *tbp* <sup>$\Delta N/\Delta N$</sup>  adults are not runted like pre-weaning animals, while both share some hepatic steatosis.

In order to find a potential age-independent role for the TBP-N, we wanted to determine if there was any intersection from lists of differentially abundant mRNA in the P11 pups and mature adults. Selecting the previously described mRNA lists and defining their overlap (Fig. 5.7A, grey area), the intersection was narrowed down to 10 genes that can be perceived as being the most influenced by the origin of TBP-N regardless of the age or the surveyed tissues.

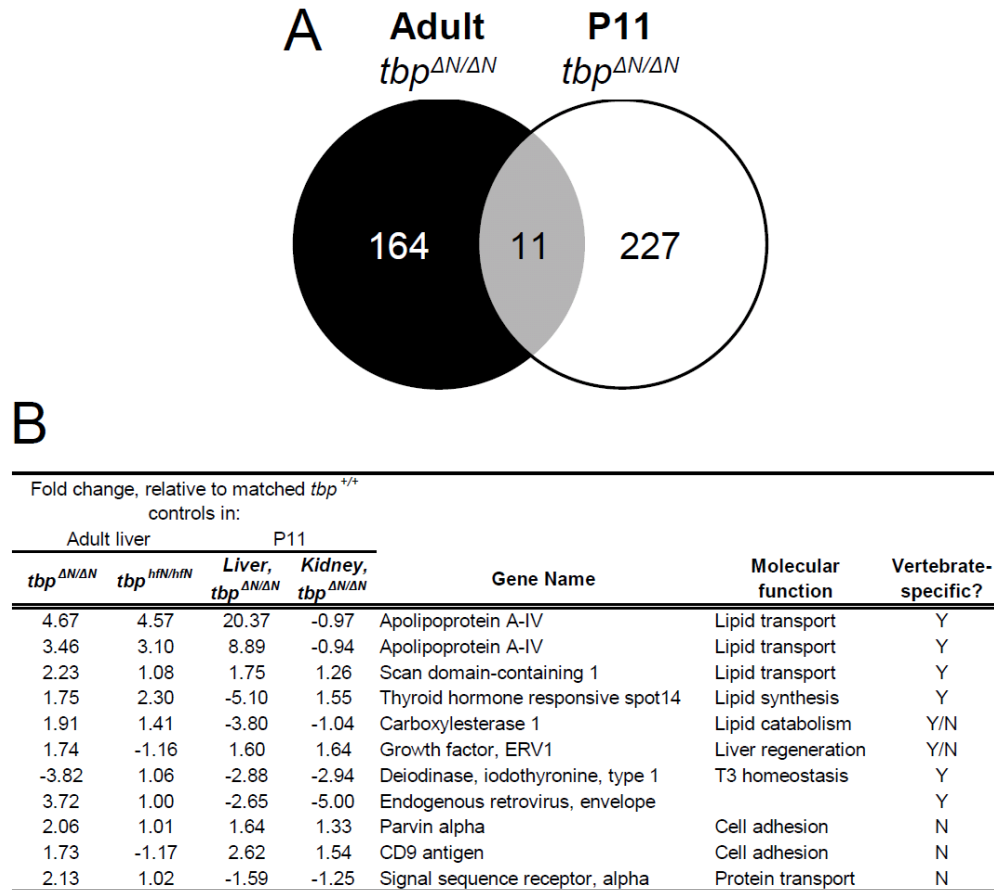


Figure 5.7. Identification of the mRNAs affected in young and adult *tbp*<sup>ΔN/ΔN</sup> and *tbp*<sup>hN/hN</sup> animals compared to age-matched *tbp*<sup>+/+</sup> animals. Venn diagrams of probe sets, which gave significantly and substantially different signals in P11 and adult animals indicated that 11 were shared between young and adult animals lacking the TBP-N. The mRNAs are listed below and magnitude by which each mRNA differed from wildtype is indicated. Also, gene function and “vertebrateness” are indicated. The “Y/N” abbreviation designated a gene present in invertebrates but that gained a novel domain in vertebrates. These new domains encompassed roughly half of the resulting vertebrate protein (data not shown). Y: yes; N: No.

Interestingly, 6 of the 11 mRNAs that were differentially abundant in both adult and P11 livers were vertebrate-specific, as no similar sequence could be found outside of the vertebrate sub-phylum (data not shown). Thus, including the two “Y/N” genes, 8 out of 10 genes (80%) that were the most affected by the lack of TBP-N were vertebrate-specific. In addition, half of those mRNAs showed a significant different abundance in

*tbp*<sup>hfN/hfN</sup>. This argued in favor of a likely role of the TBP-N in gene regulation of vertebrate-specific genes.

To test if fat composition of the adult diet had an effect on lipid accumulation in *tbp*<sup>+/+</sup> compared to *tbp*<sup>ΔN/ΔN</sup> livers, animals were subjected to either a low fat (4.5%) or a high fat (22.3%) diet. The diets were 2 day- or 38 day-long to identify immediate responses and potential long-term responses and/or compensation. We suspected we would find a correlation between a likely deficient lipid transport, based on mRNA expression, and the mutant phenotype (Table 5.2). Preliminary experiments showed that after two days on a low fat diet, CD36 mRNA abundance was not altered in *tbp*<sup>+/+</sup> while *tbp*<sup>ΔN/ΔN</sup> had more abundant CD36 mRNA than under high fat diet (Fig. 5.8). Interestingly, after 38 days, there was no apparent difference in the mRNA abundance of CD36 between *tbp*<sup>+/+</sup> and *tbp*<sup>ΔN/ΔN</sup>, regardless of the diet. This suggested that some potential long-term compensatory mechanisms could be involved to restore wildtype-like level of CD36 mRNA abundance (see below).

Histological analyses of adult livers indicated greater hepatic steatosis in *tbp*<sup>ΔN/+</sup> and *tbp*<sup>ΔN/ΔN</sup> under high fat diet (Fig. 5.8C). After 38 days of treatment under low fat and high fat diets, livers of *tbp*<sup>ΔN/+</sup> and *tbp*<sup>ΔN/ΔN</sup> had more oil red-O positive cells than wildtype controls indicating an increase in lipid accumulation primarily dependent of the mutation, and secondarily of the diet.

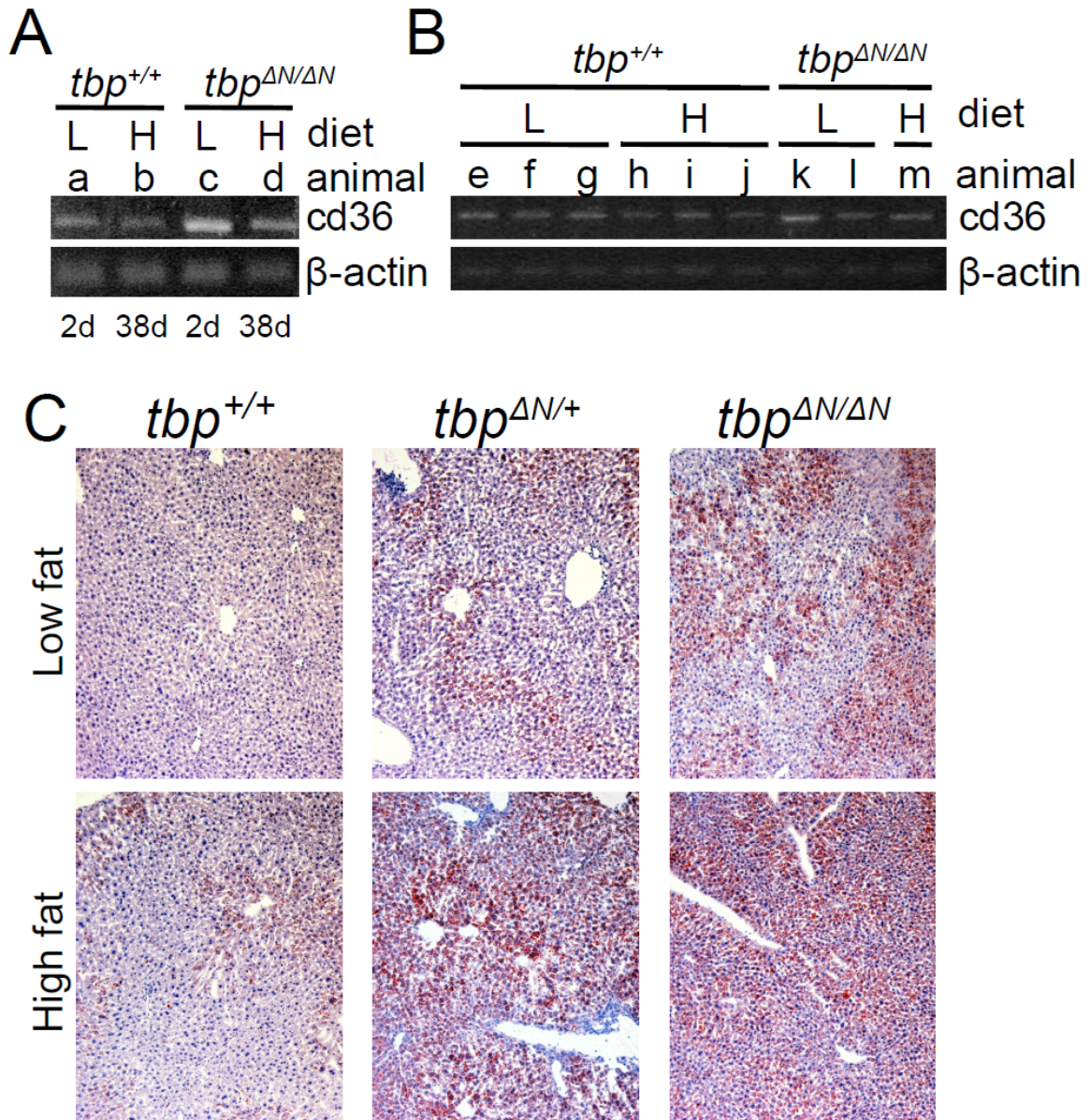


Figure 5.8. Effect of low or high fat diet on CD36 mRNA abundance and hepatic steatosis. Animals were fed a low fat (4.5%) or high fat (22.3%) diet for the indicated time. **(A, B)** CD36 RT-PCR for the indicated genotypes (n=1-3) after **(A)** 42h or **(B)** 38 days on low fat (L) or high fat (H) diet. Animals are identified with a unique letter. The lower panel indicates β-actin RT-PCR performed on the same samples as a loading control. **(C)** Liver sections from aged-matched adult males were stained with oil red-O (red stain) and counterstained with hematoxylin for the indicated genotypes after 38 days on the diet, as marked.

## Discussion

The data presented herein indicates that removal of the TBP-N ( $tbp^{\Delta N/\Delta N}$ ) or its replacement with a homologous sequence ( $tbp^{hfN/hfN}$ ) induced noticeable defects in hepatic lipid metabolism. Replacement of TBP-N by a homologous sequence did not generate an intermediate, but a distinct molecular signature, suggesting alteration of multiple pathways (Fig.5.1A, B). The identification of Pitx2 binding sites in more than 10% of the differentially expressed genes was intriguing. In chapter 2 of this dissertation, Pitx2 was shown to interact with TBP-N and regulate the expression of a vertebrate-specific gene, *nppa*. It is tempting to speculate that Pitx2 plays a role in regulating some of the mRNAs affected in  $tbp^{\Delta N/\Delta N}$  livers, possibly through interaction between Pitx2 and TBP-N. Nonetheless, Pitx2 has not yet been shown to be a major player in hepatic biology<sup>554</sup>. Nevertheless, the canonical role of Pitx2 as pituitary and left/right asymmetry transcription factor (see chapter 3 for details) has been recently extended when it was shown to be involved in gonad morphogenesis<sup>473</sup>. The increased mRNA abundance observed from the genes with Pitx2 binding sites was observed only in  $tbp^{\Delta N/\Delta N}$ . To explain this pattern, it can be hypothesized that in  $tbp^{+/+}$  and  $tbp^{hfN/hfN}$  the TBP-N is functional and involved in a protein-protein interaction with Pitx2. One explanation could be that all Pitx2-dependent genes are de-repressed when the TBP-N is lacking, as suggested for the Pitx2-TBP-N mechanism in chapter 3.

### Steatosis, *cd36*, and Putative Regulatory Elements

Pathological lipid accumulation, also known as steatosis, was observed in the liver of *tbp* <sup>$\Delta N/\Delta N$</sup>  and *tbp*<sup>*hfN/hfN*</sup>. Most metabolic diseases are not caused by a single gene. Use of multiple mouse models is required to understand complex metabolic dysfunction. For example, fatty acids and glucose utilization are closely linked and the crosstalk between fatty acids transport and insulin resistance is complex and not fully understood<sup>542</sup>. Also of note, fat accumulation in adipose tissue and the communication between adipose and hepatic tissues (see below) have not been studied here and may be part of future studies.

CD36 mRNA abundance was increased in *tbp* <sup>$\Delta N/\Delta N$</sup>  adult livers. In *tbp* <sup>$\Delta N/+$</sup> , CD36 liver mRNA levels were intermediate between *tbp* <sup>$+/+$</sup>  and *tbp* <sup>$\Delta N/\Delta N$</sup> . This suggested a possible role of TBP-N in regulation of *cd36* expression. In one possible model, interaction of TBP-N with an uncharacterized factor could modulate the expression of *cd36*. In such a model, this interaction could keep the expression of *cd36* at baseline, with TBP-N sequestering the uncharacterized factor and repressing its effects on transcription. Identification of augmented CD36 protein accumulation in the cell membrane through immunohistochemistry could be further performed. The consequence of CD36 mRNA up-regulation, and possibly increased CD36 protein, due to the lack of TBP-N (Fig. 5.2B) is interesting as known models of obesity and diabetes exhibit the same pattern of increased *cd36* gene expression<sup>537</sup>. Increased CD36 is usually found in cells secreting or storing triglycerides, like adipocytes<sup>537</sup>.

It has been shown that *tbp* <sup>$\Delta N/\Delta N$</sup>  animals die due to a placental failure where an increased trophoblast-dependent hemophagocytosis has been reported<sup>351</sup>. CD36 was

shown to be expressed on trophoblast cells <sup>555</sup>. Further studies will investigate the abundance of CD36 in *tbp*<sup>ΔN/ΔN</sup> placental tissues and a putative correlation between increased abundance of CD36, a placental defect, and erythrocyte abundance. CD36 is a scavenger receptor found on macrophages that can clear deficient erythrocytes, especially when infected with *Plasmodium* <sup>556</sup>. One may hypothesize that increased abundance of CD36 on the placental resident macrophages may increase their propensity for clearing aging as well as healthy red blood cells, which could correlate to the dysfunctional placenta reported in *tbp*<sup>ΔN/ΔN</sup> animals <sup>351</sup>. In addition, *cd36*<sup>-/-</sup> mice have a reduced (50-80%) uptake of long chain fatty acids <sup>523</sup>. Thus, as we observe a pathological lipid accumulation of lipids in *tbp*<sup>ΔN/ΔN</sup> animals in conjunction with an increased CD36 mRNA abundance, generating *cd36*<sup>-/-</sup> *tbp*<sup>ΔN/ΔN</sup> animals might restore wildtype levels of hepatic lipid accumulation. Thus, if the increased CD36 abundance were to be a cause of *tbp*<sup>ΔN/ΔN</sup> lethality, *tbp*<sup>ΔN/ΔN</sup> animals on a *cd36*<sup>-/-</sup> background might rescue them.

### Insulin Signaling

Hypoglycemia, as observed chronically in *tbp*<sup>ΔN/ΔN</sup>, is known to trigger glucagon secretion, which in turns restore a normal glycemia <sup>545</sup>. Hypersinsulinemia inhibits glucagon secretion <sup>547</sup>. Given the rapid and exacerbated increase in circulating glucose levels observed in the *tbp*<sup>ΔN/ΔN</sup> insulin tolerance tests 60 min after insulin injection, it is possible that compensatory levels of glucagon could be involved in the fast abnormal increase of glycemia in *tbp*<sup>ΔN/ΔN</sup>. Determination of insulin levels as well as measure of glucagon levels during the insulin tolerance test could be used to assess the precise physiological nature of the response observed. Moreover, ongoing glucose tolerance tests

and analysis of immediate insulin target genes during the insulin tolerance test will indicate if the animals can dispose of an increased glucose load faster, through induced hyperinsulinemia.

Insulin tolerance tests are also used to assess the pituitary and adrenal functions as growth hormone and cortisol can antagonize insulin action<sup>557</sup>. It is interesting to note that Pitx2, an interaction partner of TBP-N (see chapter 3), has been shown to be required for the development of the cell lineage responsible for producing growth hormone in the pituitary<sup>558</sup>.

#### Growth Retardation and Cholesterol Metabolism in P11 Pups

Recently, in humans, low birth weight has been positively correlated to hypomethylation of the imprinted *Igf2* locus<sup>559</sup>. IGF2 is a growth factor only expressed from the paternal allele<sup>560</sup>. Lack of *Igf2* results in a dwarf phenotype<sup>560</sup>. In P11 animals, which are more than 25% smaller than their wildtype littermates, *Igf2* was downregulated 2.1 fold. In adult *tbp*<sup>ΔN/ΔN</sup>, the mRNA abundance of *Igf2* was similar to the *tbp*<sup>+/+</sup> animals. Moreover, *tbp*<sup>ΔN/ΔN</sup> adults do not appear smaller than *tbp*<sup>+/+</sup> (data not shown). Therefore, *Igf2* could be an interesting gene to follow-up in further studies, especially through its methylation state in *tbp*<sup>+/+</sup> compared to *tbp*<sup>ΔN/ΔN</sup> animals.

In humans, the Optiz syndrome, a rare (1:20,000-1:70,000 infants<sup>561</sup>; <http://ghr.nlm.nih.gov>) autosomal recessive disease manifested by mental retardation, morphologic abnormalities, and a smaller size<sup>562,563</sup> is diagnosed by a low cholesterol level. The low cholesterol level is the result of non-functional 7-dehydrocholesterol reductase<sup>564</sup>. 7-dehydrocholesterol reductase mRNA was 1.8 time less abundant in

*tbp*<sup>ΔN/ΔN</sup>. Also, infants with Opitz syndrome tend to have a weak muscle tone, a difficulty to feed, and an overall slower growth. Infants with Opitz syndrome can be placed under cholesterol supplementation with relative success<sup>561</sup>. Cholesterol synthesis is tightly regulated through multiple feedback mechanisms<sup>550</sup>. Here we observed a down-regulated cholesterol synthesis pathway in P11 *tbp*<sup>ΔN/ΔN</sup> animals (Fig. 5.6) and a lower blood cholesterol level in adults *tbp*<sup>ΔN/ΔN</sup> animals (Fig. 5.4). Such an unusual phenotype will require measurement of circulating cholesterol in pups and mechanistic studies. In addition, as infants affected by the Opitz syndrome can be treated, with mixed results<sup>565-567</sup>, by cholesterol supplementation, *tbp*<sup>ΔN/ΔN</sup> pups might be treated by cholesterol supplementation through the dam's diet or via injection. In humans, milk cholesterol level appears to be unaffected by the diet<sup>568</sup>. Thus cholesterol injection to the pups could be more suited in order to compensate for their likely low circulating cholesterol. In human infants, cholesterol supplementation in the dietary milk was reflected by an increased circulating cholesterol<sup>569</sup>. Therefore, expect if, in addition to an altered synthesis pathway, cholesterol uptake from the blood is chronically impaired, supplementation through oral gavage, intraveinuous, intramuscular or intraperitoneal supplementation of cholesterol in the pups might restore their cholesterol metabolism and maybe rescue their failure to thrive. The marked decrease of hmg-CoA reductase mRNA abundance hinted a potential defect in signaling through SREBP. Therefore, it is possible that cholesterol sensing through SREBP, its translocation to the nucleus, or its signaling to the basal transcription machinery possibly through unknown interactions with TBP-N, is impaired.

### Vertebrate-Specific Genes and TBP-N

A recent mouse genome analysis identified 24,264 genes in which 4,065 (17%) are present in vertebrates only <sup>280</sup>. Here we observed that 8 out of 10 genes (80%) that were the most affected by the lack of TBP-N were vertebrate-specific. Such an overrepresentation of vertebrate-specific genes affected by the alteration of the TBP-N was unlikely random (Chi<sup>2</sup>, two-tailed,  $p < 0.0001$ ). Interestingly, the lack of the vertebrate-specific TBP-N seemed to affect expression of genes that appear as pure vertebrate innovations, suggesting a likely role for the TBP-N in vertebrate-specific gene regulation.

### Adipose and Hepatic Tissue Crosstalk

Adipose and hepatic tissues are intertwined in the regulation of energy and especially lipid metabolism <sup>244</sup>. Adipocytokines are adipocyte-specific hormones that have a major role in lipid metabolism <sup>545,546</sup>. The adipocytokine family includes resistin, which mediates insulin resistance while leptin and adiponectin are involved in insulin sensitization <sup>570,571</sup>. More recently, apelin, another adipocytokine, has been shown to have a very powerful glucose-lowering effect. Its administration brings down normal glycemia by 25% <sup>572</sup>, comparable to the observed chronic 21% glycemia reduction in *tbp* <sup>$\Delta N/\Delta N$</sup>  (median are 159mg/dL for *tbp*<sup>+/+</sup> and 128.5mg/dL for *tbp* <sup>$\Delta N/\Delta N$</sup> , Fig. 5.4A). Regulation of apelin, which appeared to be a vertebrate-specific hormone (Ensemble and NCBI queries, January 2009, Lucas, data not shown), is not fully understood. Interestingly, apelin was first described, more than 10 years ago, as being a hormone involved in cardiovascular regulation, mostly through anti-hypertensive effects, food intake, and fluid homeostasis <sup>573 574,575</sup>. It is interesting to note that atrial natriuretic factor, which we showed its mRNA

is likely regulated by the TBP-N-Pitx2 interaction (see chapter 3), is also a cardiac hormone.

Further studies will investigate the nature of the lipids accumulated in *tbp* <sup>$\Delta N/\Delta N$</sup>  livers and the possible crosstalk between hepatic and adipose tissues. In order to better understand the non-canonical profile observed during insulin tolerance tests, measurement of insulin and glucagon levels and investigations of candidate gene mRNA levels will be performed.

We have shown that the TBP-N is involved in lipid metabolism regulation and that the lack of TBP-N correlates with pup runting and a down-regulated cholesterol synthesis pathway. Also, adult *tbp* <sup>$\Delta N/\Delta N$</sup>  animals have an alteration in lipid transport and uptake and exhibit hepatic steatosis.

## SUMMARY AND CONCLUSION

TBP is central to nearly all transcription initiation events from all three RNA polymerases. Its role as an interface between DNA (promoter) and other proteins (transcription factors and other components of the basal transcription machinery) makes it a key protein for regulation of basal transcription. In mouse and yeast, inactivation of TBP is lethal<sup>368,576</sup>. In addition to the conserved core domain found throughout eukaryotes, metazoan phyla possess unrelated N-terminal sequences of unknown functions. Removal of the vertebrate-specific TBP-N strongly compromises fetal development, likely due to defect in placenta development<sup>351</sup>.

In the work presented here we hypothesized that the function of the TBP-N was conserved among vertebrates. To test this, the TBP-N of mouse was substituted for the one of hagfish. Survival of *tbp*<sup>hfN/hfN</sup> at a wildtype level supported the hypothesis of an ancestral and essential function that likely appeared at the base of the vertebrate lineage and that was conserved ever since. Further, the partial genetic complementation of the *tbp*<sup>hfN</sup> allele over the *tbp*<sup>ΔN</sup> allele suggested that most underlying molecular pathways were complemented, allowing a 10 fold increase in survival of *tbp*<sup>ΔN/hfN</sup> compared to *tbp*<sup>ΔN/ΔN</sup>. Investigation of the molecular mechanisms involving the TBP-N involved yeast two-hybrids screens, molecular pathways identification through global gene expression analysis, and functional assays.

Studies presented in chapter 3 showed that the TBP-N interacted with Pitx transcription factor family members. In vertebrates, Pitx proteins are essential to many vertebrate-specific processes during development and after birth. We showed that the

TBP-N-Pitx interactions likely co-evolved in vertebrates. We further showed that the interactions were functional and that, on a known target gene of Pitx2 (*nppa*), the lack of TBP-N correlated with an increased activity of Pitx2-dependent transcription initiation, suggesting that the TBP-N-Pitx2 might modulate Pitx2 activity.

Chapter 5 further characterized the *tbp*<sup>h<sup>fN</sup></sup> allele at the molecular level along with a refined functional analysis of the *tbp*<sup>ΔN</sup> allele. Transcriptome analyses showed that different pathways were affected by the different alleles but that a common set of mRNAs was affected. Many of these mRNAs exhibited an intermediate abundance in *tbp*<sup>h<sup>fN</sup>/h<sup>fN</sup></sup> when compared to wildtype and *tbp*<sup>ΔN/ΔN</sup>, reinforcing the partial complementation data gained through genetics. In particular, abundance of mRNA for the *cd36* gene, encoding a long chain fatty acid transporter, was affected in all allelic variants of *tbp* tested. Unexpectedly, TBP-N also appeared to modulate the expression of endogenous retroviruses, which are known to be involved in the appearance and development of the placenta. Both *cd36* and endogenous retroviruses expression will require further studies, notably in the placenta.

In addition, *tbp*<sup>ΔN/ΔN</sup> adult animals had low levels of circulating cholesterol. Juvenile *tbp*<sup>ΔN/ΔN</sup> showed a unique defect in the pathway leading to cholesterol synthesis and a failure to thrive. The smaller size of pre-weaning *tbp*<sup>ΔN/ΔN</sup> animals could be due to the defect observed in the cholesterol synthesis pathway. This hints that *tbp*<sup>ΔN/ΔN</sup> may be a model for the Opitz syndrome, where cholesterol synthesis is impaired. The glycemia response of *tbp*<sup>ΔN/ΔN</sup> to an increased insulin level exhibited a defect with a pattern not previously reported in the literature, which will require further investigation of the

immediate target genes of insulin during insulin response. Also, *tbp* <sup>$\Delta N/\Delta N$</sup>  animals exhibited abnormal hepatic lipid storage as juveniles and as older adults, but apparently not as young adults, reminiscent of an age-dependent fatty liver disease. Taken together, the data suggested that TBP-N is involved in energy metabolism homeostasis. This raises the exciting possibility that *tbp* <sup>$\Delta N/\Delta N$</sup>  is potentially a novel mouse model to investigate some aspects of the metabolic syndrome.

The work presented here uncovered two seemingly unrelated novel roles for the TBP-N. The interaction of TBP-N with Pitx2c, a major vertebrate-specific transcription factor with multiple developmental roles, had a functional role in gene expression. A whole animal perspective suggested major defects, at the transcriptional level, in lipid metabolism pathways. These studies expand our understanding of the TBP-N and, for the first time, clearly identify this domain as involved in multiple gene regulatory networks and as essential because of its role in embryonic development and energy metabolism in mouse.

REFERENCES CITED

1. Maynard-Smith J. On evolution. Edinburgh: Edinburgh University Press; 1972.
2. Huxley A. Evolution, the modern synthesis. London: Allen and Unwin; 1942.
3. Fisher R. The genetical theory of natural selection. Oxford: Clarendon Press; 1930.
4. Gould S. Wonderful Life: The Burgess Shale and the Nature of History. New-York: WW Norton & Co; 1989. 347 p.
5. Miyata T, Suga H. Divergence pattern of animal gene families and relationship with the Cambrian explosion. *Bioessays* 2001;23(11):1018-27.
6. Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC and others. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol* 2007;5(4):e101.
7. Furlong RF, Younger R, Kasahara M, Reinhardt R, Thorndyke M, Holland PW. A degenerate ParaHox gene cluster in a degenerate vertebrate. *Mol Biol Evol* 2007;24(12):2681-6.
8. Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ and others. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 2008;18(7):1100-11.
9. Holland LZ, Short S. Gene duplication, co-option and recruitment during the origin of the vertebrate brain from the invertebrate chordate brain. *Brain Behav Evol* 2008;72(2):91-105.
10. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK and others. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 2008;453(7198):1064-71.
11. Durand D. Vertebrate evolution: doubling and shuffling with a full deck. *Trends Genet* 2003;19(1):2-5.
12. Eichler EE. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* 2001;17(11):661-9.
13. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 2005;3(10):e314.

14. Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol* 2006;7(5):R43.
15. Nakatani Y, Takeda H, Kohara Y, Morishita S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* 2007;17(9):1254-65.
16. Aburomia R, Khaner O, Sidow A. Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. *J Struct Funct Genomics* 2003;3(1-4):45-52.
17. Donoghue PC, Purnell MA. Genome duplication, extinction and vertebrate evolution. *Trends Ecol Evol* 2005;20(6):312-9.
18. Duboc V, Lapraz F, Besnardeau L, Lepage T. Lefty acts as an essential modulator of Nodal activity during sea urchin oral-aboral axis formation. *Dev Biol* 2008;320(1):49-59.
19. Boorman CJ, Shimeld SM. The evolution of left-right asymmetry in chordates. *Bioessays* 2002;24(11):1004-11.
20. Rast JP, Messier-Solek C. Marine invertebrate genome sequences and our evolving understanding of animal immunity. *Biol Bull* 2008;214(3):274-83.
21. Gans C, Northcutt RG. Neural Crest and the Origin of Vertebrates: A New Head. *Science* 1983;220(4594):268-273.
22. Darwin C. The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray; 1859.
23. Needham J. On the dissociability of the fundamental processes in ontogenesis. *Biol. Rev.* 1933;8:180-233.
24. Schlosser G, Wagner G. Modularity in development and evolution. Chicago, Illinois: University of Chicago Press; 2004.
25. Waxman D, Peck JR. Pleiotropy and the preservation of perfection. *Science* 1998;279(5354):1210-3.
26. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W and others. Comparative genomics of the eukaryotes. *Science* 2000;287(5461):2204-15.
27. Donoghue PC, Graham A, Kelsh RN. The origin and evolution of the neural crest. *Bioessays* 2008;30(6):530-41.

28. Sauka-Spengler T, Bronner-Fraser M. A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol* 2008;9(7):557-68.
29. Sauka-Spengler T, Bronner-Fraser M. Development and evolution of the migratory neural crest: a gene regulatory perspective. *Curr Opin Genet Dev* 2006;16(4):360-6.
30. Sauka-Spengler T, Meulemans D, Jones M, Bronner-Fraser M. Ancient evolutionary origin of the neural crest gene regulatory network. *Dev Cell* 2007;13(3):405-20.
31. Raible DW. Development of the neural crest: achieving specificity in regulatory pathways. *Curr Opin Cell Biol* 2006;18(6):698-703.
32. Meulemans D, Bronner-Fraser M. Central role of gene cooption in neural crest evolution. *J Exp Zool B Mol Dev Evol* 2005;304(4):298-303.
33. Wada H, Makabe K. Genome duplications of early vertebrates as a possible chronicle of the evolutionary history of the neural crest. *Int J Biol Sci* 2006;2(3):133-41.
34. Martinez-Morales JR, Henrich T, Ramialison M, Wittbrodt J. New genes in the evolution of the neural crest differentiation program. *Genome Biol* 2007;8(3):R36.
35. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 2006;439(7079):965-8.
36. Blair JE, Hedges SB. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol* 2005;22(11):2275-84.
37. Vienne A, Pontarotti P. Metaphylogeny of 82 gene families sheds a new light on chordate evolution. *Int J Biol Sci* 2006;2(2):32-7.
38. Holland ND, Chen J. Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and palaeontology. *Bioessays* 2001;23(2):142-51.
39. Russo MT, Donizetti A, Locascio A, D'Aniello S, Amoroso A, Aniello F, Fucci L, Branno M. Regulatory elements controlling *Ci-msxb* tissue-specific expression during *Ciona intestinalis* embryonic development. *Dev Biol* 2004;267(2):517-28.
40. Jeffery WR, Strickler AG, Yamamoto Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* 2004;431(7009):696-9.
41. Mazet F, Hutt JA, Millard J, Shimeld SM. Pax gene expression in the developing central nervous system of *Ciona intestinalis*. *Gene Expr Patterns* 2003;3(6):743-5.

42. Hong CS, Saint-Jeannet JP. Sox proteins and neural crest development. *Semin Cell Dev Biol* 2005;16(6):694-703.
43. Kelsh RN. Sorting out Sox10 functions in neural crest development. *Bioessays* 2006;28(8):788-98.
44. Meulemans D, McCauley D, Bronner-Fraser M. Id expression in amphioxus and lamprey highlights the role of gene cooption during neural crest evolution. *Dev Biol* 2003;264(2):430-42.
45. Meulemans D, Bronner-Fraser M. Gene-regulatory interactions in neural crest evolution and development. *Dev Cell* 2004;7(3):291-9.
46. Holland ND, Panganiban G, Henyey EL, Holland LZ. Sequence and developmental expression of *AmphiDll*, an amphioxus *Distal-less* gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest. *Development* 1996;122(9):2911-20.
47. Holland LZ, Schubert M, Kozmik Z, Holland ND. *AmphiPax3/7*, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. *Evol Dev* 1999;1(3):153-65.
48. Holland LZ, Schubert M, Holland ND, Neuman T. Evolutionary conservation of the presumptive neural plate markers *AmphiSox1/2/3* and *AmphiNeurogenin* in the invertebrate chordate amphioxus. *Dev Biol* 2000;226(1):18-33.
49. Langeland JA, Tomsa JM, Jackman WR, Jr., Kimmel CB. An amphioxus snail gene: expression in paraxial mesoderm and neural plate suggests a conserved role in patterning the chordate embryo. *Dev Genes Evol* 1998;208(10):569-77.
50. Olson EN. Gene regulatory networks in the evolution and development of the heart. *Science* 2006;313(5795):1922-7.
51. Davidson E. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*: Academic Press; 2006. 289 p.
52. McCauley DW, Bronner-Fraser M. Neural crest contributions to the lamprey head. *Development* 2003;130(11):2317-27.
53. Vickaryous MK, Hall BK. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol Rev Camb Philos Soc* 2006;81(3):425-55.
54. Richman JM, Lee SH. About face: signals and genes controlling jaw patterning and identity in vertebrates. *Bioessays* 2003;25(6):554-68.

55. Stone JR, Hall BK. Latent homologues for the neural crest as an evolutionary novelty. *Evol Dev* 2004;6(2):123-9.
56. Yu JK, Meulemans D, McKeown SJ, Bronner-Fraser M. Insights from the amphioxus genome on the origin of vertebrate neural crest. *Genome Res* 2008;18(7):1127-32.
57. Baker CV, Bronner-Fraser M. Establishing neuronal identity in vertebrate neurogenic placodes. *Development* 2000;127(14):3045-56.
58. Begbie J, Graham A. The ectodermal placodes: a dysfunctional family. *Philos Trans R Soc Lond B Biol Sci* 2001;356(1414):1655-60.
59. Baker CV, Bronner-Fraser M. Vertebrate cranial placodes I. Embryonic induction. *Dev Biol* 2001;232(1):1-61.
60. Schlosser G. Evolutionary origins of vertebrate placodes: insights from developmental studies and from comparisons with other deuterostomes. *J Exp Zool B Mol Dev Evol* 2005;304(4):347-99.
61. Schlosser G. Induction and specification of cranial placodes. *Dev Biol* 2006;294(2):303-51.
62. Couly GF, Le Douarin NM. Mapping of the early neural primordium in quail-chick chimeras. I. Developmental relationships between placodes, facial ectoderm, and prosencephalon. *Dev Biol* 1985;110(2):422-39.
63. Chuah MI, Au C. Olfactory Schwann cells are derived from precursor cells in the olfactory epithelium. *J Neurosci Res* 1991;29(2):172-80.
64. Norgren RB, Jr., Ratner N, Brackenbury R. Development of olfactory nerve glia defined by a monoclonal antibody specific for Schwann cells. *Dev Dyn* 1992;194(3):231-8.
65. Ramon-Cueto A, Plant GW, Avila J, Bunge MB. Long-distance axonal regeneration in the transected adult rat spinal cord is promoted by olfactory ensheathing glia transplants. *J Neurosci* 1998;18(10):3803-15.
66. Fritzsche B, Dubuc R, Ohta Y, Grillner S. Efferents to the labyrinth of the river lamprey (*Lampetra fluviatilis*) as revealed with retrograde tracing techniques. *Neurosci Lett* 1989;96(3):241-6.
67. Schlosser G. Development and evolution of lateral line placodes in amphibians. - II. Evolutionary diversification. *Zoology (Jena)* 2002;105(3):177-93.

68. Schlosser G. Development and evolution of lateral line placodes in amphibians I. Development. *Zoology (Jena)* 2002;105(2):119-46.
69. Streit A. Early development of the cranial sensory nervous system: from a common field to individual placodes. *Dev Biol* 2004;276(1):1-15.
70. Christensen KL, Patrick AN, McCoy EL, Ford HL. The six family of homeobox genes in development and cancer. *Adv Cancer Res* 2008;101:93-126.
71. Duncan MK, Kos L, Jenkins NA, Gilbert DJ, Copeland NG, Tomarev SI. Eyes absent: a gene family found in several metazoan phyla. *Mamm Genome* 1997;8(7):479-85.
72. Xu PX, Woo I, Her H, Beier DR, Maas RL. Mouse Eya homologues of the *Drosophila* eyes absent gene require Pax6 for expression in lens and nasal placode. *Development* 1997;124(1):219-31.
73. Zimmerman JE, Bui QT, Steingrimsson E, Nagle DL, Fu W, Genin A, Spinner NB, Copeland NG, Jenkins NA, Bucan M and others. Cloning and characterization of two vertebrate homologs of the *Drosophila* eyes absent gene. *Genome Res* 1997;7(2):128-41.
74. Borsani G, DeGrandi A, Ballabio A, Bulfone A, Bernard L, Banfi S, Gattuso C, Mariani M, Dixon M, Donnai D and others. EYA4, a novel vertebrate gene related to *Drosophila* eyes absent. *Hum Mol Genet* 1999;8(1):11-23.
75. Mazet F, Shimeld SM. Molecular evidence from ascidians for the evolutionary origin of vertebrate cranial sensory placodes. *J Exp Zool B Mol Dev Evol* 2005;304(4):340-6.
76. Rebay I, Silver SJ, Tootle TL. New vision from Eyes absent: transcription factors as enzymes. *Trends Genet* 2005;21(3):163-71.
77. Lacalli TC. Tunicate tails, stolons, and the origin of the vertebrate trunk. *Biol Rev Camb Philos Soc* 1999;74(2):177-98.
78. Sharman AC, Shimeld SM, Holland PW. An amphioxus Msx gene expressed predominantly in the dorsal neural tube. *Dev Genes Evol* 1999;209(4):260-3.
79. Neidert AH, Virupannavar V, Hooker GW, Langeland JA. Lamprey Dlx genes and early vertebrate evolution. *Proc Natl Acad Sci U S A* 2001;98(4):1665-70.
80. Stock DW, Ellies DL, Zhao Z, Ekker M, Ruddle FH, Weiss KM. The evolution of the vertebrate Dlx gene family. *Proc Natl Acad Sci U S A* 1996;93(20):10858-63.

81. Di Gregorio A, Spagnuolo A, Ristoratore F, Pischetola M, Aniello F, Branno M, Cariello L, Di Lauro R. Cloning of ascidian homeobox genes provides evidence for a primordial chordate cluster. *Gene* 1995;156(2):253-7.
82. Burighel P, Lane NJ, Fabio G, Stefano T, Zaniolo G, Carnevali MD, Manni L. Novel, secondary sensory cell organ in ascidians: in search of the ancestor of the vertebrate lateral line. *J Comp Neurol* 2003;461(2):236-49.
83. Manni L, Lane NJ, Joly JS, Gasparini F, Tiozzo S, Caicci F, Zaniolo G, Burighel P. Neurogenic and non-neurogenic placodes in ascidians. *J Exp Zool B Mol Dev Evol* 2004;302(5):483-504.
84. Wada H, Saiga H, Satoh N, Holland PW. Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian Pax-2/5/8, Hox and Otx genes. *Development* 1998;125(6):1113-22.
85. Yasui K, Zhang S, Uemura M, Saiga H. Left-right asymmetric expression of BbPtx, a Ptx-related gene, in a lancelet species and the developmental left-sidedness in deuterostomes. *Development* 2000;127(1):187-95.
86. Bassham S, Postlethwait JH. The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate *Oikopleura dioica*. *Development* 2005;132(19):4259-72.
87. Manni L, Lane NJ, Sorrentino M, Zaniolo G, Burighel P. Mechanism of neurogenesis during the embryonic development of a tunicate. *J Comp Neurol* 1999;412(3):527-41.
88. Schubert M, Holland ND, Escriva H, Holland LZ, Laudet V. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). *Proc Natl Acad Sci U S A* 2004;101(28):10320-5.
89. Wicht H. The brains of lampreys and hagfishes: characteristics, characters, and comparisons. *Brain Behav Evol* 1996;48(5):248-61.
90. Lumsden A. Segmentation and compartmentation in the early avian hindbrain. *Mech Dev* 2004;121(9):1081-8.
91. Tao W, Lai E. Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork head gene family, in the developing rat brain. *Neuron* 1992;8(5):957-66.
92. Toresson H, Martinez-Barbera JP, Bardsley A, Caubit X, Krauss S. Conservation of BF-1 expression in amphioxus and zebrafish suggests evolutionary ancestry of

anterior cell types that contribute to the vertebrate telencephalon. *Dev Genes Evol* 1998;208(8):431-9.

93. Lacalli TC. Sensory systems in amphioxus: a window on the ancestral chordate condition. *Brain Behav Evol* 2004;64(3):148-62.
94. Shimeld SM, Purkiss AG, Dirks RP, Bateman OA, Slingsby C, Lubsen NH. Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens. *Curr Biol* 2005;15(18):1684-9.
95. Bally-Cuif L, Boncinelli E. Transcription factors and head formation in vertebrates. *Bioessays* 1997;19(2):127-35.
96. Venkatesh TV, Holland ND, Holland LZ, Su MT, Bodmer R. Sequence and developmental expression of amphioxus *AmphiNk2-1*: insights into the evolutionary origin of the vertebrate thyroid gland and forebrain. *Dev Genes Evol* 1999;209(4):254-9.
97. Holland LZ, Kene M, Williams NA, Holland ND. Sequence and embryonic expression of the amphioxus engrailed gene (*AmphiEn*): the metameric pattern of transcription resembles that of its segment-polarity homolog in *Drosophila*. *Development* 1997;124(9):1723-32.
98. Jiang D, Smith WC. An ascidian engrailed gene. *Dev Genes Evol* 2002;212(8):399-402.
99. Jackman WR, Langeland JA, Kimmel CB. *islet* reveals segmentation in the Amphioxus hindbrain homolog. *Dev Biol* 2000;220(1):16-26.
100. Ferrier DE, Brooke NM, Panopoulou G, Holland PW. The *Mnx* homeobox gene class defined by *HB9*, *MNR2* and amphioxus *AmphiMnx*. *Dev Genes Evol* 2001;211(2):103-7.
101. Jackman WR, Kimmel CB. Coincident iterated gene expression in the amphioxus neural tube. *Evol Dev* 2002;4(5):366-74.
102. Mazet F, Shimeld SM. The evolution of chordate neural segmentation. *Dev Biol* 2002;251(2):258-70.
103. Klein R. Eph/ephrin signaling in morphogenesis, neural development and plasticity. *Curr Opin Cell Biol* 2004;16(5):580-9.
104. Wilkinson DG. Multiple roles of EPH receptors and ephrins in neural development. *Nat Rev Neurosci* 2001;2(3):155-64.

105. Kozmik Z, Holland ND, Kalousova A, Paces J, Schubert M, Holland LZ. Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region. *Development* 1999;126(6):1295-304.
106. Seitanidou T, Schneider-Maunoury S, Desmarquet C, Wilkinson DG, Charnay P. *Krox-20* is a key regulator of rhombomere-specific gene expression in the developing hindbrain. *Mech Dev* 1997;65(1-2):31-42.
107. Knight RD, Panopoulou GD, Holland PW, Shimeld SM. An amphioxus *Krox* gene: insights into vertebrate hindbrain evolution. *Dev Genes Evol* 2000;210(10):518-21.
108. Williams NA, Holland PW. Molecular evolution of the brain of chordates. *Brain Behav Evol* 1998;52(4-5):177-85.
109. Hall BK. *Bones and Cartilage: Developmental and Evolutionary Skeletal Biology*: Academic Press; 2005. 760 p.
110. Felisbino SL, Carvalho HF. Ectopic mineralization of articular cartilage in the bullfrog *Rana catesbeiana* and its possible involvement in bone closure. *Cell Tissue Res* 2002;307(3):357-65.
111. Hall BK. Consideration of the neural crest and its skeletal derivatives in the context of novelty/innovation. *J Exp Zool B Mol Dev Evol* 2005;304(6):548-57.
112. Kawasaki K, Weiss KM. Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *J Exp Zool B Mol Dev Evol* 2006;306(3):295-316.
113. Kawasaki K, Buchanan AV, Weiss KM. Gene duplication and the evolution of vertebrate skeletal mineralization. *Cells Tissues Organs* 2007;186(1):7-24.
114. Sire JY, Davit-Beal T, Delgado S, Gu X. The origin and evolution of enamel mineralization genes. *Cells Tissues Organs* 2007;186(1):25-48.
115. Kawasaki K, Suzuki T, Weiss KM. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. *Proc Natl Acad Sci U S A* 2005;102(50):18063-8.
116. Hamilton H. *Lillie's development of the chick: an introduction to embryology*. New-York: Holt, Rinehart, and Winston; 1965.
117. Couly GF, Coltey PM, Le Douarin NM. The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. *Development* 1993;117(2):409-29.

118. Yan YL, Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. *Development* 2005;132(5):1069-83.
119. Duboule D. Vertebrate hox gene regulation: clustering and/or colinearity? *Curr Opin Genet Dev* 1998;8(5):514-8.
120. Noramly S, Grainger RM. Determination of the embryonic inner ear. *J Neurobiol* 2002;53(2):100-28.
121. Fritzsche B, Beisel KW. Evolution and development of the vertebrate ear. *Brain Res Bull* 2001;55(6):711-21.
122. Shimeld SM, Holland PW. Vertebrate innovations. *Proc Natl Acad Sci U S A* 2000;97(9):4449-52.
123. Lewis AK, Frantz GD, Carpenter DA, de Sauvage FJ, Gao WQ. Distinct expression patterns of notch family receptors and ligands during development of the mammalian inner ear. *Mech Dev* 1998;78(1-2):159-63.
124. Kernan M, Zuker C. Genetic approaches to mechanosensory transduction. *Curr Opin Neurobiol* 1995;5(4):443-8.
125. Bermingham NA, Hassan BA, Price SD, Vollrath MA, Ben-Arie N, Eatock RA, Bellen HJ, Lysakowski A, Zoghbi HY. Math1: an essential gene for the generation of inner ear hair cells. *Science* 1999;284(5421):1837-41.
126. Ben-Arie N, Hassan BA, Bermingham NA, Malicki DM, Armstrong D, Matzuk M, Bellen HJ, Zoghbi HY. Functional conservation of atonal and Math1 in the CNS and PNS. *Development* 2000;127(5):1039-48.
127. Benito-Gutierrez E, Nake C, Llovera M, Comella JX, Garcia-Fernandez J. The single AmphiTrk receptor highlights increased complexity of neurotrophin signalling in vertebrates and suggests an early role in developing sensory neuroepidermal cells. *Development* 2005;132(9):2191-202.
128. Hallbook F, Wilson K, Thorndyke M, Olinski RP. Formation and evolution of the chordate neurotrophin and Trk receptor genes. *Brain Behav Evol* 2006;68(3):133-44.
129. Satou Y, Imai KS, Levine M, Kohara Y, Rokhsar D, Satoh N. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. I. Genes for bHLH transcription factors. *Dev Genes Evol* 2003;213(5-6):213-21.

130. Hallbook F, Lundin LG, Kullander K. Lampetra fluviatilis neurotrophin homolog, descendant of a neurotrophin ancestor, discloses the early molecular evolution of neurotrophins in the vertebrate subphylum. *J Neurosci* 1998;18(21):8700-11.
131. Sperber G. Craniofacial development. Hamilton, CA: BC Decker; 2001.
132. Graham A, Okabe M, Quinlan R. The role of the endoderm in the development and evolution of the pharyngeal arches. *J Anat* 2005;207(5):479-87.
133. Veitch E, Begbie J, Schilling TF, Smith MM, Graham A. Pharyngeal arch patterning in the absence of neural crest. *Curr Biol* 1999;9(24):1481-4.
134. Muller W. Über die Hypobranchialrinne der Tunicaten und deren Vorhandensein bei Amphioxus und den Cyclostomen. *Jena Z Med* 1873;7:327-332.
135. Fujita H, Honma Y. Iodine metabolism of the endostyle of larval lampreys, *Ammocoetes* of *Lampetra japonica*. Electron microscopic autoradiography of 125-I. *Z Zellforsch Mikrosk Anat* 1969;98(4):525-37.
136. Fujita H, Sawano F. Fine structural localization of endogeneous peroxidase in the endostyle of ascidians, *Ciona intestinalis*. A part of phylogenetic studies of the thyroid gland. *Arch Histol Jpn* 1979;42(3):319-26.
137. Suzuki S, Kondo Y. Thyroidal morphogenesis and biosynthesis of thyroglobulin before and after metamorphosis in the lamprey, *Lampetra reissneri*. *Gen Comp Endocrinol* 1973;21(3):451-60.
138. Wright GM, Filosa MF, Youson JH. Immunocytochemical localization of thyroglobulin in the endostyle of the anadromous sea lamprey, *Petromyzon marinus* L. *Am J Anat* 1978;152(2):263-8.
139. Wright GM, Youson JH. Transformation of the endostyle of the anadromous sea lamprey, *Petromyzon marinus* L., during metamorphosis. I. Light microscopy and autoradiography with 125I1. *Gen Comp Endocrinol* 1976;30(3):243-57.
140. Ogasawara M, Satoh N. Isolation and characterization of endostyle-specific genes in the ascidian *Ciona intestinalis*. *Biol Bull* 1998;195(1):60-9.
141. Ogasawara M, Di Lauro R, Satoh N. Ascidian homologs of mammalian thyroid peroxidase genes are expressed in the thyroid-equivalent region of the endostyle. *J Exp Zool* 1999;285(2):158-69.
142. Yu JK, Holland LZ, Jamrich M, Blitz IL, Hollan ND. *AmphiFoxE4*, an amphioxus winged helix/forkhead gene encoding a protein closely related to vertebrate thyroid transcription factor-2: expression during pharyngeal development. *Evol Dev* 2002;4(1):9-15.

143. Ogasawara M. Overlapping expression of amphioxus homologs of the thyroid transcription factor-1 gene and thyroid peroxidase gene in the endostyle: insight into evolution of the thyroid gland. *Dev Genes Evol* 2000;210(5):231-42.
144. Eagleson GW, Jenks BG, Van Overbeeke AP. The pituitary adrenocorticotropes originate from neural ridge tissue in *Xenopus laevis*. *J Embryol Exp Morphol* 1986;95:1-14.
145. Eagleson G, Ferreiro B, Harris WA. Fate of the anterior neural ridge and the morphogenesis of the *Xenopus* forebrain. *J Neurobiol* 1995;28(2):146-58.
146. Ooi GT, Tawadros N, Escalona RM. Pituitary cell lines and their endocrine applications. *Mol Cell Endocrinol* 2004;228(1-2):1-21.
147. Asa SL, Ezzat S. Molecular basis of pituitary development and cytogenesis. *Front Horm Res* 2004;32:1-19.
148. Gorbman A. Olfactory origins and evolution of the brain-pituitary endocrine system: facts and speculation. *Gen Comp Endocrinol* 1995;97(2):171-8.
149. Christiaen L, Burighel P, Smith WC, Vernier P, Bourrat F, Joly JS. Pitx genes in Tunicates provide new molecular insight into the evolutionary origin of pituitary. *Gene* 2002;287(1-2):107-13.
150. Bollner T, Holmberg K, Olsson R. A rostral sensory mechanism in *Oikopleura dioica* (Appendicularia). *Acta Zool.* 1986;67:235-241.
151. Gorbman A, Nozaki M, Kubokawa K. A brain-Hatschek's pit connection in amphioxus. *Gen Comp Endocrinol* 1999;113(2):251-4.
152. Lanctot C, Lamolet B, Drouin J. The bicoid-related homeoprotein Ptx1 defines the most anterior domain of the embryo and differentiates posterior from anterior lateral mesoderm. *Development* 1997;124(14):2807-17.
153. Drouin J, Lamolet B, Lamonerie T, Lanctot C, Tremblay JJ. The PTX family of homeodomain transcription factors during pituitary developments. *Mol Cell Endocrinol* 1998;140(1-2):31-6.
154. Olsson T, Asplund K, Hagg E. Pituitary-thyroid axis, prolactin and growth hormone in patients with acute stroke. *J Intern Med* 1990;228(3):287-90.
155. Burighel P, Lane NJ, Zaniolo G, Manni L. Neurogenic role of the neural gland in the development of the ascidian, *Botryllus schlosseri* (Tunicata, Urochordata). *J Comp Neurol* 1998;394(2):230-41.

156. Scully KM, Rosenfeld MG. Pituitary development: regulatory codes in mammalian organogenesis. *Science* 2002;295(5563):2231-5.
157. Slack JM. Developmental biology of the pancreas. *Development* 1995;121(6):1569-80.
158. Kim SK, MacDonald RJ. Signaling and transcriptional control of pancreatic organogenesis. *Curr Opin Genet Dev* 2002;12(5):540-7.
159. Olinski RP, Lundin LG, Hallbook F. Conserved synteny between the Ciona genome and human paralogs identifies large duplication events in the molecular evolution of the insulin-relaxin gene family. *Mol Biol Evol* 2006;23(1):10-22.
160. Jonsson J, Carlsson L, Edlund T, Edlund H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 1994;371(6498):606-9.
161. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM and others. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 2002;298(5601):2157-67.
162. Fernald RD. Casting a genetic light on the evolution of eyes. *Science* 2006;313(5795):1914-8.
163. Halder G, Callaerts P, Gehring WJ. New perspectives on eye evolution. *Curr Opin Genet Dev* 1995;5(5):602-9.
164. Halder G, Callaerts P, Gehring WJ. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* 1995;267(5205):1788-92.
165. Gehring WJ, Ikeo K. Pax 6: mastering eye morphogenesis and eye evolution. *Trends Genet* 1999;15(9):371-7.
166. Riyahi K, Shimeld SM. Chordate beta-gamma-crystallins and the evolutionary developmental biology of the vertebrate lens. *Comp Biochem Physiol B Biochem Mol Biol* 2007;147(3):347-57.
167. Wistow GJ, Piatigorsky J. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu Rev Biochem* 1988;57:479-504.
168. Lacalli T. Evolutionary biology: light on ancient photoreceptors. *Nature* 2004;432(7016):454-5.
169. Nordstrom K, Larsson TA, Larhammar D. Extensive duplications of phototransduction genes in early vertebrate evolution correlate with block (chromosome) duplications. *Genomics* 2004;83(5):852-72.

170. Koyanagi M, Terakita A, Kubokawa K, Shichida Y. Amphioxus homologs of G<sub>o</sub>-coupled rhodopsin and peropsin having 11-cis- and all-trans-retinals as their chromophores. *FEBS Lett* 2002;531(3):525-8.
171. Kusakabe T, Tsuda M. Photoreceptive systems in ascidians. *Photochem Photobiol* 2007;83(2):248-52.
172. Nakagawa M, Orii H, Yoshida N, Jojima E, Horie T, Yoshida R, Haga T, Tsuda M. Ascidian arrestin (Ci-ar), the origin of the visual and nonvisual arrestins of vertebrate. *Eur J Biochem* 2002;269(21):5112-8.
173. Rahr H. The ultrastructure of the blood-vessels of *branchiostoma-lanceolatum* (Pallas) (Cephalochordata): 1. Relations between blood-vessels, epithelia, basal laminae, and connective-tissue. *Zoomorphology* 1981;97:53-74.
174. Stokes D, Holland ND. The lancelet: Also known as "amphioxus," this curious creature has returned to the limelight as a player in the phylogenetic history of the vertebrates. *American Scientist* 1998;86(6):552-560.
175. Alanis J, Benitez D, Lopez E, Martinez-Palomo A. Impulse propagation through the cardiac junctional regions of the axolotl and the turtle. *Jpn J Physiol* 1973;23(2):149-64.
176. Black BL. Transcriptional pathways in second heart field development. *Semin Cell Dev Biol* 2007;18(1):67-76.
177. Christoffels VM, Habets PE, Franco D, Campione M, de Jong F, Lamers WH, Bao ZZ, Palmer S, Biben C, Harvey RP and others. Chamber formation and morphogenesis in the developing mammalian heart. *Dev Biol* 2000;223(2):266-78.
178. Moorman AF, Christoffels VM. Cardiac chamber formation: development, genes, and evolution. *Physiol Rev* 2003;83(4):1223-67.
179. Graham A. The evolution of the vertebrates--genes and development. *Curr Opin Genet Dev* 2000;10(6):624-8.
180. Chen JN, Fishman MC. Genetics of heart development. *Trends Genet* 2000;16(9):383-8.
181. Satou Y, Satoh N. Gene regulatory networks for the development and evolution of the chordate heart. *Genes Dev* 2006;20(19):2634-8.
182. Holland ND, Venkatesh TV, Holland LZ, Jacobs DK, Bodmer R. *AmphiNk2-tin*, an amphioxus homeobox gene expressed in myocardial progenitors: insights into evolution of the vertebrate heart. *Dev Biol* 2003;255(1):128-37.

183. Satou Y, Imai KS, Satoh N. The ascidian *Mesp* gene specifies heart precursor cells. *Development* 2004;131(11):2533-41.
184. Kitajima S, Takagi A, Inoue T, Saga Y. *MesP1* and *MesP2* are essential for the development of cardiac mesoderm. *Development* 2000;127(15):3215-26.
185. Becker DL, Cook JE, Davies CS, Evans WH, Gourdie RG. Expression of major gap junction connexin types in the working myocardium of eight chordates. *Cell Biol Int* 1998;22(7-8):527-43.
186. Hoffman JJ, Kaplan S, Libersohn RR. Prevalence of congenital heart disease. *Am Heart J* 2004;147(3):425-39.
187. Marguerie A, Bajolle F, Zaffran S, Brown NA, Dickson C, Buckingham ME, Kelly RG. Congenital heart defects in *Fgfr2-IIIb* and *Fgf10* mutant mice. *Cardiovasc Res* 2006;71(1):50-60.
188. Doby T. Discoveries of blood circulation: from Aristotle to the times of da Vinci and Harvey. London/New York/Toronto: Abelard-Schuman; 1963. 285 p.
189. Munoz-Chapuli R, Carmona R, Guadix JA, Macias D, Perez-Pomares JM. The origin of the endothelial cells: an evo-devo approach for the invertebrate/vertebrate transition of the circulatory system. *Evol Dev* 2005;7(4):351-8.
190. Casley-Smith JR. The fine structure and functioning of tissue channels and lymphatics. *Lymphology* 1980;13(4):177-83.
191. Nishikawa M, Takeda K, Sato EF, Kuroki T, Inoue M. Nitric oxide regulates energy metabolism and *Bcl-2* expression in intestinal epithelial cells. *Am J Physiol* 1998;274(5 Pt 1):G797-801.
192. Watabe T, Nishihara A, Mishima K, Yamashita J, Shimizu K, Miyazawa K, Nishikawa S, Miyazono K. TGF-beta receptor kinase inhibitor enhances growth and integrity of embryonic stem cell-derived endothelial cells. *J Cell Biol* 2003;163(6):1303-11.
193. Ruppert E, Carle K. Morphology of metazoan circulatory systems. *Zoomorphology* 1983;103:193-208.
194. Moncada S, Higgs EA. Nitric oxide and the vascular endothelium. *Handb Exp Pharmacol* 2006(176 Pt 1):213-54.
195. Heino TI, Karpanen T, Wahlstrom G, Pulkkinen M, Eriksson U, Alitalo K, Roos C. The *Drosophila* VEGF receptor homolog is expressed in hemocytes. *Mech Dev* 2001;109(1):69-77.

196. Tettamanti G, Grimaldi A, Valvassori R, Rinaldi L, de Eguileor M. Vascular endothelial growth factor is involved in neoangiogenesis in *Hirudo medicinalis* (Annelida, Hirudinea). *Cytokine* 2003;22(6):168-79.
197. Davidson EH, McClay DR, Hood L. Regulatory gene networks and the properties of the developmental process. *Proc Natl Acad Sci U S A* 2003;100(4):1475-80.
198. Hartenstein V, Mandal L. The blood/vascular system in a phylogenetic perspective. *Bioessays* 2006;28(12):1203-10.
199. Lyons MS, Bell B, Stainier D, Peters KG. Isolation of the zebrafish homologues for the tie-1 and tie-2 endothelium-specific receptor tyrosine kinases. *Dev Dyn* 1998;212(1):133-40.
200. Metchnikoff E. *Zakon zhizni. Po- povodu nektotorykh proizvedenii. gr. L. Tolstogo. Vest. Evropy* 1891;9:228-260.
201. Muller WE, Muller IM. Analysis of the sponge [Porifera] gene repertoire: implications for the evolution of the metazoan body plan. *Prog Mol Subcell Biol* 2003;37:1-33.
202. Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch TC. The innate immune repertoire in cnidaria--ancestral complexity and stochastic gene loss. *Genome Biol* 2007;8(4):R59.
203. Miller DJ, Ball EE. Cryptic complexity captured: the *Nematostella* genome reveals its secrets. *Trends Genet* 2008;24(1):1-4.
204. Medzhitov R, Janeway CA, Jr. Innate immunity: impact on the adaptive immune response. *Curr Opin Immunol* 1997;9(1):4-9.
205. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, Birdwell D, Alejos E, Silva M, Galanos C and others. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene. *Science* 1998;282(5396):2085-8.
206. Pancer Z, Cooper MD. The evolution of adaptive immunity. *Annu Rev Immunol* 2006;24:497-518.
207. Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, Buckley KM, Brockton V, Nair SV, Berney K and others. The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol* 2006;300(1):349-65.
208. Smith AB, Pisani D, Mackenzie-Dodds JA, Stockley B, Webster BL, Littlewood DT. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol Biol Evol* 2006;23(10):1832-51.

209. Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, Angerer LM, Arnone MI, Burgess DR, Burke RD and others. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 2006;314(5801):941-52.
210. Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *Proc Natl Acad Sci U S A* 2000;97(24):13156-61.
211. Loker ES, Adema CM, Zhang SM, Kepler TB. Invertebrate immune systems--not homogeneous, not simple, not well understood. *Immunol Rev* 2004;198:10-24.
212. Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. An ancient evolutionary origin of the Rag1/2 gene locus. *Proc Natl Acad Sci U S A* 2006;103(10):3728-33.
213. Yu XQ, Kanost MR. Binding of hemolin to bacterial lipopolysaccharide and lipoteichoic acid. An immunoglobulin superfamily member from insects as a pattern-recognition receptor. *Eur J Biochem* 2002;269(7):1827-34.
214. Cannon JP, Haire RN, Rast JP, Litman GW. The phylogenetic origins of the antigen-binding receptors and somatic diversification mechanisms. *Immunol Rev* 2004;200:12-22.
215. Cannon JP, Haire RN, Schnitker N, Mueller MG, Litman GW. Individual protochordates have unique immune-type receptor repertoires. *Curr Biol* 2004;14(12):R465-6.
216. Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, Marino R, Shida K, Ikeda M, Ikeda M and others. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics* 2003;55(8):570-81.
217. Shida K, Terajima D, Uchino R, Ikawa S, Ikeda M, Asano K, Watanabe T, Azumi K, Nonaka M, Satou Y and others. Hemocytes of *Ciona intestinalis* express multiple genes involved in innate immune host defense. *Biochem Biophys Res Commun* 2003;302(2):207-18.
218. Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 2004;430(6996):174-80.
219. Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, Cooper MD. Variable lymphocyte receptors in hagfish. *Proc Natl Acad Sci U S A* 2005;102(26):9224-9.

220. Meyerowitz EM. Plants compared to animals: the broadest comparative study of development. *Science* 2002;295(5559):1482-5.
221. Choe J, Kelker MS, Wilson IA. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* 2005;309(5734):581-5.
222. Hahn MW, Han MV, Han SG. Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet* 2007;3(11):e197.
223. Hughes AL, Friedman R. Loss of ancestral genes in the genomic evolution of *Ciona intestinalis*. *Evol Dev* 2005;7(3):196-200.
224. Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. *Dev Biol* 2006;300(1):90-107.
225. Hamada H, Meno C, Watanabe D, Saijoh Y. Establishment of vertebrate left-right asymmetry. *Nat Rev Genet* 2002;3(2):103-13.
226. Bone Q. Observations upon the living larva of amphioxus. *Publ. Staz. Zool. Napoli* 1958;30:458-471.
227. Morokuma J, Ueno M, Kawanishi H, Saiga H, Nishida H. HrNodal, the ascidian nodal-related gene, is expressed in the left side of the epidermis, and lies upstream of HrPitx. *Dev Genes Evol* 2002;212(9):439-46.
228. Takahashi S, Yokota C, Takano K, Tanegashima K, Onuma Y, Goto J, Asashima M. Two novel nodal-related genes initiate early inductive events in *Xenopus* Nieuwkoop center. *Development* 2000;127(24):5319-29.
229. Schier AF. Nodal signaling in vertebrate development. *Annu Rev Cell Dev Biol* 2003;19:589-621.
230. Shiratori H, Sakuma R, Watanabe M, Hashiguchi H, Mochida K, Sakai Y, Nishino J, Saijoh Y, Whitman M, Hamada H. Two-step regulation of left-right asymmetric expression of Pitx2: initiation by nodal signaling and maintenance by Nkx2. *Mol Cell* 2001;7(1):137-49.
231. Swalla BJ, Cameron CB, Corley LS, Garey JR. Urochordates are monophyletic within the deuterostomes. *Syst Biol* 2000;49(1):52-64.
232. Hudson C, Yasuo H. Patterning across the ascidian neural plate by lateral Nodal signalling sources. *Development* 2005;132(6):1199-210.

233. Duboc V, Rottinger E, Lapraz F, Besnardeau L, Lepage T. Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. *Dev Cell* 2005;9(1):147-58.
234. Hibino T, Nishino A, Amemiya S. Phylogenetic correspondence of the body axes in bilaterians is revealed by the right-sided expression of Pitx genes in echinoderm larvae. *Dev Growth Differ* 2006;48(9):587-95.
235. Yu JK, Holland LZ, Holland ND. An amphioxus nodal gene (AmphiNodal) with early symmetrical expression in the organizer and mesoderm and later asymmetrical expression associated with left-right axis formation. *Evol Dev* 2002;4(6):418-25.
236. Campione M, Steinbeisser H, Schweickert A, Deissler K, van Bebber F, Lowe LA, Nowotschin S, Viebahn C, Haffter P, Kuehn MR and others. The homeobox gene Pitx2: mediator of asymmetric left-right signaling in vertebrate heart and gut looping. *Development* 1999;126(6):1225-34.
237. Boorman CJ, Shimeld SM. Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. *Evol Dev* 2002;4(5):354-65.
238. Duboc V, Rottinger E, Besnardeau L, Lepage T. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. *Dev Cell* 2004;6(3):397-410.
239. Gage PJ, Suh H, Camper SA. The bicoid-related Pitx gene family in development. *Mamm Genome* 1999;10(2):197-200.
240. Boorman CJ, Shimeld SM. Cloning and expression of a Pitx homeobox gene from the lamprey, a jawless vertebrate. *Dev Genes Evol* 2002;212(7):349-53.
241. Zhang XM, Ramalho-Santos M, McMahon AP. Smoothed mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R symmetry by the mouse node. *Cell* 2001;106(2):781-92.
242. Shimeld SM. The evolution of the hedgehog gene family in chordates: insights from amphioxus hedgehog. *Dev Genes Evol* 1999;209(1):40-7.
243. Takatori N, Satou Y, Satoh N. Expression of hedgehog genes in Ciona intestinalis embryos. *Mech Dev* 2002;116(1-2):235-8.
244. Canbay A, Gieseler RK, Gores GJ, Gerken G. The relationship between apoptosis and non-alcoholic fatty liver disease: an evolutionary cornerstone turned pathogenic. *Z Gastroenterol* 2005;43(2):211-7.

245. Guillou H, Martin PG, Pineau T. Transcriptional regulation of hepatic fatty acid metabolism. *Subcell Biochem* 2008;49:3-47.
246. Ruppert E. Hemichordata, chaetognatha, and the invertebrate chordates. In: Harrison FW RE, editor. *Microscopic anatomy of invertebrates*. Volume 15. New-York: Wiley-Liss; 1997. p 349-504.
247. Liang Y, Zhang S, Lun L, Han L. Presence and localization of antithrombin and its regulation after acute lipopolysaccharide exposure in amphioxus, with implications for the origin of vertebrate liver. *Cell Tissue Res* 2006;323(3):537-41.
248. Wang Y, Zhang Y, Zhang S, Tian J, Jiang S. Tissue- and stage-specific expression of a fatty acid binding protein-like gene from amphioxus *Branchiostoma belcheri*. *Acta Biochim Pol* 2008;55(1):27-34.
249. Chmurzynska A. The multigene family of fatty acid-binding proteins (FABPs): function, structure and polymorphism. *J Appl Genet* 2006;47(1):39-48.
250. Reschly EJ, Bainy AC, Mattos JJ, Hagey LR, Bahary N, Mada SR, Ou J, Venkataramanan R, Krasowski MD. Functional evolution of the vitamin D and pregnane X receptors. *BMC Evol Biol* 2007;7:222.
251. Verslycke T, Goldstone JV, Stegeman JJ. Isolation and phylogeny of novel cytochrome P450 genes from tunicates (*Ciona* spp.): a CYP3 line in early deuterostomes? *Mol Phylogenet Evol* 2006;40(3):760-71.
252. Gutierrez E, Wiggins D, Fielding B, Gould AP. Specialized hepatocyte-like cells regulate *Drosophila* lipid metabolism. *Nature* 2007;445(7125):275-80.
253. Whitfield GK, Jurutka PW, Haussler CA, Haussler MR. Steroid hormone receptors: evolution, ligands, and molecular basis of biologic function. *J Cell Biochem* 1999;Suppl 32-33:110-22.
254. Bertrand S, Brunet FG, Escriva H, Parmentier G, Laudet V, Robinson-Rechavi M. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. *Mol Biol Evol* 2004;21(10):1923-37.
255. Schwabe JW, Teichmann SA. Nuclear receptors: the evolution of diversity. *Sci STKE* 2004;2004(217):pe4.
256. Baker ME. Steroid receptor phylogeny and vertebrate origins. *Mol Cell Endocrinol* 1997;135(2):101-7.
257. Thornton JW. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc Natl Acad Sci U S A* 2001;98(10):5671-6.

258. Baker ME. Evolution of adrenal and sex steroid action in vertebrates: a ligand-based mechanism for complexity. *Bioessays* 2003;25(4):396-400.
259. Hall JM, Couse JF, Korach KS. The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J Biol Chem* 2001;276(40):36869-72.
260. Klotz DM, Hewitt SC, Ciana P, Raviscioni M, Lindzey JK, Foley J, Maggi A, DiAugustine RP, Korach KS. Requirement of estrogen receptor- $\alpha$  in insulin-like growth factor-1 (IGF-1)-induced uterine responses and in vivo evidence for IGF-1/estrogen receptor cross-talk. *J Biol Chem* 2002;277(10):8531-7.
261. Kuiper GG, Carlsson B, Grandien K, Enmark E, Haggblad J, Nilsson S, Gustafsson JA. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors  $\alpha$  and  $\beta$ . *Endocrinology* 1997;138(3):863-70.
262. Brivanlou AH, Darnell JE, Jr. Signal transduction and the control of gene expression. *Science* 2002;295(5556):813-8.
263. Shang Y, Myers M, Brown M. Formation of the androgen receptor transcription complex. *Mol Cell* 2002;9(3):601-10.
264. Bridgham JT, Carroll SM, Thornton JW. Evolution of hormone-receptor complexity by molecular exploitation. *Science* 2006;312(5770):97-101.
265. Thornton JW, DeSalle R. Gene family evolution and homology: genomics meets phylogenetics. *Annu Rev Genomics Hum Genet* 2000;1:41-73.
266. Paris M, Pettersson K, Schubert M, Bertrand S, Pongratz I, Escriva H, Laudet V. An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. *BMC Evol Biol* 2008;8:219.
267. Bridgham JT, Brown JE, Rodriguez-Mari A, Catchen JM, Thornton JW. Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS Genet* 2008;4(9):e1000191.
268. Prockop DJ, Kivirikko KI. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* 1995;64:403-34.
269. Myllyharju J, Kivirikko KI. Collagens and collagen-related diseases. *Ann Med* 2001;33(1):7-21.
270. Boot-Handford RP, Tuckwell DS. Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *Bioessays* 2003;25(2):142-51.

271. Lees JF, Tasab M, Bulleid NJ. Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. *Embo J* 1997;16(5):908-16.
272. Ohno S. Evolution from fish to mammals by gene duplication. *Hereditas* 1968;59:169-187.
273. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 2006;440(7082):341-5.
274. Orengo CA, Thornton JM. Protein families and their evolution-a structural perspective. *Annu Rev Biochem* 2005;74:867-900.
275. Bornberg-Bauer E, Beaussart F, Kummerfeld SK, Teichmann SA, Weiner J, 3rd. The evolution of domain arrangements in proteins and interaction networks. *Cell Mol Life Sci* 2005;62(4):435-45.
276. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. *Genome Biol* 2002;3(2):RESEARCH0008.
277. Conant GC, Wagner A. Asymmetric sequence divergence of duplicate genes. *Genome Res* 2003;13(9):2052-8.
278. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV and others. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 2007;317(5834):86-94.
279. Hillier LW, Fulton RS, Fulton LA, Graves TA, Pepin KH, Wagner-McPherson C, Layman D, Maas J, Jaeger S, Walker R and others. The DNA sequence of human chromosome 7. *Nature* 2003;424(6945):157-64.
280. Prachumwat A, Li WH. Gene number expansion and contraction in vertebrate genomes with respect to invertebrate genomes. *Genome Res* 2008;18(2):221-32.
281. Liu M, Grigoriev A. Protein domains correlate strongly with exons in multiple eukaryotic genomes--evidence of exon shuffling? *Trends Genet* 2004;20(9):399-403.
282. Schmidt EE, Davies CJ. The origins of polypeptide domains. *Bioessays* 2007;29(3):262-70.
283. Sorek R. The birth of new exons: mechanisms and evolutionary consequences. *Rna* 2007;13(10):1603-8.

284. Babushok DV, Ostertag EM, Kazazian HH, Jr. Current topics in genome evolution: molecular mechanisms of new gene formation. *Cell Mol Life Sci* 2007;64(5):542-54.
285. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999;151(4):1531-45.
286. Chothia C, Gough J, Vogel C, Teichmann SA. Evolution of the protein repertoire. *Science* 2003;300(5626):1701-3.
287. Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T and others. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 1998;282(5396):2022-8.
288. Aravind L, Subramanian G. Origin of multicellular eukaryotes - insights from proteome comparisons. *Curr Opin Genet Dev* 1999;9(6):688-94.
289. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W and others. Initial sequencing and analysis of the human genome. *Nature* 2001;409(6822):860-921.
290. Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T and others. Ensembl 2006. *Nucleic Acids Res* 2006;34(Database issue):D556-61.
291. Vogel C, Chothia C. Protein family expansions and biological complexity. *PLoS Comput Biol* 2006;2(5):e48.
292. Vogel C, Teichmann SA, Chothia C. The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. *Development* 2003;130(25):6317-28.
293. McLysaght A, Hokamp K, Wolfe KH. Extensive genomic duplication during early chordate evolution. *Nat Genet* 2002;31(2):200-4.
294. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* 2004;38:615-43.
295. Garcia-Fernandez J, Holland PW. Archetypal organization of the amphioxus Hox gene cluster. *Nature* 1994;370(6490):563-6.
296. Leveugle M, Prat K, Popovici C, Birnbaum D, Coulier F. Phylogenetic analysis of *Ciona intestinalis* gene superfamilies supports the hypothesis of successive gene expansions. *J Mol Evol* 2004;58(2):168-81.

297. Holland LZ, Gibson-Brown JJ. The *Ciona intestinalis* genome: when the constraints are off. *Bioessays* 2003;25(6):529-32.
298. Long M, Rosenberg C, Gilbert W. Intron phase correlations and the evolution of the intron/exon structure of genes. *Proc Natl Acad Sci U S A* 1995;92(26):12495-9.
299. Patthy L. Modular assembly of genes and the evolution of new functions. *Genetica* 2003;118(2-3):217-31.
300. Kipersztok S, Osawa GA, Liang LF, Modi WS, Dean J. POM-ZP3, a bipartite transcript derived from human ZP3 and a POM121 homologue. *Genomics* 1995;25(2):354-9.
301. Kennerson ML, Nassif NT, Nicholson GA. Genomic structure and physical mapping of C17orf1: a gene associated with the proximal element of the CMT1A-REP binary repeat. *Genomics* 1998;53(1):110-2.
302. Courseaux A, Nahon JL. Birth of two chimeric genes in the Hominidae lineage. *Science* 2001;291(5507):1293-7.
303. Koonin EV, Aravind L. The NACHT family - a new group of predicted NTPases implicated in apoptosis and MHC transcription activation. *Trends Biochem Sci* 2000;25(5):223-4.
304. Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol Biol Evol* 2002;19(3):256-62.
305. Tordai H, Nagy A, Farkas K, Banyai L, Patthy L. Modules, multidomain proteins and organismic complexity. *Febs J* 2005;272(19):5064-78.
306. Brosius J, Gould SJ. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc Natl Acad Sci U S A* 1992;89(22):10706-10.
307. Zhang XH, Chasin LA. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc Natl Acad Sci U S A* 2006;103(36):13427-32.
308. Kim E, Magen A, Ast G. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* 2007;35(1):125-31.
309. Johnson JM, Castle J, Garrett-Engle P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 2003;302(5653):2141-4.

310. Brookfield JF. The ecology of the genome - mobile DNA elements and their hosts. *Nat Rev Genet* 2005;6(2):128-36.
311. Deininger PL, Moran JV, Batzer MA, Kazazian HH, Jr. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 2003;13(6):651-8.
312. Brosius J. Retroposons--seeds of evolution. *Science* 1991;251(4995):753.
313. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 2008;9(5):397-405.
314. Miller WJ, McDonald JF, Nouaud D, Anxolabehere D. Molecular domestication--more than a sporadic episode in evolution. *Genetica* 1999;107(1-3):197-207.
315. Jordan IK, Rogozin IB, Glazko GV, Koonin EV. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet* 2003;19(2):68-72.
316. Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL. Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res* 2005;110(1-4):342-52.
317. Thornburg BG, Gotea V, Makalowski W. Transposable elements as a significant source of transcription regulating signals. *Gene* 2006;365:104-10.
318. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP and others. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 2004;430(6995):88-93.
319. Fraser HB. Modularity and evolutionary constraint on proteins. *Nat Genet* 2005;37(4):351-2.
320. Evlampiev K, Isambert H. Modeling protein network evolution under genome duplication and domain shuffling. *BMC Syst Biol* 2007;1:49.
321. Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, Betel D, Bobechko B, Boutilier K, Burgess E and others. The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res* 2005;33(Database issue):D418-24.
322. Kleinjan DA, Bancewicz RM, Gautier P, Dahm R, Schonhaler HB, Damante G, Seawright A, Hever AM, Yeyati PL, van Heyningen V and others. Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. *PLoS Genet* 2008;4(2):e29.

323. Woolfe A, Elgar G. Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. *Genome Biol* 2007;8(4):R53.
324. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W and others. Global variation in copy number in the human genome. *Nature* 2006;444(7118):444-54.
325. Ionita-Laza I, Laird NM, Raby BA, Weiss ST, Lange C. On the frequency of copy number variants. *Bioinformatics* 2008;24(20):2350-5.
326. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 2007;8(8):639-46.
327. Frith MC, Pheasant M, Mattick JS. The amazing complexity of the human transcriptome. *Eur J Hum Genet* 2005;13(8):894-7.
328. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 2007;29(3):288-99.
329. Sempere LF, Cole CN, McPeck MA, Peterson KJ. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zool B Mol Dev Evol* 2006;306(6):575-88.
330. Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ. MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci U S A* 2008;105(8):2946-50.
331. Spitz F, Gonzalez F, Peichel C, Vogt TF, Duboule D, Zakany J. Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev* 2001;15(17):2209-14.
332. Yekta S, Tabin CJ, Bartel DP. MicroRNAs in the Hox network: an apparent link to posterior prevalence. *Nat Rev Genet* 2008;9(10):789-96.
333. Thornton JW, Need E, Crews D. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* 2003;301(5640):1714-7.
334. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 2007;317(5844):1544-8.
335. Basu MK, Carmel L, Rogozin IB, Koonin EV. Evolution of protein domain promiscuity in eukaryotes. *Genome Res* 2008;18(3):449-61.

336. Wagner GP, Lynch VJ. The gene regulatory logic of transcription factor evolution. *Trends Ecol Evol* 2008;23(7):377-85.
337. Turner GF. Adaptive radiation of cichlid fish. *Curr Biol* 2007;17(19):R827-31.
338. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science* 1975;188(4184):107-16.
339. Consortium TCSaA. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 2005;437(7055):69-87.
340. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 2004;428(6984):717-23.
341. Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 2007;61(5):995-1016.
342. Hershberg R, Margalit H. Co-evolution of transcription factors and their targets depends on mode of regulation. *Genome Biol* 2006;7(7):R62.
343. Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* 2008;134(4):610-23.
344. Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V and others. Human-specific gain of function in a developmental enhancer. *Science* 2008;321(5894):1346-50.
345. Hernandez N. TBP, a universal eukaryotic transcription factor? *Genes Dev* 1993;7(7B):1291-308.
346. Green MR. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci* 2000;25(2):59-63.
347. Hernandez N. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* 2001;276(29):26733-6.
348. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. *Genes Dev* 1996;10(21):2657-83.
349. Dvir A, Conaway JW, Conaway RC. Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr Opin Genet Dev* 2001;11(2):209-14.
350. Thomas MC, Chiang CM. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* 2006;41(3):105-78.

351. Hobbs NK, Bondareva AA, Barnett S, Capecchi MR, Schmidt EE. Removing the vertebrate-specific TBP N terminus disrupts placental beta2m-dependent interactions with the maternal immune system. *Cell* 2002;110(1):43-54.
352. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005;6(3):197-208.
353. Camps M, Herman A, Loh E, Loeb LA. Genetic constraints on protein evolution. *Crit Rev Biochem Mol Biol* 2007;42(5):313-26.
354. Ranganayakulu G, Elliott DA, Harvey RP, Olson EN. Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice. *Development* 1998;125(16):3037-48.
355. Galant R, Carroll SB. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 2002;415(6874):910-3.
356. Ronshaugen M, McGinnis N, McGinnis W. Hox protein mutation and macroevolution of the insect body plan. *Nature* 2002;415(6874):914-7.
357. Schmidt EE, Bondareva AA, Radke JR, Capecchi MR. Fundamental cellular processes do not require vertebrate-specific sequences within the TATA-binding protein. *J Biol Chem* 2003;278(8):6168-74.
358. Pugh BF. Control of gene expression through regulation of the TATA-binding protein. *Gene* 2000;255(1):1-14.
359. Kraemer SM, Ranallo RT, Ogg RC, Stargell LA. TFIIA interacts with TFIID via association with TATA-binding protein and TAF40. *Mol Cell Biol* 2001;21(5):1737-46.
360. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* 2008;20(3):253-9.
361. Martinez E, Chiang CM, Ge H, Roeder RG. TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *Embo J* 1994;13(13):3115-26.
362. Albright SR, Tjian R. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* 2000;242(1-2):1-13.
363. Kaiser K, Meisterernst M. The human general co-factors. *Trends Biochem Sci* 1996;21(9):342-5.
364. Ptashne M, Gann A. Transcriptional activation by recruitment. *Nature* 1997;386(6625):569-77.

365. Martinez E. Multi-protein complexes in eukaryotic gene transcription. *Plant Mol Biol* 2002;50(6):925-47.
366. Malik S, Roeder RG. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci* 2005;30(5):256-63.
367. Hochheimer A, Tjian R. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev* 2003;17(11):1309-20.
368. Cormack BP, Struhl K. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell* 1992;69(4):685-96.
369. Schultz MC, Reeder RH, Hahn S. Variants of the TATA-binding protein can distinguish subsets of RNA polymerase I, II, and III promoters. *Cell* 1992;69(4):697-702.
370. Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 1996;21(9):327-35.
371. Persengiev SP, Zhu X, Dixit BL, Maston GA, Kittler EL, Green MR. TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc Natl Acad Sci U S A* 2003;100(25):14887-91.
372. Sharkey M, Graba Y, Scott MP. Hox genes in evolution: protein surfaces and paralog groups. *Trends Genet* 1997;13(4):145-51.
373. Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. *Adv Protein Chem* 2002;62:25-49.
374. Uversky VN. Cracking the folding code. Why do some proteins adopt partially folded conformations, whereas other don't? *FEBS Lett* 2002;514(2-3):181-3.
375. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 1999;293(2):321-31.
376. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci* 2002;27(10):527-33.
377. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 2005;579(15):3346-54.
378. Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 2008;33(1):2-8.

379. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 2004;337(3):635-45.
380. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins* 2001;42(1):38-48.
381. Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. *Proteins* 2003;52(4):573-84.
382. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 2008;9 Suppl 1:S1.
383. Nikolov DB, Hu SH, Lin J, Gasch A, Hoffmann A, Horikoshi M, Chua NH, Roeder RG, Burley SK. Crystal structure of TFIID TATA-box binding protein. *Nature* 1992;360(6399):40-6.
384. Burley SK, Roeder RG. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* 1996;65:769-99.
385. Buratowski RM, Downs J, Buratowski S. Interdependent interactions between TFIIB, TATA binding protein, and DNA. *Mol Cell Biol* 2002;22(24):8735-43.
386. Asturias FJ. RNA polymerase II structure, and organization of the preinitiation complex. *Curr Opin Struct Biol* 2004;14(2):121-9.
387. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* 2004;11(5):394-403.
388. Peterson MG, Tanese N, Pugh BF, Tjian R. Functional domains and upstream activation properties of cloned human TATA binding protein. *Science* 1990;248(4963):1625-30.
389. Hancock JM. Evolution of sequence repetition and gene duplications in the TATA-binding protein TBP (TFIID). *Nucleic Acids Res* 1993;21(12):2823-30.
390. Gazdag E, Rajkovic A, Torres-Padilla ME, Tora L. Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* 2007;134(1):51-62.
391. Bondareva AA, Schmidt EE. Early vertebrate evolution of the TATA-binding protein, TBP. *Mol Biol Evol* 2003;20(11):1932-9.

392. Neduva V, Russell RB. Linear motifs: evolutionary interaction switches. *FEBS Lett* 2005;579(15):3342-5.
393. Mayer BJ. SH3 domains: complexity in moderation. *J Cell Sci* 2001;114(Pt 7):1253-63.
394. Ladbury JE, Lemmon MA, Zhou M, Green J, Botfield MC, Schlessinger J. Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. *Proc Natl Acad Sci U S A* 1995;92(8):3199-203.
395. Karlin S, Burge C. Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. *Proc Natl Acad Sci U S A* 1996;93(4):1560-5.
396. Mar Alba M, Santibanez-Koref MF, Hancock JM. Amino acid reiterations in yeast are overrepresented in particular classes of proteins and show evidence of a slippage-like mutational process. *J Mol Evol* 1999;49(6):789-97.
397. Young ET, Sloan JS, Van Riper K. Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics* 2000;154(3):1053-68.
398. Fondon JW, 3rd, Garner HR. Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A* 2004;101(52):18058-63.
399. Alba MM, Guigo R. Comparative analysis of amino acid repeats in rodents and humans. *Genome Res* 2004;14(4):549-54.
400. Andreeva A, Prlic A, Hubbard TJ, Murzin AG. SISYPHUS--structural alignments for proteins with non-trivial relationships. *Nucleic Acids Res* 2007;35(Database issue):D253-9.
401. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN and others. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* 2007;35(Database issue):D786-93.
402. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT, Jr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 1996;35(43):13709-15.
403. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 2002;323(3):573-84.
404. Friedman MJ, Shah AG, Fang ZH, Ward EG, Warren ST, Li S, Li XJ. Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration. *Nat Neurosci* 2007;10(12):1519-28.

405. Turjanski AG, Gutkind JS, Best RB, Hummer G. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput Biol* 2008;4(4):e1000060.
406. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* 1996;93(21):11504-9.
407. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW and others. Intrinsically disordered protein. *J Mol Graph Model* 2001;19(1):26-59.
408. Bustos DM, Iglesias AA. Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins* 2006;63(1):35-42.
409. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 2002;12(1):54-60.
410. Remenyi A, Good MC, Lim WA. Docking interactions in protein kinase and phosphatase networks. *Curr Opin Struct Biol* 2006;16(6):676-85.
411. Radhakrishnan I, Perez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* 1997;91(6):741-52.
412. Patil A, Nakamura H. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett* 2006;580(8):2041-5.
413. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* 2006;2(8):e100.
414. Ekman D, Light S, Bjorklund AK, Elofsson A. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol* 2006;7(6):R45.
415. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 2006;5(11):2985-95.
416. James LC, Tawfik DS. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem Sci* 2003;28(7):361-8.

417. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 2005;18(5):343-84.
418. Hoshiyama D, Kuma K, Miyata T. Extremely reduced evolutionary rate of TATA-box binding protein in higher vertebrates and its evolutionary implications. *Gene* 2001;280(1-2):169-73.
419. Horikoshi M, Yamamoto T, Ohkuma Y, Weil PA, Roeder RG. Analysis of structure-function relationships of yeast TATA box binding factor TFIID. *Cell* 1990;61(7):1171-8.
420. Lieberman PM, Schmidt MC, Kao CC, Berk AJ. Two distinct domains in the yeast transcription factor IID and evidence for a TATA box-induced conformational change. *Mol Cell Biol* 1991;11(1):63-74.
421. Kuddus R, Schmidt MC. Effect of the non-conserved N-terminus on the DNA binding activity of the yeast TATA binding protein. *Nucleic Acids Res* 1993;21(8):1789-96.
422. Lescure A, Lutz Y, Eberhard D, Jacq X, Krol A, Grummt I, Davidson I, Chambon P, Tora L. The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters. *Embo J* 1994;13(5):1166-75.
423. Pugh BF, Tjian R. Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* 1990;61(7):1187-97.
424. Trivedi A, Vilalta A, Gopalan S, Johnson DL. TATA-binding protein is limiting for both TATA-containing and TATA-lacking RNA polymerase III promoters in *Drosophila* cells. *Mol Cell Biol* 1996;16(12):6909-16.
425. Campbell KM, Ranallo RT, Stargell LA, Lumb KJ. Reevaluation of transcriptional regulation by TATA-binding protein oligomerization: predominance of monomers. *Biochemistry* 2000;39(10):2633-8.
426. Goppelt A, Meisterernst M. Characterization of the basal inhibitor of class II transcription NC2 from *Saccharomyces cerevisiae*. *Nucleic Acids Res* 1996;24(22):4450-5.
427. Goppelt A, Stelzer G, Lottspeich F, Meisterernst M. A mechanism for repression of class II gene transcription through specific binding of NC2 to TBP-promoter complexes via heterodimeric histone fold domains. *Embo J* 1996;15(12):3105-16.

428. Zhou Q, Berk AJ. The yeast TATA-binding protein (TBP) core domain assembles with human TBP-associated factors into a functional TFIID complex. *Mol Cell Biol* 1995;15(1):534-9.
429. Mittal V, Hernandez N. Role for the amino-terminal region of human TBP in U6 snRNA transcription. *Science* 1997;275(5303):1136-40.
430. Yamada M, Tsuji S, Takahashi H. Pathology of CAG repeat diseases. *Neuropathology* 2000;20(4):319-25.
431. Okazawa H. Polyglutamine diseases: a transcription disorder? *Cell Mol Life Sci* 2003;60(7):1427-39.
432. Shatunov A, Fridman EA, Pagan FI, Leib J, Singleton A, Hallett M, Goldfarb LG. Small de novo duplication in the repeat region of the TATA-box-binding protein gene manifest with a phenotype similar to variant Creutzfeldt-Jakob disease. *Clin Genet* 2004;66(6):496-501.
433. van Roon-Mom WM, Reid SJ, Faull RL, Snell RG. TATA-binding protein in neurodegenerative disease. *Neuroscience* 2005;133(4):863-72.
434. Das D, Scovell WM. The binding interaction of HMG-1 with the TATA-binding protein/TATA complex. *J Biol Chem* 2001;276(35):32597-605.
435. Prigge JR, Schmidt EE. HAP1 can sequester a subset of TBP in cytoplasmic inclusions via specific interaction with the conserved TBP(CORE). *BMC Mol Biol* 2007;8:76.
436. Stewart JR, Thompson MB. Evolution of placentation among squamate reptiles: recent research and future directions. *Comp Biochem Physiol A Mol Integr Physiol* 2000;127(4):411-31.
437. Vogel P. The current molecular phylogeny of Eutherian mammals challenges previous interpretations of placental evolution. *Placenta* 2005;26(8-9):591-6.
438. Haines AN, Flajnik MF, Wourms JP. Histology and immunology of the placenta in the Atlantic sharpnose shark, *Rhizoprionodon terraenovae*. *Placenta* 2006;27(11-12):1114-23.
439. Cross JC, Baczyk D, Dobric N, Hemberger M, Hughes M, Simmons DG, Yamamoto H, Kingdom JC. Genes, development and evolution of the placenta. *Placenta* 2003;24(2-3):123-30.
440. Arck P, Hansen PJ, Mulac Jericevic B, Piccinni MP, Szekeres-Bartho J. Progesterone during pregnancy: endocrine-immune cross talk in mammalian species and the role of stress. *Am J Reprod Immunol* 2007;58(3):268-79.

441. Hemberger M, Cross JC, Ropers HH, Lehrach H, Fundele R, Himmelbauer H. UniGene cDNA array-based monitoring of transcriptome changes during mouse placental development. *Proc Natl Acad Sci U S A* 2001;98(23):13126-31.
442. Cross JC, Nakano H, Natale DR, Simmons DG, Watson ED. Branching morphogenesis during development of placental villi. *Differentiation* 2006;74(7):393-401.
443. Hoffman A, Sinn E, Yamamoto T, Wang J, Roy A, Horikoshi M, Roeder RG. Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature* 1990;346(6282):387-90.
444. Lee M, Struhl K. Multiple functions of the nonconserved N-terminal domain of yeast TATA-binding protein. *Genetics* 2001;158(1):87-93.
445. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P and others. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;420(6915):520-62.
446. Consortium TCGSaA. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 2005;437(7055):69-87.
447. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. *Nature* 2004;430(6995):85-8.
448. Carroll SB. Evolution at two levels: on genes and form. *PLoS Biol* 2005;3(7):e245.
449. Wray GA. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 2007;8(3):206-16.
450. Gaffney DJ, Blekhman R, Majewski J. Selective constraints in experimentally defined primate regulatory regions. *PLoS Genet* 2008;4(8):e1000157.
451. Carroll SB. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 2000;101(6):577-80.
452. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* 2003;20(9):1377-419.
453. Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet* 2008;4(11):e1000271.

454. Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* 2006;440(7087):1050-3.
455. Shimada M, Ohbayashi T, Ishida M, Nakadai T, Makino Y, Aoki T, Kawata T, Suzuki T, Matsuda Y, Tamura T. Analysis of the chicken TBP-like protein(tlp) gene: evidence for a striking conservation of vertebrate TLPs and for a close relationship between vertebrate tbp and tlp genes. *Nucleic Acids Res* 1999;27(15):3146-52.
456. Horton AC, Mahadevan NR, Minguillon C, Osoegawa K, Rokhsar DS, Ruvinsky I, de Jong PJ, Logan MP, Gibson-Brown JJ. Conservation of linkage and evolution of developmental function within the Tbx2/3/4/5 subfamily of T-box genes: implications for the origin of vertebrate limbs. *Dev Genes Evol* 2008;218(11-12):613-28.
457. Caburet S, Cocquet J, Vaiman D, Veitia RA. Coding repeats and evolutionary "agility". *Bioessays* 2005;27(6):581-7.
458. Chasman DI, Flaherty KM, Sharp PA, Kornberg RD. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc Natl Acad Sci U S A* 1993;90(17):8174-8.
459. DeDecker BS, O'Brien R, Fleming PJ, Geiger JH, Jackson SP, Sigler PB. The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J Mol Biol* 1996;264(5):1072-84.
460. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M and others. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 2008;36(Web Server issue):W465-9.
461. Jolley KA, Feil EJ, Chan MS, Maiden MC. Sequence type analysis and recombinational tests (START). *Bioinformatics* 2001;17(12):1230-1.
462. Angotzi AR, Ersland KM, Mungpakdee S, Stefansson S, Chourrout D. Independent and dynamic reallocation of pitx gene expression during vertebrate evolution, with emphasis on fish pituitary development. *Gene* 2008;417(1-2):19-26.
463. Tiozzo S, Christiaen L, Deyts C, Manni L, Joly JS, Burighel P. Embryonic versus blastogenetic development in the compound ascidian *Botryllus schlosseri*: insights from Pitx expression patterns. *Dev Dyn* 2005;232(2):468-78.
464. Tremblay JJ, Goodyer CG, Drouin J. Transcriptional properties of Ptx1 and Ptx2 isoforms. *Neuroendocrinology* 2000;71(5):277-86.

465. Shapiro MD, Bell MA, Kingsley DM. Parallel genetic origins of pelvic reduction in vertebrates. *Proc Natl Acad Sci U S A* 2006;103(37):13753-8.
466. Chang WY, Khosrowshahian F, Wolanski M, Marshall R, McCormick W, Perry S, Crawford MJ. Conservation of Pitx1 expression during amphibian limb morphogenesis. *Biochem Cell Biol* 2006;84(2):257-62.
467. Coyle SM, Huntingford FA, Peichel CL. Parallel evolution of Pitx1 underlies pelvic reduction in Scottish threespine stickleback (*Gasterosteus aculeatus*). *J Hered* 2007;98(6):581-6.
468. Picard C, Azeddine B, Moldovan F, Martel-Pelletier J, Moreau A. New emerging role of pitx1 transcription factor in osteoarthritis pathogenesis. *Clin Orthop Relat Res* 2007;462:59-66.
469. Lin CR, Kiousi C, O'Connell S, Briata P, Szeto D, Liu F, Izpisua-Belmonte JC, Rosenfeld MG. Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. *Nature* 1999;401(6750):279-82.
470. Liu C, Liu W, Palie J, Lu MF, Brown NA, Martin JF. Pitx2c patterns anterior myocardium and aortic arch vessels and is required for local cell movement into atrioventricular cushions. *Development* 2002;129(21):5081-91.
471. Kiousi C, Briata P, Baek SH, Rose DW, Hamblet NS, Herman T, Ohgi KA, Lin C, Gleiberman A, Wang J and others. Identification of a Wnt/Dvl/beta-Catenin --> Pitx2 pathway mediating cell-type-specific proliferation during development. *Cell* 2002;111(5):673-85.
472. Ganga M, Espinoza HM, Cox CJ, Morton L, Hjalt TA, Lee Y, Amendt BA. PITX2 isoform-specific regulation of atrial natriuretic factor expression: synergism and repression with Nkx2.5. *J Biol Chem* 2003;278(25):22437-45.
473. Rodriguez-Leon J, Rodriguez Esteban C, Marti M, Santiago-Josefat B, Dubova I, Rubiralta X, Izpisua Belmonte JC. Pitx2 regulates gonad morphogenesis. *Proc Natl Acad Sci U S A* 2008;105(32):11242-7.
474. Shi X, Bosenko DV, Zinkevich NS, Foley S, Hyde DR, Semina EV, Vihtelic TS. Zebrafish pitx3 is necessary for normal lens and retinal development. *Mech Dev* 2005;122(4):513-27.
475. Khosrowshahian F, Wolanski M, Chang WY, Fujiki K, Jacobs L, Crawford MJ. Lens and retina formation require expression of Pitx3 in *Xenopus* pre-lens ectoderm. *Dev Dyn* 2005;234(3):577-89.

476. Semina EV, Reiter RS, Murray JC. Isolation of a new homeobox gene belonging to the Pitx/Rieg family: expression during lens development and mapping to the aphakia region on mouse chromosome 19. *Hum Mol Genet* 1997;6(12):2109-16.
477. Lamba P, Hjalt TA, Bernard DJ. Novel forms of Paired-like homeodomain transcription factor 2 (PITX2): generation by alternative translation initiation and mRNA splicing. *BMC Mol Biol* 2008;9:31.
478. Prigge JR, Schmidt EE. Interaction of protein inhibitor of activated STAT (PIAS) proteins with the TATA-binding protein, TBP. *J Biol Chem* 2006;281(18):12260-9.
479. Gietz RD, Schiestl RH, Willems AR, Woods RA. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* 1995;11(4):355-60.
480. Hoffman CS, Winston F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 1987;57(2-3):267-72.
481. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403-10.
482. Celis JE. *Biology: A laboratory handbook*: Academic Press; 1994.
483. Bondareva AA, Capecchi MR, Iverson SV, Li Y, Lopez NI, Lucas O, Merrill GF, Prigge JR, Siders AM, Wakamiya M and others. Effects of thioredoxin reductase-1 deletion on embryogenesis and transcriptome. *Free Radic Biol Med* 2007;43(6):911-23.
484. Ota KG, Kuraku S, Kuratani S. Hagfish embryology with reference to the evolution of the neural crest. *Nature* 2007;446(7136):672-5.
485. Kuraku S, Kuratani S. Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zoolog Sci* 2006;23(12):1053-64.
486. Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Klein J. Molecular phylogeny of early vertebrates: monophyly of the agnathans as revealed by sequences of 35 genes. *Mol Biol Evol* 2003;20(2):287-92.
487. DeJong J, Bernstein R, Roeder RG. Human general transcription factor TFIIA: characterization of a cDNA encoding the small subunit and requirement for basal and activated transcription. *Proc Natl Acad Sci U S A* 1995;92(8):3313-7.
488. Tang H, Sun X, Reinberg D, Ebright RH. Protein-protein interactions in eukaryotic transcription initiation: structure of the preinitiation complex. *Proc Natl Acad Sci U S A* 1996;93(3):1119-24.

489. Mital R, Kobayashi R, Hernandez N. RNA polymerase III transcription from the human U6 and adenovirus type 2 VAI promoters has different requirements for human BRF, a subunit of human TFIIB. *Mol Cell Biol* 1996;16(12):7031-42.
490. Kassavetis GA, Soragni E, Driscoll R, Geiduschek EP. Reconfiguring the connectivity of a multiprotein complex: fusions of yeast TATA-binding protein with Brf1, and the function of transcription factor IIIB. *Proc Natl Acad Sci U S A* 2005;102(43):15406-11.
491. Dantonel JC, Murthy KG, Manley JL, Tora L. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* 1997;389(6649):399-402.
492. Rodriguez CI, Girones N, Fresno M. Cha, a basic helix-loop-helix transcription factor involved in the regulation of upstream stimulatory factor activity. *J Biol Chem* 2003;278(44):43135-45.
493. Prigge JR. Identification and characterization of novel protein-protein interactions with the basal transcription factor, TATA-Binding Protein. Bozeman: Montana State University; 2006. 117 p.
494. Millevoi S, Geraghty F, Idowu B, Tam JL, Antoniou M, Vagner S. A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. *EMBO Rep* 2002;3(9):869-74.
495. Hjalt TA, Amendt BA, Murray JC. PITX2 regulates procollagen lysyl hydroxylase (PLOD) gene expression: implications for the pathology of Rieger syndrome. *J Cell Biol* 2001;152(3):545-52.
496. Green PD, Hjalt TA, Kirk DE, Sutherland LB, Thomas BL, Sharpe PT, Snead ML, Murray JC, Russo AF, Amendt BA. Antagonistic regulation of Dlx2 expression by PITX2 and Msx2: implications for tooth development. *Gene Expr* 2001;9(6):265-81.
497. Quentien MH, Manfroid I, Moncet D, Gunz G, Muller M, Grino M, Enjalbert A, Pellegrini I. Pitx factors are involved in basal and hormone-regulated activity of the human prolactin promoter. *J Biol Chem* 2002;277(46):44408-16.
498. Chaney BA, Clark-Baldwin K, Dave V, Ma J, Rance M. Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome. *Biochemistry* 2005;44(20):7497-511.
499. Ruskoaho H. Atrial natriuretic peptide: synthesis, release, and metabolism. *Pharmacol Rev* 1992;44(4):479-602.

500. Houweling AC, van Borren MM, Moorman AF, Christoffels VM. Expression and regulation of the atrial natriuretic factor encoding gene *Nppa* during development and disease. *Cardiovasc Res* 2005;67(4):583-93.
501. Modi N, Betremieux P, Midgley J, Hartnoll G. Postnatal weight loss and contraction of the extracellular compartment is triggered by atrial natriuretic peptide. *Early Hum Dev* 2000;59(3):201-8.
502. Vollmar AM. The role of atrial natriuretic peptide in the immune system. *Peptides* 2005;26(6):1086-94.
503. Shiojima I, Komuro I, Oka T, Hiroi Y, Mizuno T, Takimoto E, Monzen K, Aikawa R, Akazawa H, Yamazaki T and others. Context-dependent transcriptional cooperation mediated by cardiac transcription factors *Csx/Nkx-2.5* and *GATA-4*. *J Biol Chem* 1999;274(12):8231-9.
504. Toro R, Saadi I, Kuburas A, Nemer M, Russo AF. Cell-specific activation of the atrial natriuretic factor promoter by *PITX2* and *MEF2A*. *J Biol Chem* 2004;279(50):52087-94.
505. Buckingham M, Meilhac S, Zaffran S. Building the mammalian heart from two sources of myocardial cells. *Nat Rev Genet* 2005;6(11):826-35.
506. Schweickert A, Campione M, Steinbeisser H, Blum M. *Pitx2* isoforms: involvement of *Pitx2c* but not *Pitx2a* or *Pitx2b* in vertebrate left-right asymmetry. *Mech Dev* 2000;90(1):41-51.
507. Franco D, Campione M. The role of *Pitx2* during cardiac development. Linking left-right signaling and congenital heart diseases. *Trends Cardiovasc Med* 2003;13(4):157-63.
508. Uversky VN. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell Mol Life Sci* 2003;60(9):1852-71.
509. Liu QX, Nakashima-Kamimura N, Ikeo K, Hirose S, Gojobori T. Compensatory change of interacting amino acids in the coevolution of transcriptional coactivator *MBF1* and *TATA-box-binding protein*. *Mol Biol Evol* 2007;24(7):1458-63.
510. Yeang CH, Haussler D. Detecting coevolution in and among protein domains. *PLoS Comput Biol* 2007;3(11):e211.
511. Jørgensen JM. *The Biology of Hagfishes*: Chapman & Hall; 1998. 578 p.
512. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 2006;22(23):2971-2.

513. Tvrdik P, Capecchi MR. Reversal of Hox1 gene subfunctionalization in the mouse. *Dev Cell* 2006;11(2):239-50.
514. Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 1987;51(3):503-12.
515. Schmidt EE, Schibler U. High accumulation of components of the RNA polymerase II transcription machinery in rodent spermatids. *Development* 1995;121(8):2373-83.
516. Koonin EV, Novozhilov AS. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* 2009;61(2):99-111.
517. Billiar TR, Curran, R.D. Antitumor defense system of the liver. In: Press C, editor. *Hepatocyte and Kupffer Cell Interactions*: CRC Press; 1992. p 79.
518. Canbay A, Bechmann L, Gerken G. Lipid metabolism in the liver. *Z Gastroenterol* 2007;45(1):35-41.
519. Sheehan DCaH, B.B. Chapter 8: Nuclear and cytoplasmic stains. Theory and practice of Histotechnology: C.V. Mosby; 1980.
520. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2008.
521. Loots GG, Ovcharenko I, Pachter L, Dubchak I, Rubin EM. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res* 2002;12(5):832-9.
522. Gotea V, Ovcharenko I. DiRE: identifying distant regulatory elements of co-expressed genes. *Nucleic Acids Res* 2008;36(Web Server issue):W133-9.
523. Coburn CT, Hajri T, Ibrahimi A, Abumrad NA. Role of CD36 in membrane transport and utilization of long-chain fatty acids by different tissues. *J Mol Neurosci* 2001;16(2-3):117-21; discussion 151-7.
524. Hajri T, Abumrad NA. Fatty acid transport across membranes: relevance to nutrition and metabolic pathology. *Annu Rev Nutr* 2002;22:383-415.
525. Villareal L. *Viruses And The Evolution Of Life*: ASM Press; 2005. 450 p.
526. Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J Virol* 2000;74(7):3321-9.

527. Lavillette D, Marin M, Ruggieri A, Mallet F, Cosset FL, Kabat D. The envelope glycoprotein of human endogenous retrovirus type W uses a divergent family of amino acid transporters/cell surface receptors. *J Virol* 2002;76(13):6442-52.
528. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S and others. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 2000;403(6771):785-9.
529. Dunlap KA, Palmarini M, Varela M, Burghardt RC, Hayashi K, Farmer JL, Spencer TE. Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proc Natl Acad Sci U S A* 2006;103(39):14390-5.
530. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT and others. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25(1):25-9.
531. Henderson SC, Kamdar MM, Bamezai A. Ly-6A.2 expression regulates antigen-specific CD4<sup>+</sup> T cell proliferation and cytokine production. *J Immunol* 2002;168(1):118-26.
532. Chitu V, Pixley FJ, Macaluso F, Larson DR, Condeelis J, Yeung YG, Stanley ER. The PCH family member MAYP/PSTPIP2 directly regulates F-actin bundling and enhances filopodia formation and motility in macrophages. *Mol Biol Cell* 2005;16(6):2947-59.
533. Baillie AG, Coburn CT, Abumrad NA. Reversible binding of long-chain fatty acids to purified FAT, the adipose CD36 homolog. *J Membr Biol* 1996;153(1):75-81.
534. Nicholson AC, Frieda S, Pearce A, Silverstein RL. Oxidized LDL binds to CD36 on human monocyte-derived macrophages and transfected cell lines. Evidence implicating the lipid moiety of the lipoprotein as the binding site. *Arterioscler Thromb Vasc Biol* 1995;15(2):269-75.
535. Calvo D, Gomez-Coronado D, Suarez Y, Lasuncion MA, Vega MA. Human CD36 is a high affinity receptor for the native lipoproteins HDL, LDL, and VLDL. *J Lipid Res* 1998;39(4):777-88.
536. Braun JE, Severson DL. Regulation of the synthesis, processing and translocation of lipoprotein lipase. *Biochem J* 1992;287 (Pt 2):337-47.
537. Greenwalt DE, Scheck SH, Rhinehart-Jones T. Heart CD36 expression is increased in murine models of diabetes and in mice fed a high fat diet. *J Clin Invest* 1995;96(3):1382-8.

538. Van Dyck F, Braem CV, Chen Z, Declercq J, Deckers R, Kim BM, Ito S, Wu MK, Cohen DE, Dewerchin M and others. Loss of the PlagL2 transcription factor affects lacteal uptake of chylomicrons. *Cell Metab* 2007;6(5):406-13.
539. Young B, Lowe JS, Stevens A, Heath JW, Deakin PJ. *Wheater's Functional Histology: A Text and Colour Atlas*: Churchill Livingstone; 2006. 448 p.
540. Adams LA, Angulo P. Recent concepts in non-alcoholic fatty liver disease. *Diabet Med* 2005;22(9):1129-33.
541. Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC, Grundy SM, Hobbs HH. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004;40(6):1387-95.
542. Hajri T, Han XX, Bonen A, Abumrad NA. Defective fatty acid uptake modulates insulin responsiveness and metabolic responses to diet in CD36-null mice. *J Clin Invest* 2002;109(10):1381-9.
543. Tilg H, Moschen AR. Insulin resistance, inflammation, and non-alcoholic fatty liver disease. *Trends Endocrinol Metab* 2008;19(10):371-9.
544. Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *Jama* 2002;287(3):356-9.
545. Lee NK, Sowa H, Hinoi E, Ferron M, Ahn JD, Confavreux C, Dacquin R, Mee PJ, McKee MD, Jung DY and others. Endocrine regulation of energy metabolism by the skeleton. *Cell* 2007;130(3):456-69.
546. Pospisilik JA, Knauf C, Joza N, Benit P, Orthofer M, Cani PD, Ebersberger I, Nakashima T, Sarao R, Neely G and others. Targeted deletion of AIF decreases mitochondrial oxidative phosphorylation and protects from obesity and diabetes. *Cell* 2007;131(3):476-91.
547. Raju B, Cryer PE. Maintenance of the postabsorptive plasma glucose concentration: insulin or insulin plus glucagon? *Am J Physiol Endocrinol Metab* 2005;289(2):E181-6.
548. Fox JG, Barthold S, Davisson M, Newcomer CE, Quimby FW, Smith A. *The Mouse in Biomedical Research*: Academic Press; 2006. 2192 p.
549. Jousse C, Averous J, Bruhat A, Carraro V, Mordier S, Fafournoux P. Amino acids as regulators of gene expression: molecular mechanisms. *Biochem Biophys Res Commun* 2004;313(2):447-52.

550. Goldstein JL, Brown MS. Regulation of the mevalonate pathway. *Nature* 1990;343(6257):425-30.
551. Espenshade PJ, Hughes AL. Regulation of sterol synthesis in eukaryotes. *Annu Rev Genet* 2007;41:401-27.
552. Huang da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC and others. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* 2007;35(Web Server issue):W169-75.
553. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27-30.
554. Gage PJ, Suh H, Camper SA. Dosage requirement of Pitx2 for development of multiple organs. *Development* 1999;126(20):4643-51.
555. Dutta-Roy AK. Cellular uptake of long-chain fatty acids: role of membrane-associated fatty-acid-binding/transport proteins. *Cell Mol Life Sci* 2000;57(10):1360-72.
556. Serghides L, Smith TG, Patel SN, Kain KC. CD36 and malaria: friends or foes? *Trends Parasitol* 2003;19(10):461-9.
557. Greenwood FC LJ, Stamp TCB. The plasma sugar, free fatty acid, cortisol and growth hormone response to insulin. *J Clin Invest* 1965;45(4):429-436.
558. Suh H, Gage PJ, Drouin J, Camper SA. Pitx2 is required at multiple stages of pituitary organogenesis: pituitary primordium formation and cell specification. *Development* 2002;129(2):329-37.
559. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* 2008;105(44):17046-9.
560. DeChiara TM, Efstratiadis A, Robertson EJ. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* 1990;345(6270):78-80.
561. Porter FD. Smith-Lemli-Opitz syndrome: pathogenesis, diagnosis and management. *Eur J Hum Genet* 2008;16(5):535-41.
562. Smith DW LL, Opitz JM: 64210-217. A newly recognized syndrome of multiple congenital anomalies. *J Pediatr* 1964;64:210-217.

563. Opitz JM. RSH/SLO ("Smith-Lemli-Opitz") syndrome: historical, genetic, and developmental considerations. *Am J Med Genet* 1994;50(4):344-6.
564. Tint GS, Irons M, Elias ER, Batta AK, Frieden R, Chen TS, Salen G. Defective cholesterol biosynthesis associated with the Smith-Lemli-Opitz syndrome. *N Engl J Med* 1994;330(2):107-13.
565. Starck L, Lovgren-Sandblom A, Bjorkhem I. Cholesterol treatment forever? The first Scandinavian trial of cholesterol supplementation in the cholesterol-synthesis defect Smith-Lemli-Opitz syndrome. *J Intern Med* 2002;252(4):314-21.
566. Azurdia RM, Anstey AV, Rhodes LE. Cholesterol supplementation objectively reduces photosensitivity in the Smith-Lemli-Opitz syndrome. *Br J Dermatol* 2001;144(1):143-5.
567. Sikora DM, Ruggiero M, Petit-Kekel K, Merkens LS, Connor WE, Steiner RD. Cholesterol supplementation does not improve developmental progress in Smith-Lemli-Opitz syndrome. *J Pediatr* 2004;144(6):783-91.
568. Lonnerdal B. Effects of maternal dietary intake on human milk composition. *J Nutr* 1986;116(4):499-513.
569. Demmers TA, Jones PJ, Wang Y, Krug S, Creutzinger V, Heubi JE. Effects of early cholesterol intake on cholesterol biosynthesis and plasma lipids among infants until 18 months of age. *Pediatrics* 2005;115(6):1594-601.
570. Friedman JM, Halaas JL. Leptin and the regulation of body weight in mammals. *Nature* 1998;395(6704):763-70.
571. Yamauchi T, Kamon J, Waki H, Terauchi Y, Kubota N, Hara K, Mori Y, Ide T, Murakami K, Tsuboyama-Kasaoka N and others. The fat-derived hormone adiponectin reverses insulin resistance associated with both lipodystrophy and obesity. *Nat Med* 2001;7(8):941-6.
572. Dray C, Knauf C, Daviaud D, Waget A, Boucher J, Buleon M, Cani PD, Attane C, Guigne C, Carpenne C and others. Apelin stimulates glucose utilization in normal and obese insulin-resistant mice. *Cell Metab* 2008;8(5):437-45.
573. Carpenne C, Dray C, Attane C, Valet P, Portillo MP, Churrua I, Milagro FI, Castan-Laurell I. Expanding role for the apelin/APJ system in physiopathology. *J Physiol Biochem* 2007;63(4):359-73.
574. Tatemoto K, Hosoya M, Habata Y, Fujii R, Kakegawa T, Zou MX, Kawamata Y, Fukusumi S, Hinuma S, Kitada C and others. Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. *Biochem Biophys Res Commun* 1998;251(2):471-6.

- 575. Masri B, Knibiehler B, Audigier Y. Apelin signalling: a promising pathway from cloning to pharmacology. *Cell Signal* 2005;17(4):415-26.
- 576. Martianov I, Viville S, Davidson I. RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science* 2002;298(5595):1036-9.

## APPENDICES

APPENDIX A

LUCIFERASE ASSAY PROTOCOL

### Introduction

The following protocol describes how to perform a transformation of mouse embryonic fibroblasts (MEF) followed by a luciferase assay in a 24-well dish format. Each condition should be performed in triplicate or quadruplicate and at least 4 independent clones in each genotype should be used. Each assay requires  $500 \times 10^3$  cells and 10 ug total of plasmid.

#### Note:

A great amount of DNA carrier (KS+ or pBS+ plasmid) is often required and adequate stocks should be readily available or should be prepared before starting the procedure.

### Equipment

Nucleofector II apparatus (Amaxa, Gaithersburg, MD).

2 mm gap cuvettes (can be reused >5 times).

Luminometer (Lumat LB 9507, EG&G Berthold).

Tubes for luminometer.

Pipettes, tips, rack, tubes.

Horizontal/orbital shaker.

Cell culture equipment and Bradford assay reagent (includes a plate reader for 595 nm reading and 96-well flat bottom well plates).

### Major reagents

D-luciferin (GoldBio, St. Louis, MO)

1 M DTT

1 M glycylglycine

Tris, EDTA, Triton X-100

Accutase

Bradford reagent

### Day 1: Set-up

Split/thin out cells (optional).

Make buffers (see below) if needed.

Determine total number of cells needed.

Prepare spreadsheet determining the amount of each plasmid needed for day 2 for a total of 10 ug/transformation.

Prepare a sheet to write down luciferase intensity.

### Day 2: Electroporation

1- Make sure all the needed cuvettes are lined up, plasmids thawed and electroporator is available.

Warm up (37°C) 500 uL of culture medium (see below) per transformation.

Make sure TM (see below) is at room temperature, not cold.

Add 2 mL of media + serum in each well of 24 well plate, put at 37C.

2- When the plates are at the desired density:

- Wash cells with PBS twice

- Add 1-2 mL accutase/10cm plate
- 37°C ~5 min
- Add 8 mL culture medium (with serum)
- Pipet up and down 8 times
- Count the cells, determine what volume is needed

=== from now on, no need to be sterile, everything is done on the bench top ===

- Spin in table top centrifuge (“mushroom”) setting 3 for 5-10 min
- Set up electroporator to program A24
- Resuspend cells at  $5 \times 10^6$  cells/ml in TM media
- Add plasmids to each cuvette
- Add 100 uL of cells/cuvette, on top of plasmids
- Add cap and tap twice
- Electroporate
- Recover ASAP with 500 uL of warm culture medium
- Transfer cell suspension to plate preincubated at 37°C with complete medium
- Incubate 20-24 h
- Rinse cuvettes well with diH<sub>2</sub>O followed by 70% EtOH and overnight under UV.

### Day 3: Cell Harvests and Luciferase Assay

#### 1-Luciferase assay

- Pick up the cuvettes left overnight under UV
- Prepare buffers, luminometer (2sec reading)

- Wash cells with PBS twice
- Lyse cells in 150 uL lysis buffer (see below)
- Rotate 20 min at RT (You should see the white dead cells floating by 15 min)
- Prepare assays tubes: add 250 uL luciferase assay buffer/tube
- Add 20 uL of cell lysate
- Place tube in luminometer
- Squirt 100 uL of 125  $\mu$ M D-luciferin
- Read and write down number

## 2-Bradford assay

- In a 96 well plate, dispense 1,2,4,8 uL of cell lysate from a random sample
- Add 100 uL Bradford reagent, wait 5min
- Determine the mid-range
- Expand the Bradford assay for each well
- Measure the 595 nm absorbance
- Normalize luciferase reading to total protein content

## Buffers

### Cell culture media

1-MEF culture media

10% heat inactivated fetal bovine serum

1X Penicillin and streptomycin

Media made in Dublecco's Modified Eagle Medium (DMEM)

2-TM media

0.5% newborn calf serum

0.1 mM  $\beta$ -mercaptoethanol

Media made in DMEM

### Luciferase assay

#### 1-Lysis buffer

| Reagents                                        | Conditions | Stocks | mL to add for 50mL |
|-------------------------------------------------|------------|--------|--------------------|
| Triton X-100                                    | 0.20%      | 20%    | 0.5                |
| KH <sub>2</sub> PO <sub>4</sub> pH 7.8          | 15 mM      | 1 M    | 0.75               |
| H <sub>2</sub> O                                | to 50 mL   | -      | 49.5               |
| <b>!!! Add DTT to 1 mM final before use !!!</b> |            |        |                    |
| <u>keep at 4°C</u>                              |            |        |                    |

#### 2-Luciferase Assay Buffer

| Reagents                                        | Conditions | Stocks | mL to add for 50mL |
|-------------------------------------------------|------------|--------|--------------------|
| Glycylglycine pH 7.8                            | 25mM       | 1M     | 1.25               |
| MgSO <sub>4</sub>                               | 15mM       | 1M     | 0.75               |
| KH <sub>2</sub> PO <sub>4</sub> pH 7.8          | 15mM       | 1M     | 0.75               |
| EGTA (add fresh!)                               | 4mM        | 0.5M   | 0.4                |
| H <sub>2</sub> O                                | to 50mL    | -      | 46.85              |
| <b>!!! Add DTT to 1 mM final before use !!!</b> |            |        |                    |
| <b>!!! Add ATP to 2mM final before use !!!</b>  |            |        |                    |
| <u>keep at 4°C</u>                              |            |        |                    |

#### 3-D-Luciferin

Stock should be kept the dark at -80°C.

Just before use, dilute D-luciferin stock from -80°C (1000x; 125 mM in H<sub>2</sub>O) in luciferase assay buffer without EGTA or ATP to 1X final.

Add DTT to 1 mM final.

APPENDIX B

GENERATION OF MOUSE EMBRYONIC FIBROBLAST CULTURES

1. Young female *tbp* <sup>$\Delta N/+$</sup>  were mated to *tbp* <sup>$\Delta N/+$</sup>  males.
2. Pregnant dams were euthanized (E11.5-12.5).
3. Whole pregnant uterus was dissected and placed in ice-cold phosphate-buffered saline (pH 7.4), transferred under laminar flow hood where the whole uterus was quickly dipped in 70% EtOH and back into sterile 1X PBS. Fetuses were isolated from the placenta and amniotic sac. A limb was used for genotyping.
4. Embryos were individually minced and collected in 2X trypsin (Mediatech, Manassas, VA) supplemented with 1x penicillin-streptomycin (pen/strep, Mediatech) and incubated 5min at 37°C with intermittent flicking.
5. The digestion was quenched by addition of 5mL of Dubelcco's Modified Eagle's Medium (DMEM; Mediatech, Manassas, VA), 10% fetal bovine serum (FBS, Hyclone, Logan, UT) supplemented with penicillin-streptomycin.
6. The cells were then gently spun 5 min in a diagnostic centrifuge and resuspended in 10 ml of DMEM supplemented with 10% FBS and 1x pen/strep.
7. After two days, the media was changed.
8. When the cells covered ~90% of the plate, cells were harvested with Accutase (Innovative Cell Technologies, San Diego, CA), spun and resuspended in freezing media (15% DMSO, 10% FBS, 1x pen/strep in DMEM) before being frozen at -80°C and further stored in liquid nitrogen.

APPENDIX C

WHOLE MOUNT IN SITU HYBRIDIZATION PROTOCOL

### Introduction

The following protocol describes how to perform a whole mount in situ hybridization on a whole mouse embryo. Whole mount in situ hybridizations were performed as previously described<sup>483</sup>.

### Probe Generation

A 676 bp mouse nppa cDNA fragment (table Sx) was cloned into pBS+ plasmid. The plasmid was linearized with *Eco* R I and transcribed with T3 RNA polymerase in presence of 2.5 mM each NTP (0.875 mM Dig-UTP and 1.625 mM UTP) to generate an antisense digoxigenin-UTP labeled RNA. The residual DNA template was digested with 10 U of RNase-free DNase I. The probe was purified on a Sephadex<sup>TM</sup> G-50 5 mL column (GE Healthcare, Uppsala, Sweden), the first radioactive peak was collected and incorporation was calculated.

### Embryos Harvest and Preparation

For each embryonic stage, staining was done simultaneously. Fetuses at the indicated developmental stages were harvested and dissected in PBST (PBS + 0.1% Tween-20) supplemented with 1% bovine serum albumin (BSA). The number of somites was determined.

1. Fix overnight in 4% paraformaldehyde (PFA) in PBST.

2. Dehydrate through a gradual 10 min series of MeOH/PBST (25% MeOH in PBST, 50% MeOH/PBST, 75% MeOH/PBST followed by two 100% MeOH) incubation on ice
3. Keep at -20°C for no more than 4 weeks.
4. The day of the whole mount in situ hybridization procedure, rehydrate embryo 5-10min through the inverse series of MeOH/PBST and wash twice in PBST, on ice.

#### Day 1

1. Incubate embryos 2 h at room temperature in 4:1 PBST:30% $H_2O_2$
2. Wash three times 5 min with PBST.
3. The proteinase K treatment (10  $\mu$ g/ml) was performed in PBST at room temperature for 6 min for E9.5, 22 min for E 10.5 or 30 min for E11.5 embryos.
4. Stop the reaction with washes of fresh 2mg/ml glycine in PBST.
5. Wash twice in PBST.
6. Incubate 20 min in 4% PFA/PBST, 0.2% glutaraldehyde fixation.
7. Wash three times 5-10 min.
8. Prehybridization: before being prehybridized for 2 h at 65°C in hybridization buffer (50% formamide, 5X SSC, 1% SDS, 50 $\mu$ g/ml heparin, 50 $\mu$ g/ml yeast RNA and 1X Denhardt's).
9. Hybridization: Change hybridization buffer, add 400 ng of linearized digoxigenin-UTP probe
10. Incubate overnight at 65°C.

Day 2

1. Wash embryos twice, 30 min in pre-warmed solution I (50% formamide, 5X SSC; 1% SDS) at 65°C.
2. Wash 10 min wash in 50% solution I/50% solution II (0.5 M NaCl; 10 mM Tris-HCl, pH7.5; 0.1% Tween-20).
3. Wash three times 5 min in solution II at room temperature.
4. Incubate twice 30 min with RNases A and T1 (100 ug/ml and 100 U/ml, respectively in solution II) at 37°C.
5. Wash twice at 65°C for 30 min in solution III (50% formamide; 2X SSC).
6. Wash three times 5min in TBST (0.137 M NaCl; 2.7 mM KCl; 25 mM Tris-HCl, pH7.6; 0.1% Tween-20)
7. Block for at least 2.5 h at room temperature in TBST with 10% heat inactivated sheep serum (HISS).
8. Prepare the anti-digoxigenin-AP antibody (Roche Diagnostics, Indianapolis, IN). Incubate 1 particule for of mouse acetone powder for >1h at 4°C in TBST with 1% anti-digoxigenin-AP antibody and 10% HISS. Spin 5min in microfuge at maximum speed.
9. Replace blocking solution with 10% HISS in TBST supplemented with 1:2,000 dilution of the subtracted antibody (for 3 mL: 300 µL of 10X TBST, 300 µL of HISS, 150 µL of blocked antibody from 8.) and incubate overnight at 4°C.

Day 3

1. Wash three times 5 min in TBST, at room temperature.
2. Wash five times for one-hour TBST, at room temperature.
3. Equilibrate twice for 20 min in freshly made alkaline phosphate buffer (100 mM tris-HCl, pH9.5; 50 mM MgCl<sub>2</sub>; 100 mM NaCl; 0.1% Tween-20 and 0.5 mg/mL levamisole).
4. Determine the anti-digoxigenin associated alkaline phosphatase activity by chromogenic reaction by adding 338ug/mL of nitro blue tetrazolium (NBT) and 175 ug/ml of 5-bromo-4-chloro-3-indolyl phosphate (BCIP), at room temperature, in the dark, on a rocker.
5. At the desired time, quench the reaction with 1mL PBST supplemented with 5mM EDTA.
6. Fix the embryos in 4% PFA/PBST overnight at 4°C.
7. Dehydrate through series of MeOH/PBST (25%, 50%, 75% and 100%) and rehydrate through the inverse series and washed twice in PBST with 5 mM EDTA for 5 min at room temperature.
8. Clear the samples in 1:4 glycerol:PBST with 5 mM EDTA for 1h followed by a 1:1 glycerol:PBST with 2 mM EDTA for 1 h and a final 4:1 glycerol:PBST with 1 mM EDTA.

Photograph under dissecting microscope.

APPENDIX D

ADULT LIVERS MICROARRAY DATA

| Gene list | Ontology <sup>a</sup> | Gene name | Fold change                        |                                      | Wildtype<br>signal <sup>b</sup> | Gene<br>ID <sup>c</sup> | Genbank<br>accession<br>number |
|-----------|-----------------------|-----------|------------------------------------|--------------------------------------|---------------------------------|-------------------------|--------------------------------|
|           |                       |           | <i>tbp</i> <sup><i>ΔN/ΔN</i></sup> | <i>tbp</i> <sup><i>hfN/hfN</i></sup> |                                 |                         |                                |
| Common    | Lipid mB              | Cyp4a14   | -49.10                             | -7.33                                | 26.4                            | 13119                   | AI327006                       |
| Common    | Lipid mB              | Ehhadh    | -1.82                              | -1.82                                | 639.6                           | 74147                   | NM_023737                      |
| Common    | Lipid mB              | St3gal5   | -1.80                              | 1.72                                 | 336.5                           | 20454                   | BB829192                       |
| Common    | Lipid mB              | Thrsp     | 1.75                               | 2.30                                 | 38.0                            | 21835                   | NM_009381                      |
| Common    | Lipid mB              | Acad9     | 2.02                               | 1.53                                 | 166.9                           | 229211                  | AV235115                       |
| Common    | Lipid mB              | Akr1d1    | 2.13                               | 1.52                                 | 88.2                            | 208665                  | BC018333                       |
| Common    | Lipid mB              | Acad9     | 2.15                               | 1.03                                 | 18.6                            | 229211                  | AV235115                       |
| Common    | Lipid mB              | Cd36      | 2.78                               | -2.00                                | 35.4                            | 12491                   | BB534670                       |
| Common    | Lipid mB              | Cd36      | 3.16                               | -1.57                                | 674.4                           | 12491                   | BB534670                       |
| Common    | Lipid mB              | Aldh1b1   | 3.17                               | 1.72                                 | 110.9                           | 72535                   | BC020001                       |
| Common    | Lipid mB              | Cd36      | 3.37                               | -2.24                                | 243.7                           | 12491                   | BB534670                       |
| Common    | Lipid mB              | Apoa4     | 3.46                               | 3.10                                 | 2726.5                          | 11808                   | AV027367                       |
| Common    | Lipid mB              | Apoa4     | 4.67                               | 4.57                                 | 275.5                           | 11808                   | BC010769                       |
| Common    | Retroviruses          | Env       | -3.74                              | 2.47                                 | 44.8                            | 67527                   | BG297038                       |
| Common    | Retroviruses          | Env       | -2.44                              | 3.41                                 | 54.8                            | 67527                   | BG297038                       |
| Common    | Retroviruses          | Int       | -1.81                              | 2.03                                 | 30.2                            | 320717                  | BB794742                       |
| Common    | Retroviruses          | vinc      | -1.60                              | -2.71                                | 761.8                           | 66961                   | AK018202                       |
| Common    | Retroviruses          | Il6ra     | -1.05                              | 2.09                                 | 115.2                           | 16194                   | X53802                         |
| Common    | Retroviruses          |           | 2.72                               | -2.00                                | 154.1                           | n/a                     | AW763751                       |
| Common    | Insulin signaling     | Igfbp2    | -3.74                              | -1.83                                | 12.5                            | 16008                   | AK011784                       |
| Common    | Insulin signaling     | Socs2     | -1.59                              | -3.18                                | 108.0                           | 216233                  | NM_007706                      |
| Common    | Antiproliferation     | Hspa1b    | -4.65                              | -2.63                                | 70.0                            | 15511                   | M12573                         |
| Common    | Antiproliferation     | Hspa1b    | -4.35                              | -2.85                                | 22.2                            | 15511                   | M12573                         |
| Common    | Antiproliferation     | Ddit4     | -2.19                              | 3.00                                 | 243.7                           | 74747                   | AK017926                       |
| Common    | Antiproliferation     | Tob2      | 3.36                               | 2.18                                 | 495.2                           | 57259                   | AV174616                       |
| Common    | Signal transduction   | Mt2       | -1.91                              | 2.66                                 | 157.6                           | 17750                   | AA796766                       |
| Common    | Signal transduction   | Rgs16     | -1.65                              | 1.83                                 | 168.9                           | 19734                   | U94828                         |
| Common    | Signal transduction   | Rgs16     | -1.59                              | 1.86                                 | 20.4                            | 19201                   | BB100249                       |
| Common    | Signal transduction   | Fkbp5     | -1.54                              | 1.91                                 | 1288.6                          | 14229                   | U16959                         |
| Common    | Signal transduction   | Gna14     | 3.60                               | 1.78                                 | 52.7                            | 14675                   | NM_008137                      |
| Common    | Transcription         | Hist1h1c  | -2.78                              | -2.50                                | 39.9                            | 50708                   | NM_015786                      |
| Common    | Transcription         | Est       | -2.24                              | 2.09                                 | 20.4                            | 102538                  | AI467657                       |
| Common    | Transcription         | Mbd1      | -2.01                              | -1.62                                | 75.7                            | 17190                   | AK007371                       |
| Common    | Transcription         | Hmgnbd2   | 2.43                               | 1.51                                 | 2567.0                          | 15331                   | BE553881                       |
| Common    | Others                |           | -3.05                              | -1.83                                | 23.2                            | 67442                   | BB775176                       |
| Common    | Others                | Tmem98    | -2.61                              | -1.52                                | 217.7                           | 103743                  | BC011208                       |
| Common    | Others                | Sult5a1   | 1.74                               | 2.43                                 | 38.2                            | 57429                   | NM_020564                      |
| Common    | Others                | Aldh1l1   | 1.78                               | 1.57                                 | 836.8                           | 107747                  | AK007822                       |
| Common    | Others                | Slc46a3   | 1.87                               | 1.66                                 | 414.6                           | 71706                   | BC006902                       |
| Common    | Others                | Reck      | 2.17                               | 1.60                                 | 197.7                           | 53614                   | NM_016678                      |
| Common    | Others                | Sulf2     | 2.30                               | 1.61                                 | 105.6                           | 72043                   | AK008108                       |
| Common    | Others                |           | 2.38                               | 4.40                                 | 64.4                            | 13034                   | BF580235                       |
| Common    | Others                | Tpmt      | 2.60                               | 2.65                                 | 42.7                            | 22017                   | AK002335                       |

| Gene list                                            | Ontology <sup>a</sup> | Gene name | Fold change                                          |                                            | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> | Genbank accession number |
|------------------------------------------------------|-----------------------|-----------|------------------------------------------------------|--------------------------------------------|------------------------------|----------------------|--------------------------|
|                                                      |                       |           | <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | <i>tbp</i> <sup><math>hfN/hfN</math></sup> |                              |                      |                          |
| Common                                               | Others                | Eg226654  | 2.61                                                 | 1.55                                       | 119.4                        | 432549               | AW111083                 |
| Common                                               | Others                | Apcs      | 3.47                                                 | 1.83                                       | 63.2                         | 20219                | NM_011318                |
| Common                                               | Others                | Loc215866 | 4.27                                                 | -2.45                                      | 115.9                        | n/a                  | BC002257                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Energy mB             | Tsc22d3   | -1.19                                                | 2.36                                       | 377.7                        | 14605                | AF201289                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Energy mB             | Sds       | -1.07                                                | 1.83                                       | 596.3                        | 231691               | BC021950                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Energy mB             | Pptc7     | -1.02                                                | 1.88                                       | 2841.1                       | 226654               | AI881989                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Energy mB             | Hsd3b2    | 1.07                                                 | 1.78                                       | 77.0                         | 15493                | BC026757                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Energy mB             | Hexb      | 1.33                                                 | 2.37                                       | 1560.2                       | 15212                | AV225808                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Energy mB             | Ppargc1a  | 1.34                                                 | 2.84                                       | 122.7                        | 19017                | BB745167                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Ion transport         | Scara5    | -1.40                                                | 4.18                                       | 33.3                         | 71145                | BC016096                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Ion transport         | Mcart1    | -1.11                                                | 1.72                                       | 93.9                         | 230125               | AW212577                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Ion transport         | Slco1a4   | -1.05                                                | 2.61                                       | 306.4                        | 28250                | NM_030687                |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Ion transport         | Kcnk5     | 1.23                                                 | 2.14                                       | 11.6                         | 16529                | AF319542                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Ion transport         | Accn5     | 1.30                                                 | 6.54                                       | 186.0                        | 58170                | NM_021370                |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Ion transport         | Kcnk5     | 1.35                                                 | 2.42                                       | 512.4                        | 16529                | AF319542                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Other                 | Fkbp5     | -1.42                                                | 2.11                                       | 16.7                         | 14229                | U16959                   |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Other                 | Tbc1d8    | -1.31                                                | 1.89                                       | 18.1                         | 54610                | BC005421                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Other                 | Eva1      | -1.10                                                | 1.78                                       | 284.3                        | 14012                | BC015076                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Other                 | Gtl2      | 1.43                                                 | 1.90                                       | 61.3                         | 17263                | BB093563                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Other                 | N4bp2l1   | 1.47                                                 | 2.00                                       | 1089.0                       | 100637               | NM_133898                |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Signal transduction   | Hspa1b    | -4.00                                                | -2.33                                      | 27.5                         | 15511                | M12573                   |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Signal transduction   | Cish      | 1.10                                                 | -2.36                                      | 1042.4                       | 12700                | NM_009895                |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Signal transduction   | Eva1      | 1.20                                                 | 1.83                                       | 107.4                        | 14012                | BC015076                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Signal transduction   |           | 1.39                                                 | -2.12                                      | 92.7                         | n/a                  | BC021831                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Transcription         | Got1      | -1.31                                                | 1.93                                       | 266.5                        | 14718                | AA792094                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Transcription         | Tsc22d3   | -1.17                                                | 2.18                                       | 10.0                         | 14605                | NM_010286                |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Transcription         | Bmyc      | 1.14                                                 | 1.78                                       | 2323.7                       | 107771               | AK012810                 |
| <i>tbp</i> <sup><math>hfN/hfN</math></sup>           | Transcription         | Map2k6    | 1.41                                                 | 1.85                                       | 210.7                        | 26399                | BB261602                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Amino acid mB         | Ddc       | 2.06                                                 | -1.14                                      | 1366.2                       | 13195                | AF071068                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Amino acid mB         | Pdxdc1    | 14.27                                                | -1.02                                      | 194.4                        | 94184                | NM_053181                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Gck       | -2.60                                                | -1.37                                      | 1095.4                       | 103988               | BC011139                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Abhd6     | -2.12                                                | -1.39                                      | 143.4                        | 66082                | NM_025341                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Pfkfb3    | -2.11                                                | 1.11                                       | 423.9                        | 170768               | NM_133232                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Pfkfb3    | -1.98                                                | -1.23                                      | 64.9                         | 170768               | AV282911                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Insig2    | 1.75                                                 | 1.13                                       | 37.0                         | 72999                | AV257512                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Acat3     | 1.78                                                 | 1.09                                       | 45.2                         | 224530               | AF356876                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Siae      | 1.82                                                 | 1.42                                       | 69.3                         | 22619                | NM_011734                |

| Gene list                                            | Ontology <sup>a</sup> | Gene name | Fold change                                          |                                            | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> | Genbank accession number |
|------------------------------------------------------|-----------------------|-----------|------------------------------------------------------|--------------------------------------------|------------------------------|----------------------|--------------------------|
|                                                      |                       |           | <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | <i>tbp</i> <sup><math>hfN/hfN</math></sup> |                              |                      |                          |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Akr1b3    | 1.85                                                 | 1.30                                       | 471.3                        | 432549               | AV127085                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       |           | 1.90                                                 | 1.24                                       | 143.0                        | 215929               | BC018406                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Inpp5a    | 2.11                                                 | 1.02                                       | 155.6                        | 212111               | BB307136                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Eg667410  | 2.85                                                 | 1.36                                       | 135.5                        | 19734                | BB777344                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Carbohydrate mB       | Akr1b3    | 3.08                                                 | 1.39                                       | 1344.0                       | 11677                | BB469763                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | Fam-32L   | 1.70                                                 | -1.05                                      | 208.0                        | 67922                | AV109006                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | Tusc3     | 1.76                                                 | -1.01                                      | 51.0                         | 80286                | NM_030254                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | Fanc1     | 1.81                                                 | 1.13                                       | 627.7                        | 67030                | BB751093                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | xlmmg1    | 1.89                                                 | -1.41                                      | 40.9                         | 17772                | BG976607                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | Loc66857  | 1.90                                                 | 1.05                                       | 61.6                         | 66857                | `                        |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | pebp1     | 2.59                                                 | 1.02                                       | 47.3                         | 23980                | AV207950                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            |           | 2.70                                                 | 1.26                                       | 89.0                         | 320770               | BB246700                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | Pebp1     | 2.70                                                 | 1.00                                       | 63.7                         | 23980                | NM_018858                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell cycle            | Drctnnb1a | 3.24                                                 | 1.07                                       | 176.4                        | 84652                | NM_053090                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Ccnf      | -2.69                                                | 1.02                                       | 974.8                        | 12449                | NM_007634                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Ccl9      | -1.87                                                | -1.02                                      | 136.0                        | 20308                | AF128196                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Cd9       | 1.73                                                 | -1.17                                      | 77.8                         | 12527                | NM_007657                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Tmem19    | 1.80                                                 | 1.13                                       | 108.0                        | 67226                | AK018383                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Prune     | 1.82                                                 | -1.19                                      | 985.6                        | 229589               | BQ176370                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Habp2     | 1.86                                                 | 1.05                                       | 90.4                         | 226243               | AI035669                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Cd207     | 2.03                                                 | 1.10                                       | 38.9                         | 246278               | AY026050                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Loc192136 | 2.28                                                 | 1.17                                       | 112.1                        | 192136               | NM_138654                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Htatip2   | 2.30                                                 | -1.19                                      | 100.0                        | 53415                | AF061972                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Asah2     | 2.46                                                 | 1.28                                       | 109.7                        | 54447                | NM_018830                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Olfm3     | 2.81                                                 | 1.33                                       | 350.2                        | 229759               | AF442824                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Cml4      | 3.44                                                 | -1.01                                      | 15.2                         | 68396                | NM_023455                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Cell communication    | Rbm39     | 4.63                                                 | 1.02                                       | 13.8                         | 96982                | BB436856                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Qprt      | -3.69                                                | 1.15                                       | 315.0                        | 67375                | AI195046                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Msrp2     | 1.79                                                 | 1.04                                       | 20.2                         | 76467                | BC021619                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Aldh7a1   | 1.81                                                 | 1.08                                       | 134.6                        | 110695               | BC012407                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Cyp2a5    | 1.82                                                 | -1.03                                      | 31.5                         | 13086                | NM_007812                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Fmo4      | 1.93                                                 | 1.03                                       | 224.6                        | 226564               | AF461145                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Sult1c2   | 1.96                                                 | 1.21                                       | 163.8                        | 69083                | NM_026935                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Pter      | 2.02                                                 | -1.06                                      | 358.5                        | 19212                | NM_008961                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Gstm6     | 2.08                                                 | 1.27                                       | 57.4                         | 14867                | NM_008184                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Fmo1      | 2.13                                                 | 1.02                                       | 40.2                         | 14261                | BC011229                 |

| Gene list                                            | Ontology <sup>a</sup> | Gene name   | Fold change                                          |                                            | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> | Genbank accession number |
|------------------------------------------------------|-----------------------|-------------|------------------------------------------------------|--------------------------------------------|------------------------------|----------------------|--------------------------|
|                                                      |                       |             | <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | <i>tbp</i> <sup><math>hfN/hfN</math></sup> |                              |                      |                          |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Cyp2dx      | 4.39                                                 | 1.02                                       | 280.3                        | 380997               | AV372365                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Detoxification        | Gstm6       | 4.80                                                 | 1.28                                       | 2165.2                       | 14863                | NM_008183                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         |             | -2.83                                                | -1.46                                      | 3587.6                       | 67442                | BB775176                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         |             | -1.84                                                | -1.14                                      | 288.4                        | 234677               | BC026374                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         | Elovl5      | 1.70                                                 | -1.22                                      | 191.3                        | 68801                | NM_134255                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         | Ces1        | 1.91                                                 | 1.41                                       | 304.5                        | 12623                | NM_021456                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         | Ppt1        | 2.02                                                 | 1.23                                       | 94.7                         | 19063                | AF326558                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         | Galc        | 2.32                                                 | -1.24                                      | 829.7                        | 14420                | AK010101                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         | Ppt1        | 2.42                                                 | 1.04                                       | 34.9                         | 19063                | AF326558                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         | Fabp2       | 2.54                                                 | -1.33                                      | 442.3                        | 14079                | NM_007980                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Fatty acid mB         | Acnat1      | 7.07                                                 | -1.04                                      | 264.2                        | 209186               | BC010829                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Serpina12   | -3.43                                                | 1.13                                       | 624.7                        | 68054                | AK014346                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | C6          | -1.95                                                | 1.04                                       | 241.5                        | 12274                | NM_016704                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Cadm4       | -1.79                                                | 1.24                                       | 75.1                         | 260299               | AY059394                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Serpina6    | 1.76                                                 | 1.19                                       | 105.9                        | 12401                | NM_007618                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Serpina3n   | 1.77                                                 | 1.38                                       | 1459.5                       | 20716                | NM_009252                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Pstpip2     | 2.18                                                 | 1.26                                       | 147.7                        | n/a                  | AV229693                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Pstpip2     | 2.84                                                 | 1.41                                       | 279.4                        | 19201                | BC002123                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Serpine2    | 2.85                                                 | 1.00                                       | 1739.4                       | 20720                | NM_009255                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Lect2       | 2.98                                                 | 1.26                                       | 69.8                         | 16841                | NM_010702                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Ly6a        | 3.56                                                 | -1.34                                      | 99.9                         | 110454               | BC002070                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Immunity              | Tinag       | 7.67                                                 | -1.20                                      | 1363.1                       | 26944                | BC010745                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Lipid mB              | St3gal5     | -1.79                                                | 1.48                                       | 100.0                        | 20454                | BB829192                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Env         | -26.33                                               | 1.80                                       | 56.3                         | n/a                  | BC004065                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Dio1        | -3.82                                                | 1.06                                       | 23.3                         | 13370                | NM_007860                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Isoc2b      | -2.82                                                | 1.14                                       | 175.7                        | 67441                | NM_026158                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Sdf2l1      | -2.00                                                | -1.21                                      | 337.0                        | 64136                | NM_022324                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Ube2s       | -1.87                                                | -1.25                                      | 436.1                        | 77891                | NM_133777                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 |             | -1.87                                                | 1.08                                       | 256.6                        | 14455                | NM_013525                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Herpud1     | -1.69                                                | 1.09                                       | 26.5                         | 64209                | AI835088                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Nudt2       | 1.71                                                 | 1.19                                       | 407.4                        | 66401                | NM_025539                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | D1Erttd622e | 1.84                                                 | 1.26                                       | 31.5                         | 52392                | NM_133825                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Mkrn1       | 3.17                                                 | 1.37                                       | 382.1                        | 54484                | AA717142                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Snrpn       | 3.47                                                 | 1.38                                       | 745.6                        | 20646                | NM_033174                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Other                 | Col4a5      | 7.09                                                 | -1.02                                      | 57.1                         | 12830                | BM250666                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Proliferation         | Fars2       | -3.06                                                | 1.21                                       | 146.6                        | 69955                | BB530332                 |

| Gene list                                            | Ontology <sup>a</sup> | Gene name | Fold change                                          |                                            | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> | Genbank accession number |
|------------------------------------------------------|-----------------------|-----------|------------------------------------------------------|--------------------------------------------|------------------------------|----------------------|--------------------------|
|                                                      |                       |           | <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | <i>tbp</i> <sup><math>hfN/hfN</math></sup> |                              |                      |                          |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Proliferation         | Plk3      | -2.64                                                | -1.16                                      | 127.0                        | 12795                | BM947855                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Proliferation         | Xlkd1     | -2.62                                                | 1.04                                       | 28.3                         | 114332               | AV124537                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Proliferation         | Bid       | -2.10                                                | 1.02                                       | 2204.5                       | 12122                | NM_007544                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Proliferation         | Myd116    | -1.87                                                | 1.19                                       | 19.3                         | 17872                | NM_008654                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Proliferation         | Gfer      | -1.72                                                | -1.06                                      | 61.1                         | 11692                | BI901126                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Proliferation         | Gfer      | 1.74                                                 | -1.16                                      | 77.7                         | 11692                | BI901126                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Retrovirus            | Pim3      | -2.87                                                | -1.06                                      | 27.2                         | 223775               | BC017621                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Retrovirus            | Env       | 3.72                                                 | 1.00                                       | 126.2                        | 433855               | BI438039                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Irs2      | -1.98                                                | 1.06                                       | 9.4                          | 384783               | BE199054                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Trip12    | -1.77                                                | 1.19                                       | 1487.3                       | 14897                | BG923744                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Rapgef4   | -1.77                                                | 1.12                                       | 78.2                         | 56508                | AK004874                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Sri       | 1.74                                                 | 1.12                                       | 122.4                        | 109552               | AK008404                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Traf3     | 1.77                                                 | -1.01                                      | 38.3                         | 22031                | U21050                   |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | march2    | 1.79                                                 | 1.28                                       | 61.2                         | 381101               | BE949497                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Wdfy1     | 1.81                                                 | 1.43                                       | 57.9                         | 69368                | BM233251                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Rab32     | 1.86                                                 | -1.11                                      | 1254.9                       | 67844                | NM_026405                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Gbp2      | 1.92                                                 | 1.38                                       | 553.1                        | 14468                | NM_010259                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Ssr1      | 2.13                                                 | 1.02                                       | 390.1                        | 107513               | BG077348                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Prlr      | 2.20                                                 | 1.16                                       | 126.7                        | 19116                | NM_008932                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Prlr      | 2.88                                                 | 1.46                                       | 932.8                        | 19116                | NM_008932                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | GPR91     | 3.05                                                 | -1.13                                      | 41.0                         | 84112                | NM_032400                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Blnk      | 3.11                                                 | 1.22                                       | 38.4                         | 17060                | AF068182                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Cap1      | 5.11                                                 | 1.01                                       | 50.2                         | 12331                | NM_007598                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Rbm39     | 5.18                                                 | -1.02                                      | 13.6                         | 96982                | BB436856                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Signal transduction   | Cap1      | 9.48                                                 | 1.18                                       | 75.5                         | 12331                | NM_007598                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Cebpd     | -2.23                                                | -1.02                                      | 32.3                         | 12609                | BB831146                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Gtf2a1    | -2.19                                                | 1.30                                       | 301.5                        | 108667               | BI689897                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Hod       | -2.00                                                | -1.12                                      | 856.0                        | 74318                | AK009007                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Hod       | -1.95                                                | -1.07                                      | 198.0                        | 74318                | BC024546                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Hes6      | -1.75                                                | -1.15                                      | 84.5                         | 55927                | AF247040                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Nsbp1     | 1.76                                                 | 1.09                                       | 2195.0                       | 50887                | NM_016710                |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Loc231238 | 1.79                                                 | -1.15                                      | 988.4                        | 231238               | BC026655                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Prim1     | 1.82                                                 | -1.09                                      | 198.3                        | 19075                | J04620                   |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Myot      | 1.94                                                 | 1.12                                       | 18.9                         | 14489                | BB124537                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Prva      | 2.06                                                 | 1.01                                       | 10.1                         | 57342                | BI690209                 |
| <i>tbp</i> <sup><math>\Delta N/\Delta N</math></sup> | Transcription         | Scand1    | 2.23                                                 | 1.08                                       | 93.9                         | 19018                | NM_020255                |

| Gene list                   | Ontology <sup>a</sup> | Gene name | Fold change                 |                               | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> | Genbank accession number |
|-----------------------------|-----------------------|-----------|-----------------------------|-------------------------------|------------------------------|----------------------|--------------------------|
|                             |                       |           | <i>tbp</i> <sup>ΔN/ΔN</sup> | <i>tbp</i> <sup>hfN/hfN</sup> |                              |                      |                          |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transcription         | Ttc8      | 2.94                        | 1.13                          | 96.0                         | 76260                | BC017523                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transcription         | Rwdd3     | 3.15                        | 1.27                          | 163.6                        | 73170                | AF439556                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Slc7a2    | -1.90                       | 1.24                          | 100.2                        | 11988                | BF533509                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Slco2a1   | -1.79                       | -1.04                         | 137.9                        | 24059                | NM_033314                |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             |           | 1.70                        | 1.02                          | 16.5                         | 215015               | BG060641                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Ccdc91    | 1.71                        | 1.12                          | 315.2                        | 67015                | AA067702                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Slc17a3   | 1.73                        | 1.00                          | 189.8                        | 105355               | NM_134069                |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | slc47a1   | 1.86                        | 1.27                          | 501.4                        | 67473                | NM_026183                |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Ap3m1     | 1.89                        | 1.49                          | 135.6                        | 55946                | AF242857                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Slc44a3   | 2.20                        | 1.06                          | 202.8                        | 213603               | BC025548                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Abcb1a    | 2.21                        | 1.14                          | 527.9                        | 18671                | M30697                   |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Slc3a1    | 2.24                        | 1.10                          | 25.9                         | 20532                | NM_009205                |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Abca3     | 3.23                        | -1.02                         | 38.6                         | 27410                | AK007703                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Abca3     | 3.34                        | -1.01                         | 103.0                        | 27410                | AK007703                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Aqp4      | 3.74                        | -1.02                         | 1438.2                       | 11829                | U48399                   |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Aqp8      | 3.92                        | 1.35                          | 31.1                         | 11833                | NM_007474                |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Slc22a20  | 4.46                        | -1.25                         | 70.9                         | 236293               | AB056443                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Aqp4      | 7.59                        | -1.03                         | 216.5                        | 11829                | BB193413                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Transport             | Slc22a20  | 10.89                       | -1.15                         | 28.1                         | 319800               | AB056443                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Ubiquitin pathway     | Wwtr1     | 1.90                        | 1.21                          | 38.2                         | 97064                | BC014727                 |
| <i>tbp</i> <sup>ΔN/ΔN</sup> | Ubiquitin pathway     | Wwtr1     | 1.91                        | 1.01                          | 28.9                         | 97064                | BG066193                 |

<sup>a</sup>: see material and methods, Chapter 5

<sup>b</sup>: Average signal from all three replicates of wildtype animals in liver given; only mRNAs in which the average signal was  $\geq 50$  units were included

<sup>c</sup>: Gene IDs were retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>)

APPENDIX E

P11 MICROARRAY DATA

| Ontology <sup>a</sup> | mRNA     | Fold change |        | P-value | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> |
|-----------------------|----------|-------------|--------|---------|------------------------------|----------------------|
|                       |          | Liver       | Kidney |         |                              |                      |
| Lipid mB              | idi1     | -10.86      | -1.4   | 0.017   | 1541                         | 319554               |
| Lipid mB              | sc4mol   | -6.4        | -1.85  | 0.013   | 2243                         | 66234                |
| Lipid mB              | fdps     | -5.12       | -1.48  | 0.005   | 1913                         | 110196               |
| Lipid mB              | pnliprp1 | -4.43       | -1.02  | 0.047   | 474                          | 18946                |
| Lipid mB              | sqle     | -4.4        | -1.25  | 0.013   | 1116                         | 20775                |
| Lipid mB              | ces1     | -4.26       | -1.04  | 0.047   | 2257                         | 12623                |
| Lipid mB              | abca1    | -4.03       | -1.33  | 0.012   | 404                          | 11303                |
| Lipid mB              | hmgcs1   | -3.86       | -1.16  | 0.013   | 3532                         | 208715               |
| Lipid mB              | cyp3a16  | -3.48       | -1     | 0.045   | 1557                         | 13114                |
| Lipid mB              | rnf14    | -3.16       | -1.84  | 0.011   | 138                          | 56736                |
| Lipid mB              | ak3      | -2.95       | -1.48  | 0.005   | 579                          | 56248                |
| Lipid mB              | dnajb6   | -2.64       | -1.39  | 0.039   | 427                          | 23950                |
| Lipid mB              | nfia     | -2.55       | -1.26  | 0.013   | 107                          | 18027                |
| Lipid mB              | mapk14   | -2.35       | -1.21  | 0.008   | 323                          | 26416                |
| Lipid mB              | cybb     | -2.3        | -1.04  | 0.036   | 188                          | 13058                |
| Lipid mB              | srr      | -2.13       | -1.07  | 0.045   | 276                          | 27364                |
| Lipid mB              | hsd17b7  | -2          | -1.15  | 0.011   | 411                          | 15490                |
| Lipid mB              | acaa2    | -1.89       | -1.05  | 0.047   | 7158                         | 52538                |
| Lipid mB              | acsm3    | -1.86       | -1.24  | 0.048   | 747                          | 20216                |
| Lipid mB              | rnf14    | -1.84       | -1.21  | 0.034   | 310                          | 56736                |
| Lipid mB              | lin7c    | -1.75       | -1.68  | 0.036   | 138                          | 22343                |
| Lipid mB              | sar1a    | -1.73       | -1.44  | 0.006   | 1070                         | 20224                |
| Lipid mB              | hdlbp    | -1.66       | -1.54  | 0.043   | 138                          | 110611               |
| Lipid mB              | dnaja1   | -1.59       | -1.17  | 0.030   | 4622                         | 15502                |
| Lipid mB              | scamp4   | -1.5        | -1.21  | 0.005   | 327                          | 56214                |
| Lipid mB              | dci      | 1.2         | 1.96   | 0.049   | 3216                         | 13177                |
| Lipid mB              | c1qtnf4  | 1.36        | 1.57   | 0.044   | 99                           | 67445                |
| Lipid mB              | ptges2   | 1.51        | 1.18   | 0.012   | 407                          | 96979                |
| Lipid mB              | pitpnm1  | 1.54        | 1.28   | 0.018   | 169                          | 18739                |
| Lipid mB              | pnpla2   | 1.62        | 1.72   | 0.037   | 516                          | 66853                |
| Lipid mB              | pou6f1   | 1.62        | 1.26   | 0.013   | 233                          | 19009                |
| Lipid mB              | c1qtnf1  | 1.63        | 1.65   | 0.018   | 311                          | 56745                |
| Lipid mB              | abca7    | 1.86        | 1.4    | 0.049   | 129                          | 27403                |
| Lipid mB              | mgst3    | 1.91        | 1.33   | 0.026   | 768                          | 66447                |
| Lipid mB              | prpf19   | 1.93        | 1.13   | 0.020   | 1418                         | 28000                |
| Lipid mB              | elovl6   | 2.15        | -0.92  | 0.039   | 190                          | 170439               |
| Lipid mB              | pitpnm1  | 2.19        | 1.31   | 0.013   | 143                          | 18739                |
| Lipid mB              | lepr     | 2.23        | -1.12  | 0.038   | 188                          | 16847                |
| Lipid mB              | klf4     | 2.61        | 1.61   | 0.015   | 108                          | 16600                |
| Lipid mB              | apoa4    | 8.48        | -0.94  | 0.046   | 2224                         | 11808                |
| Lipid mB              | apoa4    | 14.94       | -0.97  | 0.048   | 1041                         | 11808                |
| Carbohydrate mB       | clec2h   | -9.03       | -2.85  | 0.038   | 176                          | 94071                |
| Carbohydrate mB       | clec2h   | -5.42       | -1.81  | 0.025   | 346                          | 94071                |

| Ontology <sup>a</sup> | mRNA    | Fold change |        | P-value | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> |
|-----------------------|---------|-------------|--------|---------|------------------------------|----------------------|
|                       |         | Liver       | Kidney |         |                              |                      |
| Carbohydrate mB       | pmm2    | -3.3        | -1.36  | 0.047   | 192                          | 54128                |
| Carbohydrate mB       | sord    | -3.01       | -1.7   | 0.032   | 8086                         | 20322                |
| Carbohydrate mB       | clec2h  | -2.47       | -1.61  | 0.038   | 296                          | 94071                |
| Carbohydrate mB       | galns   | 1.63        | -0.92  | 0.012   | 217                          | 50917                |
| Carbohydrate mB       | angptl4 | 1.92        | 3.86   | 0.049   | 1071                         | 57875                |
| Carbohydrate mB       | h6pd    | 1.95        | 1.61   | 0.030   | 2333                         | 100198               |
| Carbohydrate mB       | hk1     | 2.13        | -0.95  | 0.011   | 76                           | 15275                |
| Amino acid mB         | rbbp4   | -2.65       | -1.66  | 0.011   | 649                          | 19646                |
| Amino acid mB         | mccc2   | -2.39       | -1.79  | 0.010   | 409                          | 78038                |
| Amino acid mB         | etnk1   | -2.23       | -2.06  | 0.034   | 69                           | 75320                |
| Amino acid mB         | ywhag   | -1.92       | -1.43  | 0.017   | 105                          | 22628                |
| Amino acid mB         | tpmt    | -1.88       | -1.18  | 0.017   | 288                          | 22017                |
| Amino acid mB         | rabl3   | -1.86       | -1.25  | 0.038   | 375                          | 67657                |
| Amino acid mB         | sec63   | -1.7        | -1.06  | 0.048   | 506                          | 140740               |
| Amino acid mB         | hdc     | 1.08        | -2.81  | 0.012   | 147                          | 15186                |
| Amino acid mB         | hdc     | 1.32        | -3.54  | 0.038   | 69                           | 15186                |
| Amino acid mB         | eng     | 1.52        | 1.23   | 0.036   | 1180                         | 13805                |
| Amino acid mB         | tnnt1   | 1.74        | -0.93  | 0.044   | 109                          | 21955                |
| Amino acid mB         | rps11   | 1.81        | 1.32   | 0.048   | 4620                         | 27207                |
| Transport             | aqp9    | -5.7        | -1.15  | 0.031   | 2449                         | 64008                |
| Transport             | slc19a2 | -3.66       | -1.11  | 0.047   | 264                          | 116914               |
| Transport             | slco1b2 | -3.58       | -1     | 0.048   | 2496                         | 28253                |
| Transport             | ap3m1   | -2.91       | -1.52  | 0.030   | 823                          | 55946                |
| Transport             | slc40a1 | -2.69       | -1.53  | 0.043   | 2303                         | 53945                |
| Transport             | kdelr2  | -1.59       | -1.37  | 0.020   | 1703                         | 66913                |
| Transport             | atp4a   | -1.36       | -1.64  | 0.016   | 146                          | 11944                |
| Transport             | kctd17  | 1.88        | 1.2    | 0.025   | 161                          | 72844                |
| Cell communication    | cldn2   | -4          | -1.37  | 0.006   | 447                          | 12738                |
| Cell communication    | ptplad1 | -3.99       | -1.5   | 0.039   | 1554                         | 57874                |
| Cell communication    | mpzl2   | -3.06       | -1.39  | 0.019   | 622                          | 14012                |
| Cell communication    | met     | -2.88       | -1.55  | 0.012   | 170                          | 17295                |
| Cell communication    | habp2   | -2.82       | -1.4   | 0.034   | 1976                         | 226243               |
| Cell communication    | cmah    | -2.68       | -1.21  | 0.031   | 337                          | 12763                |
| Cell communication    | arhgap9 | -2.27       | -1.13  | 0.047   | 1050                         | 216445               |
| Cell communication    | cxadr   | -2.23       | -1.52  | 0.037   | 494                          | 13052                |
| Cell communication    | ceacam1 | -2.08       | -1.58  | 0.015   | 980                          | 26365                |
| Cell communication    | stambp  | -1.96       | -1.27  | 0.030   | 172                          | 70527                |
| Cell communication    | gna12   | -1.77       | -1.09  | 0.036   | 227                          | 14673                |
| Cell communication    | prkacb  | -1.67       | -1.44  | 0.043   | 63                           | 18749                |
| Cell communication    | synj2bp | -1.64       | -1.48  | 0.029   | 844                          | 24071                |
| Cell communication    | stk3    | -1.54       | -1.24  | 0.031   | 137                          | 56274                |
| Cell communication    | mlx     | -1.53       | -1.27  | 0.011   | 330                          | 21428                |
| Cell communication    | rpl28   | 1.52        | 1.25   | 0.036   | 8409                         | 19943                |
| Cell communication    | pxn     | 1.55        | 1.3    | 0.017   | 386                          | 19303                |
| Cell communication    | arhgdia | 1.56        | 1.22   | 0.017   | 1400                         | 192662               |

| Ontology <sup>a</sup> | mRNA          | Fold change |        | P-value | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> |
|-----------------------|---------------|-------------|--------|---------|------------------------------|----------------------|
|                       |               | Liver       | Kidney |         |                              |                      |
| Cell communication    | plekhg2       | 1.58        | 1.3    | 0.048   | 350                          | 101497               |
| Cell communication    | parva         | 1.65        | 1.33   | 0.013   | 599                          | 57342                |
| Cell communication    | zyx           | 1.7         | 1.46   | 0.034   | 479                          | 22793                |
| Cell communication    | mlxip         | 1.73        | 1.37   | 0.017   | 209                          | 208104               |
| Cell communication    | dlgap4        | 1.8         | 1.26   | 0.018   | 251                          | 228836               |
| Cell communication    | nlgn2         | 1.82        | 1.36   | 0.044   | 181                          | 216856               |
| Cell communication    | edg2          | 2.19        | 1.29   | 0.044   | 126                          | 14745                |
| Cell communication    | col6a1        | 2.3         | 1.56   | 0.042   | 975                          | 12833                |
| Cell communication    | lgals4        | 2.41        | 2.65   | 0.047   | 367                          | 16855                |
| Cell cycle            | anapc5        | -2.05       | -1.45  | 0.005   | 727                          | 59008                |
| Cell cycle            | son           | -1.89       | -1.47  | 0.020   | 258                          | 20658                |
| Cell cycle            | smc4          | -1.65       | -1.7   | 0.026   | 192                          | 70099                |
| Cell cycle            | cdc5l         | -1.55       | -1.39  | 0.006   | 430                          | 71702                |
| Cell cycle            | gfer          | 1.4         | 1.64   | 0.048   | 460                          | 11692                |
| Cell cycle            | dab2ip        | 1.52        | 1.24   | 0.011   | 208                          | 69601                |
| Cell cycle            | centg3        | 1.57        | 1.16   | 0.042   | 142                          | 213990               |
| Cell cycle            | cdc25a        | 1.64        | 1.24   | 0.020   | 215                          | 12530                |
| Cell cycle            | park2         | 1.71        | 1.73   | 0.035   | 104                          | 50873                |
| Cell cycle            | mapre3        | 1.78        | 1.61   | 0.024   | 200                          | 100732               |
| Immune                | h2-d1         | -23.2       | -5.08  | 0.000   | 1791                         | 14964                |
| Immune                | h2-d1         | -11.32      | -2.87  | 0.005   | 1598                         | 14964                |
| Immune                | h2-d1         | -10.68      | -3.31  | 0.005   | 2807                         | 14964                |
| Immune                | h2-ea         | -4.96       | -3.54  | 0.005   | 224                          | 14968                |
| Immune                | twsg1         | -2.42       | -1.76  | 0.013   | 329                          | 65960                |
| Immune                | pbef1         | -2.23       | -1.02  | 0.006   | 229                          | 59027                |
| Immune                | tapbp         | -2.08       | -1.24  | 0.017   | 679                          | 21356                |
| Immune                | cab39l        | -2.05       | -1.19  | 0.017   | 818                          | 69008                |
| Immune                | ccl27         | -1.88       | -1.15  | 0.036   | 238                          | 20301                |
| Immune                | abcf1         | 1.3         | 1.64   | 0.038   | 800                          | 224742               |
| Insulin mB            | ide           | -4.83       | -1.48  | 0.005   | 241                          | 15925                |
| Insulin mB            | igf1          | -2.65       | -1.61  | 0.049   | 2879                         | 16000                |
| Insulin mB            | igf1r         | 1.5         | 1.52   | 0.045   | 125                          | 16001                |
| IP pathway            | pi4k2b        | -5.64       | -2.17  | 0.019   | 413                          | 67073                |
| IP pathway            | pi4k2b        | -5.4        | -1.57  | 0.005   | 523                          | 67073                |
| IP pathway            | 2610529c04rik | -3.07       | -1.14  | 0.048   | 1068                         | 67075                |
| IP pathway            | pitpnb        | -3          | -1.4   | 0.026   | 614                          | 56305                |
| IP pathway            | ahcyl1        | -2.51       | -1.37  | 0.036   | 314                          | 229709               |
| IP pathway            | pik3r1        | -2.17       | -1.2   | 0.013   | 603                          | 18708                |
| IP pathway            | bpnt1         | -2.13       | -1.39  | 0.042   | 782                          | 23827                |
| IP pathway            | rad21         | -1.71       | -1.75  | 0.038   | 296                          | 19357                |
| IP pathway            | pmpca         | -1.54       | -1.12  | 0.036   | 1025                         | 64436                |
| Receptor              | ghr           | -4.28       | -1.2   | 0.006   | 907                          | 14600                |
| Receptor              | tfrc          | -2.03       | -1.38  | 0.026   | 322                          | 22042                |
| Receptor              | ogfr          | -1.74       | -1.09  | 0.020   | 346                          | 72075                |
| Receptor              | ap2b1         | 1.53        | 1.26   | 0.014   | 644                          | 71770                |

| Ontology <sup>a</sup> | mRNA    | Fold change |        | P-value | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> |
|-----------------------|---------|-------------|--------|---------|------------------------------|----------------------|
|                       |         | Liver       | Kidney |         |                              |                      |
| Receptor              | gabbr1  | 1.61        | 1.27   | 0.047   | 270                          | 54393                |
| Receptor              | il17rb  | 1.8         | 1.22   | 0.008   | 377                          | 50905                |
| Receptor              | nrbp2   | 2.62        | 1.74   | 0.034   | 261                          | 223649               |
| Signal transduction   | gdi2    | -4.32       | -1.74  | 0.045   | 1732                         | 14569                |
| Signal transduction   | aplp2   | -3.12       | -1.92  | 0.005   | 316                          | 11804                |
| Signal transduction   | shmt1   | -2.92       | -1.19  | 0.045   | 1699                         | 20425                |
| Signal transduction   | caprin1 | -2.54       | -1.78  | 0.012   | 471                          | 53872                |
| Signal transduction   | hbs1l   | -2.32       | -1.27  | 0.013   | 296                          | 56422                |
| Signal transduction   | rnf4    | -2.2        | -1.25  | 0.006   | 859                          | 19822                |
| Signal transduction   | lancl2  | -1.85       | -1.45  | 0.045   | 59                           | 71835                |
| Signal transduction   | supt16h | -1.65       | -1.27  | 0.048   | 76                           | 114741               |
| Signal transduction   | sirpa   | -1.6        | -1.12  | 0.018   | 567                          | 19261                |
| Signal transduction   | hmg20b  | 1.52        | 1.15   | 0.030   | 221                          | 15353                |
| Signal transduction   | capn5   | 1.54        | 1.46   | 0.024   | 112                          | 12337                |
| Signal transduction   | sema3f  | 1.6         | 1.35   | 0.048   | 143                          | 20350                |
| Signal transduction   | cd9     | 2.39        | 1.54   | 0.032   | 808                          | 12527                |
| Trafficking           | tmed2   | -2.78       | -1.47  | 0.020   | 1483                         | 56334                |
| Trafficking           | arcn1   | -2.32       | -1.36  | 0.005   | 1020                         | 213827               |
| Trafficking           | golp3   | -2.22       | -1.34  | 0.013   | 433                          | 66629                |
| Trafficking           | spg20   | -2          | -1.2   | 0.029   | 170                          | 229285               |
| Trafficking           | gorasp2 | -1.77       | -1.28  | 0.012   | 609                          | 70231                |
| Trafficking           | mrpl9   | -1.76       | -1.2   | 0.017   | 630                          | 78523                |
| Trafficking           | vps25   | -1.7        | -1.02  | 0.035   | 1280                         | 28084                |
| Trafficking           | ykt6    | -1.7        | -1.33  | 0.030   | 376                          | 56418                |
| Trafficking           | snx9    | -1.55       | -1.14  | 0.012   | 657                          | 66616                |
| Trafficking           | trim3   | 1.54        | 1.23   | 0.048   | 238                          | 55992                |
| Transcription         | trim34  | -4.28       | -1.48  | 0.031   | 193                          | 434218               |
| Transcription         | hbp1    | -2.89       | -1.65  | 0.028   | 698                          | 73389                |
| Transcription         | zdhhc14 | -2.29       | -2.09  | 0.005   | 370                          | 224454               |
| Transcription         | slu7    | -2.2        | -1.09  | 0.020   | 149                          | 193116               |
| Transcription         | pcaf    | -2.12       | -1.18  | 0.026   | 130                          | 18519                |
| Transcription         | cxxc1   | -2.12       | -1.56  | 0.042   | 129                          | 74322                |
| Transcription         | sf3a3   | -2.05       | -1.4   | 0.030   | 407                          | 75062                |
| Transcription         | zdhhc14 | -2.02       | -2.08  | 0.009   | 338                          | 224454               |
| Transcription         | top2b   | -1.86       | -1.64  | 0.006   | 141                          | 21974                |
| Transcription         | zc3h11a | -1.86       | -1.22  | 0.016   | 511                          | 70579                |
| Transcription         | fbxl3   | -1.71       | -1.63  | 0.048   | 89                           | 50789                |
| Transcription         | spop    | -1.68       | -1.09  | 0.014   | 1152                         | 20747                |
| Transcription         | prpf40a | -1.62       | -1.4   | 0.047   | 189                          | 56194                |
| Transcription         | tsc22d4 | 1.5         | 1.25   | 0.011   | 420                          | 78829                |
| Transcription         | med28   | 1.56        | -0.98  | 0.048   | 892                          | 66999                |
| Transcription         | scand1  | 1.7         | 1.26   | 0.034   | 1109                         | 19018                |
| Other                 | fmo3    | -8.8        | 1.24   | 0.039   | 1367                         | 14262                |
| Other                 | thrsp   | -5.54       | 1.58   | 0.040   | 846                          | 21835                |
| Other                 | abhd1   | -5.23       | -3.34  | 0.006   | 545                          | 57742                |

| Ontology <sup>a</sup> | mRNA          | Fold change |        | P-value | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> |
|-----------------------|---------------|-------------|--------|---------|------------------------------|----------------------|
|                       |               | Liver       | Kidney |         |                              |                      |
| Other                 | cp            | -4.22       | -0.94  | 0.038   | 1972                         | 12870                |
| Other                 | tug1          | -3.18       | -1.43  | 0.006   | 277                          | 544752               |
| Other                 | rrm1          | -3.1        | -1.77  | 0.011   | 582                          | 20133                |
| Other                 | ai506816      | -2.85       | -5.12  | 0.006   | 179                          | 433855               |
| Other                 | ab056442      | -2.72       | -1.23  | 0.038   | 1803                         | 171405               |
| Other                 | 4833439l19rik | -2.66       | -1.41  | 0.012   | 325                          | 97820                |
| Other                 | bzw1          | -2.65       | -1.5   | 0.039   | 1188                         | 66882                |
| Other                 | dio1          | -2.57       | -2.89  | 0.040   | 3475                         | 13370                |
| Other                 | ttc3          | -2.52       | -1.41  | 0.012   | 319                          | 22129                |
| Other                 | 2010305c02rik | -2.45       | -1.39  | 0.017   | 3600                         | 380712               |
| Other                 | hnrnpr        | -2.45       | -1.66  | 0.020   | 342                          | 74326                |
| Other                 | d4wsu53e      | -2.4        | -1.11  | 0.019   | 1470                         | 27981                |
| Other                 | immt          | -2.33       | -1.81  | 0.005   | 479                          | 76614                |
| Other                 | msh5          | -2.23       | -1.93  | 0.025   | 281                          | 17687                |
| Other                 | glo1          | -2.21       | -1.54  | 0.009   | 6872                         | 109801               |
| Other                 | fmo2          | -2.1        | -1.01  | 0.042   | 468                          | 55990                |
| Other                 | sgpl1         | -2.04       | -1.19  | 0.005   | 202                          | 20397                |
| Other                 | 2410002o22rik | -2          | -1.5   | 0.021   | 447                          | 66975                |
| Other                 | faf1          | -1.95       | -1.33  | 0.005   | 465                          | 14084                |
| Other                 | acly          | -1.93       | -1.25  | 0.049   | 640                          | 104112               |
| Other                 | mobkl1b       | -1.89       | -1.29  | 0.042   | 213                          | 232157               |
| Other                 | rsad2         | -1.87       | -2.46  | 0.048   | 163                          | 58185                |
| Other                 | aldh8a1       | -1.85       | -1.36  | 0.036   | 2555                         | 237320               |
| Other                 | 4931406c07rik | -1.8        | -0.93  | 0.032   | 2350                         | 70984                |
| Other                 | cpd           | -1.76       | -1.35  | 0.011   | 454                          | 12874                |
| Other                 | wrn           | -1.75       | -1.16  | 0.048   | 124                          | 22427                |
| Other                 | reep3         | -1.74       | -1.13  | 0.040   | 958                          | 28193                |
| Other                 | pdxdc1        | -1.73       | -1.13  | 0.049   | 284                          | 94184                |
| Other                 | cct3          | -1.64       | -1.25  | 0.035   | 721                          | 12462                |
| Other                 | wdr77         | -1.62       | -1.28  | 0.014   | 225                          | 70465                |
| Other                 | ssr1          | -1.59       | -1.25  | 0.015   | 1670                         | 107513               |
| Other                 | sri           | -1.58       | -1.3   | 0.034   | 379                          | 109552               |
| Other                 | psmd7         | -1.55       | -1.1   | 0.012   | 665                          | 17463                |
| Other                 | ak3           | -1.54       | -1.07  | 0.034   | 1964                         | 56248                |
| Other                 | pelo          | 1.53        | 1.18   | 0.047   | 197                          | 105083               |
| Other                 | 2310066e14rik | 1.54        | 1.23   | 0.021   | 253                          | 75687                |
| Other                 | ccdc95        | 1.55        | 1.42   | 0.036   | 64                           | 233875               |
| Other                 | lsm14b        | 1.55        | 1.33   | 0.032   | 403                          | 241846               |
| Other                 | wwc1          | 1.57        | 1.5    | 0.048   | 1504                         | 211652               |
| Other                 | kifc3         | 1.62        | 1.23   | 0.012   | 233                          | 16582                |
| Other                 | rpl23a        | 1.65        | 1.3    | 0.041   | 6650                         | 268449               |
| Other                 | vasp          | 1.67        | 1.15   | 0.019   | 692                          | 22323                |
| Other                 | phc2          | 1.68        | 1.2    | 0.035   | 526                          | 54383                |
| Other                 | st3gal5       | 1.7         | 1.23   | 0.036   | 1942                         | 20454                |
| Other                 | unc119b       | 1.74        | 1.29   | 0.017   | 472                          | 106840               |

| Ontology <sup>a</sup> | mRNA          | Fold change |        | P-value | Wildtype signal <sup>b</sup> | Gene ID <sup>c</sup> |
|-----------------------|---------------|-------------|--------|---------|------------------------------|----------------------|
|                       |               | Liver       | Kidney |         |                              |                      |
| Other                 | pim1          | 1.79        | 1.53   | 0.034   | 407                          | 18712                |
| Other                 | mkl1          | 1.86        | 1.25   | 0.031   | 183                          | 223701               |
| Other                 | podxl         | 1.86        | 1.2    | 0.049   | 270                          | 27205                |
| Other                 | 2700099c18rik | 1.86        | 1.37   | 0.044   | 42                           | 77022                |
| Other                 | gal3st1       | 1.89        | -0.94  | 0.042   | 181                          | 53897                |
| Other                 | d19wsu162e    | 1.9         | 1.4    | 0.031   | 1793                         | 226178               |
| Other                 | loc624295     | 1.93        | 1.74   | 0.048   | 58                           | 624295               |
| Other                 | plod1         | 1.99        | 1.44   | 0.034   | 801                          | 18822                |
| Other                 | hba-a1        | 2.02        | 2.23   | 0.001   | 6544                         | 15122                |
| Other                 | 4933426m11rik | 2.02        | 1.15   | 0.017   | 410                          | 217684               |
| Other                 | bcl6          | 2.15        | 1.48   | 0.048   | 121                          | 12053                |
| Other                 | specc1        | 2.3         | 1.37   | 0.016   | 95                           | 432572               |
| Other                 | evc           | 2.54        | -0.98  | 0.048   | 110                          | 59056                |
| Other                 | susd4         | 2.58        | -1.07  | 0.036   | 75                           | 96935                |
| Other                 | spink3        | 18.79       | 1.18   | 0.005   | 80                           | 20730                |

<sup>a</sup>: see material and methods, Chapter 5

<sup>b</sup>: Average signal from all three replicates of wildtype animals in liver given; only mRNAs in which the average signal was  $\geq 50$  units were included

<sup>c</sup>: Gene IDs were retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>)