Enterocyte-specific regulation of the apical nutrient transporter SLC6A19 (B^0AT1) by transcriptional and epigenetic networks*

Emrah Tümer, Angelika Bröer, Sarojini Balkrishna, Torsten Jülich, Stefan Bröer

From the Research School of Biology, Australian National University, Canberra, Australian Capital Territory 0200, Australia

*Running title: Regulation of intestinal nutrient uptake

To whom the correspondence should be addressed: Stefan Bröer, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia, Phone: +61-2-6125-2540, Fax: +61-2-6125-0313, E-mail: Stefan.broeer@anu.edu.au

Keywords: Amino acids transport, chromatin histone modification, DNA methylation, gene expression, intestinal epithelium, transcription factor

Background: Intestinal epithelial cells express genes for nutrient absorption.
Results: Expression of the nutrient transporters is suppressed in the crypt and promoted in the villus by opposing transcription factors, changes of DNA methylation and histone modification.
Conclusion: Gene expression in enterocytes is regulated by three different mechanisms.
Significance: Removal of repressive transcriptional control elements is required to express enterocyte specific genes in the villus.

ABSTRACT

Enterocytes are specialized to absorb nutrients from the lumen of the small intestine by expressing a select set of genes to maximize the uptake of nutrients. They develop from stem cells in the crypt and differentiate into mature enterocytes while moving along the crypt-villus axis. Using, as an example, the *Slc6a19* gene, encoding the neutral amino acid transporter B^0AT1, we studied regulation of the gene by transcription factors and epigenetic factors in the intestine. To investigate this question we used a fractionation method to separate mature enterocytes from crypt cells and analysed gene expression. Transcription factors HNF1a and HNF4a activate transcription of the *Slc6a19* gene in villus enterocytes, while high levels of SOX9 repress expression in the crypts. CpG dinucleotides in the proximal promoter were highly methylated in the crypt and fully de-methylated in the villus. Furthermore, histone modification H3K27Ac, indicating an active promoter, was prevalent in villus cells but barely detectable in crypt cells. The results suggest that *Slc6a19* expression in the intestine is regulated at three different levels involving promoter methylation, histone modification and opposing transcription factors.

The small intestine is a highly specialized organ for nutrient digestion and absorption. Its anatomical structure is adapted to these tasks, including a highly differentiated crypt-villus structure forming a large surface to maximize uptake of digested nutrients. All epithelial cells originate from stem cells within the crypt, resulting in differentiated cell types for mucus production (Goblet cells), hormone production (enteroendocrine cells), host defense (Paneth cells), and absorption (enterocytes) (1). Enterocytes make up over 80% of all intestinal epithelial cells. They are highly polarized, possessing a brush border structure at their apical side, across which nutrients are absorbed (1). Therefore, the differentiation of crypt stem cells into fully differentiated absorptive enterocytes located in the villi of the small intestine is an excellent model for studying the transcriptional control mechanisms of protein sets that are required for nutrient absorption.

At the enterocyte brush-border of the small intestine, peptides resulting from protein digestion by pancreatic proteases are further

hydrolysed into individual amino acids, as well as di- and tripeptides (2). The majority of neutral amino acids are taken up through an active transport process involving the neutral amino acid transporter B^0AT1 (encoded by the *Slc6a19* gene, throughout the manuscript we will use *Slc6a19* when referring to the gene and B^0AT1 to identify the protein). The B^0AT1 transporter is a Na^+-amino acid symporter, and is located in the apical membrane of the absorptive enterocytes (3). Apart from being expressed in the small intestine, it is also detected in the apical membrane of proximal kidney tubule cells, and to a smaller degree in the pancreas (4) and the skin, but is not detected in any other tissue. B^0AT1 requires auxiliary proteins for expression at the cell surface, namely collectrin (TMEM27) in the kidney (5,6), and ACE2 (Angiotensin-Converting Enzyme 2) in the small intestine (7). For enterocytes to transport neutral amino acids, it is thus necessary to express both ACE2 and B^0AT1 simultaneously. Due to its activity as a carboxypeptidase, ACE2 contributes to peptide digestion and feeds, together with aminopeptidase N, amino acids to the B^0AT1 transporter (8).

The Hepatocyte Nuclear Factors HNF1a, HNF1b and HNF4a are part of an autoregulatory transcriptional network in mammalian pancreas, liver, kidney and gut (9). HNF1a and HNF1b are homeobox-containing transcriptional activators. HNF1a was first identified as a tissue-specific regulator of liver function (10), but has been shown to be expressed across a small number of tissues, including kidney, pancreas, and intestine. Mutations in the DNA-binding domain of HNF1a are the cause of type 3 maturity-onset diabetes of the young (MODY3), which is characterized by progressive loss of insulin secretory capacity (11). HNF1a is assumed to act as a master regulator for the transcription of transporters in the kidney, because HNF1a deficient mice manifest a general defect in the renal re-absorption of metabolites (12). Additionally, most of the renal amino-acid transporters in murines contain HNF1a-binding motifs within their proximal promoter regions (13). While HNF1a is expressed in three different isoforms in humans, rodents only express a single isoform, most likely resulting in species-specific regulatory differences (9). The absence of epithelial amino acid transporter expression in the liver despite HNF1a expression, has been suggested to be due to DNA methylation, thus preventing the interaction of HNF1a with the transport promoter binding sites (13). However, the observed DNA methylation was at a significant distance from the proximal promoter.

Genome-wide expression profiling carried out with liver tissue of HNF1a deficient mice showed that HNF1a is involved in the regulation of several important hepatic functions, such as bile acid and cholesterol metabolism (14). A comparative study with liver and pancreatic island tissue revealed that HNF1a-dependent genes encode a broad range of metabolic functions in both tissues, including those for amino acid transport and steroid, lipid, and xenobiotic metabolism (15). Pancreatic tissue of HNF1a deficient mice exhibits down-regulation of the non-epithelial amino acid transporter SLC38A4, ACE2, collectrin, and the Nuclear factor HNF4a (15,16). In liver HNF4a appears to control expression of *Hnf1a*, while the opposite is observed in pancreas due to the use of an alternate *Hnf4a* promoter, which is controlled by HNF1a (16). Both promoters are used in the duodenum. HNF1b is a paralogue of HNF1a with indistinguishable *in vitro* DNA-binding properties (17). HNF1b is highly expressed in pancreatic islets, yet only expressed at low levels in adult liver (15). Interestingly, HNF1b expression has been shown to be increased in HNF1a-/- liver tissue (15). Mutations in HNF1b are the principal cause for the MODY5 phenotype, exhibiting atrophy in the pancreas as well as several forms of renal disease (18).

HNF4a is a highly conserved, zinc-finger containing transcriptional activator of the steroid hormone nuclear receptor superfamily. In contrast to most nuclear receptors however, HNF4a does not appear to be activated by exogenous ligands (19). HNF4a is expressed at high levels in liver, kidney, small intestine, colon, and pancreatic beta cells. Loss-of-function mutations in HNF4a result in MODY1, a monogenic autosomal dominant non-insulin-dependent diabetes mellitus type II phenotype, exhibiting a decrease in glucose-induced insulin release (20). As mentioned above HNF4a controls the transcription of the *Hnf1a* gene in liver (16), thus indirectly regulating a vast array of genes vital for hepatocyte function (21). It was reported previously that HNF1a can down-regulate HNF4a-mediated activation of transcription via a direct interaction of these transcription factors (22). In contrast, others observed that HNF4a enhances the

transcriptional activity of HNF1a independent of its DNA-binding activity, possibly through recruitment of other co-activating regulators (23). HNF4a appears to play an important role early in embryonic development, as indicated by its presence as a maternal component of the egg, and its expression in mouse embryonic stem cells (24). During development, HNF4a expression precedes the expression of HNF1a and HNF1b, implying that HNF4a is a transcription factor higher up in the hierarchy of regulatory proteins, thus establishing the importance of HNF4a for proper terminal differentiation of hepatocytes (24).

HNF1a and HNF1b appear to play important roles in terminal differentiation and cell fate commitment in the intestine, particularly for the Paneth cell lineage. HNF1b is expressed throughout development from the primitive gut stage through organogenesis, while HNF1a is expressed later during organogenesis (25). HNF4a has been suggested to be a major transcriptional regulator in the intestine. Genome-wide ChIP-chip analysis in the human Caco-2 cell line indicates that HNF4a is associated with regulatory regions of over a thousand genes (26). While these included some genes involved in digestion and absorption, many key proteins were missing, such as the glucose transporter SGLT1, the cationic amino acid transporter rbat/$b^{0,+}$AT, sucrose/isomaltase, B^0AT1, ACE2 etc. The analysis revealed transcription factor networks, in which HNF4a regulates the expression of CDX2. A model has been proposed in which HNF4a initiates the differentiation-dependent transcription by activating *Hnf1a* and *Cdx2*, both encoding transcription factors which are required simultaneously for the expression of many genes specifically expressed in the intestine (26,27).

In order to shed light onto the regulatory mechanisms responsible for the co-expression of ACE2 and B^0AT1 in enterocytes of the small intestine, we have enriched crypt stem cells and absorptive villi enterocytes using a modified Traber method (28), and subjected them to a genome-wide expression analysis, combined with a bioinformatics approach to identify transcription factor candidates for the regulation of ACE2 and B^0AT1 expression. The involvement of candidate genes was then investigated utilizing both in vitro and in vivo methods, namely reporter gene assays and Chromatin Immunoprecipitation (ChIP).

Experimental Procedures

Molecular cloning and DNA manipulation

All primer pairs used in this study are listed in Supplementary Table 1 and were purchased from Sigma-Aldrich. The *Slc6a19* promoter region was amplified by nested PCR using genomic DNA from mouse kidney. Initially a fragment ranging from position -2639 to position +374 was amplified by standard PCR using Pfu polymerase. This fragment was used as a template for a second PCR, which used primers with an incorporated *Kpn*I and *Bgl*II site spanning from position -2494 to +57. The product of this reaction was cut with *Kpn*I and *Bgl*II and inserted into the reporter gene vector pGL4.12 (Promega). The *Ace2* promoter was cloned using a similar strategy. Initially an 1856bp fragment was amplified from genomic DNA (position -1602 to +254). This fragment was used as a template for a second PCR, which used primers with an incorporated *Xho*I and *Kpn*I site spanning from position -1509 to +170. The product of this reaction was cut with *Xho*I and *Kpn*I and inserted into the reporter gene vector pGL4.12 (Promega). Promoter deletions were made using the full promoter construct in pGL4.12 by inverse PCR. Reverse and forward primers flanking the desired deletion were used to amplify the complete vector minus the deletion and the resulting fragment was re-ligated. To clone transcription factors, mRNA was isolated from mouse intestine (RNeasy plus, Qiagen) and converted into cDNA using Superscript II reverse transcriptase (Invitrogen). *Sox*9 cDNA was cloned by nested PCR. Initially a larger fragment was amplified, followed by amplification of the coding sequence with primers containing a *Bam*HI (sense) and a *Eco*RI (antisense) site. The fragment was digested with both enzymes and ligated into pcDNA3.1(+) (Invitrogen). *Hnf1a* cDNA was purchased from Thermo Scientific (clone ID:30471380), the cDNA being inserted into pCMV-SPORT. *Hnf4a* cDNA was amplified from intestinal cDNA using primers incorporating *Bam*HI and *Eco*RI cutting sites. After digestion, the cDNA was ligated into the corresponding cites of pcDNA3.1(+).

Putative transcription factor binding sites in the reporter-gene constructs were verified by site-directed mutagenesis using the Quikchange II site directed mutagenesis kit (Agilent). Primers for the mutagenesis were designed using the manufacturer's

recommendations. All constructs were confirmed by sequencing.

In situ hybridization

Tissue specimens of small intestine from C57BL/6, SWR/J and DBA1/J mice were fixed in 4% paraformaldehyde/0.1 M sodium phosphate buffer (pH 7.2) for 4 hours and embedded in paraffin. Five µm tissue sections were de-waxed and hybridized as previously described (29). Hybridization probes were generated by in vitro transcription of mB^0AT1 cDNA cloned into pCR-blunt II-TOPO using SP6 polymerase (antisense) and T7 polymerase (sense). The hybridization mixture (10 mM Tris HCl, pH 7.4, 50% (vol/vol) deionized formamide, 600 mM NaCl, 1 mM EDTA, 0.02% polyvinylpyrrolidone, 0.02% Ficoll, 0.05% bovine serum albumin, 10% dextrane sulfate, 10 mM dithiothreitol, 200 µg/ml denatured sonicated salmon sperm DNA, 100 µg/ml rabbit liver tRNA) contained either the ^{35}S-labeled RNA antisense or sense control mB^0AT1 probe at a concentration of 500 ng/ml. Hybridization with RNA probes proceeded at 42ºC for 18 hr. Following washing steps, the slide preparations were dipped in NTB2 emulsion (Kodak) and exposed at 4°C for 3 weeks. After development the slides were stained with hematoxylin/eosin and photographed with a Sony DSC digital camera.

Cell fractionation of mouse intestine

A fractionation method was used to isolate cell populations enriched with crypt and villus cells. This method is based on the inverted intestinal sac method first described by Barnard et al. (30) and further optimized by Traber et al. (28). In brief, the small intestine was cut after the duodenum and before the ileocaecal boundary and flushed with PBS containing 1 mM DTT. Subsequently, 6 cm long sections of mouse jejunum were inverted onto a plastic rod. The plastic rod was inserted into a hole in the lid of a 12 ml plastic round-bottom centrifuge tube, containing 10 ml citrate buffer (96 mM NaCl, 1.5 mM KCl, 27 mM Na_3Citrate, 8 mM KH_2PO_4 and 5.6 mM Na_2HPO4, pH 7.4). After sealing, the assembly was incubated on a rotatory shaker at 37°C for 10 min. Subsequently, the citrate buffer was changed to 10 ml PBS containing 0.5mM DTT, 1.5mM EDTA and 1mg/ml bovine serum albumin at 37°C. Rotary shaking was continued for 10-20 min until the solution was opaque indicating detachment of cells. The solution was exchanged against warm (37°C) fresh solution until 5-7 fractions were collected. Cell viability was judged by trypan blue exclusion.

RNA isolation and microarray studies

RNA was isolated from intestinal cell fractions using RNeasy (Qiagen). RNA quality was measured on a bioanalyser chip. Gene expression was analysed using an Agilent SurePrint G3 Mouse Exon Microarray through the Ramaciotti Center at the University of New South Wales. Expression levels of selected mRNA's were confirmed by qPCR.

Quantitative PCR (qPCR)

Quantitative PCR was conducted using Fast SYBR Green master mix (Applied Biosystems) with the PCR being performed on 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 (Concentration, 200 nM). Subsequently, the mixture was pipetted into a 394-well plate. The PCR parameters were 50°C for 2 min, 95°C for 15 s and 60°C for 1 min running a total of 40 cycles. All samples were analysed in duplicate and experiments were repeated 3 times. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as an internal control. Primers were designed using Primer3 program (http://*frodo.wi.mit.edu/*) to amplify fragments of 170-250 bp.

Bioinformatic promoter analysis

The *Slc6a19* promoter region ranging from -2494 to +57 was analysed using the University of California, Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu). Bioinformatic analysis revealed that a region located between -228 bp +57 bp of the transcriptional start site (TSS) was highly conserved among different species (data not shown). A region of about 2.5 kb upstream of the TSS was analysed using the Genomatix MatInspector (genomatix.de) program. Relevant transcription factors in the *Slc6a19* promoter region are listed in Table 1.

Luciferase Assay (reporter gene assay)

Reporter gene experiments were performed using the Dual-Luciferase Reporter Assay System (Promega). HEK293 cells ($1x10^5$ cells per well) were seeded into 24-well plates 24 hours prior to transfection using RPMI with 5% fetal bovine calf serum (Invitrogen) at 37°C

in 5% CO_2 incubator. Transfections were performed with Lipofectamine LTX (Invitrogen). Cells were co-transfected with 500ng of different *Slc6a19* promoter constructs which were inserted into the pGL4.12[luc2CP] Vector (Promega), a total of 100 ng of transcription factor encoding plasmids (in pcDNA3.1) and 5 ng of pGL4.74[hRluc/TK] control vector (Promega). Luciferase activity was measured 24 hours after transfection using a Turner TD-20/20 luminometer according to the manufacturer's instructions.

Chromatin Immunoprecipitation Assay

Chromatin Immunoprecipitation (ChIP) was performed as per the instructions of the SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling). Briefly, for each immunoprecipitation, 25mg of fresh or snap-frozen mouse tissue was minced with a scalpel and placed in 1ml PBS containing protease inhibitors. Proteins and DNA were then cross-linked with 1.5% formaldehyde (prepared fresh) for 20 minutes at room temperature. The reaction was stopped by the addition of a 1/10 of the volume of glycine stop-solution (10x). After a 5 minute incubation the tissue was centrifuged for 5 minutes at 1500 rpm at 4°C and then washed once with 1ml PBS containing protease inhibitors. After disaggregating the tissue with a potter homogenizer using 20-25 strokes, the cell suspension was centrifuged for 5 minutes at 4°C at 1500 rpm. The supernatant was immediately removed from the cells and nuclei were prepared as per the instructions of the kit. Chromatin was digested with Micrococcal Nuclease provided with the kit. This was followed by sonication of the sample twice, 2 min each, at 80% amplitude with a 15 second pulse on and 15 second pulse off time using a Misonix S4000 sonicator. Lysates were clarified by centrifugation at 10000 rpm for 10 minutes at 4°C. The supernatant containing the cross-linked chromatin preparation was incubated overnight in 1-5 µg of the specific antibody. Samples were purified according to the instructions of the kit and analysed using qPCR. Antibodies were purchased from Abcam: H3K27ac(ab4729), H3K4me3(ab8580), SOX9(ab3697), HNF4a(ab41898), Cell signaling: IgG(#2729), H3(#4620), and Santa Cruz: HNF1a(sc-6547X)

DNA Methylation/Bisulfite Sequencing

Intestinal villus and crypt enterocytes were isolated using the fractionation protocol described above; other tissues were used after dissection. Following tissue extraction, genomic DNA was isolated using DNeasy Blood & Tissue kit (Qiagen). Isolated genomic DNA was treated with an EpiTect Bisulfite kit (Qiagen) according to the manufacturer's instructions. In this process, all unmethylated cytosines are converted to uracil nucleotides, whereas all methylated cytosines (5-methylcytosine) remain intact during the treatment. The promoter region was amplified by nested PCR from bisulfite-treated DNA. In the first round a 1639 bp fragment -1380 to +259 was amplified followed by amplification of a fragment ranging from -1131 to +18. The PCR products were sequenced and analysed for the methylation status of all CpG dinucleotides in the target region.

Results

At the protein and mRNA level, B^0AT1 is highly expressed in the villi, but absent in the crypt compartment.

In confirmation of previous results (7), B^0AT1 protein expression in the intestine shows a distinct gradient from the villus tip to the crypt (Fig. 1A). Expression is absent in the crypt and maximal in mature enterocytes at the villus tip. *In situ* hybridization confirms that this gradient is also observed at the level of mRNA (Fig. 1B). This is a typical feature of proteins involved in nutrient absorption and digestion, including angiotensin-converting enzyme 2 (data not shown), which is required for expression of B^0AT1 at the cell surface. In order to understand the mechanisms of *Slc6a19* transcriptional control, we analysed its promoter region. The proximal promoter of the mouse *Slc6a19* gene, as judged by conservation of the sequence across mammalian species, appeared to be relatively short, i.e. 150 bp long. It is a typical class II promoter, where the transcriptional start site is preceded by a TATA binding motif (31) at position -23 to -28 and a reasonably conserved transcription factor IIB recognition motif (-39 to -46), indicating the binding site of general transcription factors and of polymerase II (31). The transcriptional start site CCACTT is similar to the mammalian consensus initiator sequence (31). Relevant transcription factor binding sites and their position are listed in Table 1.

To investigate the tissue and cell specific expression of B^0AT1, we generated reporter gene constructs using a 2kb fragment to account

for possible species-specific regulatory elements that might be located outside the proximal promoter. When transfected into several different cell lines we could not detect significant reporter gene activity, suggesting that the promoter does not use common transcription factors found in a variety of cell lines. HNF1a and HNF4a (26), have been implicated as central players in the differentiation-induced expression of epithelial cells. When transfected into HEK293 cells, both transcription factors were sufficient to initiate transcription from the *Slc6a19* promoter construct, with HNF4a being a stronger activator than HNF1a (Fig. 2A,B). Combined transfection of HNF1a and HNF4a did not increase reporter gene activity further (data not shown). The *Ace2* promoter, by contrast, could be activated *in vitro* by HNF1a only (Fig. 2C). To identify the DNA binding sites for HNF1a and HNF4a, we generated promoter truncations and mutated predicted binding sites (Fig. 2). We identified two potential binding sites for HNF1a in the *Slc6a19* promoter located within 150 bp of the transcriptional start site (-126/-110 and -153/-137). Site-directed mutagenesis was performed, indicating that the more distal site was not used, while an active binding site was identified at position -111/-114 (Fig. 2A). Mutation of this binding site abolished the observed transcriptional activation by HNF1a. A more distal site at position -1205/-1189 did not contribute to activation of the promoter (Fig. 2A). Using the same constructs, it appeared that HNF4a had two functional binding sites, one located between -972 and -1378 and another within 136 bp of the transcriptional start site (Fig. 2B). Mutation of the predicted HNF4a binding site (-38/-41) that is proximal to the transcriptional start site abolished all reporter gene activity, indicating that this site is essential for recruitment of general transcription factors. Mutation of the predicted distal site (-1368/-1344) did not change the response of the reporter construct to co-transfection of HNF4a (Fig. 2B). Further sequence analysis and mutagenesis of potential binding sites at position -1108/-1105 and -1088/-1085 did not reveal an obvious HNF4a binding site between -972 to -1387. Thus the reason for the significant drop of HNF4a dependent reporter gene response after truncation of this region remains unclear. These results nevertheless demonstrate that HNF1a and HNF4a are both necessary and sufficient to initiate B^0AT1 transcription in the background of HEK293 cells.

Expression of transcription factors in crypt and villus enriched mouse small intestine cell fractions.

Given the strong activation of B^0AT1 transcription by both HNF4a and HNF1a, a feasible mechanism to establish the B^0AT1 expression pattern along the villus-crypt axis would be by way of a villus-specific expression of these key regulators. To test this hypothesis we used a fractionation method to isolate cell populations along villus-crypt axis. The method showed strong enrichment of *Slc6a19* mRNA at the villus tip fraction (ratio villus tip/crypt = ~20-fold, Fig. 3A) and of the well-established intestinal stem cell marker *Lgr5* (leucine-rich repeat containing G-protein coupled receptor 5) in the crypt-enriched fraction (ratio crypt/villus tip about 100-fold, Fig. 3B). This provided the basis for a genome-wide transcriptome comparison by microarray analysis. We analysed the crypt/villus ratio of all transcription factors that had significant expression in either crypts or villus tip (Table 1 listing those which have putative binding sites in the *Slc6a19* promoter). Surprisingly, we found that *Hnf1a* and *Hnf4a* mRNAs were expressed in both villus and crypts at similar levels. *Hnf4a* mRNA was more abundant than *Hnf1a* mRNA. Expression levels of both mRNAs were confirmed by qPCR (data not shown). Moreover, CDX2 and GATA4, both being previously implicated in enterocyte-specific expression (26,27,32), were either enriched in the crypt or expressed evenly (Table 1). Thus expression of *Slc6a19* mRNA in villus cells cannot be explained by a corresponding expression of *Hnf1a, Hnf4a, Cdx2 or Gata4*. A genome-wide survey showed that only eight transcription factors had robust expression and were significantly higher expressed in the villus than in the crypt (ratio > 3, *Creb3l3, Stat3, Mafb, Tbx3, Max, Fosl2, Foxo6, Trp63*). Only two of those (*Creb3l3, Stat3*) had putative binding sites in the *Slc6a19* promoter. Co-transfection of *Stat3* failed to activate reporter gene expression from the *Slc6a19* promoter, even when activated by addition of IL-6 (Table 1 showing qualitative data for all transcription factors). Similarly, *Creb3l3* failed to activate transcription when co-transfected (Table 1). We therefore considered negative regulation to explain B^0AT1 expression. The mRNA of transcription factor *Sox9* was highly expressed in the crypt, but

barely found at the villus tip fraction (Fig. 3C and Table 1). Co-expression of SOX9 suppressed *Slc6a19* promoter activity, regardless whether it was driven by HNF1a (Fig. 4A) or HNF4a (Fig. 4B). Suppression of reporter gene activity was observed in all but the two shortest constructs, suggesting that the *Slc6a19* core promoter contains a functional binding site for SOX9. Moreover, it also suppressed *Ace2* promoter activity driven by HNF1a (Fig. 2C), indicating that SOX9 may act in crypts as a general repressor of genes involved in intestinal nutrient absorption. A putative SOX9 binding site was identified at position -34/-37, overlapping with the HNF4a binding site. However, mutation of this binding site did not reduce SOX9 inhibition (data not shown). Further inspection of the proximal promoter revealed another putative SOX9 binding site at position -218/-215. Mutation of this site increased overall promoter activity and partially abrogated the SOX9 effect regardless whether transcription was driven by HNF1a (Fig. 5A) or HNF4a (Fig. 5B). Additional sequence analysis suggested a possible SOX9 binding sites a position +9/+12, mutation of which did not alter promoter activity. To investigate whether some of the SOX9 effect was independent of DNA binding, we co-transfected cells with the SOX9 mutants H65Y, which is unable to bind DNA (33). Expression of this mutant still partially inhibited transcriptional activity, suggesting that SOX9 can act both as a sequence-specific transcription factor and as an inhibitory cofactor. When SOX9(H65Y) was co-transfected with a promoter containing the mutated SOX9 binding site, transcription was no longer inhibited by SOX9 (Fig. 5A,B).

HNF1a, HNF4a and SOX9 bind to the B^0AT1 promoter in vivo.

Transfection and overexpression of transcription factors in cell lines provides a first clue as to whether a potential binding site exists within a regulatory region, however, it does not provide evidence for *in situ* occupation of this binding site within the environment of a differentiated tissue. To investigate which of the transcription factor candidates identified in the reporter studies interact with the *Slc6a19* promoter *in vivo*, we performed ChIP assays using freshly scraped mucosa (Fig. 6). In agreement with the reporter gene studies we found binding of both HNF1a and HNF4a in the first 300bp upstream of the transcriptional start site. An additional binding site for HNF1a was found between -984 and -812, which was not observed in the reporter gene assays. No specific second binding site of HNF4a was observed consistent with mutagenesis studies of the putative distal binding site. Binding of SOX9 was also detected in the proximal promoter, demonstrating that it can suppress transcription *in situ* (Fig. 6).

Epigenetic elements affecting tissue expression

To investigate whether epigenetic factors could contribute to the onset of B^0AT1 expression along the crypt-villus axis we compared villus and crypt fractions (Fig. 7A,B). Histone modification H3K4me3, indicating open promoter regions, was found both in the crypt and villus fractions. However, H3K27ac, also indicating active promoter regions, was highly abundant in the villus fraction, but barely detectable in the crypts. Liver cells showed a similar profile as crypts confirming that lack of H3K27ac is associated with lack of transcription (Fig. 7C). DNA methylation at CpG dinucleotides is widely known to be involved in gene silencing. Interestingly, it appears that CpG dinucleotides in the promoter region of *Slc6a19* are highly methylated in the crypt and become demethylated in the villus (Fig. 8). The *Slc6a19* promoter was also highly methylated in the liver, where B^0AT1 is not expressed. The changes of methylation were particularly apparent around the proximal promoter where HNF1a and HNF4a and polymerase II bind.

Discussion

Here we present a model for the regulation of expression of proteins relevant for nutrient absorption in mature differentiated enterocytes. The model involves expression of an activating transcription factor in both crypt and villus compartment, which is counteracted by a transcriptional repressor whose expression is high in the crypt and low in the villus. The main activators are either HNF4a and to a lesser extent HNF1a, whilst SOX9, which shows maximum expression levels around the crypt, appears to play the role of a transcriptional repressor.

Genome-wide screens of HNF transcription factor targets have been performed for both pancreas and liver (34), and for HNF4a in intestinal cells (26). However, these datasets have not revealed how enterocytes express a set of apical nutrient transporters. The HNF4a

screen in intestinal cells only revealed the proton amino acid transporter PAT1 as a target. However, HNF4a binds to the *Hnf1a* promoter and vice versa, thereby extending the target genes for any of the two transcription factors (34). Previous studies have found tissue-specific interactions between HNF4a and HNF1a (16). *Hnf4a* mRNA can be transcribed from two alternate promoters, one of which is HNF1a dependent. In liver and kidney *Hnf4a* transcription is independent of HNF1a, while in islets a different promoter is used resulting in HNF1a dependence. In the intestine both promoters are used, resulting in an interdependence of *Hnf1a* and *Hnf4a* expression. In agreement with these tissue differences a different model has been suggested in kidney, in which expression of epithelial transporters is driven by HNF1a alone (13). However, *in vivo*, HNF1a could activate *Hnf4a*, thereby indirectly activating *Slc6a19*. Alternatively, expression levels in the kidney are somewhat lower due to the lower efficiency with which HNF1A activates the *Slc6a19* promoter.

SOX9 has been shown to be an important regulator of differentiation in various cell types, including cells of the small intestine (35,36), where it is integral part of the Wnt signalling pathway (37). In the crypt compartment it has been shown to act as a repressor cell differentiation, inhibiting the expression of genes involved in maturation, such as *Cdx2*, and *Muc2* (37). In analogy to the results presented here for *Slc6a19*, reduced SOX9 expression in villus cells could allow CDX2 expression, which in turn could then be involved in the regulation of various other transcriptional regulators such as *Cdx1*, *Hnf1a*, and *Hnf4a* (38,39). In agreement with this model we found that CDX2 expression is higher in the crypt-enriched fraction, thus suggesting that SOX9 is not controlling gene expression indirectly via CDX2. Instead, our results suggest that SOX9 directly prevents the premature expression of genes involved nutrient absorption in undifferentiated intestinal crypt cells, despite the presence of positive regulators in this compartment. The inhibitory action of SOX9 is at least in part mediated by a specific SOX9 binding site in the *Slc6a19* promoter. In addition we found evidence for non-specific DNA binding or action as an inhibitory cofactor through protein-protein interactions. Upon down-regulation of *Sox9* in differentiated enterocytes, transcription of genes involved in nutrient absorption is activated.

This minimal model cannot explain the lack of epithelial nutrient transporter expression in hepatocytes, where HNF1a and HNF4a are abundant and SOX9 is missing. Kikuchi et al. (13) suggested that in the liver, specific CpG dinucleotides within promoters of transporter genes would be methylated. For *Slc6a19* differential methylation was observed at position -1080 upstream from the transcriptional start site. Given that the core promoter of *Slc6a19* is less than 300bp long, it appears unlikely that the proposed methylation would directly affect the promoter recruitment of activators. Our bisulfite sequencing data, by contrast, demonstrate differential methylation at the proximal promoter where binding of transcription factors and RNA polymerase II occurs. Intriguingly, the methylation pattern of the proximal B^0AT1 promoter changes rapidly, from fully methylated in the crypt to completely de-methylated in mature enterocytes. Due to the thermodynamic stability of CpG methylation, de-methylation can only occur by lack of re-methylation after mitosis or by a modified DNA excision repair mechanism (40,41). Rapid de-methylation has nevertheless been demonstrated in neurons (42) and hormone-sensitive promoters (43). Stem cells in the crypt undergo three to four cell divisions before differentiation (44), thus de-methylation could occur by both mechanisms. The complete methylation of the *Slc6a19* promoter observed in liver cells is in line with an absence of expression in this tissue. This suggests that de-methylation must occur before HNF1a or HNF4a can bind and activate transcription.

Differential histone acetylation has been reported for the sucrose-isomaltase gene and promoter along the crypt-villus axis (32). Di-acetylated histone H3 (lysine 9 and 14), which is associated with transcriptionally active genes, was found to be very low in the crypts, but much higher in the villus. Analysis of the ENCODE database (45) at the *Slc6a19* promoter confirmed an active histone signature in the intestine and kidney (Fig. 9). Histone modification H3K4me3 was highly enriched over the promoter, allowing binding of polymerase II. Histone modification H3K36me3 was present over all exons confirming that the gene was actively transcribed. These three elements were absent in other tissues including the liver, where HNF1a and HNF4a are highly expressed. This suggests

that histone modification also should occur before transcription can be initiated.

Our data suggest a model, whereby cell-specific and tissue-specific regulation of B^0AT1 expression is mediated by a concurrent combination of three mechanisms: 1) different levels of activating and repressing transcription factors in crypt and villus cells; 2) reversible histone modifications and 3) reversible promoter methylation. Stem cells in the crypt undergo 4-5 divisions during which differentiation occurs. During these divisions change of histone modification and removal of promoter methylation could occur. In addition SOX9 could become diluted allowing the onset of B^0AT1 expression, which gradually increases as cells move to the tip of the villus (Fig. 1).

References

1. van der Flier, L. G., and Clevers, H. (2009) Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium. *Annu. Rev. Physiol.* **71**, 241-260
2. Fairweather, S. J., Broer, A., O'Mara, M. L., and Broer, S. (2012) Intestinal peptidases form functional complexes with the neutral amino acid transporter B(0)AT1. *The Biochemical journal* **446**, 135-148
3. Broer, A., Klingel, K., Kowalczuk, S., Rasko, J. E., Cavanaugh, J., and Broer, S. (2004) Molecular cloning of mouse amino acid transport system B0, a neutral amino acid transporter related to Hartnup disorder. *J Biol Chem* **279**, 24467-24476
4. Broer, A., Juelich, T., Vanslambrouck, J. M., Tietze, N., Solomon, P. S., Holst, J., Bailey, C. G., Rasko, J. E., and Broer, S. (2011) Impaired Nutrient Signaling and Body Weight Control in a Na+ Neutral Amino Acid Cotransporter (Slc6a19)-deficient Mouse. *J Biol Chem* **286**, 26638-26651
5. Danilczyk, U., Sarao, R., Remy, C., Benabbas, C., Stange, G., Richter, A., Arya, S., Pospisilik, J. A., Singer, D., Camargo, S. M., Makrides, V., Ramadan, T., Verrey, F., Wagner, C. A., and Penninger, J. M. (2006) Essential role for collectrin in renal amino acid transport. *Nature* **444**, 1088-1091
6. Malakauskas, S. M., Quan, H., Fields, T. A., McCall, S. J., Yu, M. J., Kourany, W. M., Frey, C. W., and Le, T. H. (2007) Aminoaciduria and altered renal expression of luminal amino acid transporters in mice lacking novel gene collectrin. *Am J Physiol Renal Physiol* **292**, F533-544
7. Kowalczuk, S., Broer, A., Tietze, N., Vanslambrouck, J. M., Rasko, J. E., and Broer, S. (2008) A protein complex in the brush-border membrane explains a Hartnup disorder allele. *Faseb J* **22**, 2880-2887
8. Fairweather, S. J., Broer, A., O'Mara, M. L., and Broer, S. (2012) Intestinal Peptidases Form Functional Complexes with Neutral Amino Acid Transporter B0AT1. *Biochem J*
9. Harries, L. W., Brown, J. E., and Gloyn, A. L. (2009) Species-specific differences in the expression of the HNF1A, HNF1B and HNF4A genes. *PloS one* **4**, e7855
10. Courtois, G., Morgan, J. G., Campbell, L. A., Fourel, G., and Crabtree, G. R. (1987) Interaction of a liver-specific nuclear factor with the fibrinogen and alpha 1-antitrypsin promoters. *Science* **238**, 688-692
11. Yamagata, K., Oda, N., Kaisaki, P. J., Menzel, S., Furuta, H., Vaxillaire, M., Southam, L., Cox, R. D., Lathrop, G. M., Boriraj, V. V., Chen, X., Cox, N. J., Oda, Y., Yano, H., Le Beau, M. M., Yamada, S., Nishigori, H., Takeda, J., Fajans, S. S., Hattersley, A. T., Iwasaki, N., Hansen, T., Pedersen, O., Polonsky, K. S., Turner, R. C., Velho, G., Chevre, J. C., Froguel, P., and Bell, G. I. (1996) Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455-458
12. Pontoglio, M., Barra, J., Hadchouel, M., Doyen, A., Kress, C., Bach, J. P., Babinet, C., and Yaniv, M. (1996) Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome. *Cell* **84**, 575-585
13. Kikuchi, R., Yagi, S., Kusuhara, H., Imai, S., Sugiyama, Y., and Shiota, K. (2010) Genome-wide analysis of epigenetic signatures for kidney-specific transporters. *Kidney Int* **78**, 569-577

14. Shih, D. Q., Bussen, M., Sehayek, E., Ananthanarayanan, M., Shneider, B. L., Suchy, F. J., Shefer, S., Bollileni, J. S., Gonzalez, F. J., Breslow, J. L., and Stoffel, M. (2001) Hepatocyte nuclear factor-1alpha is an essential regulator of bile acid and plasma cholesterol metabolism. *Nat Genet* **27**, 375-382
15. Servitja, J. M., Pignatelli, M., Maestro, M. A., Cardalda, C., Boj, S. F., Lozano, J., Blanco, E., Lafuente, A., McCarthy, M. I., Sumoy, L., Guigo, R., and Ferrer, J. (2009) Hnf1 (MODY3) Controls Tissue-Specific Transcriptional Programs and Exerts Opposed Effects on Cell Growth in Pancreatic Islets and Liver. *Mol. Cell. Biol.* **29**, 2945-2959
16. Boj, S. F., Parrizas, M., Maestro, M. A., and Ferrer, J. (2001) A transcription factor regulatory circuit in differentiated pancreatic cells. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14481-14486
17. Cereghini, S. (1996) Liver-enriched transcription factors and hepatocyte differentiation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **10**, 267-282
18. Horikawa, Y., Iwasaki, N., Hara, M., Furuta, H., Hinokio, Y., Cockburn, B. N., Lindner, T., Yamagata, K., Ogata, M., Tomonaga, O., Kuroki, H., Kasahara, T., Iwamoto, Y., and Bell, G. I. (1997) Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet* **17**, 384-385
19. Gonzalez, F. J. (2008) Regulation of hepatocyte nuclear factor 4 alpha-mediated transcription. *Drug metabolism and pharmacokinetics* **23**, 2-7
20. Yamagata, K., Furuta, H., Oda, N., Kaisaki, P. J., Menzel, S., Cox, N. J., Fajans, S. S., Signorini, S., Stoffel, M., and Bell, G. I. (1996) Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458-460
21. Kuo, C. J., Conley, P. B., Chen, L., Sladek, F. M., Darnell, J. E., and Crabtree, G. R. (1992) A transcriptional hierarchy involved in mammalian cell-type specification. *Nature* **355**, 457-461
22. Ktistaki, E., and Talianidis, I. (1997) Modulation of hepatic gene expression by hepatocyte nuclear factor 1. *Science* **277**, 109-112
23. Eeckhoute, J., Formstecher, P., and Laine, B. (2004) Hepatocyte nuclear factor 4alpha enhances the hepatocyte nuclear factor 1alpha-mediated activation of transcription. *Nucleic Acids Res* **32**, 2586-2593
24. Ryffel, G. U. (2001) Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences. *J. Mol. Endocrinol.* **27**, 11-29
25. Richmond, C. A., and Breault, D. T. (2010) Regulation of Gene Expression in the Intestinal Epithelium. in *Progress in: Molecular Biology and Translational Science Development, Differentiation, and Disease of the Luminal Gastrointestinal Tract*, Elsevier Inc. pp 207-229
26. Boyd, M., Bressendorff, S., Møller, J., Olsen, J., and Troelsen, J. T. (2009) Mapping of HNF4α target genes in intestinal epithelial cells. *BMC Gastroenterology* **9**, 68
27. Benoit, Y. D., Pare, F., Francoeur, C., Jean, D., Tremblay, E., Boudreau, F., Escaffit, F., and Beaulieu, J. F. (2010) Cooperation between HNF-1alpha, Cdx2, and GATA-4 in initiating an enterocytic differentiation program in a normal human intestinal epithelial progenitor cell line. *Am J Physiol Gastrointest Liver Physiol* **298**, G504-517
28. Traber, P. G., Gumucio, D. L., and Wang, W. (1991) Isolation of intestinal epithelial cells for the study of differential gene expression along the crypt-villus axis. *Am J Physiol* **260**, G895-903
29. Klingel, K., Hohenadl, C., Canu, A., Albrecht, M., Seemann, M., Mall, G., and Kandolf, R. (1992) Ongoing enterovirus-induced myocarditis is associated with persistent heart muscle infection: quantitative analysis of virus replication, tissue damage, and inflammation. *Proc Natl Acad Sci U S A* **89**, 314-318
30. Barnard, J. A., Beauchamp, R. D., Coffey, R. J., and Moses, H. L. (1989) Regulation of intestinal epithelial cell growth by transforming growth factor type beta. *Proc Natl Acad Sci U S A* **86**, 1578-1582
31. Weaver, R. F. (2012) *Molecular Biology*, 5th Edition ed., McGraw-Hill, New York
32. Suzuki, T., Mochizuki, K., and Goda, T. (2008) Histone H3 modifications and Cdx-2 binding to the sucrase-isomaltase (SI) gene is involved in induction of the gene in the transition from the crypt to villus in the small intestine of rats. *Biochem Biophys Res Commun* **369**, 788-793

33. McDowall, S., Argentaro, A., Ranganathan, S., Weller, P., Mertin, S., Mansour, S., Tolmie, J., and Harley, V. (1999) Functional and structural studies of wild type SOX9 and mutations causing campomelic dysplasia. *J Biol Chem* **274**, 24023-24030
34. Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., Fraenkel, E., Bell, G. I., and Young, R. A. (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science (New York, N.Y.)* **303**, 1378-1381
35. Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., Bibeau, F., Scherer, G., Joubert, D., Hollande, F., Blache, P., and Jay, P. (2007) Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *The Journal of cell biology* **178**, 635-648
36. Gracz, A. D., and Magness, S. T. (2011) Sry-box (Sox) transcription factors in gastrointestinal physiology and disease. *AJP: Gastrointestinal and Liver Physiology* **300**, G503-G515
37. Blache, P. (2004) SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes. *The Journal of cell biology* **166**, 37-47
38. Heath, J. K. (2010) *Chapter Four - Transcriptional Networks and Signaling Pathways that Govern Vertebrate Intestinal Development*, Elsevier Inc.
39. Gao, N., White, P., and Kaestner, K. H. (2009) Establishment of Intestinal Identity and Epithelial-Mesenchymal Signaling by Cdx2. *Developmental Cell* **16**, 588-599
40. Ma, D. K., Guo, J. U., Ming, G. L., and Song, H. (2009) DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle* **8**, 1526-1531
41. Carey, N., Marques, C. J., and Reik, W. (2011) DNA demethylases: a new epigenetic frontier in drug discovery. *Drug Discov Today* **16**, 683-690
42. Ma, D. K., Jang, M. H., Guo, J. U., Kitabatake, Y., Chang, M. L., Pow-Anpongkul, N., Flavell, R. A., Lu, B., Ming, G. L., and Song, H. (2009) Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* **323**, 1074-1077
43. Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., and Salbert, G. (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45-50
44. Radtke, F., and Clevers, H. (2005) Self-renewal and cancer of the gut: two sides of a coin. *Science* **307**, 1904-1909
45. Consortium, T. E. P. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**, e1001046

Footnotes:

Acknowledgements: The authors would like to thank Karin Klingel (University of Tübingen) for carrying out in situ hybridisations for this study and Nadine Tietze for the immunofluorescence images. We would also like to thank Helen Speirs (SCI Ramaciotti Centre, University of New South Wales), Stephen Ohms (Biomolecular Resource Facility, Australian National University), Russell McInnes (Agilent, Australia), and Ines Atmosukarto (Australian National University) for their technical support. The STAT3 construct was kindly provided by Marie Bogoyevitch (University of Melbourne). The study was supported by grants from the National Health and Medical Research Council (525415) and the Australian Research Council to SB (DP0877897). ET is a recipient of a PhD scholarship from the Turkish Ministry of National Education.

Abbreviations: ChIP, Chromatin immunoprecipitation; TSS, transcriptional start site

Figure legends:

Figure 1: Expression of B^0AT1 in the intestine.
Sections from the intestine of adult mice were stained with antibodies specific for mB^0AT1 (green) (A). Probes corresponding to the sense and antisense strand of mB^0AT1 cDNA were generated by in vitro transcription. The antisense probe was hybridized to mRNA in sections from small intestine (B).

Figure 2: HNF1a and HNF4a can activate the *Slc6a19* promoter *in vitro*.
HEK293 cells were co-transfected with vectors expressing *Firefly* luciferase downstream of the *Slc6a19* promoter (A,B) or the *Ace2* promoter (C), and with control vector expressing *Renilla* luciferase from a constitutive promoter. *Firefly* luciferase activity is depicted as % of control luminescence generated by *Renilla* luciferase. The *Slc6a19* promoter deletion constructs are labeled with their 5'terminus position relative to the TSS and are depicted on the left panel, reporter gene activity is depicted in the right panel. Positions relative to the transcriptional start point are indicated. To analyse *Slc6a19* transcriptional activation, HNF1a (A) or HNF4a (B) was co-expressed in the cells (open bars) and compared to the promoter construct only (black bars). Activation of the *Ace2* promoter was analysed after co-expression of HNF1a, HNF4a and SOX9. Significant increases of promoter activity due to co-transfection of HNF1 and HNF4a are indicated (* $p<0.05$; ** $p<0.01$). P values indicate the significance derived from three experiments, each with 3 separate transfections.

Figure 3: Expression profile of *Slc6a19, Lgr5* and *Sox9* along the villus-crypt axis.
Sections of inverted intestine were mounted onto a rod and sealed in a buffer containing tube. Cells were dislodged by slow rotation of the tube. The bathing solution was exchanged seven times in regular intervals generating fractions 1-7 (F1-F7). Expression levels of *Slc6a19* (A), *Lgr5* (B) and *Sox9* (B) were determined by qPCR. The mean, upper and lower value of two experimental repeats are shown.

Figure 4: SOX9 is a negative regulator of transcription.
HEK293 cells were cotransfected with vectors expressing *Firefly* luciferase downstream of the *Slc6a19* promoter and with control vector expressing *Renilla* luciferase from a constitutive promoter. *Firefly* luciferase activity is depicted as % of control luminescence generated by *Renilla* luciferase. To analyse transcriptional activation, cells were analysed without additional transcription factors (black bar), with HNF1a (open bar, A) or HNF4a (open bar B) or with HNF1a plus SOX9 (grey bar, A) or with HNF4a plus SOX9 (grey bar, B). Constructs are labeled with their 5'terminus relative to the TSS. Significant reduction of promoter activity due to SOX9 co-transfection are indicated (* $p<0.05$; ** $p<0.01$). P values indicate the significance derived from three experiments, each with 3 separate transfections.

Figure 5: Characterization of SOX9 action at the *Slc6a19* promoter.
HEK293 cells were co-transfected with vectors expressing *Firefly* luciferase downstream of the *Slc6a19* wildtype promoter (open bars) or an *Slc6a19* promoter with mutated SOX9 binding site (black bars). All constructs were expressed together with control vector expressing *Renilla* luciferase from a constitutive promoter. *Firefly* luciferase activity is depicted as % of control luminescence generated by *Renilla* luciferase. To analyse SOX9 action, the *Slc6a19* reporter was co-expressed with the following transcription factors: (A) HNF1a, HNF1a plus SOX9, HNF1a plus SOX9(H65Y), or (B) HNF4a, HNF4a plus SOX9, HNF4a plus SOX9(H65Y). Construct -473 (see Fig. 2) was used as base for all constructs. Significant reduction of promoter activity due to SOX9 co-transfection are indicated (* $p<0.05$; ** $p<0.01$). P values indicate the significance derived from three experiments, each with 3 separate transfections.

Figure 6. *In vivo* binding analysis of SOX9, HNF1a, and HNF4a at the proximal *Slc6a19* promoter.
Chromatin immunoprecipitation was carried out using genomic DNA isolated from freshly scraped mucosa. Data are given in % relative to a quantitative PCR reaction on input DNA. Positions of primer sets along the *Slc6a19* promoter are indicated. The mean signal ± SD of three independent experiments is shown.

Figure 7: Analysis of histone modifications in the intestine and liver.
Chromatin immunoprecipitation was carried out using genomic DNA isolated from enterocytes derived from fraction 2 (villus) and fraction 5 (crypt) and from genomic DNA isolated from liver tissue. Data are given in % relative to a real-time PCR reaction on input DNA. After cross-linking, antibodies specific for histone H3, variant H3K27ac and H3K4me3 were used to precipitate DNA fragments. Positions of primer sets along the *Slc6a19* promoter are indicated. The mean signal ± SD of three independent experiments is shown.

Figure 8: DNA methylation patterns of the B^0AT1 promoter in different tissues.
Bisulfite sequencing was used to determine methylated CpG dinucleotides. DNA was isolated from liver and fractionated enteric cells using fraction 1 (villus) or fraction 5 (crypt). CpG positions (closed circle methylated, open circle unmethylated) and transcription factor binding sites are indicated. TSS, transcriptional start site. The number of sequenced independent PCR products is indicated by the number of symbols.

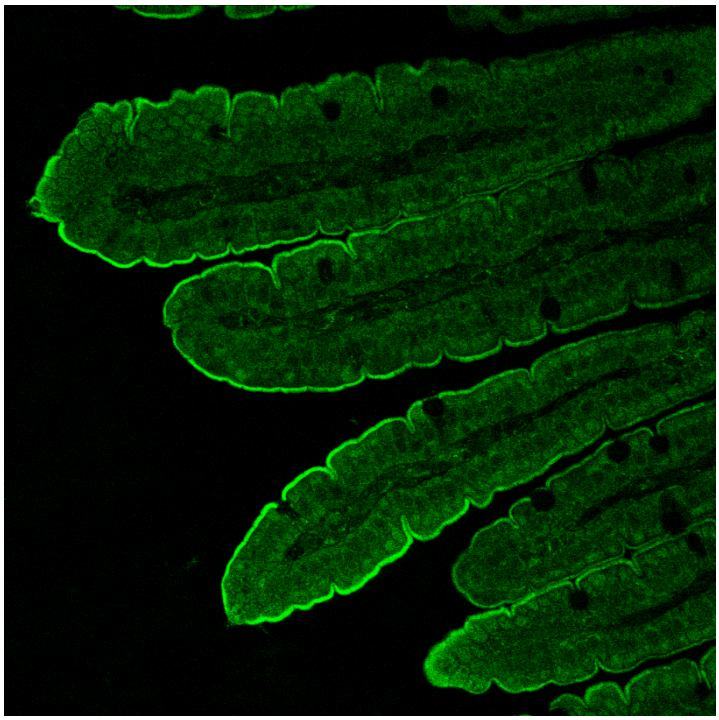
Figure 9: Histone modification and polymerase II binding to the *Slc6a19* promoter
The ENCODE database was used to extract the distribution of histone modifications H3K4me3, H3K36me3 and RNA polymerase II binding along the mouse *Slc6a19* gene. The *Slc6a19* gene shows active histone status and polymerase II binding in small intestine and kidney, but not in liver.

Table 1: Predicted Transcription factor binding sites in the *Slc6a19* promoter.
Transcription factors with robust expression in either crypt or villus were selected. GATA4 is listed because it has previously been indicated in enterocyte maturation, but does not have a predicted binding site. A matrix number of 1.00 indicates an exact match of the binding site consensus sequence. Revised location indicates results from mutagenesis data. A qualitative indicator is provided for the effect on transcription. The ratio of mRNA abundance along the villus-crypt axis was determined by microarray analysis. * Unless stated otherwise the effect of co-expression of the listed transcription factor with the promoter construct is shown.

TFs	Position	Strand	Matrix	Revised location	Effect on Transcription*	Ratio Villus/Crypt
ATF4	-1212/-1204	-	0.996		0	0.90
Cdx2	-36/-18	-	0.928		0	0.52
Cdx2	-1212/-1194	-	0.870		0	0.52
Cdx2	-1833/-1815	+	0.867		0	0.52
Creb3l3	-89/-69	+	0.954		0	8.8
Egr1	-167/-151	+	0.890		0	0.11
Egr1	-2112/-2096	+	0.863		0	0.11
FoxA2	-305/-289	-	0.991		0	0.32
FoxA2	-835/-819	-	0.995		0	0.32
GATA4	-	-	-		0	0.91
HNF1a – BS1	-126/-110	-	0.817	-126/-110	++	0.96
HNF1a – BS2	-153/-137	+	0.891		0	0.96
HNF4a – BS1	-239/-215	-	0.850	-41/-38	+++	1.1
HNF4a – BS2	-1368/-1344	+	0.937		0	1.1
Neurog3	-2352/-2340	+	0.963		0	0.18
Neurog3	-2474/-2462	+	0.988		0	0.18
Smad3	-228/-218	+	0.997		0	0.66
Smad3	-2016/-2006	+	0.997		0	0.66
Smad3	-2307/-2297	+	0.994		0	0.66
SOX9	-47/-23	+	0.771	-218/-215	0/-- (on HNF activated promoter)	0.05
SOX9	-1150/-1126	+	0.967		0	0.05
Spdef	-27/-7	+	0.868		Not tested	0.71
STAT3	-301/-283	+	0.771		0	10.6
STAT3	-303/-285	-	0.764		0	10.6
STAT3	-504/-486	-	0.977		0	10.6
STAT3	-587-569	-	0.971		0	10.6
STAT3	-718/-700	+	0.958		0	10.6
STAT3	-1044/-1026	-	0.959		0	10.6

Fig. 1

A



B

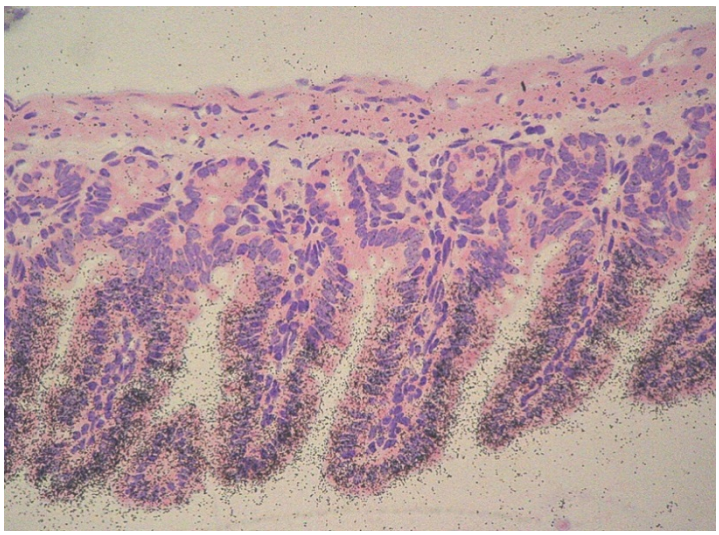


Fig. 2

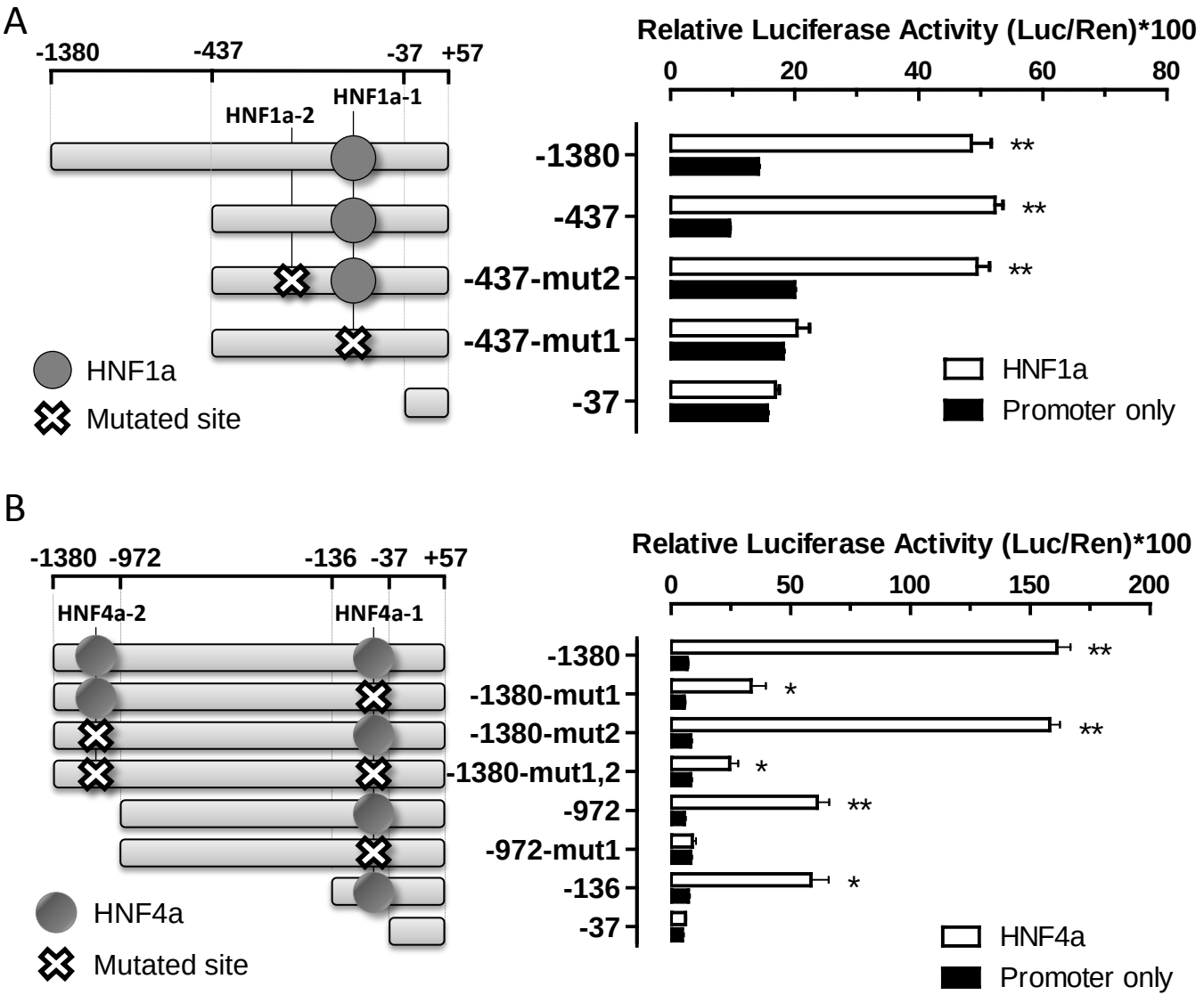

A
-1380
-437
-37
+57
HNF1a-2
HNF1a-1
HNF1a
Mutated site
Relative Luciferase Activity (Luc/Ren)*100
0
20
40
60
80
-1380
-437
-437-mut2
-437-mut1
-37
**
**
**
HNF1a
Promoter only
B
-1380
-972
-136
-37
+57
HNF4a-2
HNF4a-1
HNF4a
Mutated site
Relative Luciferase Activity (Luc/Ren)*100
0
50
100
150
200
-1380
-1380-mut1
-1380-mut2
-1380-mut1,2
-972
-972-mut1
-136
-37
**
*
**
*
**
*
HNF4a
Promoter only
C
Relative Luciferase Activity (Luc/Ren)*100
0
5
10
15
20
25
HNF4a
HNF1a + Sox9
HNF1a
Empty vector
*
**
Ace2

Fig. 3

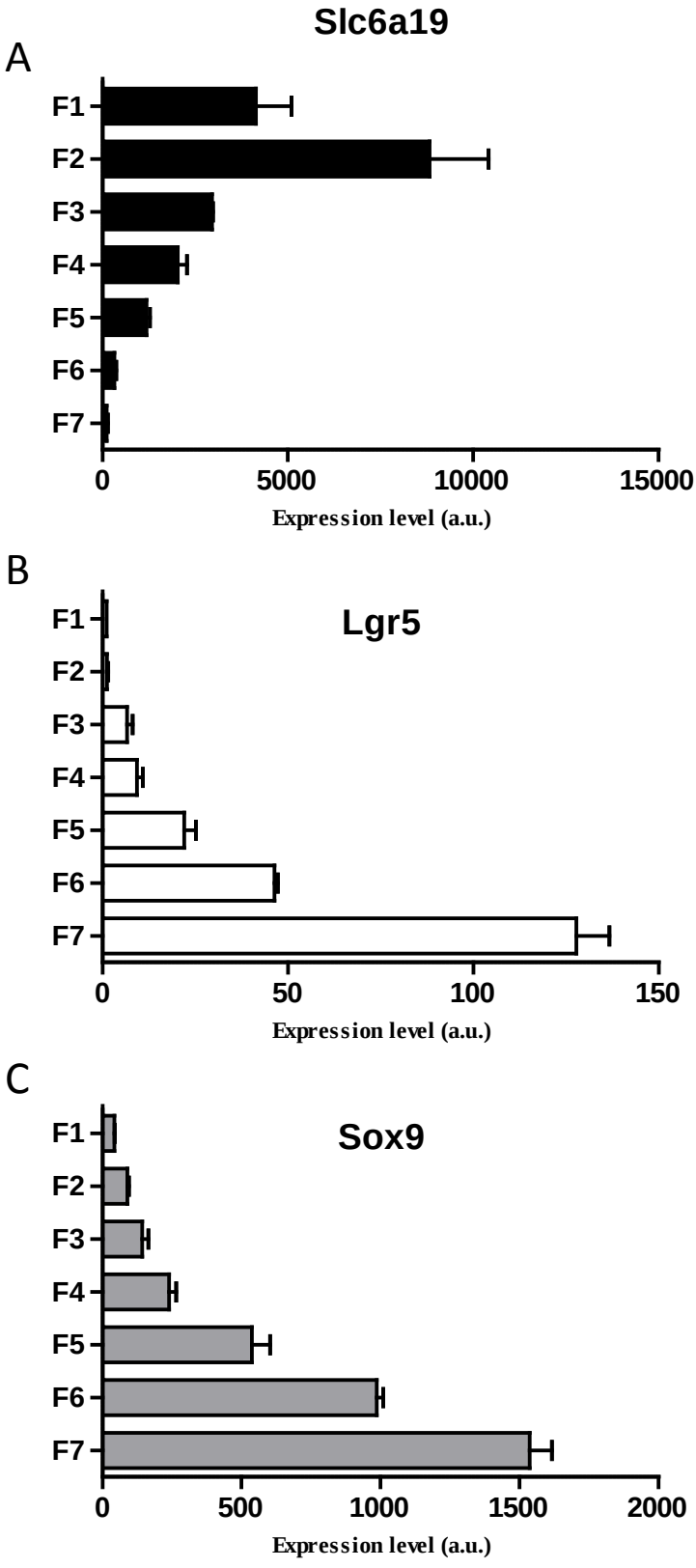

A
Slc6a19
F1
F2
F3
F4
F5
F6
F7
0
5000
10000
15000
Expression level (a.u.)
B
Lgr5
F1
F2
F3
F4
F5
F6
F7
0
50
100
150
Expression level (a.u.)
C
Sox9
F1
F2
F3
F4
F5
F6
F7
0
500
1000
1500
2000
Expression level (a.u.)

Fig. 4

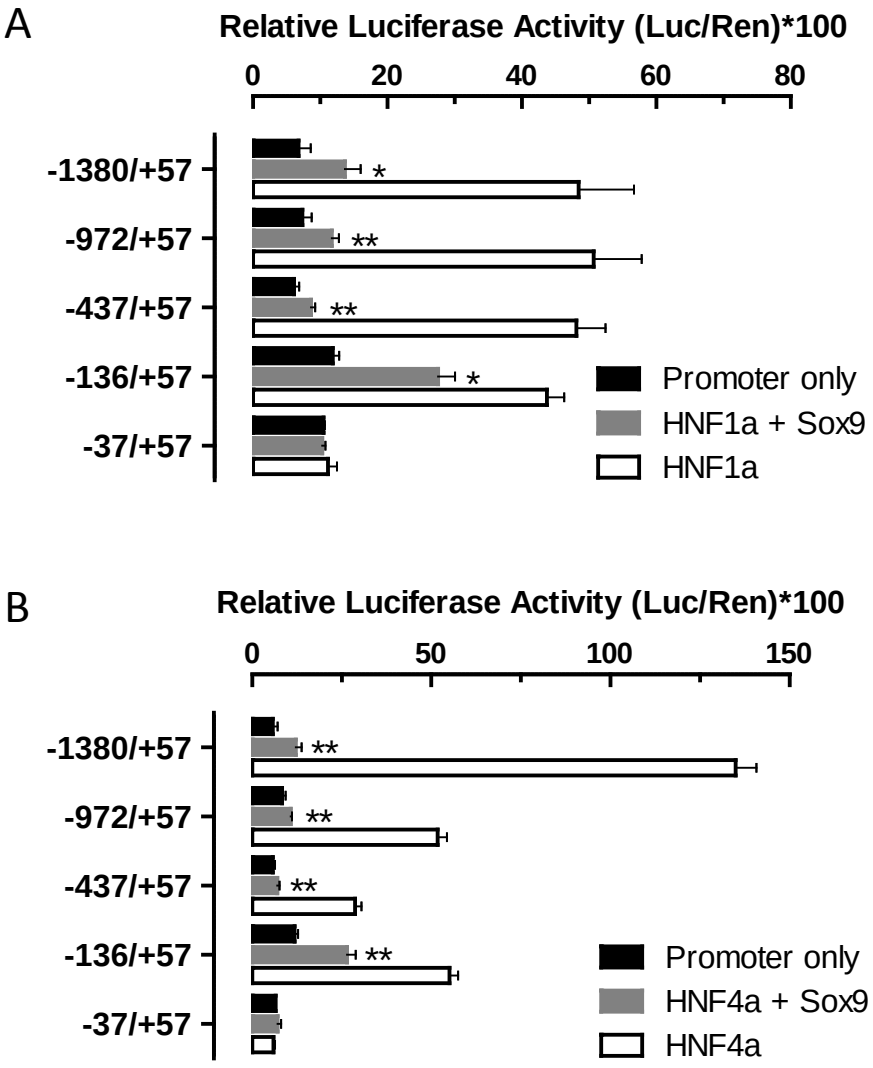

A
Relative Luciferase Activity (Luc/Ren)*100
0
20
40
60
80
-1380/+57
-972/+57
-437/+57
-136/+57
-37/+57
*
**
**
*
Promoter only
HNF1a + Sox9
HNF1a
B
Relative Luciferase Activity (Luc/Ren)*100
0
50
100
150
-1380/+57
-972/+57
-437/+57
-136/+57
-37/+57
**
**
**
**
Promoter only
HNF4a + Sox9
HNF4a

Fig. 5

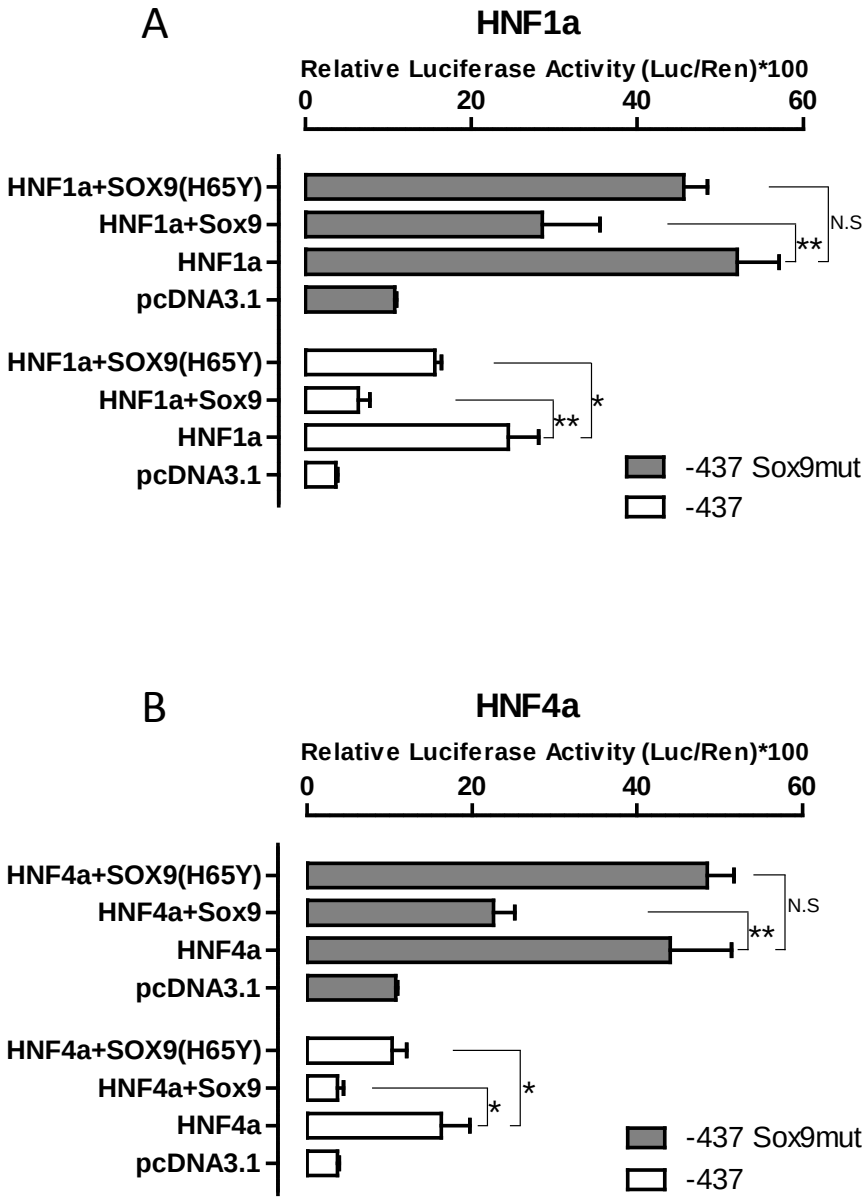

A
HNF1a
Relative Luciferase Activity (Luc/Ren)*100
0
20
40
60
HNF1a+SOX9(H65Y)
HNF1a+Sox9
HNF1a
pcDNA3.1
HNF1a+SOX9(H65Y)
HNF1a+Sox9
HNF1a
pcDNA3.1
N.S
**
*
**
-437 Sox9mut
-437
B
HNF4a
Relative Luciferase Activity (Luc/Ren)*100
0
20
40
60
HNF4a+SOX9(H65Y)
HNF4a+Sox9
HNF4a
pcDNA3.1
HNF4a+SOX9(H65Y)
HNF4a+Sox9
HNF4a
pcDNA3.1
N.S
**
*
*
-437 Sox9mut
-437

Fig. 6

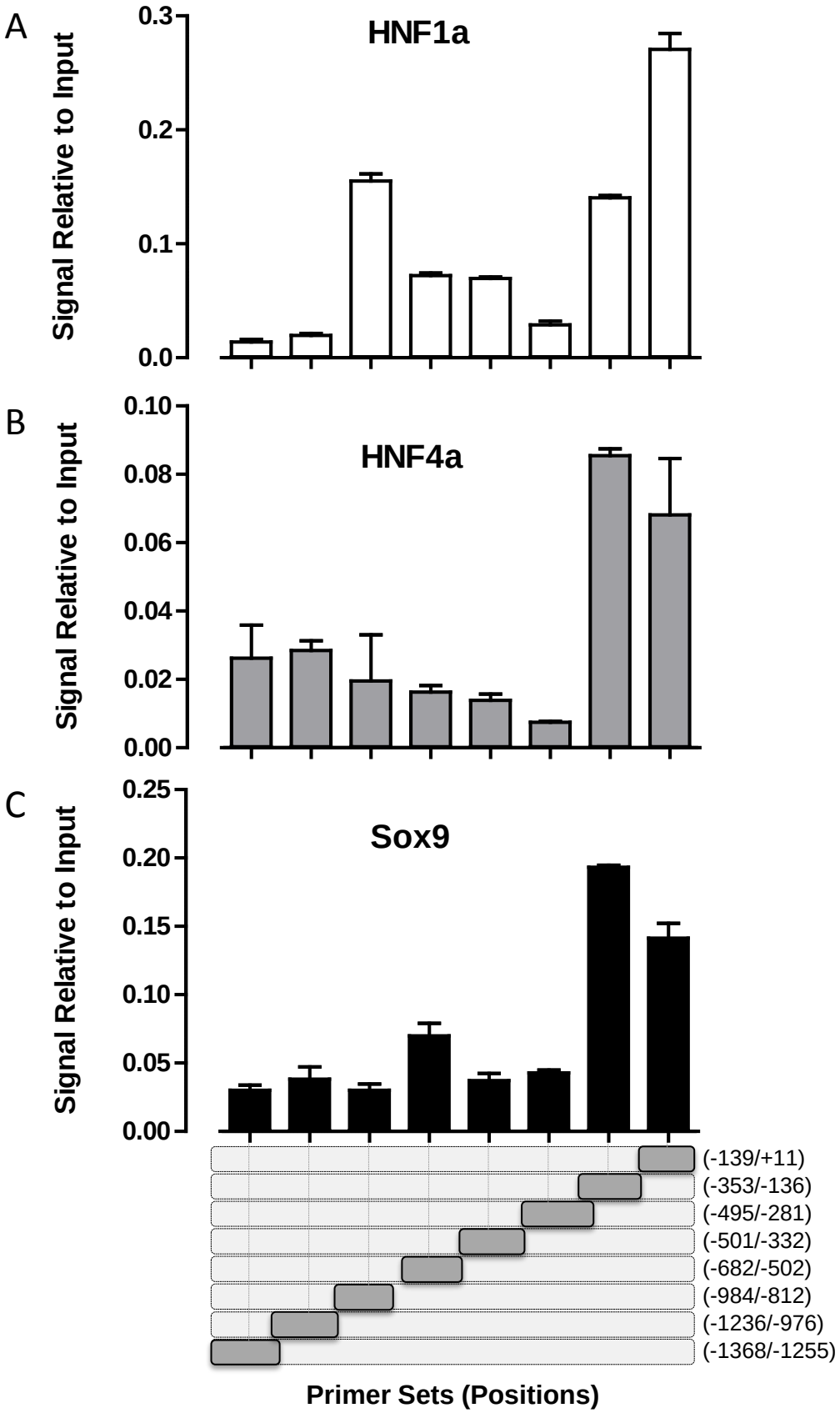

A
HNF1a
Signal Relative to Input
0.3
0.2
0.1
0.0
B
HNF4a
Signal Relative to Input
0.10
0.08
0.06
0.04
0.02
0.00
C
Sox9
Signal Relative to Input
0.25
0.20
0.15
0.10
0.05
0.00
(-139/+11)
(-353/-136)
(-495/-281)
(-501/-332)
(-682/-502)
(-984/-812)
(-1236/-976)
(-1368/-1255)
Primer Sets (Positions)

Fig. 7

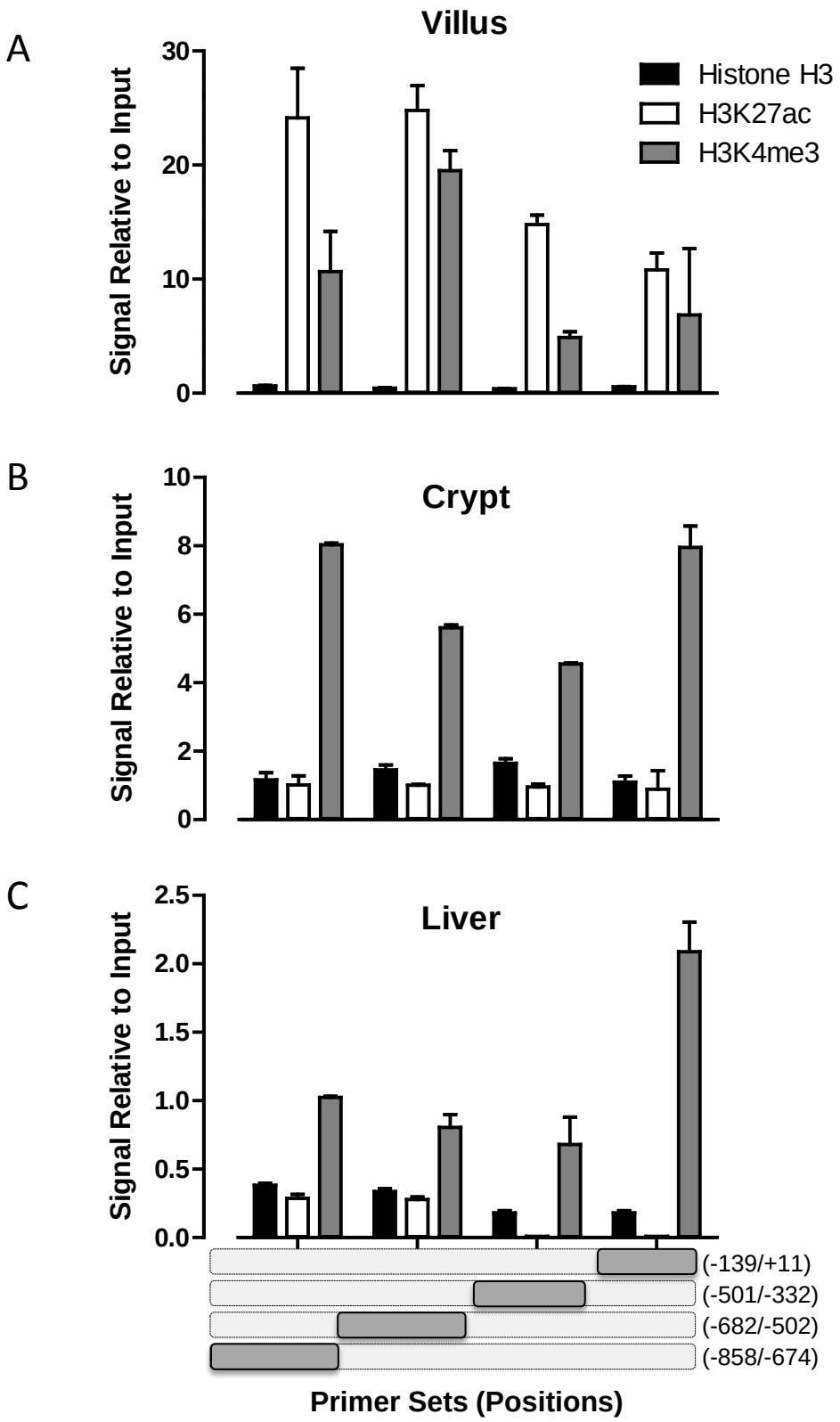

A
Villus
Signal Relative to Input
30
20
10
0
Histone H3
H3K27ac
H3K4me3
B
Crypt
Signal Relative to Input
10
8
6
4
2
0
C
Liver
Signal Relative to Input
2.5
2.0
1.5
1.0
0.5
0.0
(-139/+11)
(-501/-332)
(-682/-502)
(-858/-674)
Primer Sets (Positions)

Fig. 8

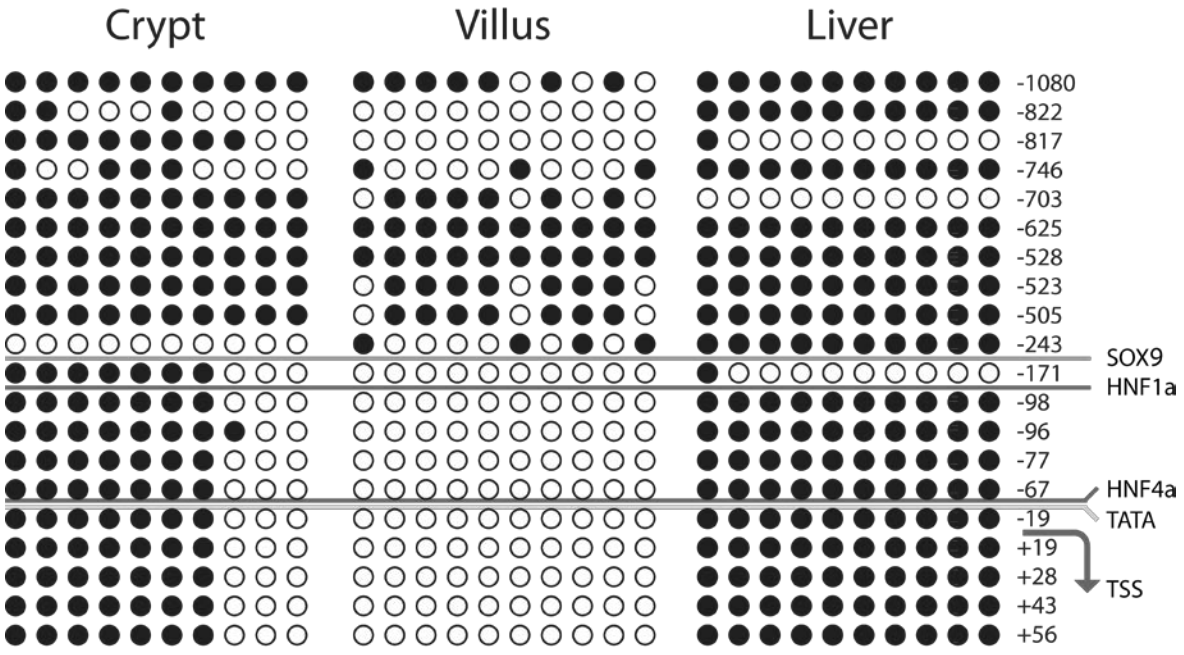
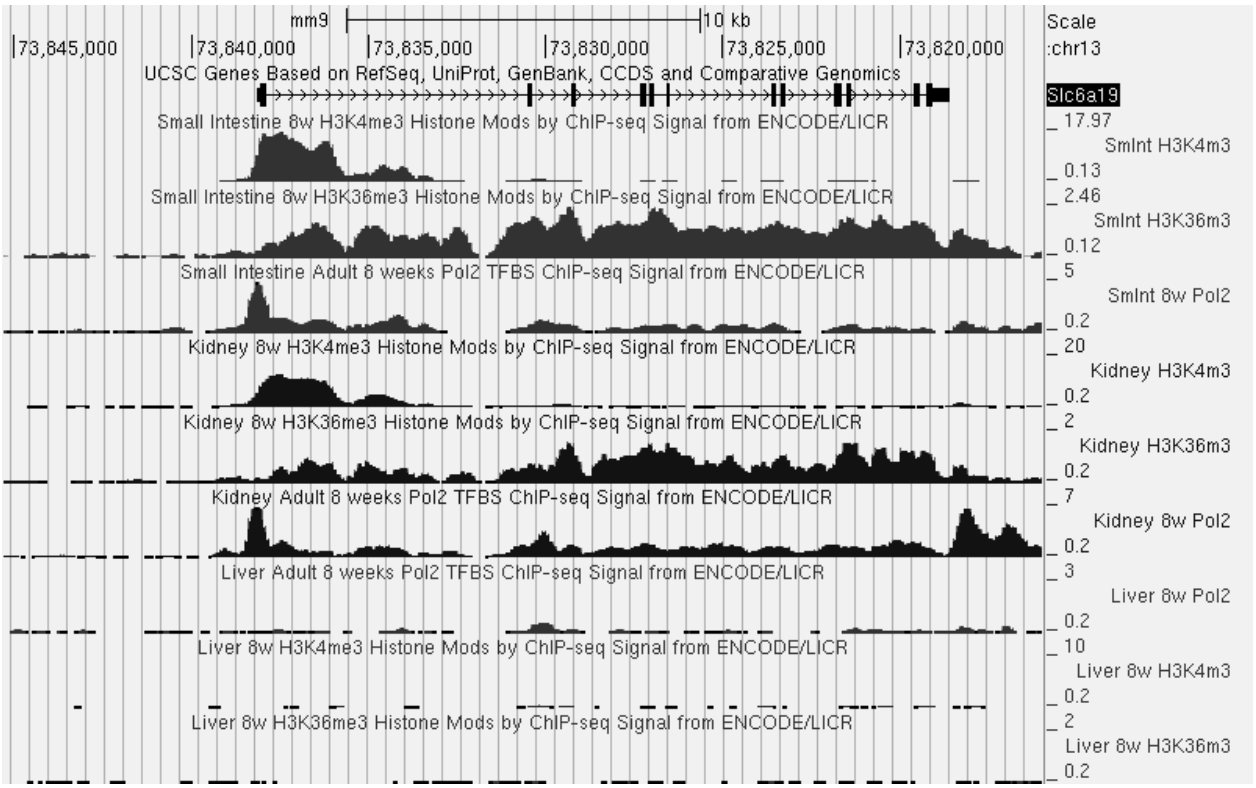

Crypt
Villus
Liver
-1080
-822
-817
-746
-703
-625
-528
-523
-505
-243
SOX9
-171
HNF1a
-98
-96
-77
-67
HNF4a
-19
TATA
+19
+28
TSS
+43
+56

Fig. 9


mm9
10 kb
73,845,000
73,840,000
73,835,000
73,830,000
73,825,000
73,820,000
Scale
:chr13
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
Slc6a19
Small Intestine 8w H3K4me3 Histone Mods by ChIP-seq Signal from ENCODE/LICR
17.97
SmInt H3K4m3
Small Intestine 8w H3K36me3 Histone Mods by ChIP-seq Signal from ENCODE/LICR
0.13
2.46
SmInt H3K36m3
0.12
Small Intestine Adult 8 weeks Pol2 TFBS ChIP-seq Signal from ENCODE/LICR
5
SmInt 8w Pol2
0.2
Kidney 8w H3K4me3 Histone Mods by ChIP-seq Signal from ENCODE/LICR
20
Kidney H3K4m3
0.2
Kidney 8w H3K36me3 Histone Mods by ChIP-seq Signal from ENCODE/LICR
2
Kidney H3K36m3
0.2
Kidney Adult 8 weeks Pol2 TFBS ChIP-seq Signal from ENCODE/LICR
7
Kidney 8w Pol2
0.2
Liver Adult 8 weeks Pol2 TFBS ChIP-seq Signal from ENCODE/LICR
3
Liver 8w Pol2
0.2
Liver 8w H3K4me3 Histone Mods by ChIP-seq Signal from ENCODE/LICR
10
Liver 8w H3K4m3
0.2
Liver 8w H3K36me3 Histone Mods by ChIP-seq Signal from ENCODE/LICR
2
Liver 8w H3K36m3
0.2