TRANSCRIPTIONAL REGULATION OF SLC6A19
ALONG THE CRYPT–VILLUS AXIS

EMRAH TÜMER

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
OF THE AUSTRALIAN NATIONAL UNIVERSITY

February 2016

© Copyright by Emrah Tümer 2016
All Rights Reserved
Statement of Originality

The research presented in this thesis was performed in the Research School of Biology, The Australian National University under supervision of Professor Stefan Bröer. The data is the result of original research, has not previously been submitted for a degree, and my own work, with all contributions from others explicitly stated.
Abstract

Digestion of proteins takes place mainly in the small intestine followed by the absorption of amino acids and peptides by enterocytes. The absorption surface area in the intestine is drastically increased by finger-like protrusions along the small intestinal lumen (called villi) and invaginations (called crypts). Enterocytes develop from stem cells in the crypt and differentiate into mature enterocytes while moving along the crypt–villus axis. The Slc6a19 mRNA and its resulting B0AT1 protein show a striking increase of expression towards the tip of the villi and are virtually absent in the crypts.

In my PhD research, I investigated transcription factors and epigenetic modulators that regulate Slc6a19 expression. The Slc6a19 gene encodes the neutral amino acid transporter B0AT1, which provides the main mechanism of neutral amino acid absorption in the intestine. My experiments demonstrate that epigenetic modifications and transcription factor distribution orchestrate B0AT1 expression along the crypt–villus axis.
Acknowledgements

Firstly, I would like to thank my supervisor Professor Stefan Bröer for his support and guidance through my PhD research and thesis editing. This thesis would have not been possible without him.

I would like to thank Angelika Bröer for sharing her knowledge and techniques on this project, and delicious pastry she made during my time here.

I would also to thank Dr. Torsten Jülich, for his patience, friendship and encouragement.

Further, I would like to thank to my advisory member Dr. Juleen Cavanaugh for her help in these years.

Most importantly I would like to thank my family for their unconditional love and support of my endeavours.

My utmost gratitude goes to Allah for everything that I am blessed with.
Dedication

To My Parents & Elif
Publication and presentations arising from this study

Peer-reviewed Journal:


Conference Presentations:

- Transcriptional regulation of B₀AT₁ (Slc6a19) gene, the gene mutated in Hartnup disorder. The 2011 Gage Ion Channels and Transporters Conference, 18th-20th April 2011, Canberra, Australia
- Does DNA methylation play a role in B₀AT₁ transcriptional regulation along crypt-villus axis? International Union of Physiological Sciences Congress 2013, 21-26 July 2013, Birmingham, UK.
- Does DNA methylation play a role in B₀AT₁ transcriptional regulation along crypt-villus axis? CHARM Canberra Health Annual Research Meeting 2013, 20-23 August 2013, Canberra, Australia.
- The role of DNA methylation in B₀AT₁ transcriptional regulation along crypt-villus axis. The American Society of Human Genetics 2013, 22-26 October 2013, Boston, USA.
Table of Contents

Statement of Originality ........................................................................................................... I
Abstract ..................................................................................................................................... II
Acknowledgements .................................................................................................................. III
Dedication ................................................................................................................................. IV
Publication and presentations arising from this study .......................................................... V
List of Figures .......................................................................................................................... XI
List of Tables ............................................................................................................................ XIII
Abbreviations .......................................................................................................................... XIV

Chapter 1 Introduction .............................................................................................................. 16
1.1 Amino acid transport systems ......................................................................................... 16
1.2 The small intestine ........................................................................................................... 17
  1.2.1 Crypt–villus functional unit ....................................................................................... 17
1.3 Homeostatic regulation of intestinal development ....................................................... 20
  1.3.1 The Wnt signalling is involved in intestinal development ........................................ 20
    1.3.1.1 Wnt/β-catenin signalling in the intestine ......................................................... 21
  1.3.2 Transcription factors (TFs) involved in intestinal development ............................ 22
    1.3.2.1 SOX9 .............................................................................................................. 22
    1.3.2.2 Interaction between SOX9 and Wnt/β-catenin ................................................. 23
    1.3.2.3 HNF1a ........................................................................................................... 24
    1.3.2.4 HNF4a ........................................................................................................... 25
    1.3.2.5 Other TFs ....................................................................................................... 25
1.4 B₀AT1 protein .................................................................................................................. 26
  1.4.1 Structure of B₀AT1 protein ....................................................................................... 26
  1.4.2 Tissue distribution and cellular localization of B₀AT1 .......................................... 27
  1.4.3 Interaction of B₀AT1 with other proteins ............................................................... 28
    1.4.3.1 Hartnup disorder ............................................................................................... 29
1.5 SLC6A19 gene structure ................................................................................................... 30
1.6 Transcriptional regulation ............................................................................................... 30
  1.6.1 Promoter recognition ................................................................................................. 31
    1.6.1.1 Promoter elements ........................................................................................... 32
      1.6.1.1.1 Core promoter elements .......................................................................... 32
      1.6.1.1.2 Proximal promoter elements .................................................................... 33
    1.6.1.2 Enhancers ......................................................................................................... 33
Chapter 2  Material and Methods ................................................................. 43

2.1  Materials ..................................................................................... 43
    2.1.1  Solutions ................................................................................. 43
    2.1.2  Reagents and commercial kits ............................................. 44
    2.1.3  Antibodies ............................................................................. 45
    2.1.4  Vectors .................................................................................. 45
    2.1.5  Animals ................................................................................ 46
    2.1.6  Antibiotics ............................................................................. 47

2.2  Bacterial experiments ...................................................................... 47
    2.2.1  Bacterial strains ..................................................................... 47
    2.2.2  Transformation of electrocompetent bacteria ....................... 48
    2.2.3  Transformation of chemically competent bacteria ................ 48
    2.2.4  Bacterial growth .................................................................... 49

2.3  Eukaryotic cell culture ................................................................. 49
    2.3.1  Cell lines and cell culture ..................................................... 49
    2.3.2  Counting of cell numbers ..................................................... 50

2.4  Isolation and manipulation of DNA .............................................. 50
    2.4.1  DNA isolation ........................................................................ 50
        2.4.1.1  DNA extraction from mouse tissues .................................. 50
        2.4.1.2  Plasmid DNA isolation ..................................................... 50
        2.4.1.3  Rapid screening of transformed colonies ......................... 50
    2.4.2  DNA experiments ................................................................. 51
        2.4.2.1  Restriction enzyme digestion ........................................... 51

1.6.2  Transcription factors .............................................................. 33
    1.6.2.1  Regulation of TFs .............................................................. 35
1.6.3  Epigenetics .............................................................................. 35
1.6.4  Chromatin modifications .......................................................... 35
    1.6.4.1  Histone modifications ....................................................... 36
    1.6.4.2  Chromatin remodelling ................................................... 37
1.6.5  DNA methylation ...................................................................... 38
    1.6.5.1  CpG islands ....................................................................... 39
1.6.6  DNA methylation and transcription ........................................ 39
1.7  Project aims .................................................................................. 42
1.8  Research questions ....................................................................... 42
Chapter 4  Transcriptional regulation of Slc6a19 gene along the crypt–villus axis. 81

4.1 Introduction ............................................................................................................ 81
  4.1.1.1 Hepatocyte nuclear factors ........................................................................ 82
  4.1.1.2 SPDEF ........................................................................................................ 83
  4.1.1.3 SOX9 ........................................................................................................... 83
  4.1.1.4 NEUROG3 ................................................................................................. 83
  4.1.1.5 FOXA2 ........................................................................................................ 83
  4.1.1.6 CDX2 .......................................................................................................... 83
  4.1.1.7 GATA4 ....................................................................................................... 84

4.2 Results .................................................................................................................... 84
  4.2.1 Prediction of putative nuclear receptor response elements in the Slc6a19 promoter84
  4.2.2 Analysis of possible TF candidates ................................................................. 87
  4.2.3 HNF1a can activate the Slc6a19 promoter ...................................................... 89
  4.2.4 HNF4a activates the Slc6a19 promoter ........................................................... 93
  4.2.5 Synergistic effects of HNF1a and HNF4a co-transfection on the Slc6a19 promoter activity................................................................. 100
  4.2.6 SPDEF can activate Slc6a19 transcription in vitro ........................................ 101
  4.2.7 Identification of possible repressor(s) along the crypt–villus axis ................. 103
  4.2.8 Analysis of SOX9-binding on Slc6a19 promoter .......................................... 106
  4.2.9 Analysis of mouse Collectrin and Ace2 promoters ........................................ 114
  4.2.10 Effects of HNF1a mutation on Ace 2 or Tmem27 promoter activities.......... 118

4.3 Discussion .............................................................................................................. 119
  4.3.1 Novel Hepatocyte Nuclear Factor response elements on the Slc6a19 promoter 119
  4.3.2 Negative effect of SOX9 on the Slc6a19 activation ........................................ 120
  4.3.3 Ace2 and Collectrin ....................................................................................... 121
  4.3.4 Summary ...................................................................................................... 121

Chapter 5  Methylation and histone modification of Slc6a19 promoter in mouse intestine, liver and kidney ............................................................................................................ 122
5.1 Introduction ........................................................................................................... 122
5.1.1 DNA methylation ................................................................................................. 123
5.1.2 DNase I hypersensitive sites .................................................................................. 125
5.2 Results ...................................................................................................................... 126
5.2.1 Analysis of H3K27ac and H3K4me3 histone modification associated with the 5’ flanking region of Slc6a19 promoter ........................................................................... 126
5.2.2 Modifications of H3K4m3 and H3K36m3 in the N-terminus of histone H3 ......... 128
5.2.3 DNase I hypersensitive sites on the Slc6a19 promoter ......................................... 131
5.2.4 DNA methylation profile of the Slc6a19 promoter in mouse ................................ 131
5.3 Discussion ................................................................................................................. 134
5.4 Summary .................................................................................................................. 136

Chapter 6 General discussion ....................................................................................... 137
6.1 Links between DNA methylation and transcription .............................................. 137
6.2 The current model .................................................................................................... 139
6.3 Limitations and future studies ................................................................................ 141
6.4 Summary .................................................................................................................. 142

Chapter 7 References ................................................................................................. 143

Chapter 8 Appendix .................................................................................................... 156
8.1 Plasmids .................................................................................................................... 156
8.2 Thermal cycler conditions ....................................................................................... 159
8.3 Gene ontology terms enriched in the villus (significant changes in expression of genes between crypt and villus) ...................................................................................... 160
8.4 Gene ontology terms enriched in the crypt (significant changes of genes between crypt and villus) ........................................................................................................ 163
List of Figures

Figure 1.1: The intestinal crypt–villus structure ................................................................. 19
Figure 1.2: The canonical Wnt pathway ........................................................................... 21
Figure 1.3: The SOX9 expression profile in the gastrointestinal system ....................... 23
Figure 1.4: Homology model of B4AT1 based on the structure of LeuT ....................... 27
Figure 1.5: Chromosomal locations of Slc6a19 in human and mouse ............................. 30
Figure 1.6: The central dogma of gene regulation .......................................................... 31
Figure 1.7: An eukaryotic promoter region and its regulatory elements ....................... 32
Figure 1.8: Structure of histone tails .................................................................................. 36
Figure 1.9: Chromatin remodelling and histone modifications are working together ......... 38
Figure 1.10: Maintenance methyltransferase ................................................................. 39
Figure 1.11: Collective effect of DNA methylation and histone modifications on the transcription ......................................................................................................................... 41
Figure 2.1: pGL4.12[luc2CP] plasmid and pGL4.74[hRluc/TK] plasmid were used as reporter plasmids in the Dual-Luciferase® Reporter Assay system ............................................ 55
Figure 2.2: Quality analysis of microarray samples ........................................................... 61
Figure 3.1: Schematic presentation of the crypt–villus structure and expression profiles of Slc6a19 and Lgr5 ........................................................................................................... 71
Figure 3.2: Confirmation of the microarray results by qRT-PCR ....................................... 72
Figure 3.3: Proportion of TFs along the crypt–villus axis ................................................ 78
Figure 4.1: Distribution of differentiated intestinal cells along the crypt–villus axis ......... 82
Figure 4.2: Promoter sequence alignment of Slc6a19 ....................................................... 85
Figure 4.3: Effects of the different TFs on the Slc6a19 promoter activity ....................... 88
Figure 4.4: Characterization of STAT3-mediated transcriptional activation by reporter gene assays ........................................................................................................................................... 89
Figure 4.5: Locations of HNF1a-binding sites on the mouse Slc6a19 promoter ......... 90
Figure 4.6: Identification of the Slc6a19 promoter regions responsive to the HNF1a TF using luciferase reporter experiments in vitro ......................................................... 91
Figure 4.7: Detection of HNF1a-binding sites using mutated constructs ................... 92
Figure 4.8: HNF1a directly binds to the Slc6a19 promoter ........................................... 93
Figure 4.9: Location of putative HNF4a-binding sites on the Slc6a19 promoter ... 94
Figure 4.10: HNF4a induces luciferase expression driven by the Slc6a19 promoter in vitro .............................................................. 95
Figure 4.11: Determination of HNF4a-binding sites using mutated constructs ..... 96
Figure 4.12: Reducing the distance between HNF4a-binding sites enhances in vitro luciferase expression driven by the Slc6a19 promoter .................................... 97
Figure 4.13: Detection of the second putative HNF4a-binding site using mutated constructs .............................................................. 98
Figure 4.14: HNF4a directly binds to the Slc6a19 promoter ......................................... 99
Figure 4.15: The combinatorial effect of CDX2 and HNF TFs on the Slc6a19 promoter activity ........................................................................................................................................... 101
Figure 4.16: Predicted binding sites for the mouse SPDEF ........................................ 102
Figure 4.17: SPDEF induces in vitro luciferase activity driven by the Slc6a19 promoter ............................................................................................................................................... 103
Figure 4.18: Expression profiles of putative TFs in the villus and crypt as determined by qPCR and microarray studies .................................................................................................................. 105
Figure 4.19: Predicted binding sites for SOX9 on the mouse Slc6a19 promoter ....... 106
Figure 4.20: SOX9 inhibits in vitro luciferase activity induced by HNF1a ............... 107
Figure 4.21: SOX9 inhibits luciferase expression activated by HNF4a in vitro .......... 108
Figure 4.22: Investigation of the putative SOX9-binding site using wild-type and mutated pSlc6a19(–972/+57) constructs........................................................................... 109
Figure 4.23: SOX9 directly binds to the Slc6a19 promoter ......................................... 110
Figure 4.24: Predicted binding sites for SOX9 ................................................................ 111
Figure 4.25: Investigation of the SOX9-binding sites using mutated constructs. ........ 112
Figure 4.26: The effect of SOX9(H65Y) on the pSlc6a19(–437/+57)............................... 113
Figure 4.27: The combinatorial effects of SOX9(H65Y) and SOX9-binding-site mutation on pSlc6a19(–437/+57)............................................................................................................ 114
Figure 4.28: Sequence alignment of promoters of Collectrin and Ace2 from different species ................................................................................................................................................. 115
Figure 4.29: Effects of the different TFs on Collectrin and Ace2 promoters .............. 117
Figure 4.30: The effect of HNF1a(R131W) on Slc6a19, Ace2 and Tmem27 promoters. ............................................................ 118
Figure 5.1: Possible positions of histone tail modification ........................................... 122
Figure 5.2: DNA methylation and demethylation ....................................................... 125
Figure 5.3: DNase I hypersensitive site in the chromatin context............................. 126
Figure 5.4: H3K27ac histone modification of the Slc6a19 promoter region in liver and intestine .................................................................................................................................................................................. 127
Figure 5.5: H3K4me3 histone modification of the Slc6a19 promoter region in liver tissue and crypt–villus fractions .................................................................................................................................................. 128
Figure 5.6: Histone modifications and polymerase II binding to the Slc6a19 promoter ......................................................................................................................................................................................... 130
Figure 5.7: DNase I hypersensitive sites on the mouse the Slc6a19 promoter ............ 131
Figure 5.8: Cpg dinucleotides positions in the mouse Slc6a19 promoter ................. 132
Figure 5.9: Methylation status of the Slc6a19 promoter in the liver and kidney ... 133
Figure 5.10: Methylation status of the Slc6a19 promoter region in the intestine... 134
Figure 6.1: Working model for regulation of the Slc6a19 along crypt–villus axis. 140
Figure 8.1: A plasmid map of the pGL4, the basic luciferase reporter vector, detailing key constructs and restriction enzyme sites .......................................................... 156
Figure 8.2: A plasmid map of the pGL4, the basic luciferase reporter vector, detailing key constructs and restriction enzyme sites ........................................................................ 157
Figure 8.3: A plasmid map of the pcDNA3.1 ................................................................ 158
List of Tables

Table 2.1: Solutions used in this study ................................................................. 43
Table 2.2: Antibodies used in this study ............................................................... 45
Table 2.3: Vectors used in this study ................................................................. 46
Table 2.4: Bacterial strains were used in this study ............................................ 47
Table 2.5: Cell lines used in this study ............................................................... 49
Table 2.6: Conditions of a PCR cycling .............................................................. 52
Table 2.7: Primer sequences used for site-directed mutagenesis ......................... 53
Table 2.8: PCR conditions for site-directed mutagenesis .................................... 54
Table 2.9: Primers used for amplification of the Slc6a19 promoter ...................... 56
Table 2.10: Oligonucleotides used for the synthesis of B0AT1 promoter deletion constructs ........................................................................................................... 57
Table 2.11: Oligonucleotides used for the synthesis of Ace2 and Collectrin promoter constructs ................................................................. 58
Table 2.12: Primers used for cloning of TFs .......................................................... 59
Table 2.13: Primer sets used for RT-PCR and qRT-PCR amplifications ............... 62
Table 2.14: Primers used for sequencing ............................................................. 65
Table 2.15: The Sequencing PCR conditions ...................................................... 65
Table 2.16: Primers used in DNA methylation analysis ....................................... 66
Table 2.17: ChIP primer sets used for B0AT1 promoter analysis ......................... 68
Table 3.1: List of gene ontology (GO) categories significantly up-regulated in cells of the villus tip .................................................................................... 73
Table 3.2: List of gene ontology (GO) categories significantly up-regulated in crypt cells ............................................................................................................. 74
Table 3.3: Overrepresented KEGG pathways in the villus ................................... 76
Table 3.4: KEGG pathways enriched in the crypt ............................................... 77
Table 4.1: Expression of candidate TFs in crypts or villi ..................................... 86
Table 8.1: Cycling conditions for the PfuUltra enzyme ........................................ 159
Table 8.2: Cycling condition for the Phusion® high-fidelity enzyme .................... 159
Table 8.3: Cycling condition for ChIP PCR ......................................................... 159
Table 8.4: Predicted TF-binding sites in the Collectrin promoter ......................... 165
Table 8.5: Predicted TF-binding sites in the Ace2 promoter ................................ 166
Table 8.6: List of TFs which were significantly expressed along crypt–villus axis (p < 0.05) ........................................................................................................... 167
**Abbreviations**

ACE2 Angiotensin-converting enzyme 2  
Caco-2 Colon adenocarcinoma cell lines  
cDNA Complementary DNA  
ChIP Chromatin immunoprecipitation  
CpG Cytosine-guanine dinucleotides  
C/EBP CCAAT/enhancer-binding protein  
DMEM Dulbecco’s Modified Eagle’s Medium  
DNA Deoxyribonucleic Acid  
dNTP Dinucleotide triphosphate  
DNMT DNA methyltransferase  
DTT Dithiothreitol  
EDTA Ethylenediaminetetraacetic acid  
*E. coli* *Escherichia coli*  
ENCODE Encyclopedia of DNA Elements  
FBS Foetal bovine serum  
FOXA2 Fork-head box A2  
GAPDH Glyceraldehyde-3-phosphate dehydrogenase  
h Hours  
H3K27ac Acetylated histone H3 tail at lysine 27  
H3K4me3 Tri-methylated histone H3 at lysine 4  
HEK293 Human Embryonic Kidney 293  
HDAC Histone deacetylase  
HNF Hepatocyte nuclear factor  
HMT Histone methylase  
KEGG Kyoto Encyclopaedia of Gene and Genomes  
LB Luria-Bertani broth for bacteria culture  
Lgr5 Leucine-rich repeat containing G-protein coupled receptor 5  
MBD Methyl-CpG-binding domain  
PBS Phosphate Buffered Saline
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>HAT</td>
<td>Histone acetyl transferases</td>
</tr>
<tr>
<td>HEPES</td>
<td>hydroxyethylpiperazineethanesulfonic acid</td>
</tr>
<tr>
<td>kb</td>
<td>Kilo bases</td>
</tr>
<tr>
<td>min</td>
<td>Minutes</td>
</tr>
<tr>
<td>MeCP2</td>
<td>Methyl-CpG-binding-protein</td>
</tr>
<tr>
<td>MODY</td>
<td>Maturity onset diabetes of the young</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>PPAR</td>
<td>Peroxisome proliferator-activated receptors</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative real-time RT-PCR</td>
</tr>
<tr>
<td>RPMI</td>
<td>Roswell park memorial institute</td>
</tr>
<tr>
<td>sec</td>
<td>Seconds</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecylsulfate</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>$T_m$</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>UCSC</td>
<td>University of California, Santa Cruz</td>
</tr>
<tr>
<td>V</td>
<td>Volt</td>
</tr>
<tr>
<td>Mnase</td>
<td>Micrococcal nuclease</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative PCR</td>
</tr>
</tbody>
</table>
Chapter 1  Introduction

1.1 Amino acid transport systems

Proteins are an essential part of nutrition and are required for normal functioning and maintenance of the human body. Once ingested, various gastrointestinal digestive enzymes break down the proteins into amino acids and peptides. Amino acids and peptides are then absorbed by epithelial cells (enterocytes) lining the small intestine. Amino acid absorption into enterocytes is mediated by many different amino acid transport systems (Silk et al., 1985).

After absorption, amino acids are delivered to all the organs and to the kidneys by the circulatory system. To prevent their loss into the urine, they are reabsorbed in proximal tubules by amino acid transporters after filtration by the kidney glomerulus. Different types of amino acid transport systems function to absorb amino acids and peptides across the enterocyte cell membrane, and in the proximal tubule including parts of the renal glomerulus. Renal amino acid transport systems are similar to the intestinal transporters (Young and Freedman, 1971).

The pioneering amino acid transport studies conducted in the 1960s (Oxender and Christensen, 1963) established that distinct amino acid transport systems accept different groups of amino acids, and that these systems may exhibit overlapping substrate specificities. Distinct transport systems for neutral, cationic and anionic amino acids were subsequently identified (Oxender and Christensen, 1963). Later, intestinal transporters for neutral amino acids, cationic amino acids, anionic amino acids, glycine, proline and hydroxyproline were described and characterised in the 1980s (Munck, 1980, Stevens and Wright, 1987).

Several amino acid transport systems have been characterised in epithelial cells. System b^0,+ transports cationic and neutral amino acids. It is Na^+-independent (Chillarón et al., 2001). A heterodimer formed by SLC3A1 and SLC7A9 was identified as the molecular correlate of this system (Feliubadaló et al., 1999). System B^0 is a Na^+-dependent transporter that carries all neutral amino acids, but has limited activity for glycine or proline (Stevens et al., 1982). The molecular correlate of the system B^0 is B^0AT1 (SLC6A19) which is the major neutral amino acid transporter (Bröer et al., 2004). The Na^+-dependent IMINO system preferentially transports glycine and proline (Stevens and Wright, 1987). The molecular correlate of this system is apical transporter SLC6A20.
(Kowalczyk et al., 2005). Anionic amino acids are taken up by system X\textsuperscript{\textminus AG}. It is a Na\textsuperscript{+}-dependent aspartate/glutamate transporter. It has been showed that EAAT3 the molecular correlate of this transporter (Kanai and Hediger, 2003). Proline and glycine are transported by the IMINO system (Stevens and Wright, 1985). The molecular correlate of this system has been identified as the proton amino acid transporter PAT1 (Anderson, 2003).

1.2 The small intestine

The gastrointestinal tract originates from three embryonic germ layers: the endoderm, mesoderm and ectoderm. Anatomically, the gut has been divided into two parts: the small intestine and the large intestine. The small intestine comprises the duodenum, the jejunum, and the ileum (Barker et al., 2008). Histologically, the wall of the small intestine consists of the following layers: mucosa, submucosa, muscularis propria, and serosa. The mucosal layer contains cuboidal epithelial cells, the connective tissue, and the muscular layers. The submucosa layer contains the connective tissue that supports the mucosa and the muscular layer; the latter is responsible for peristaltic movements throughout the gastrointestinal tract (Sancho et al., 2004).

1.2.1 Crypt–villus functional unit

Proteins are digested mainly in the small intestine, and the resultant protein-derived nutrients are absorbed by epithelial cells (van der Flier and Clevers, 2009). The intestinal absorptive surface area is highly expanded by finger-like microscopic protrusions into the small intestinal lumen, called villi, and by invaginations, called crypts (Heath, 2010). The bottom of an intestinal crypt houses five or six stem cells; these cells divide every 24 h (Barker et al., 2007). Such a high growth rate is necessary to replace the epithelial cells which slough off at the tip of a villus every 24 h (Sato et al., 2009). In addition to progenitor cells, there are 25-42 clonogenic stem cells which can regenerate progenitor cells if necessary (Booth and Potten 2000). The newly generated cells differentiate into four different cell types: enterocytes, goblet cells, enteroendocrine cells, and Paneth cells (Huet et al., 1987). Enterocytes (in this thesis the term enterocytes refers to the absorptive cells of the differentiated epithelium, not including goblet, enteroendocrine or Paneth cells) are the most numerous cell type of the mature intestinal epithelium and they absorb
most nutrients from the lumen. Enterocytes are crucial for the final step of protein
digestion. Peptides of 2-10 residues length, are generated by the action of gastric and
pancreatic proteases. These are further digested into individual amino acids, di- and
tripeptides by brush-border peptidases located on enterocytes (Erickson and Kim, 1990).
Goblet cells secrete the intestinal mucus into the lumen for lubrication and protection of
the epithelium. The enteroendocrine cells secrete hormones which regulate
gastrointestinal functions and communicate with the rest of the body. The Paneth cells
secrete lysozymes, antimicrobials, and defensins all of which have important functions in
antibacterial protection and innate immunity. The first three cell types migrate towards
the villus tip; the Paneth and stem cells are the only cells that reside at the crypt bottom
(Figure 1.1) (Clatworthy and Subramanian, 2001).
The stem cells residing in the crypt renew the epithelial cells on the villus. Arrows indicate migration of the differentiated cells along the crypt–villus axis. Epithelial cells migrate towards the villus tip, where they slough off into the intestinal lumen. The Paneth cells and stem cells reside within the crypt niche. From (Crosnier et al., 2006).

Figure 1.1: The intestinal crypt–villus structure
1.3 Homeostatic regulation of intestinal development

Different mechanisms regulate maturation of the intestinal cells from the progenitor cells in the crypt to mature cells in the villus. Here, I briefly describe the important signalling pathways and the transcription factors that are involved in intestinal development.

1.3.1 The Wnt signalling is involved in intestinal development

Wnt signalling is involved in many aspects of development, including proliferation and differentiation of the intestinal epithelium (van de Wetering et al., 2002). Proteins involved in Wnt signalling are conserved in animals (Cadigan and Nusse, 1997). Wnt signalling is conveyed through three different intracellular signalling cascades: the first is the canonical Wnt pathway (Wnt/β-catenin signalling) which controls β-catenin stability. β-Catenin has an important function in cell adhesion. The second is the Wnt/Ca\(^{2+}\) pathway (non-canonical Wnt pathway) that activates the calcium/calmodulin-dependent protein-kinase complex in the cytoplasm. Lastly, another non-canonical Wnt pathway (the planar cell-polarity pathway) interacts with the Rho family of GTPases, ultimately controlling cytoskeletal dynamics. These pathways are initiated by the Wnt protein, which is produced by mesenchymal cells underlying the epithelial cell layer and surrounding the intestinal crypt. The Wnt protein binds to the Frizzled cell-surface receptor (Lewis, 2007).

When Wnt signalling is inactive, the cytoplasmic β-catenin is captured by a destruction complex which is composed of at least four proteins: the scaffolding protein axin, the tumour suppressor adenomatous polyposis coli (APC), casein kinase I, and glycogen synthase kinase 3 (GSK3). This complex phosphorylates cytoplasmic β-catenin which is then ubiquitinated and degraded in the proteasome (Figure 1.2) (Scoville et al., 2008, Lewis, 2007, Sokol, 2007).

In the presence of Wnt, activated cell’s surface Frizzled receptor inhibits the destruction complex, causing β-catenin translocation from the cytoplasm into the nucleus. In the nucleus, β-catenin interacts with the T-cell-factor (TCF) to activate transcription of Wnt target genes (Figure 1.2) (Logan and Nusse, 2004).
Figure 1.2: The canonical Wnt pathway

(Left panel) In the absence of Wnt stimulation, the β-catenin is degraded by the destruction complex, which includes adenomatosis polyposis coli (APC), axin, glycogen synthase kinase 3 (GSK3), and casein kinase I (CKI). (Right panel) The binding of Wnt ligand to its receptor Frizzled (Fz) and the LRP co-receptor, inactivates the destruction complex in the cytoplasm. Non-phosphorylated β-catenin translocates to the nucleus and binds the T-cell factor (TCF) to activate the transcription of target genes. From (Clevers, 2006).

1.3.1.1 Wnt/β-catenin signalling in the intestine

The Wnt signalling pathway is important in the intestinal homeostasis, by regulating stem cell populations (Clevers, 2006).

Mutations of various members of this tightly regulated pathway lead to serious intestinal pathologies. For example, deletion of the mouse intestinal Tcf4 gene causes loss of progenitor cells in the proliferative niche of the intestine (Korinek et al., 1998). Thus, the Wnt pathway is fundamental to the maintenance of the proliferative cells. In agreement with this notion, mutations of Apc, (a tumour suppressor gene which negatively regulates the Wnt pathway), causes crypt enlargement and interruption of cell migration from the crypt to the villus (Sansom et al., 2004). Apc mutations were also
found to be associated with intestinal cancers in humans (Morin et al., 1997). In addition, the Wnt pathway stimulates maturation of the Paneth cells in the intestinal crypt (van Es et al., 2005).

1.3.2 Transcription factors (TFs) involved in intestinal development

Many transcription factors have been implicated in enterocyte maturation. Here, the most important candidates are introduced.

1.3.2.1 SOX9

The SOX (Sry-related HMG box) family of transcription factors contain the HMG (high-mobility group) DNA-binding domain. SOX and Sry (sex determining region Y) proteins share ~50% identity in their HMG domain. SOX factors are well-conserved proteins amongst vertebrates (Dong et al., 2004). At least 20 different human and murine SOX proteins exist, and these are subdivided into eight subgroups (Bowles et al., 2000). SOX proteins are mainly involved in organ development and tissue maturation, but mutated SOX proteins are also associated with malignancies (Kiefer, 2007).

SOX9 is an important SOX protein involved in crypt–villus development (Kormish et al., 2010). The main function of SOX9 in the intestine is enterocyte maturation and crypt formation (Bastide et al., 2007). SOX9 is also essential for the development of male gonads, the neural crest, and cartilages (Morais da Silva et al., 1996, Zhao et al., 1997, Cheung and Briscoe, 2003). In the mouse, SOX9 is expressed in many different organs; Sox9 mRNA and protein can be detected in the lower third of the intestinal crypts, in colon crypts, the Paneth cells (Blache et al., 2004), and in the epithelium of pancreatic and biliary ducts (Figure 1.3) (Furuyama et al., 2011). TCF proteins, which function in Wnt-targeted gene regulation, also have the same DNA-binding domain as SOX does (van Beest et al., 2000).
Figure 1.3: The SOX9 expression profile in the gastrointestinal system

SOX9 is significantly expressed in the extrahepatic biliary tract, pancreatic duct cells and intestinal crypts. From (Furuyama et al., 2011).

1.3.2.2 Interaction between SOX9 and Wnt/β-catenin

The Wnt pathway and SOX expression are important in the development of different organs. The interaction between SOX9 and β-catenin has been shown to be essential for cartilage development. β-catenin deletion, and inversely, SOX9 overexpression can interrupt chondrocyte differentiation (Akiyama et al., 2004). In the female gonad, activated SOX9 can repress Wnt/β-catenin pathway, and this leads to a male sex determination during development (Bernard et al., 2008). In the intestines, Wnt
signalling orchestrates crypt formation and promotes stem cell proliferation. SOX9, by contrast, inhibits the Wnt pathway and promotes differentiation of the stem cells into mature intestinal cell types. Therefore, interaction between SOX9 and Wnt is essential for the cell proliferation–maturation balance (Kormish et al., 2010).

SOX9 mutations have been shown to cause loss of the Paneth cells in the crypts (Bastide et al., 2007, Mori-Akiyama et al., 2007). Cdx2 and Muc2 proteins are highly expressed in the mature enterocytes in the villi and known as differentiation markers; however, in the crypts, they are transcriptionally repressed by SOX9 (Blache et al., 2004). As a result, SOX9 is thought to be a key transcription factor to repress the proliferation programs in the progenitor cells in the crypts.

In summary, these studies indicate that SOX9 down-regulates the Wnt pathway, and that, together with the Wnt/β-catenin pathway, it plays an important role in the homeostasis of intestinal development.

1.3.2.3 HNF1a

Hepatocyte nuclear factors (HNFs) encode a family of transcription factors. These were first identified as transcription factors that regulate hepatocyte specific genes. This family includes five transcription factors; HNF1, HNF4, HNF6, C/EBP (CCAAT/enhancer binding protein), and FOXA2 (fork-head box A2) (Costa et al., 2003).

Studies in mice revealed that Hepatocyte nuclear factor 1 alpha (HNF1a) is expressed in kidney, pancreas, intestine and stomach (Blumenfeld et al., 1991). In the kidney, it is exclusively expressed in the proximal tubules but not throughout the entire nephron (Pontoglio et al., 1996). HNF1a-deficient mice die around weaning time due to hepatomegaly. Also, these mice show signs of aminoaciduria (Lee et al., 1998, Pontoglio et al., 1996). The HNF4a-knockout mice show the renal Fanconi syndrome because of dysfunction of the proximal renal tubules. The renal Fanconi syndrome is characterised by diabetes, glycosuria, and aminoaciduria, indicating loss of renal and pancreatic functions (Pontoglio et al., 1996). It has been shown that HNF1a-knockout mice excrete three times more of the neutral amino acid proline in urine than wildtype mice (Bonzo et al., 2010).
Within the gastrointestinal system, HNF1a mRNA can be detected from the stomach down to the colon, and the gene is expressed along the crypt–villus axis in the intestine (Boudreau et al., 2002, Serfas and Tyner, 1993). A study showed that in the intestine of HNF1a-null mice, epithelial barrier function, Paneth cell differentiation, and crypt cell proliferation were disrupted (Lussier et al., 2010). HNF1a is also essential for pancreatic islets development (Pontoglio et al., 1996, Shih et al., 2001, Tronche et al., 1997). Mutations in the HNF1a gene are the most common reason for maturity onset diabetes of the young (MODY) (Yamagata et al., 1996b).

1.3.2.4 HNF4α

HNF4α (hepatocyte nuclear factor 4 alpha) is a member of the nuclear receptor super family (Sladek et al., 1990), and is highly conserved in vertebrates (Ryffel, 2001). In mice, the HNF4α gene is located on chromosome 2 and it consists of 11 exons. The tissue expression profile of HNF4α is similar to that of HNF1α, and it was initially identified as a transcriptional regulator of hepatic genes such as the fibrinogen, albumin, and α1-antitrypsin (Cereghini et al., 1987). HNF4α is expressed in differentiated adult cells located in the liver, pancreas, stomach, intestine, skin and kidney (Taraviras et al., 1994). HNF4α is highly expressed in blastocysts at day 4.5 and in the visceral endoderm at day 5.5. It is also expressed in the embryonic liver and gut (Duncan et al., 1994). The HNF4α knockout is embryonically lethal, demonstrating its crucial role as a transcription factor in early development (Duncan et al., 1994). Certain HNF4α gene mutations cause maturity onset diabetes of the young (Yamagata et al., 1996a). Additional studies demonstrated that HNF4α is a key regulator of crypt formation and goblet-cell maturation (Garrison et al., 2006).

Interestingly, not only do HNF4α and HNF1α share similar expression patterns, but they also appear to be regulating each other’s transcription, being linked through a transcriptional feedback loop (Shih et al., 2001).

1.3.2.5 Other TFs

A variety of TFs are involved in intestinal cell differentiation. The transcription factor CDX2 is necessary for intestinal development (Lorentz et al., 1997) and together with HNF1α, regulates sucrase-isomaltase and lactase-phlorizin hydrolase gene
expression along the crypt-villus axis (Mitchelmore et al., 2000). GATA4 is another crucial TF which is required for activation of intestinal genes involved in nutrient absorption (Bosse et al., 2007). CDX2 and GATA4 appear to be sufficient to activate many of the enterocyte-specific genes in the mammals (Benoit et al., 2010). SPDEF and FOXA2 TFs are exclusively expressed in cells residing in crypts (Besnard et al., 2004). SPDEF is required for maturation of intestinal Paneth and goblet cells (Gregorieff et al., 2009), while FOXA2 regulates goblet cell differentiation via activation of Muc2, which induces expression of mucin secreted by the goblet cells (Ye and Kaestner, 2009).

1.4 B⁰AT1 protein

System B⁰ is a Na⁺-dependent, low-affinity transport system that takes up neutral amino acids. It was initially described as neutral brush border system (NBB) by a study of rabbit intestinal brush border membrane vesicles (Stevens et al., 1982) and afterwards, it was renamed as system B⁰, to designate broad neutral amino acids (Maenz and Patience, 1992).

In 2004, the molecular correlate of this transport system was identified in mouse and named B⁰AT1 (B⁰-like amino acid transporter 1). B⁰AT1 was found to be highly expressed on the apical membrane of intestinal brush borders and kidney proximal tubules (Bröer et al., 2004). It is a member of the solute carrier (SLC) family 6, also known as the neurotransmitter transporter family. Unlike other members of the solute carrier (SLC) family 6, the B⁰AT1 transport function is Cl⁻-independent (Bohmer et al., 2005).

1.4.1 Structure of B⁰AT1 protein

Structure of the SLC6 transporters is similar to the crystal structure of the bacterial leucine transporter (LeuT). The LeuT transporter of the bacterium Aquifex aeolicus is a structural homolog of the mammalian B⁰AT1 protein (Yamashita et al., 2005). LeuT has 12 transmembrane α-helices. Particularly, the first and sixth helix of both proteins show high similarity, containing residues that are important for substrate and ion binding. Both proteins mediate Na⁺-dependent, Cl⁻-independent transport of a broad range of neutral amino acids (Broer, 2006).
According to the LeuT-based homology model, the B⁰AT1 structure is characterized by 12 transmembrane helices and intracellular N- and C- termini. Helices 1 and 6 are crucial for substrate translocation in the protein. Additionally, helix 8 has a role in substrate binding (Figure 1.4) (Broer, 2006).

![Figure 1.4: Homology model of B⁰AT1 based on the structure of LeuT](image)

The B⁰AT1 model was generated based on the crystal structure of the bacterial leucine transporter, LeuT. Helix 1 (yellow) and helix 6 (blue) are indicated among other helices (pink). From (Broer, 2008).

### 1.4.2 Tissue distribution and cellular localization of B⁰AT1

The B⁰AT1 protein is mainly present in the kidney and the intestine. In the kidney, B⁰AT1 is highly expressed in the cells of the S1 and S2 segments of the proximal tubule (Romeo et al., 2006). Here, B⁰AT1 protein expression is limited to the brush border membrane of the proximal tubules (Kleta et al., 2004). B⁰AT1 expression in the digestive system is evident predominantly in the duodenum, the jejunum and the ileum (Terada et al., 2005). In the intestine, B⁰AT1 is highly expressed in the villus tip, but not at the crypt bottom (Bröer et al., 2004). B⁰AT1 is also expressed, to some extent, in the colon, pancreas, and prostate (Kleta et al., 2004).
1.4.3 Interaction of B⁰AT1 with other proteins

Intracellular trafficking of B⁰AT1 to the cell membrane is an important step required for its proper functioning. To date, two trafficking proteins have been found in the kidney and the intestine, namely collectrin (TMEM27) and angiotensin-converting enzyme 2 (ACE2), respectively.

Collectrin was named according to its presumed location in the cells of the renal collecting ducts. Collectrin expression was later confirmed in the pancreas, stomach, jejunum, ileum, heart, lung, spleen, and liver (Akpinar et al., 2005, Danilczyk et al., 2006). In 2006, Danilczyk and colleagues showed that in collectrin-deficient mice, neutral amino acids were excreted into the urine. Further analysis revealed that expression of collectrin was similar to B⁰AT1 expression in the kidney tubules (Danilczyk et al., 2006, Malakauskas et al., 2007). B⁰AT1 cell surface expression in the proximal tubule was significantly reduced in collectrin-null mice.

However, collectrin is not expressed in the intestinal brush border where B⁰AT1 is highly expressed (Kowalczuk et al., 2008). In the intestine, B⁰AT1 requires a different cofactor to transport it to the enterocyte apical membrane. Studies in Xenopus laevis oocytes showed that co-expression of the brush-border peptidase ACE2, which shows homology with the transmembrane domain of collectrin (Zhang et al., 2001), and B⁰AT1 increased the uptake of neutral amino acids due to increased B⁰AT1 surface expression (Kowalczuk et al., 2008, Camargo et al., 2009). ACE2 also has an important regulatory role in the renin-angiotensin system, which is crucial for renal and cardiovascular functions (Burrell et al., 2004). ACE2 is associated with diseases of the cardiovascular system (Crackower et al., 2002).

In collectrin-deficient mice, B⁰AT1 was almost absent in the luminal membrane of the renal proximal tubules, thereby causing neutral aminoaciduria (Danilczyk et al., 2006). By contrast, amino acid composition in the urine of ACE2-deficient mice was not affected (Camargo et al., 2009). The B⁰AT1 (Slc6a19)-null mice show aminoaciduria similar to collectrin-null mice but in addition have reduced body size and body weight and also show weak insulin responses in the pancreas (Broer et al., 2011).
1.4.3.1 Hartnup disorder

Hartnup disorder is an autosomal recessive inherited disorder which is characterized by aminoaciduria, and a photosensitive pellagra-like skin rash. Occasionally cerebellar ataxia, nystagmus and tremor are observed. It was first described in 1956 (Baron et al., 1956). Intestinal absorption and renal absorption of most neutral amino acids were reduced in this disorder. Nowadays, protein-rich diets appear to compensate for deficient neutral amino acid uptake; therefore, in developed countries, Hartnup patients are usually asymptomatic (Broer, 2009). Newborn screening programs in Australia and North America showed an incidence of about 1 case per 30,000 newborns (Levy 2001).

Neutral aminoaciduria is the diagnostic hallmark of Hartnup disorder. Aminoaciduria is caused by the renal defect, whereas other symptoms are likely caused by the lack of intestinal neutral amino acid uptake. The pellagra-like symptoms may result from a lack of tryptophan. Pellagra is typically caused by Niacin deficiency, the vitamin involved in NAD(P)H biosynthesis. However, the body can, to some extent, produce NAD(P)H, using the essential neutral amino acid tryptophan, the uptake of which is reduced in Hartnup disorder. NADPH among other roles is necessary for fatty acid biosynthesis and glutathione reduction. Typical manifestations of Pellagra are bilateral, symmetric and photosensitive dermatitis, dementia, and diarrhoea (Broer, 2009).

Homozygosity mapping performed by a Japanese group established that Hartnup disorder is linked to chromosome band 5p15 (Nozaki et al., 2001). Subsequently, Kleta et al. (Kleta et al., 2004) and Seow et al., (Seow et al., 2004) found that SLC6A19 gene mutations cause Hartnup Disorder. So far, 21 different mutations associated with Hartnup disorder have been identified, including missense and nonsense deletions, splicing and small deletions (Broer, 2009). Functionally, all of these mutations cause loss of neutral amino acid transportation (Broer and Palacin, 2011). Among these mutations, the most frequent in caucasians is 517 G→A causing a D173N replacement in the protein (Azmanov et al., 2008). While many mutations were not informative with regard to protein function, B⁰AT1(R240Q) results in loss of transport function only when auxiliary proteins are co-expressed with B⁰AT1. No defect was observed when it is expressed without auxiliary proteins (Kowalczyk et al., 2008). Subsequently, it was demonstrated that the collectrin binding site is in close proximity to residue R240 (Fairweather et al., 2015).
1.5  **SLC6A19 gene structure**

The *SLC6A19* gene encodes for B₀AT1. *SLC6A19* is located on the chromosomal region 5p15.33 in humans; in the mouse, it is located on chromosome 13 and maps to region 13C1. The gene has 12 exons and encodes a protein of 634 amino acids in both species (Figure 1.5) (Broer, 2006).

![Figure 1.5: Chromosomal locations of Slc6a19 in human and mouse](image)

Location of human and mouse *Slc6a19* genes using the UCSC Genome Browser (http://genome.ucsc.edu). 1) The human *Slc6a19* is located at chr5p15. 2) The mouse *Slc6a19* is located at chr13qC1.

1.6  **Transcriptional regulation**

Almost every cell has an identical copy of the entire genomic material, yet each cell type expresses its unique complement of proteins (Russell, 2002). Activation of specific network of genes during differentiation and development is the basis of existence of many different cell types. Gene expression can be regulated at multiple points (Figure 1.6), the most important of which occurs at the initiation of transcription (Alberts, 2004).
Well-orchestrated interactions between DNA regulatory elements (which include promoter-binding elements), RNA polymerase enzymes, and histones ensure timely and well-coordinated transcription of different genes. Accessibility of transcription factors to DNA is regulated by chromatin modifications, thereby constituting an additional level of transcriptional regulation.

### 1.6.1 Promoter recognition

To initiate the transcriptional process, RNA polymerases must be positioned upstream of the first exon of the target gene in the promoter region. Formation of a transcription preinitiation complex, which incorporates a polymerase and several general transcription factors, is necessary for RNA polymerase to bind to its correct site with correct orientation (Wolpert, 2007).

---

**Figure 1.6: The central dogma of gene regulation**

The control points during gene expression and protein synthesis are shown. Control of transcription and translation occurs at three major checkpoints: transcriptional, post-transcriptional, and post-translational regulation.
1.6.1.1 Promoter elements

In eukaryotes, different promoter elements are defined according to their positions. These elements are the core promoter elements, proximal promoter elements, and distal promoter elements (enhancers).

1.6.1.1.1 Core promoter elements

A core promoter element is the minimum DNA sequence required to initiate proper transcription by the transcription initiation complex. This region spans ~40 base pairs (bp) upstream of the transcription start site (TSS). The main core promoter elements identified across a wide range of promoters are BRE (TFIIB recognition element), TATA (TATA box), Inr (initiator element), and DPE (downstream promoter element). Importantly, it is not necessary for all these elements to be found in each promoter region. Rather, they appear in different combinations, usually depending on the particular regulatory requirements of a specific gene (Figure 1.7) (Latchman, 2008).

![Figure 1.7: An eukaryotic promoter region and its regulatory elements](image)

Graph shows specific DNA regulatory components and regions, including the TATA box, proximal (core) promoter elements, and the enhancer. From (Lodish, 2003).

In most TATA box-containing eukaryotic promoters, the core promoter element is located 25–35 bp upstream of the TSS. This sequence is recognized by the TATA-binding protein to initiate transcription. Except for housekeeping genes, most genes require this sequence in their promoter regions for activating transcription.
1.6.1.2 Proximal promoter elements

Proximal promoter elements are located ~50–200 bp upstream of the TSS. Two frequently identified elements within this region are the GC box and CCAAT box. These DNA sequences can be recognized by transcription factor Sp1 and CCAAT-binding transcription factor (CTF), respectively (Lodish, 2003).

1.6.1.2 Enhancers

Enhancers are position-independent, positive regulatory elements that are present in eukaryotic cells, but usually absent in prokaryotic cells. The activity of enhancers does not depend on their exact location or distance from the promoter. They can be found thousands of base pairs away from the TSS, either downstream or upstream of its target gene promoter (Lodish, 2003).

1.6.1.3 Other elements

Silencers are regulatory DNA elements that have an opposite function to enhancers, i.e. they act to decrease gene transcription. There are two types: position-dependent and position-independent silencers. Their position is variable in mammalian genomes. Another regulatory element is the insulators (boundary elements) which are located up to 3 kb from TSS and mainly prevent the interactions between the enhancers and the promoters (Lodish, 2003).

1.6.2 Transcription factors

Transcription factors are regulatory proteins that non-covalently interact with specific DNA sequences via relevant DNA-binding domains. TFs can positively or negatively control the rate of gene expression. This regulation is often tissue-specific or elicited by appropriate signals. The human genome contains ~2000 transcription factor genes, amounting to roughly 10% of the genome (Wolpert, 2007).

The activity of TFs is based on two major domains. These domains are the DNA-binding domain and the activation domain. The DNA-binding domain recognises specific DNA sequences within regulatory regions of genes, whereas the activation domain
interacts with other protein complexes comprising the transcriptional machinery (Lewin, 2004).

In eukaryotes, many different types of DNA-binding motifs have been identified. The most commonly observed motifs are helix-turn-helix, zinc-finger, leucine zipper, and helix-loop-helix motifs. The most studied motif is the helix-turn-helix motif, which is found in the homeodomain family, for instance, HNF1a. This motif has two helices that are linked to each other by a short turn. The C-terminal helix of this motif fits into the major groove of the DNA helix and binds a specific DNA sequence. It is, therefore, called the recognition helix. The N-terminus helps to position the C-terminal helix (Strachan and Read, 2004).

Another DNA-binding domain is the zinc finger. This domain includes a zinc molecule inside the helical structure of the recognition helix. HNF4a contains an example of this group. Some TFs have more than one zinc-finger domain.

In leucine zipper TFs, the α-helix has regular leucine repeats, which are necessary for dimerization. This motif contains two α-helical proteins forming homodimers or heterodimers, which bind to a specific DNA region, with both helices (Watson, 2003).

The helix-loop-helix motif has a short and a longer α-helix, which are bound together by a loop. The dimer gains flexibility with the loop structure. Together with another helix-loop-helix protein, they can form a homodimeric or heterodimeric structure (Watson, 2003).

The second characteristic domain of TFs is the transcription activation domain (TAD), which is essential for transcription initiation via interaction with the transcription initiation complex. In contrast to DNA-binding domains, TADs do not interact with DNA structures. These domains mostly interact with general TFs to recruit the RNA polymerase II to the promoter (Tjian and Maniatis, 1994). General TFs such as TFIIB, TFIID, and TFIIF can interact with different activation domains (Blau et al., 1996). For instance, AF2 is the activation domain of the HNF4 protein, and it binds to TFIIB, and this interaction recruits the preinitiation complex (Malik and Karathanasis, 1996).
1.6.2.1 Regulation of TFs

Almost all cell types express a set of housekeeping genes, which are necessary for basic cell functions. In contrast, the tissue-specific genes are expressed in the specific cell types only. Therefore, regulation of TFs is an important mechanism underlying development and differentiation in organisms. This process is achieved by different steps in TF expression and activation.

TFs themselves are controlled by external signalling. Gene expression can be triggered by extracellular signals such as growth factors, hormones, chemokines, cytokines and apoptosis factors. Each of them shows its effects via different intracellular pathways and TFs. The STAT pathway, for instance, is activated by cytokines and growth factors. The activated cytokine receptor triggers STAT phosphorylation in the cytoplasm and promotes dimerization of phosphorylated STAT. The activated dimers can then enter the nucleus to start transcription of target genes (Lodish, 2003).

1.6.3 Epigenetics

Epigenetic factors can modify transcriptional regulation without affecting the DNA code. Intriguingly, these changes can be inherited through mitosis (Nussbaum et al., 2007). Epigenetics alter gene expression by modulating DNA methylation or histone modifications. Epigenetic modifications have been studied extensively over the last several decades (Razin and Riggs, 1980, Holliday, 1996).

1.6.4 Chromatin modifications

In eukaryotes, genetic information is compacted as chromatin. This compaction prevents TFs from binding to regulatory DNA regions (see Section 1.6.1.1). Therefore, chromatin unpacking is a prerequisite of transcriptional regulation. It is now established that chromatin not only plays a role in packaging the genomic material, but also plays a central role in transcriptional regulation (Turner, 2002).

Chromatin packing is regulated by histones, protein components of chromatin that compact the DNA (see Section 1.6.4.1). Inactive genes are in a highly condensed state, inaccessible to transcription factors. Chromatin remodelling and histone modifications
can reversibly unpack the compacted DNA to let activators access their corresponding binding sites (Fischle et al., 2003).

1.6.4.1 Histone modifications

Histones are positively charged proteins, which associate with DNA. Five different histone subtypes exist in eukaryotic cells, namely H1, H2A, H2B, H3, and H4. A histone octamer contains two H2A–H2B and two H3–H4 dimers. Approximately 150 bp of DNA wrap around each octamer of core histones to form a complex called the nucleosome (Kornberg and Lorch, 1999).

Histones N-terminal tails (Figure 1.8) bind DNA and stabilise the chromatin structure. Post-translational modifications in histone tails, such as methylation, acetylation, phosphorylation, and ubiquitination, regulate chromatin stability and, therefore, transcriptional regulation (Kouzarides, 2007).

![Figure 1.8: Structure of histone tails](image)

The DNA double helix wraps around histone octamers containing the H2A, H2B, H3, and H4 subunits. N-terminal tails of each histone are chemically modified to regulate nucleosome DNA accessibility. From (Watson, 2003).

Histone acetyl transferases (HATs) catalyse addition of acetyl groups to lysine residues on the N-terminal tails of histones. This causes loss of their positive charges, which in turn weakens their binding to negatively charged DNA. As a result, DNA
becomes accessible, and the transcription initiation complex can bind the promoter region to start transcription. The acetylation reaction is reversible; histone deacetylases (HDAC) remove the acetyl groups from lysine residues and thereby reduce the accessibility to the nucleosomal DNA (Richards and Elgin, 2002).

Similarly, methylation of histone H3 tails by histone methyl transferases (HMTs) can control gene transcription. HMTs can transfer a methyl group onto lysine or arginine residue in a histone. Usually, methylation of amino acid residues of histones negatively regulates gene activation by blocking TF interaction with DNA (Berger, 2002, de la Cruz et al., 2005).

1.6.4.2 Chromatin remodelling

In some situations, histone modification is not sufficient for the TFs to bind the target promoter. Often, chromatin remodelling, an ATP-dependent process, is required to increase the accessibility of the target gene for transcription.

Activators bind to their specific DNA sequences and recruit a chromatin remodelling complex that alters the chromatin structure. Chromatin remodelling systems act together with histone-modifying enzymes to assist TF binding to their cognate sequences. Chromatin complexes can make this change via sliding, transferring, or restructuring a nucleosome on the DNA (Figure 1.9) (Strachan and Read, 2004).
A chromatin remodelling complex is recruited by DNA-binding protein 1 to alter nucleosome accessibility. Action of the DNA-binding proteins stimulates formation of a more accessible chromatin structure by recruiting HATs to the target DNA region. From (Watson, 2003).

1.6.5 DNA methylation

DNA methylation can regulate transcription in eukaryotic cells, especially in mammals. Methylation of cytosine residues in cytosine–guanine dinucleotides (CpGs), which are located around the promoter region, can cause transcriptional inhibition. This modification disturbs the interaction of transcription factors with the promoter region. In general, DNA methylation correlates with an inactive gene status, but switching off a gene completely requires further modifications. This is achieved by a combination of DNA methylation and nucleosomal alterations (Cedar and Bergman, 2009).

Methylation patterns are heritable. During the replication process, methylated parental DNA strands are divided and passed onto daughter cells. Methylation of the
complementary new strand is completed by the maintenance methyltransferase which recognise the hemi-methylated pattern and methylates the CpG dinucleotide on the newly synthesized strand (Figure 1.10) (Alberts, 2004).

![Figure 1.10: Maintenance methyltransferase](image)

During DNA replication, the maintenance methyltransferase allows transfer of an existing DNA methylation pattern to a new DNA strand. From (Alberts, 2004).

### 1.6.5.1 CpG islands

In mammals, 70–80% of CpG dinucleotides are methylated across the whole genome, with the other non-methylated CpG dinucleotides concentrated in the gene promoter regions (Portela and Esteller, 2010). A region of 200 bp with at least 50% CpG content is considered a CpG island. CpG islands are commonly found as non-methylated regions in promoters of housekeeping genes and tissue-specific genes in differentiated cells.

### 1.6.6 DNA methylation and transcription

CpG methylation has a significant function in transcriptional control, chromosomal structure and organisation, and differentially expression of genes depending on their parent-of-origin, called genomic imprinting. (Robertson, 2002). Most promoter regions
have CpG-rich regulatory fragments, and some TF-binding sites contain CpG dinucleotides.

In the promoter of an inactive gene, a DNA methyltransferase (DNMT) adds a methyl group to position 5 of the cytidine generating 5-methylcytosine. Specific proteins, which recognise a methylated cytosine residue, can bind methylated sequences. These proteins can then switch off a gene completely through the action of HDACs and HMTs (Figure 1.11) (Watson, 2003).

DNA methylation in a promoter region can inhibit binding of the transcriptional initiation machinery or TFs in a number of ways. Firstly, methylation can inhibit TF binding (Tate and Bird, 1993). Secondly, some proteins specifically bind to methylated CpG residues and repress transcription by preventing TF binding to the target promoter region. For example, DNMTs can inhibit transcription via modifying histones. Studies have shown that DNMTs are associated with HDACs (Robertson, 2002).

Methylated CpGs can also be recognised by methyl-CpG-binding-protein (MeCP2) which recruits HDACs (Nan et al., 1998). Also, the DNA-methyl-binding domain (MBD) proteins interact with methylated DNA, in the same way as MeCP2, and can recruit proteins to suppress transcription.
Figure 1.11: Collective effect of DNA methylation and histone modifications on the transcription

This figure illustrates the switching-off of an active gene by changing its DNA methylation and histone modification status. (A) DNMTs catalyse methylation on target DNA CpG dinucleotides. (B) This change triggers histone deacetylation. (C) Finally, DNA becomes more compact and this compaction leads to transcriptional silencing of a gene.
1.7 Project aims

The functional properties and physiological roles of B\textsuperscript{0}AT1 in different organs have been extensively studied (Bröer et al., 2004). However, not many studies have attempted to describe the regulatory elements of the Slc6a19 gene. By bioinformatics analysis, Kikuchi et al. predicted a possible HNF1a-binding site on the Slc6a19 promoter at position –126 to –110. Except this study no single study has been focused on the B\textsuperscript{0}AT1 regulation so far. The aim of this project was to investigate the transcriptional regulation of the neutral amino acid transporter B\textsuperscript{0}AT1 and to identify key control elements involved in its transcription.

1.8 Research questions

1. Does B\textsuperscript{0}AT1 mRNA expression mirrors its protein expression reported previously?
2. What are the regulatory elements in the promoter region of B\textsuperscript{0}AT1?
3. Why is B\textsuperscript{0}AT1 not expressed in intestinal crypt cells?
4. What transcription factors are involved in the positive regulation of B\textsuperscript{0}AT1 expression in cells located in the intestinal villi?
5. Does the nuclear chromatin status across the B\textsuperscript{0}AT1 promoter differ between the crypt and the villus?
6. Is DNA methylation involved in B\textsuperscript{0}AT1 transcriptional silencing outside the villus region (e.g. cells in the crypts or liver cells)?
Chapter 2  Material and Methods

2.1  Materials

2.1.1  Solutions

Table 2.1: Solutions used in this study

<table>
<thead>
<tr>
<th>Luria-Bertani (LB) medium</th>
<th>Luria-Bertani (LB) agar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone 10 g</td>
<td>Tryptone 10 g</td>
</tr>
<tr>
<td>Yeast extract 5 g</td>
<td>Yeast extract 5 g</td>
</tr>
<tr>
<td>NaCl 10 g</td>
<td>NaCl 10 g</td>
</tr>
<tr>
<td>Add 1 L H₂O, pH = 7.0</td>
<td>Agar 15 g</td>
</tr>
<tr>
<td>Sterilized by autoclaving.</td>
<td>Add 1 L H₂O, pH = 7.0</td>
</tr>
<tr>
<td></td>
<td>Sterilized by autoclaving.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Super optimal broth (SOB) medium</th>
<th>Hanks Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone 20g</td>
<td>CaCl₂ 1.26 mM</td>
</tr>
<tr>
<td>Yeast extract 5g</td>
<td>KCl 5.4 mM</td>
</tr>
<tr>
<td>NaCl 8.5 mM</td>
<td>KH₂PO₄ 0.44 mM</td>
</tr>
<tr>
<td>KCl 2.5 mM</td>
<td>MgCl₂ 6H₂O 0.5 mM</td>
</tr>
<tr>
<td>Add 1 L H₂O, pH = 7.0</td>
<td>MgSO₄ 7H₂O 0.4 mM</td>
</tr>
<tr>
<td>Sterilized by autoclaving, then add</td>
<td>NaCl 136.6 mM</td>
</tr>
<tr>
<td>MgCl₂ 10 mM</td>
<td>HEPES 10 mM</td>
</tr>
<tr>
<td>MgSO₄ 10 mM</td>
<td>Na₂HPO₄ 2.7 mM</td>
</tr>
<tr>
<td>For SOC (also add glucose 20 mM)</td>
<td>pH 7.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TB buffer</th>
<th>Phosphate buffered saline (PBS, 10×)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEPES 10 mM, pH 6.7</td>
<td>NaCl 1.37 M</td>
</tr>
<tr>
<td>CaCl₂ 15 mM</td>
<td>KCl 26.8 mM</td>
</tr>
<tr>
<td>MnCl₂ 55 mM</td>
<td>Na₂HPO₄ 80.6 mM</td>
</tr>
<tr>
<td>KCl 250 mM</td>
<td>KH₂PO₄ 14.7 mM</td>
</tr>
<tr>
<td>Sterilized by filtration through a 0.22-μm filter</td>
<td>pH 7.4, sterilize by autoclaving</td>
</tr>
<tr>
<td>TAE buffer (50×)</td>
<td>Cracking buffer (for rapid screening, 2×)</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Tris 2 M</td>
<td>Sucrose 0.2 g</td>
</tr>
<tr>
<td>Na₂EDTA 50 mM</td>
<td>SDS 50 μL of 10% (w/v) SDS</td>
</tr>
<tr>
<td>Glacial acetic acid 5.71 %</td>
<td>NaOH 40 μL 5M</td>
</tr>
<tr>
<td>pH 8.0</td>
<td>H₂O top up to 1 ml</td>
</tr>
<tr>
<td><strong>Tris buffered saline (TBS, 10×)</strong></td>
<td><strong>Marker mix (for rapid screening)</strong></td>
</tr>
<tr>
<td>Tris-HCl 0.5 M, pH 7.5</td>
<td>KCl 4 M</td>
</tr>
<tr>
<td>NaCl 1.5 M</td>
<td>Bromphenol blue 0.4% (w/v)</td>
</tr>
<tr>
<td>Sterilized by autoclaving</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Tris-borate-EDTA buffer (TBE, 10×)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris 892 mM</td>
</tr>
<tr>
<td>Boric acid 890 mM</td>
</tr>
<tr>
<td>Na₂EDTA 20 mM</td>
</tr>
<tr>
<td>pH 8.0</td>
</tr>
</tbody>
</table>

### 2.1.2 Reagents and commercial kits

Opti-MEM®, DMEM/F-12, 0.25% trypsin–EDTA solution, 200 mM L-glutamine, Ultrapure™ 0.5 M EDTA, Ultrapure™ 10% SDS solution, heat inactivated foetal bovine serum (FBS) and 1× PBS buffer were purchased from Gibco® Life Technologies. Penicillin-streptomycin solution, DMEM, and oligonucleotides were obtained from Sigma-Aldrich®. RPMI 1640 medium with L-glutamine was purchased from PAA Laboratories. SimpleCHIP® Enzymatic Chromatin IP Kit was purchased from Cell Signaling Technology®. Power SYBR® Green PCR Master Mix, Supercoiled DNA ladder, SuperScript® III Reverse Transcriptase, T4 DNA ligase, Lipofectamine™ 2000 reagent, Lipofectamine® LTX, PureLink™ Quick Gel Extraction Kit, PureLink™ Quick Plasmid Miniprep Kit, Zero Blunt® TOPO® PCR Cloning Kit, and TOPO® XL PCR Cloning Kit were obtained from Life Technologies™. RNeasy Mini Kit, EpiTect® Bisulfite Kit, DNeasy® Blood & Tissue Kit, QIAquick® PCR Purification Kit, Rotor-Gene SYBR® green PCR Kit, and *Thermus aquaticus* (Taq) DNA polymerase were purchased from Qiagen. RNA 6000 Nano Labchip® Kit and QuikChange II Site-directed Mutagenesis Kit were purchased from Agilent. NucleoSpin® Plasmid purification Kit and NucleoBond Xtra Maxi EF were purchased from Macherey-Nagel. cOmplete, EDTA-free protease
inhibitor cocktail tablets were purchased from Roche. Phusion® high-fidelity DNA polymerase and T4 DNA Quick Ligation™ Kit were obtained from New England Biolabs® (NEB). Twenty-four-well cell culture plates were purchased from Iwaki. Dual-Luciferase® Reporter Assay System was purchased from Promega.

2.1.3 Antibodies

Antibodies used in chromatin immunoprecipitation assay (ChIP) are listed in Table 2.2.

<table>
<thead>
<tr>
<th>Antibody (anti-mouse)</th>
<th>Source</th>
<th>Supplier (catalogue no)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histone H3 (acetyl K27)</td>
<td>Rabbit polyclonal antibody</td>
<td>Abcam (ab4729)</td>
</tr>
<tr>
<td>Histone H3 (tri-methyl K4)</td>
<td>Rabbit polyclonal antibody</td>
<td>Abcam (ab8580)</td>
</tr>
<tr>
<td>SOX9</td>
<td>Rabbit polyclonal</td>
<td>Abcam (ab3697)</td>
</tr>
<tr>
<td>TATA-binding protein</td>
<td>Mouse monoclonal</td>
<td>Abcam (ab51841)</td>
</tr>
<tr>
<td>HNF4a</td>
<td>Mouse monoclonal</td>
<td>Abcam (ab41898)</td>
</tr>
<tr>
<td>IgG</td>
<td>Rabbit polyclonal antibody</td>
<td>Cell signaling (2729)</td>
</tr>
<tr>
<td>Histone H3</td>
<td>Rabbit antibody</td>
<td>Cell signaling (4620)</td>
</tr>
<tr>
<td>HNF1a</td>
<td>Goat polyclonal</td>
<td>Santa Cruz (sc-6547X)</td>
</tr>
</tbody>
</table>

2.1.4 Vectors

The pcDNA 3.1 (+) vector was used for gene expression in mammalian cells (see Section 2.6.4). This vector contains a human cytomegalovirus (CMV) immediate–early promoter that achieves a high level of expression in a wide variety of cell lines. It contains the ampicillin- and neomycin resistance genes to facilitate selection in bacteria and cell lines, respectively.
pGL4.12[luc2CP] and pGL4.74[hRluc/TK] reporter vectors (Promega) were used in the dual luciferase assay (see Section 2.6). Promoter fragments were cloned into pGL4.12 vector to measure their activity. The pGL4.74 vector was used as an internal control in the dual luciferase assay. See Table 2.3 for sources and comparison of these vectors. Detailed maps of the plasmids generated in this study are illustrated in the Appendix 8.1.

Table 2.3: Vectors used in this study

<table>
<thead>
<tr>
<th>Vector</th>
<th>Features</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA3.1 (+)</td>
<td>• CMV promoter for optimal expression in a wide range of mammalian cells</td>
<td>Invitrogen</td>
</tr>
<tr>
<td></td>
<td>• SV40 polyadenylation site upstream of the CMV promoter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Neomycin- and ampicillin-resistance genes</td>
<td></td>
</tr>
<tr>
<td>pGL4.12[luc2CP]</td>
<td>• Promoter-less</td>
<td>Promega</td>
</tr>
<tr>
<td></td>
<td>• Firefly luciferase gene</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Ampicillin-resistance gene</td>
<td></td>
</tr>
<tr>
<td>pGL4.74[hRluc/TK]</td>
<td>• Thymidine kinase promoter</td>
<td>Promega</td>
</tr>
<tr>
<td></td>
<td>• Renilla luciferase gene</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Ampicillin-resistance gene</td>
<td></td>
</tr>
</tbody>
</table>

2.1.5 Animals

Adult mice C57BL/6J (B6) mice were obtained from the Australian Phenomics Facility of The Australian National University (ANU) and housed under pathogen-free conditions. All experiments were approved by the ANU Animal Ethics and Experimentation Committee.
2.1.6 Antibiotics

Ampicillin: A stock solution of ampicillin sodium salt was dissolved in distilled water (50 mg/mL) and stored at –20 °C.

Penicillin-Streptomycin: This solution includes 10,000 IU of penicillin and 10 mg streptomycin per mL. The antibiotic aliquots were kept at 4 °C.

Kanamycin (ICN Biomedicals): A stock solution of kanamycin monosulphate was dissolved in distilled water (20 mg/mL) and aliquots were stored at –20 °C.

2.2 Bacterial experiments

2.2.1 Bacterial strains

Three strains of *Escherichia coli* (*E. coli*) were used as host bacteria for plasmid transformations (Table 2.4), namely the chemically competent DH5α, and electrocompetent XL-1 and strains. Aliquots of competent cells were stored at –80 °C.

**Table 2.4: Bacterial strains were used in this study**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> DH5α</td>
<td>F–Φ80lacZΔM15 Δ(lacZYA-argF) U169recA1 endA1 hsdR17 (rk−, mk+) phoA supE44 λ− thi-1 gyrA96 relA1</td>
<td>Life Technologies</td>
</tr>
<tr>
<td><em>E. coli</em> DH10B</td>
<td>F–mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara, leu)7697 galU galK λ− rpsL nupG /pMON14272/pMON7124</td>
<td>NEB</td>
</tr>
<tr>
<td><em>E. coli</em> XL-1 Blue</td>
<td>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F’ proAB lacIqZ ΔM15 tn10 tet’]</td>
<td>Stratagene</td>
</tr>
<tr>
<td>Strain</td>
<td>Genotype</td>
<td>Source</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><em>E. coli</em> Top10</td>
<td>F–mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) endA1 nupG</td>
<td>Invitrogen</td>
</tr>
</tbody>
</table>

2.2.2 Transformation of electrocompetent bacteria

Plasmid DNA (2.5 μL) was added to electro-competent bacteria (*E. coli* XL-1 Blue) that were thawed on ice. The mixture was incubated on ice for 30 seconds and then transferred to a pre-chilled 0.2 cm electroporation cuvette (Astral). The cells were briefly exposed to a strong electrical field using a Bio-Rad Gene Pulser II (BioRad) set to 2.5 kV, 200 Ω and 25 μF. SOC medium (0.5–1 mL) was added to the cuvette immediately after the pulse and the mixture was transferred to a 15 mL Falcon tube. The cells were then allowed to recover from the electroporation for 60 min by shaking at 220 rpm at 37 °C.

The transformed cells were then plated onto LB-agar plates containing the appropriate antibiotic (100 μg/mL for ampicillin, 50 μg/mL for kanamycin) corresponding to the resistance gene present on the plasmid. The presence of the specific antibiotic in the medium ensured the growth of those cells that had taken up the plasmid. The plates were incubated overnight at 37 °C and visible colonies were further analysed (see Section 2.4.1.3).

2.2.3 Transformation of chemically competent bacteria

Aliquots (50 μL) of chemically competent bacteria (*E. coli* DH5α, DH10B or Top10) were thawed on ice and transferred into a pre-chilled 14 mL round bottom Falcon tube (BD Biosciences). One microliter of DNA was added to the bacteria. The tube was swirled and incubated on ice for 30 min. The transformation reaction was then placed in a 42 °C water bath for 45 seconds. The tube was immediately placed on ice for 2 min, after which 0.5 mL of SOC was added to the reaction. The bacteria were incubated at 37 °C for 60 min by shaking at 220 rpm. A volume of 150-250 μL of the reaction was plated onto an LB-agar plate containing the appropriate antibiotic to select transformed colonies. The plate was incubated overnight at 37 °C.
2.2.4 Bacterial growth

To culture transformed bacterial colonies, 10 mL of LB was placed in a 50 mL Falcon tube. The appropriate antibiotic was added to the LB at a final concentration of 100 μg/mL or 50 μg/mL for Ampicillin or Kanamycin, respectively. A sterile toothpick was used to pick a single transformed colony and was placed into the Falcon tube. The cultures were grown overnight in a rotatory shaker at 220 rpm at 37 °C.

2.3 Eukaryotic cell culture

2.3.1 Cell lines and cell culture

Cell lines used in this study were maintained in appropriate culture media (Table 2.5) at 37 °C in a humidified atmosphere of 5% CO₂ in an incubator. The cells were grown in 80-cm² tissue-culture flasks (Nunc) and cell density and morphology was examined on a daily basis by light microscopy. When cells reached 85–90% confluence, the medium was removed and the cells were detached from the surface of the flask by addition of 10 mL 0.25% trypsin–EDTA. After stopping trypsinization by adding media containing 10% FBS, the cells were transferred into a 50 mL falcon tube and collected by centrifugation at 1,500 rpm for 5 minutes. The pellet was resuspended with the relevant medium and centrifugation was repeated. After the supernatant was removed, the pellet was resuspended again in fresh medium and transferred into a new culture flask at a 1:20 ratio.

Table 2.5: Cell lines used in this study

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Description</th>
<th>Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeLa</td>
<td>Human cervix carcinoma cells</td>
<td>DMEM/F12 + 2 mM Q + 10% FBS</td>
</tr>
<tr>
<td>Caco-2</td>
<td>Human epithelial colorectal adenocarcinoma cells</td>
<td>DMEM/F12 +10% FBS +1% penicillin-streptomycin</td>
</tr>
<tr>
<td>FRTC</td>
<td>Rat thyroid cancer cells</td>
<td>DMEM/F12 + 5% FBS</td>
</tr>
<tr>
<td>HEK293</td>
<td>Human embryonic kidney cells</td>
<td>RPMI + 2 mM Q+ 5% FBS</td>
</tr>
<tr>
<td>HepG2</td>
<td>Human liver carcinoma cells</td>
<td>RPMI + 10% FBS</td>
</tr>
</tbody>
</table>
2.3.2 Counting of cell numbers

Cell suspensions were analysed and counted by using an automated cell counter (Scepter, Millipore).

2.4 Isolation and manipulation of DNA

2.4.1 DNA isolation

2.4.1.1 DNA extraction from mouse tissues

Isolation of DNA from C57BL/6J mice tissues was carried out using DNeasy Blood & Tissue Kit (Qiagen). The kidneys and intestine were used to isolate genomic DNA and its concentration was measured using a NanoDrop spectrophotometer (ThermoFisher Scientific).

2.4.1.2 Plasmid DNA isolation

Plasmids were isolated from overnight cultures using the NucleoSpin Plasmid Kit according to the manufacturer’s protocol. Plasmid preparations for transient transfection of cell lines (see Section 2.6.1) were prepared using the NucleoBond Xtra Maxi Endotoxin-free Kit. The quality of the isolated plasmid DNA was checked using spectrophotometry (see Section 2.8.2) and agarose gel electrophoresis (see Section 2.8.1). The isolated plasmids were stored at −20 °C.

2.4.1.3 Rapid screening of transformed colonies

Following the transformation of the plasmid of interest, ~15 colonies were picked using a sterile toothpick or a pipet tip and re-streaked onto LB-agar plates containing appropriate antibiotic(s). After incubation at 37 °C overnight, a single colony was transferred to an Eppendorf tube using a toothpick. Forty μL of 10 mM EDTA (pH 8.0) was added to the tube and the bacteria were resuspended by vortex mixing. Fifty μL of freshly prepared 2× cracking buffer was added to lyse the bacteria and the mixture was mixed by a vortex and incubated for 5 minutes at 23 °C. After adding 10 μL of Marker mix, the tube was incubated on ice for 10 min. The samples were centrifuged at 4 °C for
3 min at 15,000×g. Of the supernatant, 20 μL was loaded onto a 1% agarose gel adjacent to a supercoiled DNA ladder (Invitrogen).

2.4.2 DNA experiments

2.4.2.1 Restriction enzyme digestion

Restriction digestion was used for cloning and linearization of plasmid DNA. Target DNA was typically digested for 2h by 10–20 U of a specific restriction enzyme in the recommended reaction buffer at the recommended temperature. The restriction enzymes were sourced from Invitrogen, NEB, or Takara Biotechnology.

2.4.2.2 Gel elution

Gel elution was used to clean PCR products, restriction enzyme digested DNA, and to separate the DNA fragments from unwanted vector impurities. Briefly, after electrophoresing the samples, the DNA fragment was cut out from the agarose gel using a scalpel and weighed. The DNA fragment was extracted from the isolated band using PureLink Quick Gel Extraction Kit. Spectrophotometry (see Section 2.8.2) was used to determine the concentration of the extracted DNA.

2.4.2.3 DNA ligation

To ligate DNA inserts into vectors, the T4 DNA Quick Ligation Kit was used. Ten μL 2× Quick Ligation Buffer and 1 μL of Quick T4 DNA ligase were added to a 1:3 molar ratio of vector to insert. The reaction mixture was incubated at 25 °C for 5 min after which the ligated DNA was precipitated by addition of 1 μL of yeast tRNA (5 μg/μL), 0.5 volumes of 7.5 M ammonium acetate, and 2.5 volumes of 100% ice-cold ethanol. The contents were mixed by a vortex and incubated for 10 min at 23 °C. The DNA was pelleted by centrifugation for 20 min at 16,000×g and the supernatant was discarded. A volume of 250 μL 70% ice-cold ethanol was added and the mixture was spun for a further 10 min at the same speed. The supernatant was discarded and the pellet was dried on the bench at 23 °C. The pellet was then dissolved in 5 μL of sterile water. A 2 μL aliquot of ligation product was later used to transform electro-component bacteria (see Section 2.2.2).
2.4.2.4 Polymerase chain reaction (PCR)

PCR reactions were performed using Taq DNA polymerase (Qiagen) according to the manufacturer’s instructions. Each reaction contained 2 units of the polymerase enzyme, 1 μL of 10 mM deoxynucleotide triphosphate (dNTP) mix, 100 ng of DNA template, 5 μL of 10× PCR buffer, and 0.5 μM of the forward primer and 0.5 μM of the reverse primer in a final volume of 50 μL in distilled water. PCR reactions were run in a 96-well Veriti® thermal cycler (ABI) at the conditions shown in Table 2.6.

Table 2.6: Conditions of a PCR cycling

<table>
<thead>
<tr>
<th>Process</th>
<th>Temperature (°C)</th>
<th>Time (min:sec)</th>
<th>Number of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>5:00</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>0:30</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>50–58 (depending on the melting temperature ($T_m$) of primers)</td>
<td>0:30</td>
<td>30</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>1:00</td>
<td></td>
</tr>
<tr>
<td>Extension (Final)</td>
<td>72</td>
<td>10:00</td>
<td>1</td>
</tr>
</tbody>
</table>

In order to obtain proofread products, the PCR reactions were performed using PfuUltra high fidelity DNA polymerase (Agilent) or Phusion high-fidelity DNA polymerases according to the manufacturers’ protocols. The products were separated on a 1% agarose gel (see Section 2.8.1). Thermal cycler conditions used in conjunction with PfuUltra and Phusion® high-fidelity DNA polymerases are listed in the Appendix 8.2.

2.4.2.5 Site-directed mutagenesis

QuikChange II Site Directed Mutagenesis Kit was used for this method. The entire plasmid was amplified by PCR using long primers (25–45 bp) with the desired mutation located in the centre of the primer (Table 2.7).
### Table 2.7: Primer sequences used for site-directed mutagenesis

Red text indicates inserted mutations.

<table>
<thead>
<tr>
<th>Mutation type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNF1α BS1(Mut1)</td>
<td>5’ TGGAGGGGGTTGGCTTAGCTTCGGAGGCTGGTGCCCTCTGAG3’ 3’ CTGCAAGGGAACCACCACCTCCGAAGCTAAGCCACCCCTCCCA5’</td>
</tr>
<tr>
<td>HNF1α BS2(Mut2)</td>
<td>5’ TGGGTGCCTCTGCAAGATAGGCAAGGCCAGTTCTGGCAAGAGC3’ 3’ GCGTCCCTGCAAGACTGAGGCTTATCTGCAAGGCAACCC5’</td>
</tr>
<tr>
<td>HNF1α mutation (R131W)</td>
<td>5’ GCACAACATCCCCCAGTGGGAGGTGGTGAC3’ 3’ GGGTGCTCCACCACCCTGCCACTGGGGATGTTGTGC5’</td>
</tr>
<tr>
<td>HNF4α BS2(Mut2)</td>
<td>5’ GGGTTGAGGTGCTGACGGTTCTCTATAAAG3’ 3’ CTTTATAGAGAACCCTCAACCTCAACCC5’</td>
</tr>
<tr>
<td>HNF4α BS4(Mut3)</td>
<td>5’ GACAACAGAACACTCCACGCACCTGTGGAATTC3’ 3’ GAATTCACAGGAGGTGGTGGACACC3’</td>
</tr>
<tr>
<td>HNF4α BS5(Mut4)</td>
<td>5’ GTGAATTCTAGTCATGTCAGCGTCTGATTCTAAAG3’ 3’ CTTGGGCAACACGTGACTAGAATTCAC5’</td>
</tr>
<tr>
<td>SOX9-BS1 Mut</td>
<td>5’ CAGTGTGTTTTGTTGACGGGTTCTCTTAAACG3’ 3’ CAGGGGCTTGACGTAACAAACACTG5’</td>
</tr>
<tr>
<td>Sox9-Mut1</td>
<td>5’ GCCTATGTCTGGGCTTTCCCTGCAGGCCAGGAGG3’ 3’ CCCTCCTCCTGGCAAGGGCCCAGACATAGGC5’</td>
</tr>
<tr>
<td>Sox9-Mut2</td>
<td>5’ CAACCACTTGGCCCAGGGTCTGTGGAGCTGC3’ 3’ GCCAGCTCGGCGAGACCCAGGGCAATGGTTG5’</td>
</tr>
<tr>
<td>SOX9 mutation (H65Y)</td>
<td>5’ GCTGCGCGTGCAGTAAGAAGGACCACCC3’ 3’ GGGTGGTCTTTCTGTAACGCGCCAGCG5’</td>
</tr>
<tr>
<td>Spdef-mut</td>
<td>5’ CAGGGCCTCACAGAGGATCTCCTGTCGG3’ 3’ CACGGACAGGAGATCCGCTGAGGCGCCTG5’</td>
</tr>
</tbody>
</table>

Template DNA (50 ng) was mixed with 5 μL of the provided 10× reaction buffer, 125 ng of the sense primer, 125 ng of the antisense primer, and 1 μL of the dNTP mix. The final volume was adjusted to 50 μL using deionized water. One μL of *PfuUltra* high-
fidelity DNA polymerase (2.5 U/μL) was added to the reaction before performing PCR amplification (Table 2.8).

Table 2.8: PCR conditions for site-directed mutagenesis

<table>
<thead>
<tr>
<th>Process</th>
<th>Temperature (°C)</th>
<th>Time (min:sec)</th>
<th>Number of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>95</td>
<td>0:30</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95</td>
<td>0:30</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>55</td>
<td>1:00/kb of plasmid</td>
<td>16–18 cycles</td>
</tr>
<tr>
<td>Extension</td>
<td>68</td>
<td>4:00</td>
<td>1</td>
</tr>
</tbody>
</table>

The amplified DNA was digested for 3 hours with *Dpn*I (10 U/μL) at 37 °C to remove the template DNA. The DNA was then transformed into either chemically or electrocompetent *E. coli*.

2.5 Bioinformatics analysis

2.5.1 Alignment of promoter region

To identify the promoter regions conserved among different species, the promoter sequences were analysed using the University of California, Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu). Using this method, regions of ~2.5 kb upstream of mouse *Slc6a19*, *Ace2*, and *Collectrin* transcription start sites were examined.

2.5.2 Prediction of transcription-factor-binding sites

The MatInspector program (http://genomatix.de) is a web-based bioinformatics tool which predicts transcription-factor-binding sites based on similarity of the input sequence to a library of known motifs (Quandt et al., 1995). A 2.5 kb fragment of *Slc6a19* promoter was analysed using this tool.

2.6 Reporter gene assay

The Dual-Luciferase® Reporter Assay system was used to identify effects of transcription factors on the mouse *Slc6a19* promoter. To this end plasmid
pGL4.12[luc2CP] containing the target promoter region, was expressed alone or was co-transfected with transcription factors of interest.

2.6.1 Principle of dual reporter assay

HEK293 cells were seeded out in 24 well plates at a density of 1×10⁵ cells per well 24 hours before transfection. Transfections were performed using the Lipofectamine LTX (Invitrogen) reagent, according to the manufacturer’s instructions.

Briefly, 500 ng of different Slc6a19 promoter constructs in pGL4.12[luc2CP] vector (Figure 2.1), plus 100 ng of a transcription factor (TF) expressing vector (typically inserted into pcDNA3.1(+)), 5 ng of pGL4.74[hRluc/TK] control vector (Figure 2.1), and 2 μL of Lipofectamine LTX were added to 100 μL Opti-MEM®. The DNA solution was mixed and incubated for 30 min at 23 °C. After incubation, the mixture was added to HEK293 cells in a 24-well plate and incubated at 37 °C for 24 h.

Figure 2.1: pGL4.12[luc2CP] plasmid and pGL4.74[hRluc/TK] plasmid were used as reporter plasmids in the Dual-Luciferase® Reporter Assay system.

pGL4.12 lacks a promoter sequence to start transcription of the Firefly Luciferase gene. Potential promoter sequences can be inserted upstream of the luciferase sequence. The Renilla luciferase, by contrast, is expressed constitutively from the Herpes Simplex thymidine kinase promoter (pGL4.74 vector). It serves as an internal control for the reporter gene assay system.

After 24 h of incubation, the cell medium was removed and, cells washed once with PBS (pH 7.4). Cell layers were lysed in 100 μL Passive Lysis Buffer (Promega) for 20 min at 23 °C using a shaker plate.
A 20 μL aliquot of the lysate was placed into a 1.5 mL Eppendorf tube containing 100 μL Luciferase Assay Reagent (LARII) and after 10 seconds, luminescence activity of the Firefly luciferase was measured using a Turner TD-20/20 illuminometer. Subsequently, 100 μL of Stop and Glo reagents was injected into the same tube to stop the Firefly luminescence and activate the Renilla luciferase. The relative luciferase activity was quantified by normalising the Firefly luciferase activity relative to the Renilla luciferase activity of the pGL4.74 control plasmid. Each transfection and measurement of luciferase activity was performed in triplicates.

2.6.2 Cloning of mouse Slc6a19 promoter

Based on bioinformatic analysis, a 2.5 kb region upstream of the Slc6a19 transcriptional start site was selected.

A 2551 bp fragment of the promoter spanning from position –2494 to +57, was generated using a two-step nested PCR amplification procedure. Initially, a 3013 bp section (–2639 to +374) of the promoter was amplified by PCR (Table 2.9). Two μL of the first PCR was used as a template for the second PCR targeting the 2551-bp promoter region. Primer Slc6a19(–2494) included a KpnI restriction site and primer Slc6a19(+57) contained a BglII restriction site (Table 2.9). The final PCR product was purified and ligated (see Section 2.4.2.3) into a KpnI/BglII digested pGL4.12 vector.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slc6a19(–2639)</td>
<td>5’ACCGGGTCATTTTTTCTGCTCG3’</td>
</tr>
<tr>
<td>Slc6a19(+374)</td>
<td>3’TCCTGAGGAACATCCATCCATCCAT5’</td>
</tr>
<tr>
<td>Slc6a19(–2494)</td>
<td>KpnI 5’ATGGTACCTTAGAGAGCTGCTGCTC3’</td>
</tr>
<tr>
<td>Slc6a19(+57)</td>
<td>BglII 3’GAAGATCTCGCTGGGCTGGCCGCCG5’</td>
</tr>
</tbody>
</table>

Table 2.9: Primers used for amplification of the Slc6a19 promoter
2.6.2.1 Preparing the truncated Slc6a19 constructs

The truncated Slc6a19 promoter constructs were generated using the full promoter construct in pGL4.12 using inverse PCR. Selection of the truncation points was based on sequence conservation. The resulting fragment sizes are listed in Table 2.10.

Table 2.10: Oligonucleotides used for the synthesis of B^0AT1 promoter deletion constructs

<table>
<thead>
<tr>
<th>Promoter construct</th>
<th>Oligonucleotide</th>
<th>Insert size</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSlc6a19(–2494/+57)</td>
<td>5’ ATGGTACCTTAGAGAGCTGCCTC3’</td>
<td>2551 bp</td>
</tr>
<tr>
<td>pSlc6a19(–1703/+57)</td>
<td>5’ ATCTCTCTTTGGTCTCTGTCC3’</td>
<td>1760 bp</td>
</tr>
<tr>
<td>pSlc6a19(–1380/+57)</td>
<td>5’ CTAGTTTCCCCAGTGTTGTTGG3’</td>
<td>1437 bp</td>
</tr>
<tr>
<td>pSlc6a19(–972/+57)</td>
<td>5’ GAGAAAGGGGATAGGGTAAC3’</td>
<td>1029 bp</td>
</tr>
<tr>
<td>pSlc6a19(–437/+57)</td>
<td>5’ CTTCAGCCACCTTTAGGAATA3’</td>
<td>494 bp</td>
</tr>
<tr>
<td>pSlc6a19(–201/+57)</td>
<td>5’ GTGCCAGGCCTTCAGCAGG3’</td>
<td>258 bp</td>
</tr>
<tr>
<td>pSlc6a19(–136/+57)</td>
<td>5’ GGTGCCCTCTGCAGATAAGG3’</td>
<td>193 bp</td>
</tr>
<tr>
<td>pSlc6a19(–108/+57)</td>
<td>5’ GGTCTCAGCCAGCAGCGCCCT3’</td>
<td>165 bp</td>
</tr>
<tr>
<td>pSlc6a19(–37/+57)</td>
<td>5’ GGTCTCTATAAGAGCCG3’</td>
<td>94 bp</td>
</tr>
</tbody>
</table>

2.6.3 Cloning of the Ace2 and Collectrin promoter constructs

The mouse Ace2 promoter sequence was amplified by nested PCR. Firstly, a fragment ranging from position –1602 to +254 was amplified. Subsequently, a shorter fragment from position –1509 to +170 was amplified by a nested PCR, using primers with incorporated XhoI and KpnI restriction sites (Table 2.11). The final product was digested using XhoI and KpnI, purified, and ligated into reporter vector pGL4.12.

The mouse Collectrin promoter DNA was generated by nested PCR. In the first PCR step, a fragment ranging from position –2125 to +266 was amplified (Table 2.11). In the second step, a smaller fragment (–1903 to +76) was generated using a primer set
that contained restriction sites for KpnI and XhoI (Table 2.11). The final product was digested using KpnI and XhoI, purified, and ligated into the reporter vector pGL4.12.

Table 2.11: Oligonucleotides used for the synthesis of Ace2 and Collectrin promoter constructs

<table>
<thead>
<tr>
<th>Promoter construct</th>
<th>Oligonucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ace2(–1602)F</td>
<td>5’ CCGGGGTACTGCTTAGTTCA3’</td>
</tr>
<tr>
<td>Ace2(+254)R</td>
<td>3’ ATTTTCTCGGTAGGGACT5’</td>
</tr>
<tr>
<td>Ace2 (–1509)XhoI</td>
<td>5’ ATCTCGAGAGCTGACTGTGAGCATCCAC3’</td>
</tr>
<tr>
<td>Ace2 (+170)KpnI</td>
<td>3’ TAGGTACGCAAGATCCATCCACTGGA5’</td>
</tr>
<tr>
<td>Collectrin(–2125)F</td>
<td>5’ CTCCTTGCGTACCTGCTTTC3’</td>
</tr>
<tr>
<td>Collectrin (+266)R</td>
<td>3’ TTTCAAGCCACATGTCCAA5’</td>
</tr>
<tr>
<td>Collectrin(–1903)KpnI</td>
<td>5’ ATGTTACCTGGTGGTGTAGGTGATT3’</td>
</tr>
<tr>
<td>Collectrin(+76)XhoI</td>
<td>3’ TACTCGAGGCCAAGACAGACAAACT5’</td>
</tr>
</tbody>
</table>

2.6.4 Transcription factor constructs

Transcription factor complementary DNAs (cDNAs) used in this study were either cloned in our lab or obtained commercially.

The coding sequences of transcription factors HNF4a, SOX9, ATF4, PAX4, CREB3L3, GATA4, FOXA2 (HNF3b), NEUROG3, SMAD3, EGR1, and HNF1b were amplified on mouse intestinal cDNA by nested PCR. Initially, larger fragments were amplified using primers listed in Table 2.12. Subsequently, the coding sequence was amplified using the primers containing appropriate restriction sites. The final fragment was digested using the corresponding enzymes, purified, and ligated into mammalian expression vector pcDNA3.1(+). All transcription factor constructs were sequenced to confirm their identities (Biomolecular Resource Facility, John Curtin School of Medical Research (JCSMR), ANU).

The mouse HNF1a cDNA was purchased from Thermo Scientific (Clone ID: 30471380). The mouse Cdx2 (BC103516) was purchased from Open Biosystems and cloned into the pcDNA3.1(+)(Invitrogen). The mouse GFP–STAT3 (embedded into pEGFP-C3) and the human Flag–STAT3 (embedded into pXJ40-FLAG) expression plasmids were kindly provided by Marie Bogoyevitch (Bio21 Institute, Melbourne).
## Table 2.12: Primers used for cloning of TFs

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
<th>NCBI Ref. Seq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNF4a</td>
<td>5’ TTGGATCCGGGAGAATGCGACTCTCTCTAAAA3’</td>
<td>NM_008261.2</td>
</tr>
<tr>
<td>mHNF4a-BamHI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>SOX9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mSox9s</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mSox9a</td>
<td>3’ ATGAATTCATGGCTGCTAGATGGC9’</td>
<td></td>
</tr>
<tr>
<td>mSox9-BamHI</td>
<td>5’ TTGGATCCATGAATCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mSox9-EcoRI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>ATF4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mAtf4-BamHI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mAtf4-EcoRI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>PAX4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mPax4s</td>
<td>5’ GTCTCTCCCGCTCTTTTTCAGTTTGGC9’</td>
<td></td>
</tr>
<tr>
<td>mPax4a</td>
<td>3’ GAGGCTCTTCTATGAGCTTTGGC9’</td>
<td></td>
</tr>
<tr>
<td>mPax4-BamHI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mPax4-EcoRI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>CREB3L3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mCrb3L3s</td>
<td>5’ AACATCCGGTACCTAGCGACG9’</td>
<td></td>
</tr>
<tr>
<td>mCrb3L3a</td>
<td>3’ GCCAGCCCTCGCTTACTAAGAG9’</td>
<td></td>
</tr>
<tr>
<td>mCrb3L3-NotI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mCrb3L3-XbaI</td>
<td>3’ ATTCTAGATACACGACCTACCAATGCG9’</td>
<td></td>
</tr>
<tr>
<td>GATA4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mGata4-EcoRI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mGata4-BamHI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>FOXA2 (HNF3b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mHnf3bs</td>
<td>5’ CCGCCGACTTTAATGTACAGC9’</td>
<td></td>
</tr>
<tr>
<td>mHnf3ba</td>
<td>3’ GTGAGACTGCTCTTCCCTGGC9’</td>
<td></td>
</tr>
<tr>
<td>mHNF3-EcoRI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mHNF3-BamHI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>NEUROG3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mNeurog3s</td>
<td>5’ GTGTGTGTTGGGGGATACCTC9’</td>
<td>BC104327.2</td>
</tr>
<tr>
<td>mNeurog3a</td>
<td>3’ TGGAGACGAGCTTGGTGGC9’</td>
<td></td>
</tr>
<tr>
<td>mNeurog3-NotI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mNeurog3-XbaI</td>
<td>3’ ATTCTAGATACACGACCTACCAATGCG9’</td>
<td></td>
</tr>
<tr>
<td>SMAD3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mSmad3s</td>
<td>5’ GCGGGAGCAAACTTCTTTCTA9’</td>
<td>NM_006754.6</td>
</tr>
<tr>
<td>mSmad3a</td>
<td>3’ GAAACAGGCTTGGTGGC9’</td>
<td></td>
</tr>
<tr>
<td>mSmad3-EcoRI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mSmad3-BamHI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>EGR1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mEgr1s</td>
<td>5’ CCACCCAACATCAGTACTCC’</td>
<td>NM_007913.5</td>
</tr>
<tr>
<td>mEgr1a</td>
<td>3’ GCCAGGGATGTTAAGTGGAA5’</td>
<td></td>
</tr>
<tr>
<td>mEgr1-EcoRI</td>
<td>5’ TTGGATCCATGAATCTCTCTGAGGCC9’</td>
<td></td>
</tr>
<tr>
<td>mEgr1-BamHI</td>
<td>3’ ATGAATTCACCTGTGCTAGATGGCTTTTGG5’</td>
<td></td>
</tr>
<tr>
<td>HNF1b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mHNF1bs</td>
<td>5’ CCCTCACCCTCCTCTCTTTTTC3’</td>
<td>NM_009330.2</td>
</tr>
</tbody>
</table>
Clone | Sequence | NCBI Ref. Seq.
--- | --- | ---
mHNF1ba | 3’ GGGTTTCTCCCTTCTCGTTG5’ |  |
mHNF1b- EcoRI | 5’ TTGGATCCATGGTGTCCAAGCTCACGT3’ |  |
mHNF1b- BamHI | 3’ ATGAATTCAACCAGCTTCAGTG5’ |  |

2.7 RNA Studies

2.7.1 RNA extraction and quantitative PCR

2.7.1.1 Isolation of epithelial cells along the intestinal crypt–villus axis

Mouse intestinal tissue was isolated from an equal number of female and male mice (C57BL/6J (B6), 6 to 8 weeks old). To isolate cells populations from crypts and villi of mouse intestines, a fractionation method was used. This method used inversion of the intestinal lumen as described by Barnard et al. (1989) and optimized by Traber et al. (1991).

Briefly, the small intestine was cut after the duodenum and before the ileocaecal valve and perfused with ice-cold PBS, pH 7.4, containing 1mM dithiothretiol (DTT). Subsequently, a 6-cm-long section of mouse jejunum was inverted onto a plastic rod. The plastic rod was inserted into a purpose made hole in the cap of a 12–mL, plastic, round-bottom centrifuge tube, containing 10 mL citrate buffer (96 mM NaCl, 1.5 mM KCl, 27 mM sodium citrate, 8 mM KH₂PO₄, and 5.6 mM Na₂HPO₄, pH 7.4). After sealing with Parafilm, the assembly was incubated on a rotatory shaker at 37 °C for 10 min. Subsequently, the citrate buffer was changed to 10 mL PBS (pH 7.4) containing 0.5 mM DTT, 1.5 mM EDTA, and 1 mg/mL bovine serum albumin at 37 °C. Rotary shaking was continued for 10–20 min until the solution turned opaque indicating detachment of cells from the intestinal lumen. The solution was removed and replaced by pre-warmed fresh solution until 5 to 7 fractions were collected. Cell viability was judged by Tryphan Blue exclusion (Strober, 2001).

2.7.1.2 RNA extraction and microarray studies

Total RNA was isolated using the RNeasy Mini Kit (Qiagen) according to the manufacturer’s protocol. RNA quality was determined by using an RNA 6000 Nano LabChip® Kit on the Agilent Bioanalyzer 2100 (Agilent Technologies) platform.
according to the manufacturer’s protocol. RNA samples with and RNA integrity number (RIN) of ≥8 were used in subsequent microarray analyses (Figure 2.2).

**Figure 2.2: Quality analysis of microarray samples**

The quality of the total RNA was examined using the Bionalyzer 2100 (Agilent). According to RIN values, the quality of the isolated cells from the crypt was slightly better than that from villus. In these electropherograms, the two peaks represent the 18s and 28s ribosomal RNA (rRNA).

Global gene expression was analysed using an Agilent SurePrint G3 Mouse Exon Microarray kit at the Ramaciotti Center, University of New South Wales, Sydney.

### 2.7.1.3 Complementary DNA synthesis

cDNA was generated from extracted RNA using the SuperScript™ III Reverse Transcriptase enzyme (Invitrogen). The reaction was prepared by adding 1 μL of oligo(dT) primer (Invitrogen), 0.5 mM of dNTPs, and 8 μL of total RNA (concentration, 250 ng/μL). The reaction was incubated at 65 °C for 5 min to denature RNA before being transferred onto ice for 1 minute. Subsequently, 4 μL of 5× First-Strand buffer, 1 μL of
0.5 mM DTT, 1 μL RNaseOut (Invitrogen), 0.5 μL SuperScript™ III, and 3.5 μL of sterile distilled water was added. The resulting mixture was incubated for 60 min at 50 °C. The reaction was stopped by heating at 85 °C for 5 min.

2.7.1.4 Quantitative real-time RT-PCR (qRT-PCR)

qRT-PCR was conducted using the Fast SYBR® Green master mix (Applied Biosystems) in conjunction with an Applied Biosystems 7900 instrument. The reaction mixture (10 μL) included 5 μL SYBR® Green master mix, 3 μL of sample cDNA (1:10 dilution) and 1 μL of each primer (final concentration, 200 nM). The samples were pipetted into a 394-well plate. The PCR cycling parameters were 50 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 15 s and 60 °C for 1 min. All samples were analysed in duplicates. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA served as an internal control. GAPDH previously used as a single reference gene to compare the gene expression in the crypt-villus axis (Anderle et al., 2005). Primers were designed using the Primer3 program (http://frodo.wi.mit.edu/) to amplify fragments of 170–250 bp. The primers used for qRT-PCR are listed in Table 2.13. To validate the quality of the mRNA for microarray analysis, mRNA expression of six selected genes was quantified by qRT-PCR.

Table 2.13: Primer sets used for RT-PCR and qRT-PCR amplifications

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lgr5</td>
<td>5’TAAAGACGACGCAACACAGTG3’&lt;br&gt;3’GATTCGGATCAGCCAGCTAC5’</td>
</tr>
<tr>
<td>Slc6a19</td>
<td>5’CCGTCGTCTACGTTATGGGA3’&lt;br&gt;3’ACCCAGTTGGGTGTGGAAT5’</td>
</tr>
<tr>
<td>Hnf1a</td>
<td>5’TCACAGACACCAACCTCAGC3’&lt;br&gt;3’CCGTTGGATCGGAACACTC5’</td>
</tr>
<tr>
<td>Hnf4a</td>
<td>5’GCCAAGATTCGGACACCTCAG3’&lt;br&gt;3’ATTCAGATTTGGGATGGAAT5’</td>
</tr>
<tr>
<td>Gata4</td>
<td>5’TCAACCAGAAACGGAAGC3’&lt;br&gt;3’GCCGGCACAGTACTGGAATG5’</td>
</tr>
<tr>
<td>Cdx2</td>
<td>5’GCAGTCCCCTAGGAAGCAAG3’&lt;br&gt;3’CTGCATGTTCTGAAACAAAT5’</td>
</tr>
<tr>
<td>GAPDH</td>
<td>5’AACCTGGCGATTGGGAAGG3’&lt;br&gt;3’CCTGTGTGCTGGCCTAT5’</td>
</tr>
<tr>
<td>Sox9</td>
<td>5’TCAAGATGCGTGGAGCAACG3’&lt;br&gt;3’CCAGCCACAGCAGTGA5’</td>
</tr>
</tbody>
</table>
2.7.2 Microarray Analysis

2.7.2.1 Bioinformatics analyses of microarray data

Microarray data were analysed using the GeneSpring GX software (Agilent Technology). The GeneSpring software has tools for normalisation, statistical analyses, statistical interpretations, and quality control of gene expression data. Duplicate samples were prepared for each cell population. The normalisation step was performed as follows: individual probe sets were first filtered on expression between the 20th and 100th percentile in at least 2 out of 4 samples. After normalisation, the remaining probe sets were filtered for a more than three-fold change of expression. Unpaired $T$-tests, which were corrected with Bonferroni multiple testing, were applied on the differentially expressed genes. After normalisation and statistical analysis, 4464 probes were identified as differentially expressed genes.

2.7.2.2 Gene Ontology (GO) Analysis

Gene ontology (GO) is a bioinformatics tool designed to organise genes in different groups based on their attributes. To identify significantly enriched GO terms, differentially expressed genes from the microarray analysis were classified using GeneSpring GX software. GeneSpring GX program identified classes of genes based on $p < 0.01$. Each significant set of genes was grouped into different categories, based on biological processes, molecular function, and cellular components according to the GO consortium database (http://www.geneontology.org).

2.7.2.3 Pathway Analysis

The GeneSpring program was also used for pathway analysis of the 4464 differentially expressed genes. For a more selective pathway analysis a >10-fold change was set. The GeneSpring program uses the Kyoto Encyclopaedia of Gene and Genomes (KEGG) pathway database to identify significant pathways ($p < 0.05$).
2.8 Analysis of DNA and RNA

2.8.1 Agarose gel electrophoresis

DNA products were analysed using an agarose gel electrophoresis system. Briefly, the agarose gel (1%) was dissolved in the TAE buffer by boiling the mixture using microwave heating. After cooling the agarose to ~55 °C, it was poured into the electrophoresis apparatus containing a comb. Subsequently, the agarose was left to set and the gel was submersed in the TAE buffer, and the comb removed.

Approximately 20 μL of DNA products were mixed with the 6× loading buffer (Invitrogen) and loaded onto the gel wells. The presence of 60% glycerol in the loading buffer helped to sink the sample to the bottom of the well. This included bromophenol blue which assisted in the visualisation of DNA migration. Either a 1-kb or a 100-bp DNA ladder (NEB) was loaded as marker to determine the electrophoresed band sizes.

The electrophoresis tank (Bio-Rad) was run at 10 V/cm to separate DNA bands. The gel was stained with ethidium bromide solution (5 μL/mL) for 15 min before being photographed using a Gel-Doc UV system and an integrated imaging program (Vision-Capture).

2.8.2 Spectrophotometry

Both DNA and RNA samples were quantified using a NanoDrop Spectrophotometer (NanoDrop Technologies) at a wavelength of 260 and 280 nm using deionized water as a control.

2.8.3 DNA sequencing

DNA sequencing was performed by the Biomolecular Resource Facility (BRF) at JCSMR, ANU. Sequencing was performed using the Big Dye Terminator sequencing kit (ABI) and the Applied Biosystems 3730XL DNA analyser.

The sequencing reaction consisted of 2 μL of Big Dye Terminator (BDT), which included the polymerase enzyme and fluorescent nucleotides, 3 μL of a 5× BDT buffer (ABI), 3.2 pmol of the amplification primer (Table 2.14), 300 ng of template DNA, and
distilled water to adjust the final reaction volume to 20 μL. The contents of the reaction were mixed and placed in a 96 well Veriti thermal cycler (ABI). Cycling conditions are shown in Table 2.15.

Table 2.14: Primers used for sequencing

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA3.1 Seq Antisense #10</td>
<td>3’CTGGCAACTAGAAGGCACAGT5’</td>
</tr>
<tr>
<td>pcDNA3.1 Seq Sense #6</td>
<td>5’GGCTAGCGGTTTTAACTTAAGC3’</td>
</tr>
<tr>
<td>pcDNA3.1 Seq Antisense #7</td>
<td>3’TAAACGGGGCCCTCTAGACTC5’</td>
</tr>
<tr>
<td>pcDNA3.1 mHNF1a Seq Sense #8</td>
<td>5’GCTTACAAACTGGTTTGCCAAC3’</td>
</tr>
<tr>
<td>pGL4.12 int100L #24</td>
<td>3’TGGTGGCTTTTACCAACAGTA5’</td>
</tr>
</tbody>
</table>

Table 2.15: The Sequencing PCR conditions

<table>
<thead>
<tr>
<th>Process</th>
<th>Temperature (°C)</th>
<th>Time (min:sec)</th>
<th>Number of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>5:00</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>0:10</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>50</td>
<td>0:10</td>
<td>30 cycles</td>
</tr>
<tr>
<td>Extension</td>
<td>60</td>
<td>4:00</td>
<td></td>
</tr>
<tr>
<td>Extension (Final)</td>
<td>60</td>
<td>8:00</td>
<td>1</td>
</tr>
</tbody>
</table>

To purify the reaction, 2 μL of 3 M sodium acetate (pH 5.2) and 50 μL of ethanol (ice cold) were added to the 20 μL of the PCR reaction. After incubating the mixture at 23 °C for 15 min, it was spun at maximum speed for 20 min in a table top centrifuge. The supernatant was discarded, 250 μL of 70% ethanol was added, and the mixture spun for 5 min at maximum speed. After discarding the supernatant, the reaction was dried at 23 °C and delivered to the BRF for sequencing.

2.8.4 DNA bisulphite sequencing

Genomic DNA was isolated using DNeasy Blood & Tissue kit (Qiagen). Up to 200 μg of isolated genomic DNA was treated with an EpiTect Bisulfite Kit (Qiagen) according to the manufacturer’s instructions. In this process, all unmethylated cytosine residues were converted to uracil nucleotides, whereas methylated cytosine residues (5-methylcytosine) remained intact during the treatment.
The promoter region was amplified by nested PCR from bisulphite-treated template DNA. In the first round, a 1639 bp fragment starting from position –1380 to +259 was amplified, followed by another amplification of a fragment ranging from position –1131 to +18. The PCR products were sequenced and analysed for the methylation status of all CpG dinucleotides in the selected region. Primers used to investigate CpG dinucleotides in the B0AT1 promoter region are listed in Table 2.16. MethPrimer (http://www.urogene.org/methprimer/index1.html) was used for selecting and designing these methylation primers.

Table 2.16: Primers used in DNA methylation analysis

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>metB0AT1(--1380)</td>
<td>5’T TAGTTTTTTTAGTGTTTG3’</td>
</tr>
<tr>
<td>metB0AT1(+259)</td>
<td>3’ AAAAAATACTATAACTTCCAC5’</td>
</tr>
<tr>
<td>metB0AT1(--1131)</td>
<td>5’ GAGGAAAAAGATAATAGAAT3’</td>
</tr>
<tr>
<td>metB0AT1(+18)</td>
<td>3’AACACAAACCTCACCATAATAATC5’</td>
</tr>
</tbody>
</table>

The amplification was performed using the Taq polymerase (Qiagen) in a 50 µL reaction volume for 35 cycles under the following conditions: 94 ºC for 5 min, 94 ºC for 30 s, 50 ºC for 30 s, 72 ºC for 1 min and a terminal extension at 72 ºC for 5 min. A 50 µL aliquot of the amplification product was run on a 1% agarose gel, and gel elution was carried out using PureLink Quick Gel Extraction Kit (Invitrogen) according to the manufacturer’s protocol (see Section 2.4.2.2). The purified fragment of the correct product size was cloned using the Zero Blunt Topo XL cloning kit (Invitrogen). A small aliquot (2 µL) of the ligation reaction was transformed into chemically competent E. coli (TOP10 cells) (see Section 2.2.3). Ten colonies were grown overnight, plasmid was isolated and subjected to sequencing for confirmation of methylation status at the BRF. Methylation results were visualised by the BIQ Analyzer software (http://biq-analyzer.bioinf.mpi-inf.mpg.de/).
2.9 Protein-DNA interaction

2.9.1 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) was performed using the SimpleChip Enzymatic Chromatin IP Kit #9003 (Cell Signaling Technology) according to the manufacturer’s instructions. Freshly prepared tissue of mouse kidney cortex, liver, mucosa, or intestinal fractionations were used in the ChIP assay. Forty mg of each of the mouse tissues was cut into small pieces using a scalpel. The cut tissues were placed into the 15-mL falcon tubes with 1 mL ice cold PBS, pH 7.4, buffer containing the protease inhibitor cocktail (Cell Signaling Technology). Nuclear proteins were then cross-linked to the DNA by addition of freshly prepared formaldehyde (to a final concentration of 1.5%), and the tubes were incubated on a rotator for 20 min at 23 °C. The cross-linking was stopped by addition of glycine to a final concentration of 125 mM followed by further incubation on a rotator for 5 min at 23 °C. The tissues were centrifuged at 4 °C for 5 min at 1,500 rpm, and washed with 1 mL ice-cold PBS (pH 7.4) containing protease inhibitors. The centrifugation and washing steps were repeated seven times for seven fractions. The tissues were treated with 20–25 strokes in a Potter homogeniser, followed by centrifugation at 4 °C for 5 min at 1500 rpm. Chromatin was digested to 150–900-bp nucleosomal fragments using Micrococcal Nuclease (Mnase; Cell Signaling Technology) for 20 min at 37 °C. The reaction was stopped by adding EDTA. For further transcription factor studies, the nuclear membrane was fragmented by 25 strokes using a Potter homogeniser. For histone protein studies, the membrane was disrupted using a Misonix S-4000 sonicator at 15- s intervals for 2 min at 4 °C. The lysates then were centrifuged at 10000 rpm for 10 min at 4 °C to obtain the cross-linked chromatin in the supernatant.

The chromatin samples were immunoprecipitated with specific antibodies (listed in Section 2.1.3) at 4 °C with overnight rotation. After immunoprecipitation (IP) with the indicated antibodies, the IP mixture was incubated with ChIP Grade Protein G Magnetic Beads (#9006; Cell Signaling Technology) at 4 °C for 2 h with rotation. The magnetic beads were then washed with low-salt and high-salt ChIP buffer (Cell Signaling Technology). Subsequently, crosslinking was reversed in elution buffer and proteins were digested by adding proteinase K at 65 °C for 2 h. The DNA was purified using a spin column provided in the kit. Purified DNA fragments were analysed further by quantitative PCR (qPCR) and PCR amplifications using specific primers (Table 2.17).
Table 2.17: ChIP primer sets used for B\textsuperscript{0}AT1 promoter analysis

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (-139/+11)</td>
<td>5’CTGGGTCCTCTGCAGATAA3’</td>
<td>3’AAAGGGCAAGTGTTGTGTC5’</td>
</tr>
<tr>
<td>P2 (-353/–136)</td>
<td>5’ATCAGATATCTGCTGGTCTG3’</td>
<td>3’CCAGCCTAACCAGCTAAGCC5’</td>
</tr>
<tr>
<td>P3 (-495/–281)</td>
<td>5’CTGGAAGAACCCAAGCATA3’</td>
<td>3’GGAGGCACTCTCCAGCAATA5’</td>
</tr>
<tr>
<td>P4 (-501/–332)</td>
<td>5’CTGTGCTGGGAAGAACCCTC3’</td>
<td>3’TACAGACCAGGACTGTAAT5’</td>
</tr>
<tr>
<td>P5a (-858/–674)</td>
<td>5’CATGCCACCTAAGTCTC3’</td>
<td>3’TACAGAAGCCAGCATGAC5’</td>
</tr>
<tr>
<td>P5 (-682/–502)</td>
<td>5’TGCTGGCTCTCTGTATCTCC3’</td>
<td>3’GCTCCGTGCTCTAAGTGTCC5’</td>
</tr>
<tr>
<td>P6 (-984/–812)</td>
<td>5’GGAGGGGAATTGTGAGAAAGG3’</td>
<td>3’TGGTCTTTCTGCTGCAAAACAT5’</td>
</tr>
<tr>
<td>P7 (-1236/–976)</td>
<td>5’ATTTGGGCTTTAGGGGTGTT3’</td>
<td>3’CCCCTACATCATGTTCTTG5’</td>
</tr>
<tr>
<td>P8 (-1368/–1255)</td>
<td>5’TCTCAGGGTCTCTTCTTCACC3’</td>
<td>3’AACACCCCTAAAGCCAAAT5’</td>
</tr>
</tbody>
</table>

2.10 Statistical analysis

All results are expressed as mean ± SD. p < 0.05 was considered to be statistically significant. For statistical comparison Student’s t-test was used. All statistical studies were performed using GraphPad Prism v.6 (GraphPad Software).
Chapter 3  Comparison of differentially expressed genes along the crypt–villus axis

3.1  Introduction

Each villus can be separated anatomically into two regions, namely crypt and villus tip. Enterocytes are continually dividing and differentiating, starting in intestinal crypts, and migrating along a villus, until they eventually slough off at the villus tip (Cosentino et al., 1996). The B⁰AT1 protein is expressed at the apical membrane of villus enterocytes but not in crypt cells. Previous studies also showed that B⁰AT1 mRNA is highly expressed at the villus tips, while absent in intestinal crypts (Bröer et al., 2004). The comparison of these two regions should provide insight about transcriptional regulation of the Slc6a19 gene on the intestinal surfaces.

Caco-2 cells are endowed with villi and have been used as a cellular model of intestinal enterocytes for transport and metabolic studies (Van Beers et al., 1995). Rodent IEC6 cells, by contrast, have been used as a model of crypt-like intestinal cells (Quaroni and Beaulieu, 1997, Quaroni and May, 1979, Trompette et al., 2004, Drago et al., 2006). Similarly, another progenitor-like cell model, the human intestinal epithelial cell line (HIEC), was used to study intestinal cell differentiation (Benoit et al., 2010, Escaffit et al., 2006).

As an alternative to cell lines, cell fractionation methods have been developed (Traber et al., 1991) and widely used to examine the function, proliferation or differentiation of intestinal epithelial cells along the crypt–villus axis. Alternatively, laser micro-dissection has been used to examine and characterise mouse or human intestinal epithelial cells (Anderle et al., 2005, Stappenbeck et al., 2003, Crosnier et al., 2006).

Initially, Olsen et al. (2004) established an online database of crypt–villus gene expression patterns in the mammalian intestine. The mammalian crypt–villus axis has also been studied by microarray analyses by a number of groups with different study objectives (Anderle et al., 2005, Mariadason et al., 2005, George et al., 2008). These studies have shown that genes associated with DNA synthesis are expressed in the crypt, whereas genes related to digestion and absorption of nutrients present in differentiated enterocytes of the villus.
3.2 Results

3.2.1 Validation of the crypt–villus cell fractionation

Using a rotatory shaking mechanism, cells were detached in different fractions (see Section 2.7.1.1). Initial fractions were enriched in enterocytes released from the tip, and later fractions were enriched in cells derived from the crypt.

Cell enrichment in fractions was assessed by qRT-PCR (see Section 2.7.1.4). For instance, fractions 1 and 2 contained high levels of Slc6a19 mRNA, while it was barely detectable in fraction 7 (Figure 3.1). Lgr5 (leucine-rich repeat containing G-protein coupled receptor 5), a crypt stem cell marker, showed an opposite trend and was highly enriched in fraction 7.

Accordingly, RNA from fraction 2 (villus) and 7 (crypt) were selected for microarray analyses. RNA samples were hybridised to Agilent arrays using 39,430 probe sets.
Figure 3.1: Schematic presentation of the crypt–villus structure and expression profiles of *Slc6a19* and *Lgr5*
Blue horizontal lines visualise fractionation (A). Expression profile of villus marker (*Slc6a19*) along the crypt–villus axis (B). Expression profile of crypt marker (*Lgr5*) along the crypt–villus axis (C).
3.2.1.1 Validation of microarray data

The microarray data were validated through qRT-PCR of six selected genes (Figure 3.2). A high level of correlation was found between qRT-PCR and microarray results (R = 0.9146).

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Microarray</th>
<th>qRT-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lgr5</td>
<td>+49.21</td>
<td>+115.69</td>
</tr>
<tr>
<td>Slc6a19</td>
<td>-18.36</td>
<td>-7.59</td>
</tr>
<tr>
<td>Hnf1a</td>
<td>+2.83</td>
<td>+4.11</td>
</tr>
<tr>
<td>Hnf4a</td>
<td>+3.42</td>
<td>+5.41</td>
</tr>
<tr>
<td>Gata4</td>
<td>+3.39</td>
<td>+4.98</td>
</tr>
<tr>
<td>Cdx2</td>
<td>+4.18</td>
<td>+3.23</td>
</tr>
</tbody>
</table>

Figure 3.2: Confirmation of the microarray results by qRT-PCR

(A) Expression levels of selected genes as measured by microarray and qRT-PCR.
(B) Correlation between microarrays and qRT-PCR.
3.2.1.2 Gene ontology (GO) analysis

Gene ontology analysis was performed to identify genes that perform villus-specific or crypt-specific functions (see Section 2.7.2.2). More than 3000 differentially expressed genes were identified. Of these, 1,854 genes were up-regulated and 2,610 genes were down-regulated along crypt–villus axis.

Robustly expressed genes were categorised based on their involvement in biological processes, molecular functions, or cellular structures/organelles according to the GO consortium database (http://www.geneontology.org). In agreement with the roles of mature enterocytes in nutrient and ion absorption, transport-associated genes were highly enriched in the villus tip. A summary of the results is shown in Table 3.1. A detailed graph of gene ontology terms enriched in the villus is shown in Appendix 8.3.

Table 3.1: List of gene ontology (GO) categories significantly up-regulated in cells of the villus tip

<table>
<thead>
<tr>
<th>GO category</th>
<th>Total genes in GO class</th>
<th>Up-regulated genes</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological process/organic substance transport</td>
<td>392</td>
<td>44</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/Na+ transport</td>
<td>145</td>
<td>23</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/carboxylic acid transport</td>
<td>179</td>
<td>24</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/nitrogen compound transport</td>
<td>147</td>
<td>21</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/lipoprotein transport</td>
<td>12</td>
<td>5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cellular component/brush border</td>
<td>61</td>
<td>13</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Molecular function/symporter activity</td>
<td>114</td>
<td>20</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Molecular function/solute–sodium symporter activity</td>
<td>47</td>
<td>12</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Molecular function/cofactor transporter activity</td>
<td>11</td>
<td>5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Molecular function/carboxylic acid transmembrane transporter activity</td>
<td>95</td>
<td>14</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
Genes that were up-regulated in crypt cells were mostly associated with cell proliferation. This result is shown in Table 3.2. A more detailed graph showing gene ontology terms enriched in the crypt is shown in Appendix 8.4.

### Table 3.2: List of gene ontology (GO) categories significantly up-regulated in crypt cells

“Total genes in GO class” stands for the number of genes included in a GO category. “Number of genes” stands for the actual number of genes that were up-regulated in the category. P <0.05 indicates statistical significance.

<table>
<thead>
<tr>
<th>GO category</th>
<th>Total genes in GO class</th>
<th>Up-regulated genes</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological process/M phase of mitotic cell cycle</td>
<td>216</td>
<td>63</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/cell cycle</td>
<td>787</td>
<td>115</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/chromosome segregation</td>
<td>105</td>
<td>30</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/DNA replication</td>
<td>152</td>
<td>25</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Biological process/DNA repair</td>
<td>291</td>
<td>33</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cellular component/chromosome</td>
<td>498</td>
<td>61</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cellular component/microtubules</td>
<td>627</td>
<td>68</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cellular component/nuclear part</td>
<td>1745</td>
<td>120</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Molecular function/microtubule motor activity</td>
<td>44</td>
<td>10</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

#### 3.2.1.3 Pathway analysis

The GeneSpring program was used to categorize genes according to the Kyoto Encyclopaedia of Gene and Genomes (KEGG) (Kanehisa and Goto, 2000) (see Section 2.7.2.3).

In addition to genes involved in digestion and absorption, cell adhesion and cytoskeleton-associated pathways were significantly up-regulated in the villus compartment. This is consistent with a role of adherence and tight junctions between enterocytes functioning as an epithelial barrier (Champe et al., 2005). Furthermore, the brush border on the surface of differentiated enterocytes requires an elaborate cytoskeletal network to support the microvillus structure (Mooseker, 1985).
Digested dietary lipids, including cholesterol, free fatty acids, and monoacylglycerol, are absorbed by fully differentiated enterocytes and distributed to the body via chylomicrons (Shiau, 1987). Chylomicrons, which contain triacylglycerol, cholesterol, and phospholipids, are assembled in the endoplasmic reticulum of enterocytes and secreted into the lymphatic system which transports them to the blood (Champe et al., 2005). Accordingly, it is not surprising that expression of genes associated with lipid uptake and lipid biosynthesis was found to be up-regulated in villus cells. Also, genes associated with the peroxisome proliferator-activated receptors (PPAR) signalling pathway, which mediate chylomicron formation and lipid metabolism (Braissant et al., 1996), were highly expressed in the villus compartment.

Pathway analyses revealed that genes associated with immune functions, including T cell and B cell receptor signalling pathways, chemokine signalling pathways, interleukin signalling pathways, were highly enriched in villus. A summary of these pathways is shown in Table 3.3. In contrast to earlier studies, genes related to pathways associated with immune cell signalling were enriched in the villus. This is consistent with the fact that the differentiated epithelium harbors lymphocytes which are important for host defence against potential pathogens (Ferguson, 1977).
Table 3.3: Overrepresented KEGG pathways in the villus

“Pathway entities” refers to the number of genes corresponding to the KEGG pathways. “Matched entities” is the number of the significantly expressed genes in a pathway. The P-value indicates statistical significance.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Pathway entities</th>
<th>Matched entities</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T cell receptor signalling pathway</td>
<td>133</td>
<td>62</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Regulation of actin cytoskeleton</td>
<td>151</td>
<td>54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>B cell receptor signalling pathway</td>
<td>156</td>
<td>54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chemokine signalling pathway</td>
<td>193</td>
<td>60</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PPAR signalling pathway</td>
<td>87</td>
<td>32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Calcium regulation in the cardiac cell</td>
<td>150</td>
<td>48</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Integrin mediated cell adhesion</td>
<td>100</td>
<td>35</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>191</td>
<td>53</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Insulin signalling</td>
<td>159</td>
<td>47</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Glycolysis and gluconeogenesis</td>
<td>48</td>
<td>21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IL-2 signalling pathway</td>
<td>76</td>
<td>27</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IL-5 signalling pathway</td>
<td>69</td>
<td>24</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IL-4 signalling pathway</td>
<td>61</td>
<td>22</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Myometrial relaxation and contraction</td>
<td>157</td>
<td>43</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>EGF1 signalling pathway</td>
<td>176</td>
<td>46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Statin pathway</td>
<td>20</td>
<td>10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Triacylglyceride synthesis</td>
<td>23</td>
<td>11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>G protein signalling pathways</td>
<td>91</td>
<td>27</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Kennedy pathway</td>
<td>14</td>
<td>8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IL-6 signalling pathway</td>
<td>99</td>
<td>28</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Nuclear receptors in lipid metabolism</td>
<td>30</td>
<td>12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Eicosanoid synthesis</td>
<td>19</td>
<td>9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Selenium metabolism</td>
<td>31</td>
<td>10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Adipogenesis</td>
<td>133</td>
<td>34</td>
<td>0.001</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>7</td>
<td>5</td>
<td>0.001</td>
</tr>
<tr>
<td>IL-3 signalling pathway</td>
<td>100</td>
<td>27</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Similar to GO analysis, KEGG pathway analysis of crypt cells confirmed enrichment of mRNAs involved in cell proliferation, including those involved in DNA replication and cell cycle. A summary of the analysis is shown in Table 3.4.

As outlined in Section 1.3.1.1, Wnt signalling is important for regulation of stem cell populations in the crypt and maturation of Paneth cell. The microarray analysis confirmed that Wnt signalling pathway was up-regulated in the crypt tissue.

### Table 3.4: KEGG pathways enriched in the crypt

“Pathway entities” refers to the number of genes corresponding to the KEGG pathways. “Matched entities” is the number of the significantly expressed genes in a pathway. The P-value indicates statistical significance.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Pathway entities</th>
<th>Matched entities</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ribosomal proteins</td>
<td>80</td>
<td>73</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>DNA replication</td>
<td>41</td>
<td>34</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>88</td>
<td>49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>G1-to-S cell-cycle control</td>
<td>62</td>
<td>40</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Translation factors</td>
<td>51</td>
<td>26</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>178</td>
<td>54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mRNA processing</td>
<td>551</td>
<td>194</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miRNA regulation of DNA damage response</td>
<td>91</td>
<td>29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Eukaryotic transcription initiation</td>
<td>41</td>
<td>21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Homologous recombination</td>
<td>13</td>
<td>10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TNFα NF-κB signalling</td>
<td>184</td>
<td>50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mismatch repair</td>
<td>9</td>
<td>8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>One carbon metabolism</td>
<td>29</td>
<td>13</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Nucleotide metabolism</td>
<td>19</td>
<td>9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mitochondrial gene expression</td>
<td>19</td>
<td>9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TGF-β receptor signalling pathway</td>
<td>150</td>
<td>35</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Non-homologous end joining</td>
<td>8</td>
<td>5</td>
<td>0.001</td>
</tr>
<tr>
<td>Oestrogen signalling</td>
<td>74</td>
<td>19</td>
<td>0.003</td>
</tr>
<tr>
<td>Histone modifications</td>
<td>5</td>
<td>3</td>
<td>0.01</td>
</tr>
<tr>
<td>Wnt signalling pathway</td>
<td>109</td>
<td>22</td>
<td>0.03</td>
</tr>
</tbody>
</table>
3.2.1.4 Transcription factors

To determine which TFs could potentially regulate Slc6a19 gene expression, microarray data for all TFs was separately analysed. TFs were classified into three groups, namely those that showed a more than 3-fold difference (up or down) between expression in the villus tip and crypts and those that were expressed with little change in expression. A total of 428 TFs showed significant expression along the crypt–villus axis. Among these, 73 TFs had higher expression in the tip and 68 TFs had higher expression in the crypts. Whilst 287 TFs were evenly expressed along the entire axis (Figure 3.3). A complete list of expressed TFs is presented in Table 8.6. A possible regulatory role of these TFs in transportation of nutrients along the crypt–villus axis was analysed further.

Figure 3.3: Proportion of TFs along the crypt–villus axis

Distribution of total 428 TF genes along the crypt–villus axis based on their expression profiles, each criteria p < 0.05.
3.3 Discussion

In this chapter, I investigated the genome-wide expression patterns of different genes using RNA derived from cells at the villus tip and cells from the crypts.

As shown in previous studies, B0AT1 expression is higher at the villus tip than in the crypt region. To further analyse these two different regions, a modified version of a fractionation method developed by Traber et al. (Traber et al., 1991) was used. This method uses mechanical force (flushing of inverted intestine) to isolate intestinal epithelial cells from the crypts through villus tips. The method was employed successfully, as demonstrated by a significant enrichment of Slc6a19 mRNA in villus tip fractions and of the well-known intestinal stem cell marker Lgr5 in crypt fractions. Some limitations of the crypt–villus fractionation technique were observed. For instance, the technique appears to underrepresent cells derived from the very bottom of the crypt. Therefore, genes expressed in the crypt region may not be highly enriched in these fractions i.e. crypt genes such as Cdx2 and Tcf4 were found to be evenly expressed along the crypt–villus axis (Table 8.6).

3.3.1 Pathway and GO analysis

To gain insight into the biological processes underlying the expression of a specific set of genes at the villus tip, GO and KEGG pathway analyses were applied to the microarray data. These analyses confirmed that crypt and villus epithelial cells express entirely separate sets of genes.

In mature enterocytes at the villus tip, many genes were expressed that are involved in nutrient and ion absorption. Accordingly, GO categories associated with transporter-related genes were identified. This observation is consistent with earlier studies indicating that most genes associated with nutrient absorption were overrepresented in differentiated enterocytes (Mariadason, Nicholas et al. 2005, George, Wehkamp et al. 2008). A previous study also demonstrated that genes associated with lipid synthesis were up-regulated during crypt to villus differentiation (Stegmann, Hansen et al. 2006), which was confirmed by our microarray analyses.

In the crypt compartment, GO and KEGG analyses, identified genes that are involved in DNA replication. It has been reported that DNA synthesis quickly shuts down during crypt to villus transition (Stegmann et al., 2006). It has also been shown that genes
related to the Wnt signalling pathways were highly expressed in crypt cells (Stegmann, Hansen et al. 2006, Sabates-Bellver, Van der Flier et al. 2007). Our analysis confirms these previous reports.

### 3.4 Summary

In summary, these results showed significant alterations in expression of genes related to TFs and signalling pathways during the crypt–villus differentiation. More importantly, the analysis provides us with TF candidates that may control villus-specific regulation of the B₀AT1 transporter in mouse intestine. Our results further showed that the fractionation method successfully isolated crypt–villus cell populations.
Chapter 4  Transcriptional regulation of *Slc6a19* gene along the crypt–villus axis

4.1 Introduction

Finger-like protrusions, called villi, expand the intestinal epithelial surface area. Crypts are located at the base between villi. The crypt–villus axis is an important structural feature of the intestinal surface. Intestinal stem cells reside at the bottom of the crypts. Mature intestinal cells localise along the crypt–villus axis. Continuous division of stem cells generates new cells, which differentiate into absorptive enterocytes or secretory cells (see Section 1.2.1) while migrating up towards the villus tip. Intestinal secretory cells are categorised as enteroendocrine cells, goblet cells, Paneth cells and tuft cells (Barker and Clevers, 2007).

Previously, it was shown that B0AT1 is highly expressed on the apical membrane of intestinal enterocytes (Bröer et al., 2004). While the functional properties of the protein have been extensively characterized and its physiological role in different organs has been elucidated, the transcriptional regulation underlying its localised expression has not been investigated. TFs play key roles in controlling cell differentiation along the crypt–villus axis (Figure 4.1). Some of the TFs that are involved in intestinal cell differentiation are summarised in the following paragraphs.
Figure 4.1: Distribution of differentiated intestinal cells along the crypt–villus axis

Transcription factors depicted in boxes control cell-type-specific gene expression required for cell differentiation; from (Gerbe et al., 2011).

4.1.1.1 Hepatocyte nuclear factors

The transcription factors HNF1α (see Section 1.3.2.3) and HNF4α (see Section 1.3.2.4) are necessary for hepatocyte development, but they are also expressed in the kidney, intestine and pancreas (Sladek et al., 1990, Pontoglio et al., 1996). In the kidney both TFs are exclusively expressed in the proximal tubules (Pontoglio et al., 1996). HNF1α and HNF4α are highly expressed in the crypts and villi, and are important for crypt formation (Sauvaget et al., 2002). HNF1α has a dimerization domain, a DNA-binding domain and a transactivation domain (Chouard et al., 1990). Another important
hepatocyte nuclear factor is HNF1b, which controls terminal cell differentiation of stem cells into mature enterocytes (D'Angelo et al., 2010).

4.1.1.2 SPDEF

SPDEF is a member of the Ets (E26 transformation-specific) family of TFs and is expressed in cells residing in crypts. SPDEF is particularly important for maturation of intestinal goblet and Paneth cells (Figure 4.1) (Gregorieff et al., 2009).

4.1.1.3 SOX9

In the intestine, SOX9 (see Section 1.3.2.1) is expressed in the progenitor cells located in the crypts. SOX9 is required for differentiation of stem cells into Paneth cells and goblet cells (Figure 4.1) (Bastide et al., 2007). As mentioned in Section 1.3.2.2, SOX9 is crucial for β-catenin/Wnt interaction, which is essential for crypt–villus formation (Sato et al., 2009). SOX9 is known to inhibit Cdx2 and Muc2 genes in the crypts; both proteins are expressed in the villus where SOX9 is not expressed (Blache et al., 2004).

4.1.1.4 NEUROG3

Neurogenin 3 (NEUROG3) is required for enteroendocrine cell development in the gastrointestinal tract and the pancreas (Figure 4.1) (Wang et al., 2006).

4.1.1.5 FOXA2

Forkhead box protein a2 (FOXA2) belongs to the forkhead class of DNA-binding proteins and is widely expressed in mouse intestinal crypts (Besnard et al., 2004). Foxa2 can activate Muc2, which encodes mucin secreted by the goblet cells. Foxa2 is important for regulation of goblet cell differentiation (Ye and Kaestner, 2009).

4.1.1.6 CDX2

CDX2 contains a homeobox domain that is required for the development and differentiation of the intestine (Lorentz et al., 1997). Sucrase–isomaltase (SI) was one of the first genes the expression of which along the crypt–villus axis was characterized in detail. These studies suggested that HNF1a and CDX2 are regulating gene activation (Traber, 1990). Similarly, HNF1a and CDX2 also regulate the lactase-phlorizin hydrolase (LPH) gene transcription in the intestine (Mitchelmore et al., 2000).
4.1.1.7 GATA4

GATA4 is a member of the zinc-finger family of TFs. GATA4 and HNF1a are necessary for expression of epithelial genes involved in nutrient absorption (Bosse et al., 2007). Together with CDX2, these TFs regulate many of the enterocyte-specific genes in the mammals (Benoit et al., 2010).

4.2 Results

4.2.1 Prediction of putative nuclear receptor response elements in the Slc6a19 promoter

To identify relevant TFs and their binding sites on the Slc6a19 (GenBank® accession number NM_028878) promoter, the UCSC Genome Browser (www.genome.ucsc.edu) was used (see Section 2.5.1). Fifteen different species were examined. Bioinformatic analysis revealed that a region located between –1960 and +57 of the TSS was conserved to some extent (Figure 4.2.A). Furthermore, another region close to the TSS was highly conserved (–228 to +57) among the different species (Figure 4.2.B). Therefore, in our studies, we targeted a 2.5 kb fragment of this region of the promoter sequence.

The proximal promoter of the mouse Slc6a19 gene, as judged by conservation of the sequence across mammalian species, was relatively short, i.e. 228 bp. It is a typical class II promoter where the TSS is preceded by a TATA-binding motif (Weaver, 2012) at position –23 to –28 and a reasonably conserved transcription factor IIB recognition motif (position –39 to –46), comprising the binding sites for general TFs and for polymerase II (Weaver, 2012). The TSS CCACCTT is similar to the mammalian consensus initiator sequence (Weaver, 2012).
Figure 4.2: Promoter sequence alignment of *Slc6a19*

The UCSC Genome Browser was used for sequence alignment. The alignments include *Slc6a19* sequences from 15 species spanning position –2494 to +57 (A) and at higher resolution position –352 to +57 (B). Black domains indicate conserved regions when compared to the rat sequence.
Subsequently, the MatInspector program was used to identify TF candidates that potentially bind to this region. The MatInspector program found 227 different TF-binding sites in a 2.5 kb region upstream of the Slc6a19 TSS. Candidates (see Section 3.2.1.4) with robust microarray expression along the crypt–villus axis (Table 4.1) were analysed further.

### Table 4.1: Expression of candidate TFs in crypts or villi

TFs with robust expression in either crypts or villi are shown. GATA4, HNF1b and PAX4 are listed because they have been previously implicated in enterocyte maturation but do not have predicted promoter binding sites. A matrix number of 1.00 indicates an exact match with the TF consensus binding site, whereas 0.75 is considered a possible binding site for a TF. Revised location indicates results from mutagenesis data (shown below). The ratio of mRNA abundance along the crypt–villus axis was determined by microarray analysis. BS# was used to discriminate multiple binding sites.

<table>
<thead>
<tr>
<th>TFs</th>
<th>Position</th>
<th>Strand</th>
<th>Matrix</th>
<th>Revised Location</th>
<th>Effect</th>
<th>Ratio Villus/Crypt (Microarray)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATF4</td>
<td>–1212/–1204</td>
<td>–</td>
<td>0.99</td>
<td>–</td>
<td>0</td>
<td>0.90</td>
</tr>
<tr>
<td>CDX2</td>
<td>–36/–18</td>
<td>–</td>
<td>0.92</td>
<td>–</td>
<td>0</td>
<td>0.52</td>
</tr>
<tr>
<td>CDX2</td>
<td>–1212/–1194</td>
<td>–</td>
<td>0.87</td>
<td>–</td>
<td>0</td>
<td>0.52</td>
</tr>
<tr>
<td>CDX2</td>
<td>–1833/–1815</td>
<td>+</td>
<td>0.86</td>
<td>–</td>
<td>0</td>
<td>0.52</td>
</tr>
<tr>
<td>CREB3L3</td>
<td>–89/–69</td>
<td>+</td>
<td>0.95</td>
<td>0</td>
<td>8.80</td>
<td></td>
</tr>
<tr>
<td>EGR1</td>
<td>–167/–151</td>
<td>+</td>
<td>0.89</td>
<td>0</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>EGR1</td>
<td>–2112/–2096</td>
<td>+</td>
<td>0.86</td>
<td>0</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>FOXA2</td>
<td>–305/–289</td>
<td>–</td>
<td>0.99</td>
<td>0</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>FOXA2</td>
<td>–835/–819</td>
<td>–</td>
<td>0.99</td>
<td>0</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>GATA4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>HNF1a–BS1</td>
<td>–126/–110</td>
<td>–</td>
<td>0.81</td>
<td>–126/–110</td>
<td>++</td>
<td>0.96</td>
</tr>
<tr>
<td>HNF1a–BS2</td>
<td>–153/–137</td>
<td>+</td>
<td>0.89</td>
<td>0</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>HNF1a–BS3</td>
<td>–1205/–1189</td>
<td>–</td>
<td>0.90</td>
<td>0</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>HNF1b</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>HNF4a–BS1</td>
<td>–239/–215</td>
<td>–</td>
<td>0.85</td>
<td>–41/–38</td>
<td>+++</td>
<td>1.10</td>
</tr>
<tr>
<td>HNF4a–BS2</td>
<td>–1368/–1344</td>
<td>+</td>
<td>0.93</td>
<td>0</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>HNF4a–BS3</td>
<td>–1701/–1677</td>
<td>+</td>
<td>0.88</td>
<td>0</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>NEUROG3</td>
<td>–2352/–2340</td>
<td>+</td>
<td>0.96</td>
<td>0</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>NEUROG3</td>
<td>–2474/–2462</td>
<td>+</td>
<td>0.98</td>
<td>0</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>PAX4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0.36</td>
</tr>
<tr>
<td>SMAD3</td>
<td>–228/–218</td>
<td>+</td>
<td>0.99</td>
<td>0</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>SMAD3</td>
<td>–2016/–2006</td>
<td>+</td>
<td>0.99</td>
<td>0</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>SOX9–BS1</td>
<td>–47/–23</td>
<td>+</td>
<td>0.77</td>
<td>–218/–215</td>
<td>0/– (on HNF activated promoter)</td>
<td>0.05</td>
</tr>
<tr>
<td>SOX9–BS2</td>
<td>–1150/–1126</td>
<td>+</td>
<td>0.967</td>
<td>0</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>SPDEF</td>
<td>–27/–7</td>
<td>+</td>
<td>0.86</td>
<td>–195/–192</td>
<td>++</td>
<td>0.71</td>
</tr>
<tr>
<td>TFs</td>
<td>Position</td>
<td>Strand</td>
<td>Matrix</td>
<td>Revised Location</td>
<td>Effect</td>
<td>Ratio Villus/Crypt (Microarray)</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>------------------</td>
<td>--------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>STAT3</td>
<td>–301/–283</td>
<td>+</td>
<td>0.77</td>
<td>0</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>–303/–285</td>
<td>–</td>
<td>0.76</td>
<td>0</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>–504/–486</td>
<td>–</td>
<td>0.97</td>
<td>0</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>–718/–700</td>
<td>+</td>
<td>0.95</td>
<td>0</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>–1044/–1026</td>
<td>–</td>
<td>0.959</td>
<td>0</td>
<td>10.6</td>
<td></td>
</tr>
</tbody>
</table>

### 4.2.2 Analysis of possible TF candidates

Twelve different TFs (Table 4.1) were identified as putative candidates regulating *Slc6a19* transcription. Their functional regulatory effects on the *Slc6a19* promoter activity were tested by luciferase reporter assays (see Section 2.6). GATA4, HNF1b and PAX4 were included because they have previously been identified as enterocyte differentiation regulators.

HEK293 cells were co-transfected with the full-length *Slc6a19* promoter pGL4(–2494/+57) (see Section 2.6.2) and each of the predicted TFs separately, namely ATF4, CDX2, CREB3L3, EGR1, FOXA2, GATA4, HNF1a, HNF1b, HNF4a, NEUROG3, PAX4, SMAD3, SOX9, SPDEF, or STAT3 (Figure 4.3).
HEK293 cells were co-transfected with three different plasmids. (i) Firefly luciferase inserted into pSlc6a19(–2494 to +57) (ii) Renilla luciferase in the control vector (iii) one of the following constructs: pcDNA3.1 carrier vector alone (blue bar), or pcDNA3.1 expressing ATF4, CDX2, CREB3L3, EGR1, FOXA2, GATA4, HNF1a, HNF4a, HNF1b, HNF4a, NEUROG3, PAX4, SMAD3, SOX9, or SPDEF. Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments. (**) p < 0.01).

STAT3 is unique it has to be activated after co-transfection. When not activated, STAT3 is located in the cytoplasm (Kishimoto, 1993); however, after phosphorylation, it translocates into the nucleus. STAT3 phosphorylation can be triggered by interleukin-6 (IL-6), a pro-inflammatory cytokine (Zhong et al., 1994). Two different constructs were used to express STAT3, namely, mouse STAT3–GFP and human STAT3, both inserted in a mammalian expression vector. Luciferase assays were hence performed after co-transfection of STAT3 and treatment with IL-6 (20 ng/mL).

Except for HNF1a, HNF4a, and SPDEF, none of the other candidates (Table 4.1) significantly activated the Slc6a19 promoter (Figure 4.3 and Figure 4.4).
Figure 4.4: Characterization of STAT3-mediated transcriptional activation by reporter gene assays

HEK293 cells were co-transfected with three different plasmids. (i) Firefly luciferase inserted into pSlc6a19(–1703/+57) (ii) Renilla luciferase in the control vector (iii) one of the following constructs: mouse STAT3-GFP or human STAT3 empty pcDNA3.1 expression vector. Cells were treated with IL-6 to activate STAT3. Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments.

4.2.3 HNF1a can activate the Slc6a19 promoter

HNF1a is important for the differentiation of stem cells into epithelial cells in the intestine (D'Angelo et al., 2010). The MatInspector program predicted three different binding sites for HNF1a in the proximal region of the Slc6a19 promoter (Figure 4.5). These putative binding sites were located at positions; –111 to –114 (HNF1-BS1), –144 to –147 (HNF1-BS2), and –1205 to –1189 (HNF1-BS3). To validate these binding sites, the Slc6a19 promoter deletion constructs were co-transfected with HNF1a expression vectors into HEK293 cells. Seven different deletion constructs were used: pSlc6a19(–2494/+57), pSlc6a19(–1703/+57), pSlc6a19(–1380/+57), pSlc6a19(–972/+57), pSlc6a19(–437/+57), pSlc6a19(–136/+57), or pSlc6a19(–37/+57).
Figure 4.5: Locations of HNF1a-binding sites on the mouse Slc6a19 promoter

Predicted HNF1a-binding sites are underlined and core binding sites indicated in red colour. The TATA box is marked by a box. The arrow indicates the TSS, designated as +1.

Figure 4.6 shows that except for the shortest construct pSlc6a19 (−37/+57), HNF1a coexpression caused a three to fivefold higher activity than the background activity measured in the absence of HNF1a. Even a construct comprising just 136 bp upstream of the TSS was fully active in the presence of HNF1a. This suggests that an HNF1a-binding site is located between positions −136 to −37 bp on the Slc6a19 promoter.
**Figure 4.6: Identification of the Slc6a19 promoter regions responsive to the HNF1a TF using luciferase reporter experiments in vitro**

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the Slc6a19 promoter deletion constructs and with the Renilla luciferase as control. HNF1a expression vector (red bars) or empty pcDNA3.1 expression vector (blue bars) was co-expressed to activate transcription or serve as control, respectively. The following promoter deletion constructs were used: pSlc6a19(–2494/+57), pSlc6a19(–1703/+57), pSlc6a19(–1380/+57), pSlc6a19(–972/+57), pSlc6a19(–437/+57), pSlc6a19(–136/+57), or pSlc6a19(–37/+57). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant increases of promoter activity due to co-transfection of HNF1a are indicated by asterisks (** p < 0.01).

To further examine the putative HNF1a-binding sites, we mutated the sites identified by the MatInspector program. Site-directed mutagenesis (see Section 2.4.2.5) was used to mutate two predicted HNF1a-binding sites in the pSlc6a19(–136/+57) and pSlc6a19(–437/+57) promoter constructs. Mutation of the HNF1a-BS1 at (–111/–114) (Hnf1a-Mut1) diminished activity compared with the activity of the corresponding wild-type promoter. The second mutation, HNF1-BS2 at (–144/–147) (HNF1a-Mut2), did not affect luciferase activity when co-transfected with HNF1a (Figure 4.7). The in vitro studies showed that the binding site at position –111 to –114 was responsible for expression of Slc6a19 in the presence by HNF1a. No evidence was found for
MatInspector program predicted binding sites (Table 4.1) at positions –153 to –137 (HNF1-BS2) or –1205 to –1189 (HNF1a-BS3).

Figure 4.7: Detection of HNF1a-binding sites using mutated constructs

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the Slc6a19 promoter constructs and with the Renilla luciferase as control. HNF1a expression vector (red bars) or empty pcDNA3.1 expression vector (blue bars) was co-expressed to activate transcription or serve as control, respectively. The following promoter constructs were used: pSlc6a19(–1380/+57), pSlc6a19(–972/+57), pSlc6a19(–437/+57), pSlc6a19(–136/+57), pSlc6a19(–437/+57)Mut2, pSlc6a19(–437/+57)Mut1, pSlc6a19(–136/+57)Mut1, or pSlc6a19(–37/+57). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant increases of promoter activity due to co-transfection of HNF1a are indicated by asterisks (* p < 0.05; ** p < 0.01).

To determine whether HNF1a could bind to the Slc6a19 promoter within living cells, chromatin immunoprecipitation (ChIP) assay was used. The ChIP assay was performed using chromatin fragments isolated from intestinal enterocytes (see Section 2.9.1) and an antibody against HNF1a. Real-time PCR was subsequently performed to generate amplicons of 100-200 bp spanning the Slc6a19 promoter from position –1368
to +11. ChIP analysis revealed that the promoter region from –139 to +11 could bind to HNF1a (Figure 4.8). This result was consistent with the luciferase assay, which indicated the presence of an HNF1a-binding site at position –111 to –114. An additional binding site for HNF1a was detected between –984 and –812; this site could not be observed by the reporter gene assays.

**Figure 4.8: HNF1a directly binds to the Slc6a19 promoter**

ChIP assay was performed to identify proteins cross-linked to chromatin extracted from the mouse small intestinal epithelium. The chromatin was immunoprecipitated using an antibody against HNF1a. After DNA purification, the precipitated DNA was quantified by qPCR using primer sets specific for –139/+11, –353/–136, –495/–281, –501/–332, –682/–502, –984/–812, –1236/–976, or –1368/–1255 regions. The lower part of the graph shows the location of primer pairs used in the real-time PCR. The upper part of the graph indicates that ChIP signals relative to the input signal.

### 4.2.4 HNF4a activates the Slc6a19 promoter

Our initial promoter experiment indicated that HNF4a is another activator of the Slc6a19 promoter (Figure 4.3). HNF4a was highly expressed along the crypt–villus axis.
(Figure 4.18). Three HNF4a-binding sites were predicted using the MatInspector program at positions –239 to –215 (HNF4a-BS1), –1368 to –1344 (HNF4a-BS2), and –1701 to –1677 (HNF4a-BS3). According to HNF4a consensus binding site (CAAAGT) (Wallerman et al., 2009), three additional binding sites were predicted by visual inspection at locations –1108 to –1105 (HNF4a-BS4), –1088 to –1085 (HNF4a-BS5), and –38 to –41 (HNF4a-BS6) (Figure 4.9). To confirm the possible binding sites in vitro, HNF4a was co-expressed with truncated Slc6a19 promoter constructs: pSlc6a19(–2494/+57), pSlc6a19(–1703/+57), pSlc6a19(–1380/+57), pSlc6a19(–972/+57), pSlc6a19(–437/+57), pSlc6a19(–136/+57), or pSlc6a19(–37/+57).

**Figure 4.9: Location of putative HNF4a-binding sites on the Slc6a19 promoter**

Predicted binding sites for HNF4a are indicated on the Slc6a19 promoter sequence. The TATA box is highlighted by a box. The arrow indicates the TSS designated as +1.
Co-expression of HNF4a activated the promoter about 20-fold in the longest construct pSlc6a19(--2494/+57) (Figure 4.10). This activation dropped to fourfold in construct pSlc6a19(--972/+57). The smallest construct did not show any significant promoter activity. These results suggested the presence of at least two HNF4a-binding sites.

Figure 4.10: HNF4a induces luciferase expression driven by the Slc6a19 promoter in vitro

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the Slc6a19 promoter constructs and with the Renilla luciferase as control. HNF4a expression vector (red bars) or empty pcDNA3.1 expression vector (blue bars) was co-transfected to activate transcription or serve as control, respectively. The following promoter constructs were used: pSlc6a19(--2494/+57), pSlc6a19(--1703/+57), pSlc6a19(--1380/+57), pSlc6a19(--972/+57), pSlc6a19(--437/+57), pSlc6a19(--136/+57), or pSlc6a19(--37/+57). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant increases of the promoter activity due to co-transfection with HNF4a are indicated by asterisks (* p < 0.05; ** p < 0.01). N.S designates non-significant results.

Functional expression analysis did not show a significant activity difference between the pSlc6a19(--1703/+57) and the pSlc6a19(--1380/+57) deletion constructs. Similarly, luciferase activity did not show any significant decrease when we deleted the
region spanning –437 to –136. Therefore, we concluded that the predicted HNF4a-BS3 and HNF4a-BS1 sites were inactive in the *in vitro* assay system. Consequently, HNF4a-BS2 is most likely the actual binding site of HNF4a. In addition, the results suggested that there should be another binding site located between –136 and –37, which was not identified by the MatInspector program.

To determine the potential HNF4a-binding regions, we mutated their corresponding binding sites in the pSlc6a19(–1380/+57) and the pSlc6a19(–972/+57) reporter constructs (Figure 4.11). This included the atypical HNF4a-binding site between –136 and –37 (HNF4a-BS6), which were accidentally discovered while searching for putative SOX9-binding sites (Figure 4.19) on the *Slc6a19* promoter.

**Figure 4.11: Determination of HNF4a-binding sites using mutated constructs**

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the *Slc6a19* promoter constructs and with the *Renilla* luciferase as control. HNF4a expression vector (red bars) or empty pcDNA3.1 expression vector (blue bars) was co-expressed to activate transcription or serve as control, respectively. The following promoter constructs were used: pSlc6a19(–1380/+57), pSlc6a19(–1380/+57)Mut1, pSlc6a19(–972/+57)Mut1, pSlc6a19(–1380/+57)Mut1,2, pSlc6a19(–972/+57)Mut1, pSlc6a19(–136/+57), or pSlc6a19(–37/+57). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant increases of promoter activity due to co-transfection of HNF4a are indicated by asterisks (* p < 0.05; ** p < 0.01). N.S designates non-significant results.
Mutation of the HNF4a-BS6 binding site on the shorter construct pSlc6a19(−972/+57), completely suppressed HNF4a-induced activation of the Slc6a19 promoter. A very short promoter deleted upstream of position −37, did not induce luciferase activity. However, mutating the predicted HNF4a-binding site, HNF4a-BS2, on pSlc6a19(−1380/+57) did not reduce the reporter gene activity (Figure 4.11).

When we deleted the region from −1380 to −972, the luciferase activity was significantly reduced (Figure 4.11). Furthermore, reducing the distance between both HNF4a-binding sites increased reporter gene activity, suggesting that both sites were functional (Figure 4.12). These suggest the presence of another functional binding site for HNF4a between −1380 to −972.

![Diagram](image)

**Figure 4.12: Reducing the distance between HNF4a-binding sites enhances in vitro luciferase expression driven by the Slc6a19 promoter**

HEK293 cells were co-transfected with vectors expressing the *Firefly* luciferase downstream of the *Slc6a19* promoter constructs and with the *Renilla* luciferase as control. HNF4a expression vector (red bars) or empty pcDNA3.1 expression vector (blue bars) was co-transfected to activate transcription or serve as control, respectively. The following promoter constructs were used: pSlc6a19(−2494/+57), pSlc6a19(−972/+57), or pSlc6a19(−2494/+57)[−812/−136]. Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the *Firefly* luciferase activity to the control *Renilla* luciferase activity. The experiment was performed three times, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant increases of promoter activity due to co-transfection of HNF4a are indicated by asterisks (* p < 0.05; ** p < 0.01).
As a result, we predicted and mutated two further putative sites, namely HNF4a-BS4 (–1105/–1108) and HNF4a-BS5 (–1085/–1088) (Figure 4.9), and prepared more targeted deletion constructs to identify the second HNF4a-binding site. The deletion constructs showed a significant drop in reporter gene activity between –1226 to –1078. However, mutating HNF4a-BS4 and HNF4a-BS5 (HNF4a-Mut3 and HNF4-Mut4) did not reduce the luciferase activity (Figure 4.13).

Figure 4.13: Detection of the second putative HNF4a-binding site using mutated constructs

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the Slc6a19 promoter constructs and with the Renilla luciferase as control. HNF4a expression vector (red bars) or empty pcDNA3.1 expression vector (blue bars) was co-expressed to activate transcription or serve as control, respectively. The following promoter constructs were used: pSlc6a19 (–1380/+57), pSlc6a19 (–1316/+57), pSlc6a19 (–1226/+57), pSlc6a19 (–1078/+57), pSlc6a19 (–1380/+57)Mut4, pSlc6a19 (–1380/+57)Mut3, pSlc6a19 (–972/+57), pSlc6a19 (–136/+57), or pSlc6a19 (–37/+57). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments. Significant increases of the promoter activity due to co-transfection of HNF4 are indicated by asterisks. (* p < 0.05; ** p < 0.01)
To verify whether HNF4a binds to the putative binding sites on the Slc6a19 promoter in intact tissue, ChIP assays were performed. Cross-linked chromatin, isolated from scraped mouse intestinal mucosa, was used for ChIP assays. The intestinal enterocyte chromatin was immunoprecipitated using an antibody against HNF4a. Immunoprecipitated DNA was analysed by real-time PCR amplification generating amplicons of 100-250bp spanning position –1368 to +11 of the Slc6a19 promoter, as shown in Figure 4.14. ChIP analysis of this DNA fragment showed that the two primer sets covering the region –353 to +11 showed the strongest signal for HNF4a-binding (Figure 4.14). This result confirmed the putative HNF4a-binding site at position –38 to –41, but failed to confirm any other strong binding site between –1226 and –1078.

Figure 4.14: HNF4a directly binds to the Slc6a19 promoter

ChIP assay was performed on cross-linked chromatin isolated from the mouse small intestine epithelium. Chromatin was immunoprecipitated using an antibody against HNF4a. After purification, the precipitated DNA was quantified by qPCR using primer sets specific for –139/+11, –353/–136, –495/–281, –501/–332, –682/–502, –984/–812, –1236/–976, or –1368/–1255 regions. The lower part of the graph shows the corresponding location of primer pairs used in the real-time PCR. The upper part of the graph presents that the ChIP signal relative to the input signal.
4.2.5 Synergistic effects of HNF1a and HNF4a co-transfection on the \textit{Slc6a19} promoter activity

It has previously been reported that HNF1a and HNF4a can act synergistically to transcriptionally activate human \(\alpha\)-antitrypsin (\(\alpha\)-AT), UDP glucuronosyltransferase 1 family-polypeptide A9 (\(UGT1A9\)), and dihydrodiol dehydrogenase (\(DD4\)) (Eekhoute et al., 2004, Ozeki et al., 2001, Hu and Perlmutter, 1999).

To determine whether HNF1a and HNF4a could also synergistically activate the \textit{Slc6a19} promoter, both TFs were co-expressed in HEK293 cells (Figure 4.15). The experiment showed that co-expression of HNF1a and HNF4a did not increase the luciferase activity more than HNF4a-induced promoter activity alone.

Another study suggested synergistic action between HNF1a and Cdx2 at the lactase-phlorizin hydrolase (\textit{LPH}) promoter (Mitchelmore et al., 2000). However, co-expression of HNF1a and Cdx2 using our \textit{Slc6a19} reporter gene constructs did not significantly increase the luciferase activity compared to HNF1a-induced activity alone (Figure 4.15). CDX2 alone could not activate \textit{Slc6a19} transcription (Figure 4.3).
Figure 4.15: The combinatorial effect of CDX2 and HNF TFs on the Slc6a19 promoter activity

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the promoter construct pSlc6a19(−2494/+57), together with various TFs and with the Renilla luciferase as control. The following combinations were used: HNF1a expression vector, HNF1a plus CDX2, HNF4a, HNF1a plus HNF4a, or empty pcDNA3.1 expression vector (blue bar). Twenty-four hours after transfection, the luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments.

4.2.6 SPDEF can activate Slc6a19 transcription in vitro

According to the microarray data, SPDEF was highly expressed in the intestinal villus (Figure 4.18) and could therefore be a candidate TF to activate the Slc6a19 promoter. Furthermore, the MatInspector program predicted that the Slc6a19 promoter contains a putative SPDEF-binding site located between −27 bp to −7 bp (SPDEF-BS1) (Figure 4.16).

Luciferase assays, however, indicated that SPDEF-induced reporter gene activity dropped to background activity when deletions passed beyond position −201 (Figure 4.17). No further loss could be detected in the smallest construct pGL4(−136/+57), which contained the predicted SPDEF-binding region. The pGL4(−136/+57) construct also
could not be activated by co-expression of SPDEF, suggesting that the potential binding site was likely located upstream of the predicted site.

The consensus binding site for mouse SPDEF has previously been confirmed (Wei et al., 2010) to be the sequence CCGGAT. A similar site (CAGGAT) was identified in the Slc6a19 promoter at position −183 to −180 (SPDEF-BS2) (Figure 4.16).

**Figure 4.16: Predicted binding sites for the mouse SPDEF**

Two putative SPDEF-binding sites are shown in red letters. Both sites were mutated to test their relevance.

Luciferase activity was measured to investigate the two putative SPDEF-binding sites. The assays showed a sevenfold and sixfold increase in activity of the pSlc6a19(−437/+57) and pSlc6a19(−201/+57) promoter constructs, respectively, when SPDEF was co-expressed (Figure 4.17). Slc6a19 promoter activity was abolished by mutation of the second proposed SPDEF-BS2 binding site. Accordingly, a deletion upstream of position −136 bp abolished the effect of SPDEF on Slc6a19 promoter activity. This experiment demonstrates that SPDEF, similar to HNF1α and HNF4α is sufficient to activate the Slc6a19 promoter in the HEK293 cells.
Figure 4.17: SPDEF induces \textit{in vitro} luciferase activity driven by the \textit{Slc6a19} promoter

HEK293 cells were co-transfected with vectors expressing the \textit{Firefly} luciferase downstream of the \textit{Slc6a19} promoter constructs and with the \textit{Renilla} luciferase as control. SPDEF expression vector (red bars) or empty pcDNA3.1 expression vector (blue bars) was co-transfected to activate transcription or serve as control, respectively. The following promoter constructs were used: pSlc6a19(−437/+57), pSlc6a19(−201/+57), pSlc6a19(−201/+57)[mutated], or pSlc6a19(−136/+57). Twenty-four hours after transfection, the luciferase activity was measured. Relative luciferase activity was determined by normalising the \textit{Firefly} luciferase activity to the control \textit{Renilla} luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant increases of promoter activity due to co-transfection of SPDEF are indicated by asterisks (* p < 0.05; ** p < 0.01).

4.2.7 Identification of possible repressor(s) along the crypt–villus axis

The \textit{in vitro} and \textit{in vivo} experiments indicated that SPDEF, HNF1\textalpha{} and HNF4\textalpha{} strongly activated B\textsuperscript{0}AT1 transcription in the mouse intestine. qPCR was used to confirm the relative expression of TFs in the villus and the crypt as determined by microarray for some of the most prominent TFs (\textbf{Table 4.1}). Fractionation experiments followed by qPCR showed robust transcript levels of HNF1\textalpha{} and HNF4\textalpha{} along the crypt-villus axis (\textbf{Figure 4.18A}). HNF1\textalpha{} and HNF4\textalpha{} expression was high in the crypt region, but marked expression remained in villus cells.

All TF candidates tested thus far had higher levels of expression in crypts as ascertained by qPCR and microarray analysis (\textbf{Table 4.1}). Thus, \textit{Slc6a19} mRNA expression in the villus cells could not be explained by a corresponding expression of SPDEF, HNF1\textalpha{}, or HNF4\textalpha{}.
Genome-wide analyses showed that 68 TFs had robust expression in the intestine and were expressed at significantly higher levels (ratio > 3) in the villus than in the crypt (Table 8.6). Of these, only STAT3 and CREB3L3 have putative binding sites on the Slc6a19 promoter. However, co-transfection of STAT3 and CREB3L3 failed to activate reporter gene expression (Figure 4.3). Consequently, we hypothesized that other TF(s) might suppress Slc6a19 expression in the crypts.
Figure 4.18: Expression profiles of putative TFs in the villus and crypt as determined by qPCR and microarray studies

(A) Expression levels of *Hnf1a*, *Hnf4a* and *Sox9* are shown in consecutive fractions derived from mouse small intestine (F1 villus tip – F7 crypts). (B) Microarray results showing expression levels in crypt and villus fractions.
4.2.8 Analysis of SOX9-binding on Slc6a19 promoter

The TF Sox9 was particularly lowly expressed in the villus compared to crypt fractions (Figure 4.18.A). Microarray results confirmed expression patterns similar to those obtained by qPCR analyses (Figure 4.18.B). Moreover, it was markedly expressed in hepatic and pancreatic tissues (Mori-Akiyama et al., 2007, Kanai and Koopman, 1999), which do not express Slc6a19 similarly to intestinal crypts.

The pGL4(–2494/+57) promoter construct contains two predicted SOX9-binding sites at positions between –37 and –34 (SOX9-BS1), and between –1150 and –1126 (SOX9-BS2) (Figure 4.19). To elucidate the effects of SOX9 on the regulation of Slc6a19 promoter activity, SOX9 was co-expressed in HEK293 cells.

<table>
<thead>
<tr>
<th>SOX9– BS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAAA7TACCAGAGACAGAATTAAAAAGTG?ATCTA1TCA1TCA1TCTA1TCTA1TTTA1AAAAGAGAGAGCAGCAATGTA -1139</td>
</tr>
<tr>
<td>AAAGGAAGAGGAAAGACACACAACAGAACTAAAGGCTCTGTGAATTCTAGGAGAGTCACTGTTGCTGCA -1072</td>
</tr>
<tr>
<td>AG7TTG7CTGAACTGTCGCTCCCTGAGGCCAAGCCCTGAGCTTGCTATTGTGTGCTTATTGTGCAGAGAGAGG -201</td>
</tr>
<tr>
<td>TGCCCGAGGCCCTGGAGCAGAATCTCCTGAGCCTGTAATGAGGGGGGGGTGCTTAGCTGGTAGCTGGGATTAGGCTGGGT -134</td>
</tr>
<tr>
<td>GCTCT7GCAAGMIAAGGCATTAAACAGTTTGCAGGAGCGGCCCTGAGGATCTGCTGACGCTGCTTCTTTC -67</td>
</tr>
</tbody>
</table>

Figure 4.19: Predicted binding sites for SOX9 on the mouse Slc6a19 promoter

The two predicted SOX9-binding sites on the Slc6a19 proximal promoter region are indicated in red. The predicted TATA box and TSS (+1) are also shown.

Luciferase reporter assays showed that SOX9 did not activate Slc6a19 transcription (Figure 4.3). However, when this TF was co-transfected with either HNF1a (Figure 4.20) or HNF4a (Figure 4.21), it strongly inhibited the promoter activity as long as the predicted SOX9-binding site BS1 was included in the tested construct.
Figure 4.20: SOX9 inhibits *in vitro* luciferase activity induced by HNF1a

HEK293 cells were co-transfected with vectors expressing the *Firefly* luciferase downstream of the *Slc6a19* promoter constructs, with the *Renilla* luciferase as control. HNF1α expression vector (purple bars) or empty pcDNA3.1 expression vector (red bars) was co-expressed to activate transcription or serve as control, respectively. In a separate experiment, SOX9 was expressed together with HNF1α (orange bars). The following promoter constructs were used: pSlc6a19(--1380/+57), pSlc6a19(--972/+57), pSlc6a19(--437/+57), pSlc6a19(--136/+57), or pSlc6a19(--37/+57). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the *Firefly* luciferase activity to the control *Renilla* luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant changes in promoter activity due to co-transfection of SOX9 are indicated by asterisks (* p < 0.05; ** p < 0.01). N.S designates non-significant results.
HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the Slc6a19 promoter constructs and with the Renilla luciferase as control. HNF4a expression vector (purple bars) or empty pcDNA3.1 expression vector (blue bars) was co-expressed to activate transcription or serve as control, respectively. In a separate experiment, SOX9 was co-expressed with HNF4a (green bars). The following promoter constructs were used: pSlc6a19(–1380/+57), pSlc6a19(–972/+57), pSlc6a19(–437/+57), pSlc6a19(–136/+57), or pSlc6a19(–37/+57). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was calculated by normalising the Firefly luciferase activity relative to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments. Significant changes of the promoter activity due to co-transfection of SOX9 are indicated by asterisks (* p < 0.05; ** p < 0.01). N.S designates non-significant results.

We were unable to analyse mutation of SOX9-binding site SOX9-B51 (–37 to –34), because the proposed SOX9-B51 binding site overlaps with the HNF4a-binding site (Figure 4.22); mutation of this site also abolishes activation by HNF4a. As shown in Figure 4.20 and Figure 4.21, deletion of the SOX9-B52 region (–1150 to –1126) did not abolish the inhibitory effect of SOX9 on Slc6a19 promoter constructs –972/+57, –437/+57, and –136/+57, respectively.
Figure 4.22: Investigation of the putative SOX9-binding site using wild-type and mutated pSlc6a19(–972/+57) constructs

HEK293 cells were co-transfected with Renilla luciferase as control and with vectors expressing the Firefly luciferase downstream of the promoter construct pSlc6a19(–972/+57) or the same construct with mutated SOX9-BS1. In addition different transcription factors were co-expressed. The following constructs were used: HNF1a expression vector (orange bars), HNF4a (green bars), HNF1a plus SOX9 (grey bars), or empty pcDNA3.1 expression vector (red bars). Twenty-four hours after transfection, the luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments. (* p < 0.05; ** p < 0.01).

To further validate the reporter gene results, SOX9-binding to the Slc6a19 promoter was tested in isolated mouse intestinal cells by ChIP assay using cross-linked mouse intestinal chromatin. Chromatin fragments were immunoprecipitated using a polyclonal antibody against SOX9 and the Slc6a19 promoter was detected by real-time PCR generating amplicons of 150-200 bp spanning the promoter region from position –1368 to +11.

These experiments (Figure 4.23) suggested that the SOX9 binding site was located between position –353 to –136 confirming findings by luciferase assays. The precise location, however, could not be determined.
Figure 4.23: SOX9 directly binds to the Slc6a19 promoter

ChIP assay was used to identify SOX9-binding sites on cross-linked chromatin isolated from mouse small intestinal epithelium. The chromatin was immunoprecipitated using an antibody against SOX9. After purification, the precipitated DNA was quantified by qPCR using primer sets specific for –139/+11, –353/–136, –495/–281, –501/–332, –682/–502, –984/–812, –1236/–976, or –1368/–1255. The lower part of the graph shows the location of primer pairs used in the real-time PCR. The upper part of the graph shows the ChIP signal relative to the input signal.

SOX9 has been shown to recognise an ACAA motif to regulate transcription (Harley et al., 1994) (Figure 4.24.A). Further inspection of the proximal promoter revealed possible binding sites at positions +10/+13 (SOX9-BS3) and –218/–215 (SOX9-BS4) (Figure 4.24.B).
Figure 4.24: Predicted binding sites for SOX9

(A) Weblogo presentation of the SOX9 consensus sequence. (B) Location of predicted SOX9-binding on the Slec6a19 promoter.

To check whether these binding sites could be targeted by SOX9, site-directed mutagenesis was performed. SOX9-BS3 mutation did not change the effect of SOX9 on luciferase activity. However, mutation of the SOX9-BS4 site increased the overall promoter activity and partially abrogated the effect of SOX9 regardless of whether transcription was driven by HNF1a or HNF4a (Figure 4.25).
HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the promoter constructs, together with various transcription factors and with the Renilla luciferase as control. The following constructs were used: pSlc6a19(−437/+57), SOX9-BS3 mutated pSlc6a19(−437/+57), SOX9-BS4 mutated pSlc6a19(−437/+57), together with HNF1α expression vector (orange bars), HNF1α plus SOX9 (grey bars), HNF4α (green bars), HNF4α plus SOX9 (red bars), or empty pcDNA3.1 expression vector (black bars). The Renilla luciferase was used as control vector. Twenty-four hours after transfection, the luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments. (* p < 0.05; ** p < 0.01).

Mutation of SOX9-BS4 did not completely abolish the inhibitory effect of SOX9, suggesting that this TF may affect transcription indirectly, for instance by binding to other TFs. To test whether the SOX9 effect required direct binding to the promoter region, the SOX9 DNA-binding domain was mutated. To this end, a well-known SOX9 mutation, H65Y, was chosen (McDowall et al., 1999), which affects its target sequence binding. The experiments indicated that the inhibitory effect of SOX9 was weakened but not completely abolished (Figure 4.26). This suggested that SOX9 may inhibit transcription through interaction with the promoter region and through protein-protein interaction with other TFs, such as HNF1α and HNF4α.
Figure 4.26: The effect of SOX9(H65Y) on the pSlc6a19(–437/+57)

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of the promoter construct pSlc6a19(–437/+57), together with different TFs and with the Renilla luciferase as control. The following combinations were used: HNF4a expression vector, HNF4a plus SOX9, HNF4a plus SOX9(H65Y), HNF1a, HNF1a plus SOX9, HNF1a plus SOX9(H65Y), or empty pcDNA3.1 expression vector (red bar). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant changes of the promoter activity due to co-transfection of HNF1a and HNF4a are indicated by asterisks (* p < 0.05; ** p < 0.01).

To further investigate whether some of the SOX9 effects are independent of DNA-binding, SOX9(H65Y) was co-transfected with a promoter containing a mutated SOX9-binding site 4 which completely abolished the inhibitory effect of SOX9 (Figure 4.27). This result suggested that SOX9 can act both as a sequence-specific transcription factor and as an inhibitory cofactor. In summary SOX9 acts as powerful inhibitor of HNF4a and HNF1a-driven Slc6a19 expression.
Figure 4.27: The combinatorial effects of SOX9(H65Y) and SOX9-binding-site mutation on pSlc6a19(–437/+57)

HEK293 cells were co-transfected with vectors expressing the Firefly luciferase downstream of promoter construct SOX9-BS4 mut/pSlc6a19(–437/+57), together with different TFs and with the Renilla luciferase as control. The combinations used were: HNF1α expression vector, HNF1α plus SOX9, HNF1α plus SOX9(H65Y), HNF4α, HNF4α plus SOX9, HNF4α plus SOX9(H65Y), or empty pcDNA3.1 expression vector (red bar). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. P values indicate statistical significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. Significant changes of promoter activity due to co-transfection of HNF4α and HNF1α are indicated by asterisks (* p < 0.05; ** p < 0.01).

4.2.9 Analysis of mouse Collectrin and Ace2 promoters

To explain regulatory mechanisms driving ACE2 (Ace2) and TMEM27 (Collectrin) co-expression (see Section 1.4.3) with B0AT1 in kidney and intestine, we analysed their promoter regions. To determine the length and position of the Ace2 and Collectrin promoters and to determine highly conserved regions, bioinformatics was used to analyse promoter regions of the genes encoding Collectrin (NM_020626) and Ace2 (NM_027286) upstream of their corresponding TSSs. As a result, a sequence 1450 kb upstream of the Ace2 TSS (Figure 4.28.A) and another sequence 1800 bp upstream of the Collectrin TSS (Figure 4.28.B) were chosen.
Figure 4.28: Sequence alignment of promoters of Collectrin and Ace2 from different species

Alignments were generated using the UCSC Genome Browser. Black domains indicate conserved regions in different species. Regions 1.5 kb upstream of the Ace2 TSS (A) and 2 kb upstream of the Collectrin TSS (B) were analysed.
The distributions of putative TF-binding sites on the Collectrin and Ace2 promoters were examined by the MatInspector program. Possible binding sites for HNF1a, HNF1b and HNF4a, but not SOX9, were found on the Collectrin promoter (Table 8.4). The Ace2 promoter had binding sites for HNF1a and HNF4a, but not HNF1b or SOX9 (Table 8.5).

To test whether the Tmem27 and Ace2 promoters were controlled by similar regulatory elements as the Slc6a19 promoter, a region 2 kb upstream of the Tmem27 TSS and another 1.5 kb upstream of the Ace2 TSS were inserted separately into the pGL4.12 firefly luciferase reporter plasmid (see Section 2.6.3). Subsequently, HEK293 cells were co-transfected with reporter gene constructs and TFs previously identified to regulate the Slc6a19 promoter (Figure 4.29).

Similar to Slc6a19, Tmem27 and Ace2 promoters were markedly transactivated by HNF1a and HNF1b transcription factors and these effects were suppressed by SOX9. However, HNF4a, did not activate Tmem27 and Ace2 promoters. While HNF1b did not affect the Slc6a19 promoter. Despite these differences Slc6a19 and Ace2 are co-expressed in enterocytes (Kowalczuk et al., 2008), confirming the synergistic action of HNF1a and HNF4a in the intestinal epithelium (Shih et al., 2001).
Figure 4.29 Effects of the different TFs on Collectrin and Ace2 promoters

HEK293 cell were co-transfected with the Firefly luciferase inserted in (A) pCollectrin(−1903/+76) and (B) pAce2(−1509/+170) together with the Renilla luciferase vector as control. In addition, the following constructs were co-transfected separately: empty pcDNA3.1 (black bar), or pcDNA3.1 expressing HNF1a, HNF1a plus SOX9, HNF4a, HNF1b, SPDEF, CDX2, ATF4, PAX4, GATA4, CREB3L3, EGR1, FOXA2, NEUROG3, or SMAD3. Twenty-four hours after transfection, the luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. Transfections were repeated at least three times using triplicate samples. The data are means ± SD of triplicate experiments. (* p < 0.05; ** p < 0.01).
4.2.10 Effects of HNF1a mutation on Ace 2 or Tmem27 promoter activities

One hundred and ninety-three different HNF1a mutations have been described in MODY patients (Ellard and Colclough, 2006). The HNF1a(R131W) mutation is the most common cause of type 3 MODY disorder. The R131W mutation is located in the DNA-binding domain of HNF1a; this mutation abrogates HNF1a-binding to its target sequence (Chi et al., 2002). When we co-transfected this HNF1a variant with Ace2 and Tmem27 promoters (Figure 4.30), transcriptional activation was largely abolished. While co-expression with the Slc6a19 promoter reduced activation by about 50%.

**Figure 4.30: The effect of HNF1a(R131W) on Slc6a19, Ace2 and Tmem27 promoters.**

HEK293 cells were co-transfected with the Firefly luciferase inserted in pSlc6a19(−1380/+57), pCollectrin(−1903/+76), or pAce2(−1509/+170) together with the Renilla luciferase vector as control. In addition, the following constructs were co-transfected: empty pcDNA3.1 (blue bars), and pcDNA3.1 expressing HNF1a (red bars) or HNF1a(R131W). Twenty-four hours after transfection, luciferase activity was measured. Relative luciferase activity was determined by normalising the Firefly luciferase activity to the control Renilla luciferase activity. P values indicate the significance derived from three experiments, each using triplicate transfections. The data are means ± SD of triplicate experiments. (* p < 0.05; ** p < 0.01).
4.3 Discussion

Previously, $Slc6a19$ was shown to be highly expressed in intestinal enterocytes and epithelial cells of the kidney proximal tubule (Bröer et al., 2004). While the transport activity of $B^0$AT1 and its physiological roles in different organs have been described in some detail, the transcriptional regulation underlying its localised expression has not been investigated. In this chapter, I have examined the transcriptional regulation of the $Slc6a19$ gene in detail.

4.3.1 Novel Hepatocyte Nuclear Factor response elements on the $Slc6a19$ promoter

Hepatocyte Nuclear Factors are necessary for hepatocyte development. However, these TFs are not only expressed in the liver, but they are also expressed in the kidney and intestine. Accordingly, HNF1a and HNF4a have been shown to be important regulators of differentiation in intestinal epithelial cells (Pontoglio et al., 1996, Sladek et al., 1990).

Kikuchi et al. (2010) used bioinformatics and epigenetic analyses to investigate the regulation of amino acid transporters’ genes in kidney (Kikuchi, Yagi et al. 2010). The bioinformatics analysis showed an HNF1a-binding site at position –111 to –114. The study also showed possible HNF1a-binding sites on all major amino acid transporter genes. To our knowledge, this was the first and only report suggesting HNF1a-binding motifs in the $Slc6a19$ promoter region. In agreement with that study, our in vitro analysis confirmed that HNF1a binds to the $Slc6a19$ promoter at this predicted murine promoter site. Additionally, our ChIP analysis using mouse intestinal chromatin verified that HNF1a binds the promoter region, in vivo. As shown in human patients with MODY disorder, HNF1a and HNF4a are also critical for glucose metabolism (Fajans et al., 2001). We replicated an HNF1a mutation, which causes MODY disorder, and co-transfected this mutated form of the TF with the $Slc6a19$ promoter, but it did not completely abolish transcriptional activation.

To our knowledge, no other studies have investigated the regulatory function of HNF4a on the $Slc6a19$ gene in the mammals. With similar methodology, an HNF4a-binding site was confirmed by in vitro and in vivo studies. This is the first study showing that HNF4a can regulate $Slc6a19$ gene in the mammalian intestine. As discussed earlier (see Section 1.3.2.3), HNF1a-knockout mice suffer from renal Fanconi syndrome, which
is characterised by diabetes, glycosuria and aminoaciduria (Pontoglio et al., 1996). Recent clinical studies have shown that an R76W mutation in HNF4a causes Fanconi syndrome due to dysfunction of the renal tubules (Hamilton et al., 2013, Numakura et al., 2015). This new finding indicates that not only HNF1a, but also HNF4a, is necessary for proper functioning of the renal proximal tubules. In accordance with the above studies, our results suggest that HNF4a together with HNF1a controls Slc6a19 expression along the renal proximal tubules.

Some studies have shown that a number of intestinal genes are regulated by combinatorial action of HNFs, GATA TFs and CDX2 (Boudreau et al., 2002, Coskun et al., 2010, Maher et al., 2006). In contrast, our luciferase experiments testing different combinations of these TFs did not show any synergistic regulatory effect on the Slc6a19 promoter.

4.3.2 Negative effect of SOX9 on the Slc6a19 activation

SOX9 is a well-known TF involved in transcriptional activation of genes (Bell et al., 1997, Bastide et al., 2007). Several studies have shown that SOX9 drives stem cell differentiation into mature enterocytes. Yet some genes can be negatively regulated by SOX9 in the crypt, including Cdx2 and Muc2 which are crucial for enterocyte maturation (Blache et al., 2004).

CDX2 positively regulates intestinal TFs such as Cdx1, Hnf1a and Hnf4a, which control gene expression in villus cells (Gao et al., 2009, Heath, 2010). Therefore, reduced SOX9 expression could allow CDX2 expression in the villus cells; thus, activating Slc6a19 expression indirectly. However, we revealed that CDX2 expression in crypt cells is higher than in mature enterocytes; therefore, SOX9 is unlikely to indirectly activate Slc6a19 expression via CDX2. Instead, while HNF1a and HNF4a activated the Slc6a19 promoter, their action was directly inhibited by SOX9. SOX9 has direct and indirect effects on the expression of its target genes. It has been shown that SOX9 negatively regulates the B-catenin/Tcf4 pathway indirectly via cyclin D1 and c-Myc (Bastide, Darido et al. 2007). A recent study showed that genes involved in chondrocyte maturation are regulated by SOX9 either directly or indirectly (Ohba et al., 2015).

Our data further revealed a SPDEF consensus binding site on the Slc6a19 promoter. Since SPDEF is expressed only in goblet and Paneth cells (Gregori eff et al.,
2009, Noah et al., 2010) it could cause expression of \textit{Slc6a19} in these cells. However, SOX9 has a similar expression pattern, to SPDEF and is also expressed in Paneth cells (Mori-Akiyama et al., 2007). SOX9 down-regulation causes a significant decrease in the number of goblet cells (Bastide et al., 2007). Therefore, a possible activating effect of SPDEF could be suppressed by SOX9 in Paneth and goblet cells. Further analysis should include \textit{in vivo} studies to investigate SOX9 and SPDEF interaction, on the \textit{Slc6a19} promoter of goblet and Paneth cells separately.

4.3.3 Ace2 and Collectrin

HNF1B was previously shown to transcriptionally regulate \textit{Tmem27} expression in the pancreatic β-cell and in kidney (Zhang et al., 2007). HNF1b is a member of the same TF family as HNF1a and they share almost 90% homology at their corresponding DNA-binding domains (Tronche and Yaniv, 1992). HNF1B is expressed in epithelial cells along the whole nephron (Zhang et al., 2007). Our microarray results, however, did not reveal any significant expression of the \textit{Hnf1b} mRNA in the crypt–villus fractions. In vitro luciferase assays confirmed that \textit{Ace2} and \textit{Tmem27} transcriptions were positively regulated by HNF1b co-transfection. However, no significant activation of the \textit{Slc6a19} promoter by HNF1b was observed. These results indicating that despite co-expression of these trafficking proteins with B0AT1, their transcriptions are controlled by different regulatory elements in the intestine and kidney.

Although \textit{Slc6a19} and \textit{Ace2} are in part trans-activated by different regulatory elements, B0AT1 is co-expressed with the trafficking proteins in the intestine and kidney. These results may point to yet unknown TF(s), which help to coordinate the observed overlapping expression patterns”

4.3.4 Summary

The results from this study demonstrate a possible mechanism of the transcriptional regulation of the \textit{Slc6a19} gene in the crypt–villus axis. \textit{In vitro} and \textit{in vivo} analyses revealed that HNF1a, HNF4a and SOX9 bind and orchestrate regulation of \textit{Slc6a19} promoter in the mouse intestine.
5.1 Introduction

In living cells, DNA is compacted as chromatin. The chromatin structure is an important regulator of the interaction of transcriptional factors with target regulatory DNA regions (Heintzman et al., 2007). Especially in differentiating cells, chromatin changes have been shown to play a central role in the regulation of transcription (Schübeler et al., 2004). Modifications on histone tails such as, methylation, acetylation and phosphorylation can affect transcriptional regulation (Figure 5.1).

![Possible positions of histone tail modification](image)

Figure 5.1 Possible positions of histone tail modification

Locations and types of well-known histone modifications are shown in the figure. Numbers under each amino acid residue represent location of a modification. A = acetylation, P = phosphorylation, U = ubiquitination, M = methylation. From (Portela and Esteller, 2010).
Acetylation of histone H3 tails is generally accompanied by transcriptional activation (Allfrey et al., 1964). For instance, acetylation of histone H3 tail at lysine 27 (H3K27ac) has been shown to correlate with transcription activation (Heintzman et al., 2007). However, methylation of H3 residues and its effect depend on the position. For example, methylation of the H3K36 histone tails (H3K36me3) leads to repression of transcription (Azuara et al., 2006). While, tri-methylation of the H3 histone tails at lysine 4 (H3K4me3) causes gene activation (Santos-Rosa et al., 2002).

DNA methylation and histone modification together play a critical role in the regulation of transcriptional activity. Recent studies suggest a close relationship between these two epigenetic modifications. As an example, DNA methylation can guide histone modifications, and vice versa histone modifications can affect the methylation status (Hayashi, Nagae et al. 2007). However, the mechanisms underlying these interactions are not well understood.

5.1.1 DNA methylation

DNA-methyltransferases (DNMT) are responsible for the transfer of the methyl group to cytosine nucleotides. Four major methyltransferases have been identified so far (DNMT1, DNMT3a, DNMT3b and DNMT3L) (Okano et al., 1999).

DNMT3a and DNMT3b are vital for embryonic development and de novo methylation in mammals (Chen et al., 2003). As a result, these enzymes need to access non-methylated CpG regions. Some histone modifications and specific DNA sequences can potentially initiate de novo methylation via recruitment of DNMT3a and DNMT3b (Klose and Bird, 2006).

DNMT3L lacks methyltransferase activity alone, but it has a stimulatory effect on the methylase activity of DNMT3a and DNMT3b (Suetake et al., 2004). DNMT3L is crucial for maternal imprinting at the gametogenesis stage (Bourc'his et al., 2001).

DNMT1 (maintenance methyltransferase) is the most common methyltransferases and the main enzyme for maintaining methylation during cell division (Richardson and Yung, 1999). DNMT1 has a high affinity for hemimethylated DNA strands during the S-phase of mitosis, thereby, transferring the methylation pattern to the newly synthesised DNA strand (Figure 5.2).
While methylation is well understood, demethylation remains to be fully described. During vertebrate development, DNA is globally demethylated during two stages of embryonic development. First, demethylation happens during germ cell development. Second, demethylation happens after fertilisation and implantation of an embryo. During the replication process, failing of DNMT1 function causes passive demethylation (Razin and Riggs, 1980). Active demethylation, by contrast, occurs during DNA repair processes by specific enzymes, such as Gadd45 (Barreto et al., 2007).
Figure 5.2. DNA methylation and demethylation

After global DNA demethylation, *de novo* methylation is re-established by DNMT3A and DNMT3B. Following cell division, this methylation pattern is preserved by DNMT1. Inhibition of DNMT1 during replication causes loss of methylation on the new DNA strand, this is called passive demethylation. Whereas, removal of methyl groups by enzymatic reaction is called active demethylation. From (Wu and Zhang, 2010).

5.1.2 DNase I hypersensitive sites

Deoxyribonuclease I (DNase I) is an endonuclease that catalyses cleavage of unprotected DNA. In the chromatin context, nucleosome-free DNA is more sensitive to fragmentation by DNase I; these regions are called DNase I hypersensitive sites. An average DNase I sensitive region is around 100 to 200 bp long and is a nucleosome-free DNA fragment (Figure 5.3). This region, most of the time, contains transcriptional
regulators. Accordingly, searching for DNase I hypersensitive sites on genomic material is crucial to determine regulatory elements of transcription.

Figure 5.3 DNase I hypersensitive site in the chromatin context.
An example of a DNA fragment, showing DNase I hypersensitive sites (red arrows) around TSS (blue arrow) and promoter region, and protected areas of sequence. From (Bell et al., 2011).

5.2 Results

5.2.1 Analysis of H3K27ac and H3K4me3 histone modification associated with the 5’ flanking region of Slc6a19 promoter

To examine whether histone modifications are associated with the Slc6a19 gene expression, ChIP assays were performed using two antibodies, one against acetylated histone H3 at residue K27 and the other against methylated histone H3 at residue K4.

H3K27ac is typically associated with gene activation (Heintzman et al., 2009). To verify whether H3K27ac is found in the vicinity of the Slc6a19 promoter, ChIP was performed using cross-linked chromatin isolated from intestinal villus and crypt enterocytes using the fractionation protocol (see Section 2.7.1.1). The intestinal fractions and liver were immunoprecipitated using an antibody against H3K27ac. Immunoprecipitated DNA was analysed by real-time PCR generating amplicons of 150-200 bp spanning the Slc6a19 promoter region from position –858 to +11. ChIP experiments demonstrated that acetylation of lysine 27 of histone H3 (H3K27ac) was significantly enriched in chromatin from villus tips compared to crypt and liver chromatin (Figure 5.4).
H3K27ac histone modification of the *Slc6a19* promoter region in liver and intestine

A H3K27ac-specific antibody was used in the ChIP assay. The genomic DNA was isolated from mouse villus and crypt cells, and from liver tissue. Data are given in percent signal relative to a real-time PCR reaction on input DNA. Primer pairs span the following promoter regions: \(-139/+11\), \(-501/-332\), \(-682/-502\) and \(-858/-674\). The data are means ± SD of triplicate experiments.

H3K4me3 is known to be associated with open chromatin regions (Santos-Rosa et al., 2002). ChIP assays were used to test whether H3K4me3 was bound to the promoter regions within living cells. The ChIP assay was performed using the chromatin isolated from intestinal fractionation, and from liver tissue. The enterocytes chromatin was immunoprecipitated using an H3K4me3 antibody. Real-time PCR was used to determine the localisation of the modified histones along the *Slc6a19* promoter. ChIP indicated that H3K4me3 was significantly enriched in the intestine and more than that in the liver (Figure 5.5).
Figure 5.5 H3K4me3 histone modification of the Slc6a19 promoter region in liver tissue and crypt–villus fractions

An H3K4me3-specific antibody was used in the ChIP assay. The Genomic DNA was isolated from mouse intestinal fractions and liver tissues. Data are given in percent signal relative to a real-time PCR reaction on input DNA. Primer pairs span the following promoter regions: –139/+11, –501/–332, –682/–502, –858/–674. The data are means ± SD of triplicate experiments.

5.2.2 Modifications of H3K4m3 and H3K36m3 in the N-terminus of histone H3

The ENCODE project (Rosenbloom et al., 2010) was completed by a worldwide consortium. The aim was to identify regulatory regions of the human genome in tissues and in more than 100 different cell lines. In addition to our ChIP experiments, the Encyclopedia of DNA Elements (ENCODE) database on the UCSC webpage (http://genome.ucsc.edu) was used to identify positioning of other histone modifications along the Slc6a19 gene.

The ENCODE data indicated that the histone modification H3K4me3 was highly enriched over the promoter in kidney and small intestine, allowing binding of polymerase
II. Additionally, trimethylation of Lys36 in histone H3 (H3K36me3) was present over all exons confirming that the gene was actively transcribed. These two elements were absent in other tissues (Figure 5.6).
Figure 5.6. Histone modifications and polymerase II binding to the \textit{Slc6a19} promoter

The ENCODE database was used to extract the distribution of histone modifications H3K4me3, H3K36me3, and RNA polymerase II binding along the mouse \textit{Slc6a19} gene in different tissues. The \textit{Slc6a19} gene shows active histone status and polymerase II binding in the small intestine and kidney but not in the liver.
5.2.3 DNase I hypersensitive sites on the *Slc6a19* promoter

To localise possible TF-binding sites on the *Slc6a19* gene, DNase I hypersensitive sites were identified using ENCODE data. Two sensitive regions were identified in the mouse *Slc6a19* promoter (Figure 5.7). The first hypersensitive fragment was located from +75 to –220, consistent with the core promoter region. The second hypersensitive region was located between position –1000 to –1200. The results suggested that the liver chromatin around the *Slc6a19* promoter region was more condensed than the kidney chromatin. This explains RNA polymerase may bind to DNA in the extended chromatin of kidney. ENCODE data does not supply any information about DNase I hypersensitive status of the small intestine in the mouse.

![Figure 5.7 DNase I hypersensitive sites on the mouse the *Slc6a19* promoter](image)

The histogram shows that two separate DNase I sensitive regions located around 1 kb upstream of the *Slc6a19* promoter in the kidney chromatin. In liver chromatin, no DNase I sensitive regions were identified in the same region.

5.2.4 DNA methylation profile of the *Slc6a19* promoter in mouse

In addition to showing the effects of histone modifications on the transcriptional activity, we also analysed the methylation status changes of the promoter in different tissues. Analysis of a region 1200 bp upstream of the *Slc6a19* TSS revealed 20 CpG dinucleotides (Figure 5.8). Ten CpGs were located between (–171 to +56), which covers binding sites for transcription factor SOX9, HNF1a, HNF4a and the TATA-binding protein. However, no CpG island was detected by UCSC Genome Browser (http://genome.ucsc.edu).
Locations of 20 CpG sites in the 1.2 kb upstream of the Slc6a19 promoter was examined. The CpG dinucleotides (bold letters), putative TF-binding positions and TSS are indicated on the promoter.

To examine DNA methylation effects on the Slc6a19 promoter, liver and kidney tissues were isolated from B6 mice as described in Section 2.8.4. Following DNA isolation, the methylation status of the target region was analysed using the bisulphite sequencing method. As shown in Figure 5.9, almost all CpG sites in the 5'-flanking sequence of the Slc6a19 gene were methylated in both tissues.
Figure 5.9 Methylation status of the Slc6a19 promoter in the liver and kidney

Genomic material was isolated from the mouse kidney and liver. The methylation pattern of each tissue was analysed by bisulphite conversion and sequencing methods. Twenty different CpG locations (closed circle, methylated; open circle, unmethylated) and TF-binding sites (TATA, SOX9, HNF1a and HNF4a) are indicated.

To investigate the methylation status of the Slc6a19 promoter during cell differentiation along the crypt–villus axis, the fractionation method was used. Only three of the fractions were analysed, namely F1 (villus), F4 (crypt/villus), and F5 (crypt) (Figure 5.10). Notably, DNA methylation around the core promoter region, spanning from +56 to –171, gradually decreased from crypt–villus cell fractions.
Figure 5.10 Methylation status of the Slc6a19 promoter region in the intestine

Genomic material was isolated from three different intestinal cell fractions. The methylation pattern of each fraction was analysed by the bisulphite conversion method. CpG locations (closed circle, methylated; open circle, unmethylated) and TF-binding sites (TATA, SOX9, HNF1a and HNF4a) are indicated.

5.3 Discussion

Chromatin changes during transcription are controlled by histone modification and DNA methylation (Schubeler et al., 2004). Histone modifications are one of the major epigenetic mechanisms that alter gene expression. Generally, promoter regions of active genes have hyperacetylated histones and non-methylated CpG dinucleotides (Vaissière et al., 2008).

ChIP assays were performed to determine histone modification profiles of the Slc6a19 promoter sequence. The results revealed that there was a significant difference in H3K27ac levels between crypt and villus. Acetylated histone H3K27 was highly enriched in the Slc6a19 promoter in villus enterocytes. Similarly, the sucrase-isomaltase gene (SI) was shown to be regulated by histone acetylation in the crypt–villus axis (Suzuki et al., 2008). Acetylation of histone H3 (lysine 9 and 14) was found to be highly associated with active SI gene expression in the villus.
The general transcription factor TFIID can directly interact with H3K4me3 (van Ingen et al., 2008). Therefore, enrichment of H3K4me3 in a promoter region is crucial for initiating transcription. Studies in different species have showed that H3 acetylation occurs before H3K4 methylation (Rice and Allis, 2001). In contrast to this finding, our ChIP assay indicated that in crypts before histone acetylation started, H3K4me3 was highly enriched at the promoter. How histone modification affect one another is still debated (Maltby et al., 2012).

The ENCODE database results also supported expression of the Slc6a19 gene in intestine and kidney. H3K4me3 levels were significantly higher in the intestine and kidney. Another modification H3K36me3, which is similarly found over active promoters, confirmed that the gene is actively transcribed in the intestine and kidney. However, molecular mechanisms underlying activation of genes by H3K27ac and H3K4me3 are still unclear.

In addition to histone modification, methylation of cytosine residues in genomic DNA is an important epigenetic factor that can contribute to gene silencing (Hayashi et al., 2007). Our findings indicated that DNA methylation was consistent with expression in the intestine, showing a highly methylated proximal promoter in the crypt, but complete demethylation in villus fractions. This result was contrary to a previous study (Kikuchi et al., 2010) showing differential methylation at position 1080 upstream of the TSS in intestinal cells. Our bisulphite sequencing data, by contrast, demonstrate differential methylation at the core promoter of the Slc6a19 gene, which is <300 bp long and where binding of TFs and RNA polymerase II occurs.

Surprisingly, no significant difference was found between DNA methylation patterns in liver and kidney. The Slc6a19 promoter was highly methylated in both tissues. One possible reason explaining the unexpected hypermethylation in the kidney is the fraction of Slc6a19 expressed in the epithelial cells of the proximal tubules compared to other kidney cells. The methylation pattern of the Slc6a19 core promoter region is likely to contribute to gene activation along the crypt–villus axis. The mechanism of DNA demethylation during enterocyte differentiation from crypt to villus remains to be determined. It could involve passive demethylation during cell mitosis in the crypt region or could be mediated by active demethylation.
5.4 Summary

In this chapter, I investigated whether histone modification and DNA methylation contribute to specific expression of the B^0AT1 protein in intestine and kidney. I defined histone marks and DNA methylation patterns of the Slc6a19 promoter along the crypt–villus axis and in kidney cells. The results showed that coordination of histone acetylation–methylation activities with CpG dinucleotide methylation plays important roles in the differential expression of the gene.
Chapter 6  General discussion

The crypt–villus structure has interested many researchers because it is an ideal model for conveniently studying rapid differentiation of intestinal cells. As such, undifferentiated cells are located in the crypt which has been described as a perfect model for studying of stem cell niches (Clevers, 2013). After completing differentiation in the upper third of the crypt, cells gradually move up towards the villus tips. Explicit interplays between TFs and epigenetic regulatory mechanisms together control intestinal cells maintenance and differentiation. The Slc6a19 is a perfect model for studying intestinal gene differentiation because it is only expressed in differentiated enterocytes in the crypt–villus axis.

6.1 Links between DNA methylation and transcription

Interactions between HNF1a and HNF4a have been well studied. The HNF1a promoter region includes conserved DNA-binding sites for HNF4a, and vice versa (Odom et al., 2004). Thus, HNF1a or HNF4a increase each other’s transcription. These two TFs play important roles in the regulation of Slc6a19 and some transporter genes. For example, HNF1a and HNF4a function as an important regulator controlling expression of many epithelial transporters (Kikuchi et al., 2010, Desvergne et al., 2006). In the kidney, HNF1a is expressed only in the proximal tubules (Pontoglio et al., 1996). This expression pattern mirrors that of the organic anion transporter 1 (OAT1) in mouse or human kidney. OAT1 is highly expressed in the renal proximal tubules. Oat1 expression in renal proximal tubules is regulated by HNF1a and HNF4a (Maher et al., 2006). However, in extrarenal tissues, DNA methylation of the transporter promoter regions inhibits gene expression. Further, demethylated DNA in the promoter region of Oat1 is crucial for kidney-specific expression (Kikuchi et al., 2007). Thus, HNF1a, HNF4a and DNA methylation act in concert to control organic anion transporters in a tissue-specific manner in the mouse or human kidney.

DNA methylation was long considered as a stable epigenetic mark. Inactivation of the X chromosome by DNA methylation during embryonic development was one of the first findings proving that DNA methylation was a dynamic epigenetic factor regulating gene expression (Mohandas et al., 1981). Moreover, genome-wide DNA
methylation studies have shown that DNA methylation is modified during the differentiation processes at relevant gene promoters (Ziller et al., 2013, Feldmann et al., 2013).

The DNA methylation status around the Slc6a19 core promoter region decrease during differentiation. The methylated Slc6a19 promoter in the crypt cells becomes fully demethylated after three to four cell divisions (Radtke and Clevers, 2005). Similarly, hormone-sensitive promoters and neuron-specific gene promoters have been found to undergo rapid demethylation (Ma et al., 2009b, Métivier et al., 2008). DNA demethylation could occur by two different mechanisms; first by lack of DNMT1 activity during mitosis, and second active demethylation by the DNA excision repair machinery (Ma et al., 2009a). We suggest that both mechanisms may be involved in the demethylation processes along the crypt–villus axis. Thus, DNMT1 is highly expressed in the crypt region while its expression decreases towards the villus. Ablation of DNMT1 in the intestine leads to expansion of crypts (Sheaffer et al., 2014). This finding indicates that DNA methylation is crucial in balancing cell proliferation and differentiation along crypt–villus axis.

Other reports have also demonstrated that many important intestinal TFs are controlled by epigenetic modifications. For example, DNA methylation analyses showed that the Lgr5 promoter was demethylated along the crypt–villus axis. Lgr5 is an intestinal stem cell marker. Lgr5 transcription is regulated by H3K4me3 modification, which are active in the crypts but repressed in the villi (Vincent et al., 2015). DNA methylation studies have also shown that complete DNA demethylation of SOX9 promoter occurs in the crypts, whereas it becomes gradually methylated in the villus cells except for goblet cells where SOX9 is highly expressed (Vincent et al., 2015). SOX9 expression has also been shown to be mainly regulated by the histone modification H3K4me2 in the cartilage tissues (Zhang et al., 2015).

CDX2 is one of the major transcription factors that controls cell differentiation in the intestine. Its expression is necessary for activation of other essential intestinal TFs such as, HNF4, HNF1a and CDX1. Epigenetic studies have revealed that CDX2 interplays between TFs and epigenetic modifications. CDX2 regulates histone modifications by triggering the formation of an active chromatin enhancer, H3K4me3 (Verzi et al., 2013). DNA Methylation status of the Hnf1a promoter in the liver, kidney and brain was investigated previously (Kikuchi et al., 2010). It was reported that except
in the brain, the *Hnf1a* promoter was hypomethylated in the liver and kidney where the gene was highly expressed. All these studies revealed that TFs which are essential regulators of transporter genes were also controlled by epigenetic factors.

### 6.2 The current model

Taken together, my studies propose a model ([Figure 6.1](#)) for the regulation of *Slc6a19* transcription along the crypt–villus axis. Hypermethylated CpG dinucleotides and deacetylated histone molecules cause condense chromatin state in the crypt and liver ([Figure 6.1.A](#)). DNA methylation affects expression of target genes by two mechanisms. First, methylated DNA directly interferes with TFs binding and prevents gene expression. Second, methylated cytosine residues can be recognised by methyl-CpG-binding domain (MBD) proteins which attracts chromatin remodelers to change nucleosome occupancy of TFs which in turn causes transcriptional inhibition (Salozhin et al., 2005). The inhibition mechanism of MBD proteins is not well understood. Studies showed that MBDs can interact with histone deactylases, which leads to a more condense chromatin status (Ballestar and Wolffe, 2001). Except position –243 other CpG residues were completely methylated in the core promoter region. Additionally, we located TF binding sites for HNF1a, HNF4a and SOX9 within a 250 bp region upstream of the TSS. We found eight differentially methylated CpG dinucleotides within the same region. DNA methylation at position –243 does coincide with the SOX9 binding site. This may allow direct SOX9 binding at the predicted position and lead to *Slc6a19* silencing in the crypts. Thus, SOX9-induced inhibition, hypermethylated DNA and deactylated histones around the core promoter region may lead to more condensed chromatin structures, which can inhibit the transcriptional machinery and silence the gene in the crypt and the liver.
Figure 6.1 Working model for regulation of the Slc6a19 along crypt–villus axis

The Slc6a19 promoter region in the crypt and liver tissues showing condensed chromatin conformation (A), due to methylated CpG dinucleotides and deacetylated histone molecules. By contrast, demethylation of CpG dinucleotides and histone acetylation (on the Slc6a19 promoter) allows binding of HNF1a and HNF4a TFs to activate gene expression in the villus (B). Differentially methylated CpG locations (red circle, methylated; white circle, unmethylated), acetylated histones and TFs (SOX9, HNF1a and HNF4a) are indicated.

During differentiation, demethylation of DNA and histone acetylation within the Slc6a19 promoter lead open chromatin state in the villus (Figure 6.1.B). We speculated that an active DNA demethylation mechanism could involve in the Slc6a19 activation.
Our study discovered that the marker H3K4me3 for active histones was already enriched within the hypermethylated promoter sequences. Methylation of CpG residues can change histone modifications and nucleosome distribution around the targeted promoter regions (Cedar and Bergman, 2009). Inhibitor of growth protein 1 (Ing1), which is a tumor suppressor gene and expressed in the intestine and epidermis (Pena et al., 2008), recognises and binds to H3K4me3 chromatin if this modification is enriched at methylated DNA regions. Binding of Ing1 to a methylated H3K4 recruits Gadd45a, which is a crucial player for active DNA demethylation, and triggers DNA excision repair mechanisms for DNA demethylation (Schäfer et al., 2013). Therefore, H3K4me3 histone modification may trigger active DNA demethylation during early differentiation phases. In addition to active demethylation, passive DNA demethylation may occur by lack of DNMT1 after mitosis. Furthermore, histones in the promoter receive acetylated H3K27 tails, thus increasing accessibility of the promoter region for transcriptional regulatory elements in the villus.

### 6.3 Limitations and future studies

There are some limitations to my study. We have conducted many fractionation experiments to use in ChIP, microarray and bisulphite sequencing assays. An alternative method could have been laser microdissection. However, instead of using laser microdissection, we had to use fractionation experiments to obtain sufficient amounts of DNA, and to keep fractionated samples consistent among different experiments. Furthermore, because of inherent limitations in the crypt–villus fractionation technique, our crypt fractions may not have included the cells from the very bottom of the crypt. Fractions may have also been contaminated with villus cells, but the crypt fractions were still highly enriched with undifferentiated cells originating from the crypt differentiation niches. The limitations of fractionation assays may have been compensated by studies using cell lines models of intestinal enterocytes. Caco-2, T84 and HT29 cells are resembling mature enterocytes and have been used as a cellular model of enterocytes for transport studies (Bolte et al., 1997). Ablating positive and negative regulating TF genes in these cell lines may help to support the TF effects.

Despite many attempts, we could not achieve a successful ChIP assay to immunoprecipitate TFs from fractionated intestinal cells likely due to insufficient binding of the antibodies used. This missing experiment would have further clarified the TFs’
roles in the crypt–villus cells’ differentiation. Also, further investigations should focus on genome-wide DNA methylation and histone modification studies along the crypt–villus axis to shed light on the crosstalk between genetic and epigenetic changes during intestinal differentiation.

6.4 Summary

In my PhD project, I investigated the mechanism of Slc6a19 transcriptional regulation by genetic and epigenetic mechanisms. The histone modification and DNA methylation status in the Slc6a19 core promoter region might be key reasons for lack of gene transcription in the crypt and in the liver tissue. Also, this thesis proved that multiple modifications act in combination to regulate Slc6a19 transcription dynamically along the crypt–villus axis.

In conclusion, this study contributes to our understanding of the transcriptional regulation of intestinal genes and intestinal cell differentiation. However, there is still intense debate on hierarchy of interactions between genetic and epigenetic factors driving histone modifications, regulating gene expression or repression. Although researchers have made remarkable contributions to the understanding of relationships between TFs and histone modifications, it is still unclear in which order epigenetic modifications change during intestinal cell differentiation.


ELLARD, S. & COLCLOUGH, K. 2006. Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity-onset diabetes of the young. *Hum Mutat*, 27, 854-69.


London, Garland Science ;

Taylor & Francis [distributor].


RUSSELL, P. J. 2002. iGenetics, San Francisco, Benjamin Cummings.


SAUVAGET, D., CHAUFFETON, V., CATADELLE, D., CHATELET, F. P., CYWINER-GOLENZER, C.,
Restriction of apolipoprotein A-IV gene expression to the intestine villus depends on a
hormone-responsive element and parallels differential expression of the hepatic
nuclear factor 4alpha and gamma isoforms. The Journal of biological chemistry, 277,
34540-8.

SCHÄFER, A., KARaulANOV, E., STAFF, U., DÖDERLEIN, G. & NIEHRS, C. 2013. Ing1 functions in
DNA demethylation by directing Gadd45a to H3K4me3. Genes & development, 27, 261-
273.

SCHÜBELER, D., MACALPINE, D. M., SCALZO, D., WIRBELAUER, C., KOOPERBERG, C., VAN
The histone modification pattern of active genes revealed through genome-wide
chromatin analysis of a higher eukaryote. Genes & development, 18, 1263-1271.

SCHÜBELER, D., MACALPINE, D. M., SCALZO, D., WIRBELAUER, C., KOOPERBERG, C., VAN
LEEUVEN, F., GOTTSCHLING, D. E., O’NEILL, L. P., TURNER, B. M., DELROW, J., BELL, S. P.
& GROUDINE, M. 2004. The histone modification pattern of active genes revealed
through genome-wide chromatin analysis of a higher eukaryote. Genes Dev, 18, 1263-
71.

Gastroenterology, 134, 849-64.

2004. Hartnup disorder is caused by mutations in the gene encoding the neutral amino

SERFAS, M. S. & TYNER, A. L. 1993. HNF-1 alpha and HNF-1 beta expression in mouse intestinal

SHEAFFER, K. L., KIM, R., AOKI, R., ELLIOTT, E. N., SCHUG, J., BURGER, L., SCHÜBELER, D. &
KAESTNER, K. H. 2014. DNA methylation is required for the control of stem cell

SHIAU, Y.-F. 1987. Lipid digestion and absorption. Physiology of the gastrointestinal tract, 2,
1527-1556.

SHIH, D. Q., SCREENAN, S., MUNOZ, K. N., PHILIPSON, L., PONTOGLIO, M., YANIV, M., POLONSKY,
K. S. & STOFFEL, M. 2001. Loss of HNF-1alpha function in mice leads to abnormal
expression of genes involved in pancreatic islet development and metabolism. Diabetes,
50, 2472-80.

SILK, D. B., GRIMBLE, G. K. & REES, R. G. 1985. Protein digestion and amino acid and peptide

HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes &
Development, 4, 2353-2365.


intestinal epithelial progenitors. Proceedings of the National Academy of Sciences, 100,
1004-1009.

STEGMANN, A., HANSEN, M., WANG, Y., LARSEN, J. B., LUND, L. R., RITIE, L., NICHOLSON, J. K.,
transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central
regulator of gene expression during enterocyte differentiation. Physiol Genomics, 27,
141-55.

acids in rabbit jejunal brush border vesicles. The Journal of membrane biology, 66, 213-
25.


Chapter 8 Appendix

8.1 Plasmids

Figure 8.1: A plasmid map of the pGL4, the basic luciferase reporter vector, detailing key constructs and restriction enzyme sites
Figure 8.2: A plasmid map of the pGL4, the basic luciferase reporter vector, detailing key constructs and restriction enzyme sites
Figure 8.3: A plasmid map of the pcDNA3.1

pcDNA3.1 (+) 5428/5427 bp
8.2 Thermal cycler conditions

Table 8.1: Cycling conditions for the *PfuUltra* enzyme

<table>
<thead>
<tr>
<th>Process</th>
<th>Temperature (°C)</th>
<th>Time (min:sec)</th>
<th>Number of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>95</td>
<td>5:00</td>
<td>–</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95</td>
<td>0:30</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>50–58 (depending on <em>T</em>&lt;sub&gt;m&lt;/sub&gt; of primers)</td>
<td>0:30</td>
<td>30</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>6:00</td>
<td></td>
</tr>
<tr>
<td>Extension (Final)</td>
<td>72</td>
<td>10:00</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 8.2: Cycling condition for the *Phusion*<sup>®</sup> high-fidelity enzyme

<table>
<thead>
<tr>
<th>Process</th>
<th>Temperature (°C)</th>
<th>Time (min:sec)</th>
<th>Number of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>98</td>
<td>0:30</td>
<td>–</td>
</tr>
<tr>
<td>Denaturation</td>
<td>98</td>
<td>0:10</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>50–58 (depending on <em>T</em>&lt;sub&gt;m&lt;/sub&gt; of primers)</td>
<td>0:30</td>
<td>30</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>2:00</td>
<td></td>
</tr>
<tr>
<td>Extension (Final)</td>
<td>72</td>
<td>5:00</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 8.3: Cycling condition for ChIP PCR

<table>
<thead>
<tr>
<th>Process</th>
<th>Temperature (°C)</th>
<th>Time (min:sec)</th>
<th>Number of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>5:00</td>
<td>–</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>0:30</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>50–58 (depending on <em>T</em>&lt;sub&gt;m&lt;/sub&gt; of primers)</td>
<td>0:30</td>
<td>35–40</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>1:00</td>
<td></td>
</tr>
<tr>
<td>Extension (Final)</td>
<td>72</td>
<td>5:00</td>
<td>–</td>
</tr>
</tbody>
</table>
8.3 Gene ontology terms enriched in the villus (significant changes in expression of genes between crypt and villus)

Biological Process/ Organic Substance Transport

Adjusted p value : 9.146E-5
p value : 8.728E-8
Total genes in GO class : 392
Genes in input list in GO class : 44

Biological Process/ Na+ Transport

Adjusted p value : 2.626E-4
p value : 2.94E-7
Total genes in GO class : 145
Genes in input list in GO class : 23

Biological Process/ Organic Acid Transport

Adjusted p value : 0.003
p value : 4.44E-6
Total genes in GO class : 181
Genes in input list in GO class : 24

Biological Process/ Nitrogen Compound Tr.

Adjusted p value : 0.003
p value : 5.341E-6
Total genes in GO class : 147
Genes in input list in GO class : 21
Biological Process/ Lipoprotein Transport

Adjusted p value : 0.049
p value : 1.358E-4
Total genes in GO class : 12
Genes in input list in GO class : 5

Biological Process/ Carboxylic Acid Transport

Adjusted p value : 0.002
p value : 3.65E-6
Total genes in GO class : 179
Genes in input list in GO class : 24

Cellular Component/ Brush Border

Adjusted p value : 0.002
p value : 4.03E-6
Total genes in GO class : 61
Genes in input list in GO class : 13

Molecular Function/ Symporter Activity

Adjusted p value : 2.883E-4
p value : 3.281E-7
Total genes in GO class : 114
Genes in input list in GO class : 20
**Molecular Function/ Cofactor Transporter Activity**

Adjusted p value : 0.035  
*p value : 8.24E-5*  
Total genes in GO class : 11  
Genes in input list in GO class : 5

**Molecular Function/ Carboxylic Acid Transmembrane Transporter Activity**

Adjusted p value : 0.049  
*p value : 1.363E-4*  
Total genes in GO class : 95  
Genes in input list in GO class : 14

**Molecular Function/ Solute: Sodium Symporter Activity**

Adjusted p value : 8.926E-4  
*p value : 1.229E-6*  
Total genes in GO class : 47  
Genes in input list in GO class : 12
8.4 Gene ontology terms enriched in the crypt (significant changes of genes between crypt and villus)

**Biological Process/ M phase of mitotic cell cycle**

- Adjusted p value: $3.434E^{-28}$
- p value: $9.526E^{-33}$
- Total genes in GO class: 216
- Genes in input list in GO class: 63

**Biological Process/ Cell cycle**

- Adjusted p value: $4.478E^{-24}$
- p value: $5.753E^{-28}$
- Total genes in GO class: 787
- Genes in input list in GO class: 115

**Biological Process/ DNA Replication**

- Adjusted p value: $4.658E^{-5}$
- p value: $4.36E^{-8}$
- Total genes in GO class: 152
- Genes in input list in GO class: 25

**Biological Process/ DNA Repair**

- Adjusted p value: 0.002
- p value: $2.801E^{-6}$
- Total genes in GO class: 291
- Genes in input list in GO class: 33
Molecular Function/ Microtubule Motor Activity

Adjusted p value : 0.013
p value : 2.881E-5
Total genes in GO class : 44
Genes in input list in GO class : 10

Cellular Component/ Microtubule Cytoskeleton

Adjusted p value : 2.157E-7
p value : 1.188E-10
Total genes in GO class : 627
Genes in input list in GO class : 68

Cellular Component/ Nuclear Part

Adjusted p value : 0.007
p value : 1.405E-5
Total genes in GO class : 1745
Genes in input list in GO class : 120

Cellular Component/ Chromosome

Adjusted p value : 2.443E-11
p value : 7.62E-15
Total genes in GO class : 498
Genes in input list in GO class : 67
Biological Process/ Chromosome Segregation

Adjusted p value : 1.805E-12
p value : 5.3E-16
Total genes in GO class : 105
Genes in input list in GO class : 30

Table 8.4: Predicted TF-binding sites in the Collectrin promoter.

A matrix number of 1.00 indicates an exact match with the binding site consensus sequence. Position and the core binding site of the TFs are indicated.

<table>
<thead>
<tr>
<th>TFs</th>
<th>Position</th>
<th>Strand</th>
<th>Matrix</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNF1b</td>
<td>–706/–690</td>
<td>–</td>
<td>0.866</td>
<td>taacaATTTatcaagcg</td>
</tr>
<tr>
<td></td>
<td>–711/–695</td>
<td>+</td>
<td>0.797</td>
<td>gataaATTGtactgat</td>
</tr>
<tr>
<td>HNF1a</td>
<td>–56/–40</td>
<td>+</td>
<td>0.948</td>
<td>tccaaattattaATTGc</td>
</tr>
<tr>
<td></td>
<td>–58/–42</td>
<td>–</td>
<td>0.910</td>
<td>ccgcaattaattATTTGg</td>
</tr>
<tr>
<td></td>
<td>–77/–61</td>
<td>–</td>
<td>0.849</td>
<td>gcataATTGgttgact</td>
</tr>
<tr>
<td></td>
<td>–84/–68</td>
<td>–</td>
<td>0.836</td>
<td>gataacctgcatATTGg</td>
</tr>
<tr>
<td></td>
<td>–235/–219</td>
<td>–</td>
<td>0.889</td>
<td>atcgaATTGgttggttg</td>
</tr>
<tr>
<td></td>
<td>–299/–283</td>
<td>+</td>
<td>0.858</td>
<td>gtccgATTGgttagact</td>
</tr>
<tr>
<td></td>
<td>–401/–385</td>
<td>–</td>
<td>0.931</td>
<td>taacaATTGgttggttg</td>
</tr>
<tr>
<td></td>
<td>–406/–390</td>
<td>+</td>
<td>0.882</td>
<td>aaccATTGtttaagggt</td>
</tr>
<tr>
<td></td>
<td>–408/–392</td>
<td>–</td>
<td>0.865</td>
<td>caacccttaacaATTGg</td>
</tr>
<tr>
<td></td>
<td>–431/–415</td>
<td>–</td>
<td>0.857</td>
<td>ctcaaATTGgatcggtga</td>
</tr>
<tr>
<td></td>
<td>–659/–643</td>
<td>–</td>
<td>0.928</td>
<td>gttgaATTGatttactg</td>
</tr>
<tr>
<td>HNF4a</td>
<td>–284/–260</td>
<td>+</td>
<td>0.839</td>
<td>aaccgtattagAACcgggaaggt</td>
</tr>
</tbody>
</table>
Table 8.5: Predicted TF-binding sites in the Ace2 promoter.

A matrix number of 1.00 indicates an exact match of the binding site consensus sequence. Position and the core binding site of the TFs are indicated.

<table>
<thead>
<tr>
<th>TFs</th>
<th>Position</th>
<th>Strand</th>
<th>Matrix</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNF1a</td>
<td>–189/–173</td>
<td>–</td>
<td>0.919</td>
<td>cagtaATTGctcaagt</td>
</tr>
<tr>
<td></td>
<td>–196/–180</td>
<td>–</td>
<td>0.802</td>
<td>gaaaaagcagtaATTGc</td>
</tr>
<tr>
<td></td>
<td>–213/–197</td>
<td>–</td>
<td>0.853</td>
<td>agacaATTGtagaataa</td>
</tr>
<tr>
<td></td>
<td>–218/–202</td>
<td>+</td>
<td>0.894</td>
<td>ctacaATTGtctgccca</td>
</tr>
<tr>
<td></td>
<td>–462/–446</td>
<td>+</td>
<td>0.860</td>
<td>ttaaaATTGctttggag</td>
</tr>
<tr>
<td></td>
<td>–770/–754</td>
<td>+</td>
<td>0.829</td>
<td>aatttatattttatATTTt</td>
</tr>
<tr>
<td></td>
<td>–814/–798</td>
<td>+</td>
<td>0.824</td>
<td>tagctgtctttgATTGg</td>
</tr>
<tr>
<td></td>
<td>–1222/–1206</td>
<td>+</td>
<td>0.845</td>
<td>agcttATTGatagaatt</td>
</tr>
<tr>
<td>HNF4a</td>
<td>–234/–210</td>
<td>–</td>
<td>0.778</td>
<td>cgctttattcTAAAacctgggcagac</td>
</tr>
<tr>
<td></td>
<td>–681/–657</td>
<td>+</td>
<td>0.920</td>
<td>atgactgtcttgAAAAtttaaccaag</td>
</tr>
<tr>
<td></td>
<td>–1377/–1353</td>
<td>–</td>
<td>0.770</td>
<td>tggccaaatcAAAAtctggacatt</td>
</tr>
</tbody>
</table>
Table 8.6: List of TFs which were significantly expressed along crypt–villus axis (p < 0.05)

### TFs expressed in the villus (Fold change >3)

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ev2</td>
<td>18.</td>
<td>Fosl2</td>
<td>35.</td>
<td>Maf</td>
<td>52.</td>
</tr>
<tr>
<td>2.</td>
<td>Pou2f2</td>
<td>19.</td>
<td>Mta3</td>
<td>36.</td>
<td>Epa1</td>
<td>53.</td>
</tr>
<tr>
<td>9.</td>
<td>Mafb</td>
<td>26.</td>
<td>Pax3</td>
<td>43.</td>
<td>Obfc1</td>
<td>60.</td>
</tr>
<tr>
<td>10.</td>
<td>Esr1</td>
<td>27.</td>
<td>Spib</td>
<td>44.</td>
<td>Foxi1</td>
<td>61.</td>
</tr>
<tr>
<td>13.</td>
<td>Creb3l3</td>
<td>30.</td>
<td>Barx2</td>
<td>47.</td>
<td>Erf</td>
<td>64.</td>
</tr>
<tr>
<td>16.</td>
<td>Tcf7l2</td>
<td>33.</td>
<td>Tcf7l2</td>
<td>50.</td>
<td>Hnf4g</td>
<td>67.</td>
</tr>
<tr>
<td>17.</td>
<td>Max</td>
<td>34.</td>
<td>Pura</td>
<td>51.</td>
<td>Prdm1</td>
<td>68.</td>
</tr>
</tbody>
</table>

### TFs expressed in the crypt (Fold change >3)

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td>Esrrg</td>
<td>22.</td>
<td>Ehf</td>
<td>41.</td>
<td>Orc2</td>
<td>60.</td>
</tr>
<tr>
<td>4.</td>
<td>Uhrf1</td>
<td>23.</td>
<td>Egr1</td>
<td>42.</td>
<td>Hoxb7</td>
<td>61.</td>
</tr>
<tr>
<td>6.</td>
<td>Nr2e3</td>
<td>25.</td>
<td>Dnmt1</td>
<td>44.</td>
<td>Dnmt3b</td>
<td>63.</td>
</tr>
<tr>
<td>11.</td>
<td>Sox9</td>
<td>30.</td>
<td>Tgif1</td>
<td>49.</td>
<td>Terf1</td>
<td>68.</td>
</tr>
<tr>
<td>13.</td>
<td>Foxm1</td>
<td>32.</td>
<td>Irf3</td>
<td>51.</td>
<td>Meis1</td>
<td>70.</td>
</tr>
<tr>
<td>15.</td>
<td>Zfp711</td>
<td>34.</td>
<td>Hdac2</td>
<td>53.</td>
<td>Lonp1</td>
<td>72.</td>
</tr>
<tr>
<td>17.</td>
<td>Ncl</td>
<td>36.</td>
<td>Nr5a2</td>
<td>55.</td>
<td>Nme1</td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Klf2</td>
<td>37.</td>
<td>Sox4</td>
<td>56.</td>
<td>Nme1</td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>Nr4a1</td>
<td>38.</td>
<td>Zc3h8</td>
<td>57.</td>
<td>Nfyb</td>
<td></td>
</tr>
</tbody>
</table>
TFs evenly expressed in the crypt–villus axis (Fold change <3)

1. Elf2
2. Nfyb
3. E2f5
4. Mta3
5. Elf3
6. Zfp105
7. Isl1
8. Hnrrnk
9. Fos
10. Pura
11. Zfp238
12. Nr2c1
13. Crebfz
14. Creb5
15. Zhx3
16. Bptf
17. Tcf3
18. Usf2
19. Hnrrnpa2b1
20. Hnrrnpab
21. Vax2
22. Med1
23. Cctf
24. Obfc1
25. Atf5
26. Ppara
27. Cebp
28. Crebfz
29. Cd2
30. Pbx2
31. Ppara
32. Creb3l4
33. Lrwd1
34. Drap1
35. Foxp1
36. Pole4
37. Pole3
38. Thap1
39. Erf
40. Pbx3
41. Xbp1
42. Cd2
43. Cebp
44. Pot1a
45. Atf1
46. Cebp
47. Pax4
48. Tcf3