UNIVERSITY OF COPENHAGEN

# Metabolome, transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central regulator of gene expression during enterocyte differentiation.

Stegmann, Anders; Hansen, Morten; Wang, Yulan; Larsen, Janus B; Lund, Leif R.; Ritié, Léa; Nicholson, Jeremy K; Quistorff, Bjørn; Simon-Assmann, Patricia; Troelsen, Jesper T; Olsen, Jørgen

**You might find this additional information useful...**

Supplemental material for this article can be found at:
http://physiolgenomics.physiology.org/cgi/content/full/00314.2005/DC1

This article cites 45 articles, 17 of which you can access free at:
http://physiolgenomics.physiology.org/cgi/content/full/27/2/141#BIBL

This article has been cited by 10 other HighWire hosted articles, the first 5 are:

**Hepatocyte Nuclear Factor 4{alpha}, a Key Factor for Homeostasis, Cell Architecture, and Barrier Function of the Adult Intestinal Epithelium**
A.-L. Cattin, J. Le Beyec, F. Barreau, S. Saint-Just, A. Houllier, F. J. Gonzalez, S. Robine, M. Pincon-Raymond, P. Cardot, M. Lacasa and A. Ribeiro
*Mol. Cell. Biol.*, December 1, 2009; 29 (23): 6294-6308.
[Abstract] [Full Text] [PDF]

**Meta-analysis of gene expression in human pancreatic islets after in vitro expansion**
B. Kutlu, A. G. Kayali, S. Jung, G. Parnaud, D. Baxter, G. Glusman, N. Goodman, L. A. Behie, A. Hayek and L. Hood
*Physiol Genomics*, September 1, 2009; 39 (1): 72-81.
[Abstract] [Full Text] [PDF]

**Hepatocyte nuclear factor 4{alpha} contributes to an intestinal epithelial phenotype in vitro and plays a partial role in mouse intestinal epithelium differentiation**
J.-P. Babeu, M. Darsigny, C. R. Lussier and F. Boudreau
*Am J Physiol Gastrointest Liver Physiol*, July 1, 2009; 297 (1): G124-G134.
[Abstract] [Full Text] [PDF]

**Hepatocyte nuclear factor-4{alpha} regulates human cellular retinol-binding protein type II gene expression in intestinal cells**
N. Yamaguchi, S. Miyamoto, Y. Ogura, T. Goda and K. Suruga
*Am J Physiol Gastrointest Liver Physiol*, March 1, 2009; 296 (3): G524-G533.
[Abstract] [Full Text] [PDF]

**Transcriptome response of enterocytes to dietary lipids: impact on cell architecture, signaling, and metabolism genes**
O. Beaslas, F. Torreilles, P. Casellas, D. Simon, G. Fabre, M. Lacasa, F. Delers, J. Chambaz, M. Rousset and V. Carriere
*Am J Physiol Gastrointest Liver Physiol*, November 1, 2008; 295 (5): G942-G952.
[Abstract] [Full Text] [PDF]

Updated information and services including high-resolution figures, can be found at:
http://physiolgenomics.physiology.org/cgi/content/full/27/2/141

Additional material and information about *Physiological Genomics* can be found at:
http://www.the-aps.org/publications/pg

This information is current as of November 9, 2009 .

# Metabolome, transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central regulator of gene expression during enterocyte differentiation

**Anders Stegmann,[1] Morten Hansen,[1] Yulan Wang,[2] Janus B. Larsen,[1] Leif R. Lund,[3] Léa Ritié,[4] Jeremy K. Nicholson,[2] Bjørn Quistorff,[1] Patricia Simon-Assmann,[4] Jesper T. Troelsen,[1] and Jørgen Olsen[1]**

[1]*Department of Medical Biochemistry and Genetics, The Panum Institute, University of Copenhagen, Copenhagen, Denmark;* [2]*Biological Chemistry, Division of Biomedical Sciences, Imperial College London, London, United Kingdom;* [3]*The Finsen Laboratory, Rigshospitalet, Copenhagen, Denmark; and* [4]*Institut National de la Santé et de la Recherche Médicale U682, University Louis Pasteur, Strasbourg, France*

**Stegmann, Anders, Morten Hansen, Yulan Wang, Janus B. Larsen, Leif R. Lund, Léa Ritié, Jeremy K. Nicholson, Bjørn Quistorff, Patricia Simon-Assmann, Jesper T. Troelsen, and Jørgen Olsen.** Metabolome, transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central regulator of gene expression during enterocyte differentiation. *Physiol Genomics* 27: 141–155, 2006. First published July 25, 2006; doi:10.1152/physiolgenomics.00314.2005.—DNA-binding transcription factors bind to promoters that carry their binding sites. Transcription factors therefore function as nodes in gene regulatory networks. In the present work we used a bioinformatic approach to search for transcription factors that might function as nodes in gene regulatory networks during the differentiation of the small intestinal epithelial cell. In addition we have searched for connections between transcription factors and the villus metabolome. Transcriptome data were generated from mouse small intestinal villus, crypt, and fetal intestinal epithelial cells. Metabolome data were generated from crypt and villus cells. Our results show that genes that are upregulated during fetal to adult and crypt to villus differentiation have an overrepresentation of potential hepatocyte nuclear factor (HNF)-4 binding sites in their promoters. Moreover, metabolome analyses by magic angle spinning ^{1}H nuclear magnetic resonance spectroscopy showed that the villus epithelial cells contain higher concentrations of lipid carbon chains than the crypt cells. These findings suggest a model where the HNF-4 transcription factor influences the villus metabolome by regulating genes that are involved in lipid metabolism. Our approach also identifies transcription factors of importance for crypt functions such as DNA replication (E2F) and stem cell maintenance (c-Myc).

crypt-villus axis; intestine; gene regulation; hepatocyte nuclear factor-4

IN A DIAGRAM OF A GENE REGULATORY NETWORK, the transcription factors form nodes with many connections drawn as lines and extending to the genes that the transcription factors regulate (for reviews see Refs. 6, 34). Genome-wide chromatin immunoprecipitation experiments (27) have previously showed that the hepatocyte nuclear factors-1, 4, and 6 (HNF-1, HNF-4, and HNF-6) form important nodes in hepatic gene regulatory networks. Thus HNF-1, HNF-4, and HNF-6 were shown to bind to at least 1.6, 12, and 1.7% of the assayed promoters, respectively, in hepatocytes (27). HNF-4 stands out as being partic-

ularly important because it binds to almost 10 times as many promoters in the hepatocyte than HNF-1 and HNF-6 do. HNF-4 controls genes involved in hepatic lipid metabolism (47), thereby influencing the hepatocyte metabolome.

During vertebrate embryonic development, the liver develops as an outgrowth from the anterior primitive endoderm, which also gives rise to the adult small intestinal epithelium (for a review see Ref. 33). This embryonic relationship is reflected in the adult organs, where many gene products such as genes involved in lipoprotein synthesis are expressed in both the liver and the small intestine. The small intestinal epithelium can be divided into two parts: the villus and the crypt compartments (see Fig. 1). The epithelium covers the underlying connective tissue (called the lamina propria) to form finger-like protrusions, the villi, which point outward to the gut lumen. At the base of the villi, the epithelium continues, to line the flask-shaped crypts that penetrate into the connective tissue. The cellular dynamic of the epithelium originates from the positioning of one to four stem cells, which are situated at a few cell positions above the bottom of the crypts. The stem cells give rise to a layer of committed so-called transient amplifying cells, which are positioned at the middle and upper parts of the crypts. These transient amplifying cells undergo a few cell divisions as they migrate toward the crypt openings. At the crypt-villus transition zone, proliferation ceases and the cells differentiate (for reviews see Refs. 30, 31, 36). The absorptive enterocyte is by far the most abundant cell type in the small intestinal epithelium, and the fully differentiated enterocyte is a cell type that in many ways functionally resembles the hepatocyte. It is therefore relevant to ask to what extent the HNF transcription factors might be important for the generation of the villus-specific gene expression. Two decades of work focusing on a few selected genes lends support to the idea that members of the HNF-1 and HNF-4 transcription factor families might indeed be of importance for villus-specific gene expression (for a review see Ref. 48). The order of magnitude of the number of target genes for HNF-1 and HNF-4 in the differentiated enterocyte is, however, not known at present. It is also not known whether other transcription factors might similarly drive a high number of differentiation-induced genes in the villus enterocyte and thereby be just as important for the villus gene expression.

It was the purpose of the present work to determine on a genome-wide scale which transcription factor binding sites are the most common in the promoters for differentiation-induced
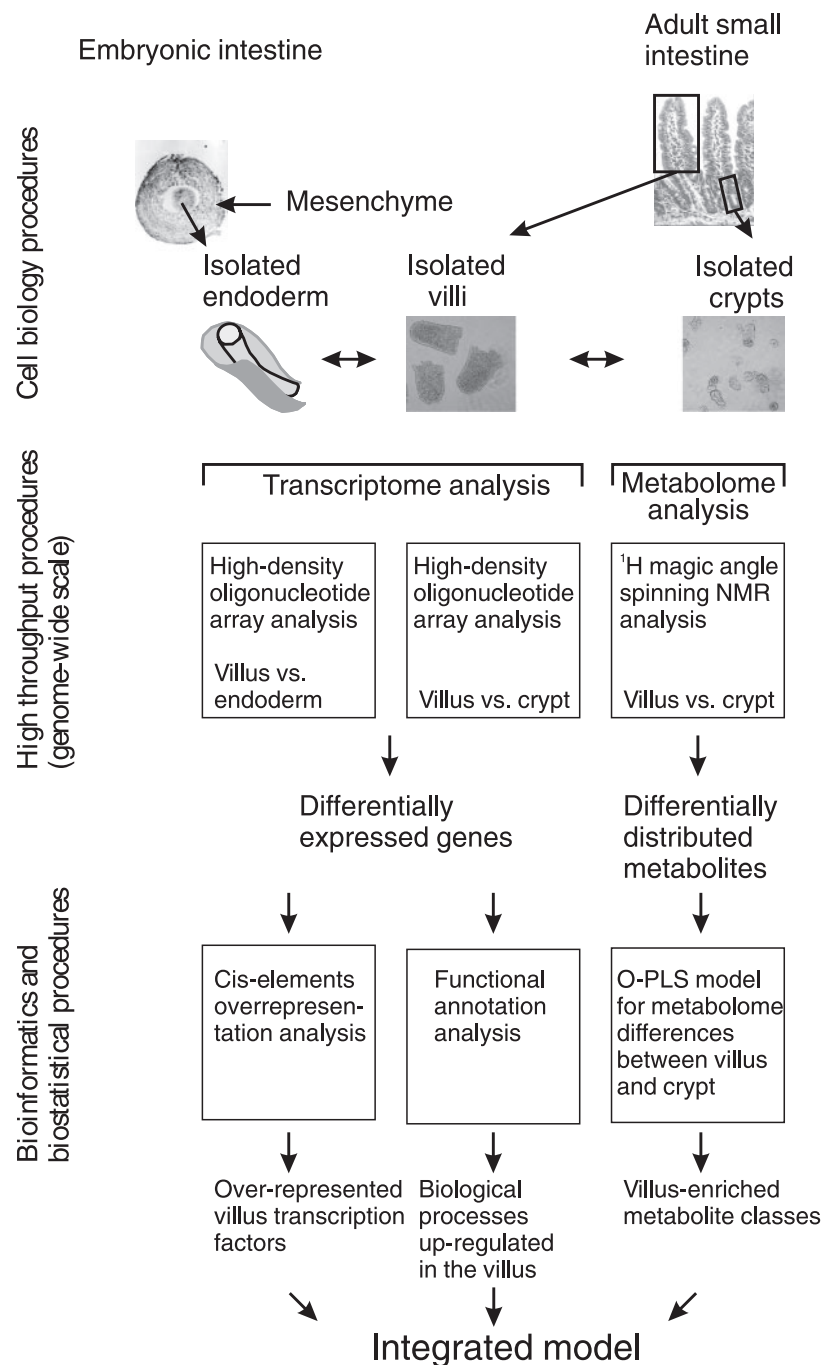
Fig. 1. Overall experimental strategy. Established cell biology procedures were used to isolate mouse embryonic *day 13* endoderm and adult mouse crypt and villus epithelium. Hybridization probes were generated from the extracted RNA and used for transcriptome analysis using Affymetrix MOE 430 A 2.0 high-density oligonucleotide arrays. Metabolome analysis was carried out on crypts and villi by $^1$H magic-angle spinning NMR spectroscopy. Genes with differential expression between villi and crypts and between villi and endoderms were identified and used to generate gene lists, which were subsequently used for promoter cis-element overrepresentation analysis and functional annotation analysis for biological processes. $^1$H NMR spectra were compared with identify metabolites with higher concentration in villi than in crypts. To evaluate the significance of the differences observed in the NMR spectra, a multivariate model was built using orthogonal partial least-squares (O-PLS) regression. The outcome of this overall systems biology approach was intended to be a model that integrates the findings.

genes during fetal-to-adult and crypt-to-villus differentiation of small intestinal epithelial cells. Another purpose was to investigate whether connections between the enterocyte metabolome and the investigated transcription factors might exist.

Transcriptome data were collected from embryonic mouse endoderm, adult mouse crypt, and adult mouse villus epithelium by high-density oligonucleotide array analysis. Metabolome data were collected from adult mouse crypt and villus epithelium by magic angle spinning $^1$H nuclear magnetic resonance (NMR) spectroscopy, a technique that can provide detailed molecular information about a wide range of metabolites in small amounts of intact tissue (45, 46). To identify overrepresentation of potential transcription factor binding

sites in the promoters controlling genes with a differentiation-dependent expression, a bioinformatic algorithm was applied.

Our results point to HNF-4 as a critical regulator of the villus specific gene expression because potential HNF-4 binding sites are found in a high fraction of the promoters that control upregulated genes during development and during crypt-to-villus differentiation. Analysis of the villus metabolome revealed the presence of higher concentrations of lipid carbon chains in the villi than in the crypts. This finding led us to formulate a model in which HNF-4 indirectly controls the concentration of lipid carbon chains in the villi by regulating genes involved in lipid metabolism. Finally, our results also provide information about transcription factors that regulate

crypt-specific functions. We have identified both a c-Myc crypt transcription factor node, which is presumably associated with epithelial stem cell maintenance, as well as an E2F gene regulatory node, which is presumably associated with crypt cell proliferation.

## METHODS

*Isolation of mouse intestinal tissues.* The protocol involving experimental animals conformed to the rules concerning review and approval by the committee for experimental animals under the Danish Ministry of Justice. C57BL/6 mice were kept on a standard rodent diet and fed ad libitum. Animals were killed by cervical dislocation. Rapid access to the abdominal cavity was achieved by use of surgical scissors, the ileum was dissected out, and a 10-cm segment was cut free and immediately placed in ice-cold PBS. The intestinal segments were flushed with ice-cold HBSS [3.3 mM $Na_2HPO_4$, 4.1 mM $NaHCO_3$, 136.8 mM NaCl, 0.44 mM $KH_2PO_4$, 5.3 mM KCl, 5.5 mM D(+)-glucose] adjusted to pH 7.2. DTT was added to 0.5 mM just before use. Isolation of crypts and villi was performed according to the procedure by Flint et al. (15) with some modifications: A plastic rod (diameter 3 mm and length 115 mm) was gently introduced ~5 mm into the lumen of the intestinal segment, which was fixed to the rod using 3-0 suture. The rest of the intestinal segment was inverted onto the remaining free part of the plastic rod and fixed at the other end with 3-0 suture. The inverted intestine on the plastic rod was incubated overnight (4°C, 15 h) in chelating buffer (27 mM Na-citrate, 5 mM $Na_2HPO_4$, 96 mM NaCl, 8 mM $KH_2PO_4$, 1.5 mM KCl, 55 mM D-sorbitol, 44 mM, 0.5 mM DTT) adjusted to pH 7.2. All of the following manipulations were performed at 4°C. The plastic rod with the inverted intestine was placed in fresh chelating buffer in a 15-ml plastic centrifuge tube with a screw cap. The tube was fixed with a clamp that inserted into a motor for a Potter-Elvehjem homogenizer. The motor was adjusted to a speed of 1–2 rpm, allowing the tube to be continuously inverted. Initially, the chelating buffer was collected every 30 min, and the released villi was inspected by phase contrast microscopy. The first fractions, which were dominated by intact villi, were pooled, washed once in PBS, pelleted (800 $g$, 5 min), snap-frozen in liquid $N_2$, and stored as the villus fraction. The rotation and collection of fractions were continued for 8–10 h until very few cells were released into the new fractions. Crypts were subsequently released by tapping the centrifuge tube hard into a lab dish three to four times. The released cells were harvested by centrifugation and washed once in PBS, and the pellets were stored frozen in liquid $N_2$.

Intestinal endoderms and mesenchymes were isolated from 13-day C57BL/6 mouse fetuses by dissection of collagenase-treated intestines as described in detail previously (12, 22).

*Histological procedures.* Mouse ileal segments were placed in 4% paraformaldehyde (4°C, 16 h) and subsequently in 60% ethanol (4°C) until embedding. The tissue segments were embedded in paraffin, sectioned, and stained with hematoxylin and eosin according to standard histological procedures. Rehydrated paraffin sections were boiled for 10 min in 10 mM Na-citrate, pH 6.0. The heating was turned off, and the buffer was allowed to reach room temperature. After the antigen retrieval procedure, the sections were incubated for 30 min in blocking buffer (50 mM Tris·HCl pH 7.4, 150 mM NaCl, 0.5% ovalbumin, 0.1% gelatine, 0.2% teleostean gelatine, 0.05% Tween 20) and incubated with a 1:50 dilution of a polyclonal anti-HNF-4 antibody (SC-8987, Santa Cruz Biotechnology) in blocking buffer overnight. The sections were washed three times for 10 min each in blocking buffer and incubated for 30 min at room temperature with a 1:100 dilution of an Alexa-488-conjugated goat anti-rabbit antibody (Invitrogen). After three washes in PBS, the sections were mounted for fluorescence microscopy.

*Chromatin immunoprecipitation.* Villus epithelial cells were isolated from the ileum of five C57BL/6 mice as described above. The villus cells were pooled, pelleted (1,000 $g$, 5 min), and resuspended in 10 ml of minimal essential medium. The resuspended cells were allowed to equilibrate to room temperature for 10 min. We added 280 µl of 37% formaldehyde, and fixation was allowed to proceed for 30 min at room temperature with gentle shaking. The fixation was stopped by the addition of 540 µl of 2.5 M glycine. After the harvest (4,000 $g$, 10 min) of fixed villus cells, sonication and immunoprecipitation with the HNF-4 antibody (SC-8987, Santa Cruz Biotechnology) were performed exactly as described previously (28). The amount of immunoprecipitated promoter DNA was measured by quantitative real-time PCR. The primers were designed to amplify 130- to 150-bp regions including the predicted HNF-4 binding site in the *Apoa4*, *Numb*, *Anpep,* and *Mep1a* promoters, respectively. In addition, primers were designed for a region in the *Cd24a* promoter, which does not have a predicted HNF-4 binding site. The primer sequences, the sequences of the amplified regions, and the predicted HNF-4 binding sites for the promoters can be found in Supplementary Table 1 (the online version of this article contains supplemental data). All amplified promoter regions were sequenced to verify their identity. For quantitative real-time PCR, the LightCycler FastStart DNA Masterplus SYBR green I system (Roche) was applied. Reactions were assembled in LightCycler capillary tubes (Roche), and 5 µl of purified immunoprecipitated DNA were used as template. Melting curves were routinely inspected to rule out the presence of unrelated amplified DNA in the real-time PCR reaction.

*Cloning and analysis of the Mep1a promoter.* The region from position −668 to +11 (from the February 2006 assembly of the mouse genome) surrounding the *Mep1a* gene was amplified using 0.5 µg of mouse (C57BL/6) tail DNA as template in a standard PCR reaction. The primers used were 5′-TTGGCTAGCACCCTTTCCCT-GCTTTGTTT-3′ and 5′-TGCAAGCTTCCTATTGGACCTTGCTC-TCA-3′ carrying 5′-extensions with *Nhe*I and *Hin*dIII restriction sites (underlined). The sequence-verified promoter fragment was cloned into the pGL3-basic vector (Promega Biotech) in front of the firefly luciferase gene using *Nhe*I and *Hin*dIII as cloning sites. To analyze the responsiveness of the *Mep1a* promoter to HNF-4a, the *Mep1a* promoter/luciferase construct was cotransfected with the CMVLacZ internal control vector, with or without the rat HNF-4a expression vector, into HeLa cells. As a positive control for HNF-4 responsiveness, the human intestinal alkaline phosphatase promoter was used (28). The culture of HeLa cells, cotransfection with the rat HNF-4a expression vector, and measurements of luciferase and β-galactosidase were performed exactly as described previously (28).

*RNA extraction, hybridization probe preparation, and GeneChip hybridization.* Total RNA was isolated using the RNeasy kit (Qiagen, Hilden, Germany). Frozen intestinal tissue pellets were lysed directly in lysis buffer, and the RNA isolated with the Qiagen column was digested, on-column by DNase I according to the manufacturer's protocol (Qiagen). First-strand cDNA was synthesized from 5 µg of total RNA by incubation (42°C, 1 h) in a 20-µl reaction volume containing 2.5 mM T7-(dT)$_{24}$ primer, 50 mM Tris·HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, and 500 mM dNTP, 10 units/ml Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA) . Second-strand cDNA was synthesized directly by adding 91 µl of RNase-free water, 30 µl of 5× second-strand reaction buffer (Invitrogen), 3 µl 10 mM dNTP, 1 µl *Escherichia coli* DNA ligase (10 U/µl), 4 µl *E. coli* DNA polymerase I (10 U/ml), and 1 µl *E. coli* RNase H (2 U/ml) followed by incubation (2 h, 16°C). The ends of the double-stranded cDNA were polished using T4 DNA polymerase (20 units, 5 min, 16°C). The cDNA was purified and concentrated by phenol-chloroform extraction and ethanol precipitation. Generation of biotin-labeled RNA was accomplished by in vitro transcription with T7 RNA polymerase using the BioArray High Yield RNA transcript labeling kit (Enzo LifeSciences). Biotin-labeled cRNA was subsequently purified from the transcription reaction using the RNeasy system (Qiagen). Hybridization of biotin-labeled cRNA to MOE430A 2.0 GeneChips, washing, staining, and scanning were performed according to the protocols published by the manufacturer (Af-

fymetrix). Six MOE430A 2.0 GeneChip hybridizations were performed with crypt- and villus-derived RNA (three with villus probes and three with crypt probes). To achieve sufficient amounts of RNA for GeneChip analysis, endoderms and mesenchymes were isolated from 79 embryos and grouped into four separate pools for RNA extraction. Four independent GeneChip hybridization experiments were performed with both the endodermal probes and with the mesenchymal probes.

*Expression level comparisons.* Summarization of probe-level data from the scanned GeneChips into single normalized gene expression measures for each probe set was performed by the robust multiarray analysis (RMA) procedure (20). The calculations were performed using the implementation of RMA provided by the open source bioconductor project (http://www.bioconductor.org) (16). The difference between the mean crypt and mean villus expression measure was calculated for each probe set, and the significance was evaluated by an unpaired Student *t*-test using standard statistical calculations (4). Similar comparisons were performed for the endoderm and villus expression measures. The calculated *P* values were stored in a table together with the mean expression measure values for each probe set. To classify genes according to the abundance of their transcripts in intestinal cells, an expression measure of eight was chosen as the upper limit for low-abundance transcripts; an expression measure of 10 was chosen as the lower limit for a high-abundance transcript. With these limits, ~8% of the probe sets had expression measures corresponding to high copy number transcripts, and probe sets with expression measures between 8 and 10 corresponding to transcripts with intermediate copy numbers constitute 17% of the probe sets. Seventy-five percent of all probe sets had expression measures corresponding to low copy number transcripts. Clearly, a large fraction of the probe sets with expression measures below eight will represent transcripts that are not expressed at all in the small intestinal epithelium. Our experience from performing RT-PCR on mRNA extracted from the mouse small intestinal epithelium suggests that an RMA calculated expression measure of five in most cases represents a gene that cannot be amplified by RT-PCR from the RNA sample that was used for the GeneChip analysis. The calculated expression measures have been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under the series accession number GSE3216.

*Functional interpretation of gene expression changes during enterocyte differentiation.* From the table with the results of the comparisons between villus and crypt expression measures (see section above), two lists of probe set IDs were generated according to the criteria: *1)* mean villus expression measure fourfold higher than the mean crypt expression measure and *P* < 0.01 for the unpaired Student's *t*-test and *2)* mean crypt expression measure fourfold higher than the mean villus expression measure and *P* < 0.01 for the unpaired Student's *t*-test. The probe set IDs were loaded into the program GoSurfer (51) and overrepresented (*P* < 0.01) gene ontology terms for biological processes visualized using the graphical output from the program. Similar calculations were performed for the villus and endoderm comparisons.

*Identification of overrepresented promoter cis-elements.* The table described above containing the mean expression measures and the *P* values form the unpaired Student's *t*-test were imported into an MySQL database server running in a 64-bit Mandrake Linux 10 environment on a personal computer equipped with an AMD 64 Athlon processor (Advanced Micro Devices, Sunnyvale, CA). Lists of genes (Supplementary Tables 2–7) with a specified significant (*P* < 0.05) difference in expression measure (for example, >10 in mean crypt expression measure and <8 in the mean villus expression measure) were generated using standard structured query language statements. To identify potential transcription factor binding sites that occur more frequently than expected by chance (i.e., they are overrepresented) in the promoters regulating the genes that change abundance classes, we used an algorithm developed by Elkon and colleagues (14). We developed our own implementation of the algorithm

in a program called PRIMO (promoter integration in microarray result organization), which is significantly faster and provides more detailed data output. In brief, the program uses a simple position weight matrix (PWM)-scoring algorithm exactly as previously described (14) to scan a target set of promoters one nucleotide at a time and on both strands in windows corresponding to the length of the transcription factor binding site described by the PWM. The target set promoters are a part of a larger promoter set of 1.1-kb sequences extracted from the mouse genome sequence (May 2004, build 33). Each promoter in the promoter set represents the mouse genome sequence from 100 bp downstream to 1,000 bp upstream of the nucleotide that aligns with the 5'-end of a transcripts from the mouse reference sequence (RefSeq) (32) collection of mouse curated transcripts. In total, 16,095 promoters were extracted using the UCSC table browser (21). Overrepresentation of promoters with hits for a given PWM in the target set in relation to the occurrence of promoters with hits in the larger promoter set was calculated by the Fisher exact test for proportions (4). For the analysis reported here, a list (Supplementary Table 8) with 65 PWMs derived from the Transfac database (50) was used. Accordingly, the *P* values reported from the PRIMO analysis have been corrected for performing 65 tests by the Bonferroni method (4). PWMs with overrepresentation of hits in the promoters for both up- and down-regulated genes (crypt vs. villus or endoderm vs. villus) were not reported. The PRIMO source codes are available upon request, and a demo version of PRIMO is available at http://gastro.imbg.ku.dk/primoweb.

*Magic-angle spinning $^1H$ NMR spectroscopy.* Fifteen samples, corresponding to eight samples of intestinal crypt cells and seven samples of intestinal villus cells, were used for $^1H$ NMR spectroscopy. Approximately 15 mg each of crypt and villus cells were packed into separate 4-mm-diameter zirconia rotors with spherical inserts and Kel-F caps. Approximately 20 μl of $D_2O$ were added to the rotor to provide filled lock. All NMR experiments were carried out on a Bruker DRX-600 spectrometer (Bruker Biospin, Rheinstetten, Germany), at 283K, operating at a $^1H$ frequency of 600.13 MHz. Samples were spun at 5 kHz at the magic angle. A total of 15 min was allowed for temperature equilibration before NMR acquisition. A standard Bruker high-resolution magic-angle spinning probe with a magic-angle gradient was employed, and the 90° pulse length was adjusted individually for each sample, having a value between 9.6 and 10 μs. A total of 128 transients were collected into 16,000 data points for each spectrum with a spectral width of 20 parts per million (ppm) and a recycle delay of 2.0 s.

Standard $^1H$ NMR spectra were acquired for each tissue using the water-suppressed NOESY1DPR (90-$t_1$-90-$t_m$-90-acq) (26). The interpulse delay ($t_1$) was 3 μs, and the mixing time ($t_m$) was 100 ms. A weak irradiation was applied on the water resonance during both the mixing time and the recycle delay.

*NMR data analysis.* $^1H$ NMR spectra were phased and baseline-corrected using XWINNMR 3.5 (Bruker). The spectra were referenced to the anomeric proton α-glucose resonance at δ5.22 (where δ = resonance interval). The continuous spectra over the range δ 0.5–8.0 were digitized into discrete resonance intervals using a MATLAB script developed in-house (Dr. O. Cloarec, Imperial College London). The region δ 4.7–5.1 was removed to avoid the effects of imperfect water suppression. In total, the digitization procedure generated 30,280 chemical shift intervals, each defining a variable. Each of these variables is referenced by its δ value (in ppm) and holds the value of the resonance signal measured. Normalization to the total sum of the spectrum was carried out on the data before data analyses. Orthogonal-partial least-squares discriminate analysis (O-PLS-DA) (41) of the NMR spectra was carried out in a MATLAB 7.0 environment with a MATLAB script developed in-house (Dr. O. Cloarec) (9). All variables were mean centered and scaled to unit variance before O-PLS-DA. The O-PLS-DA model was constructed using the NMR data as the *X*-variables and the different cell type as the *Y*-variables (9). One orthogonal component was calculated for the model to

remove the irrelevant variations in the NMR data, and one PLS component was calculated for the model. The quality of the model was described by the cross-validation parameters ($R^2 = 0.69$ and $Q^2 = 0.51$), indicating the predictability of the $Y$-matrix and the total explained variation, respectively. To visualize metabolites that discriminate crypts and villi, the average villus to crypt difference for each variable was calculated and plotted as a function of the chemical shift. In this plot, villus-enriched metabolites are represented by peaks with positive values on the ordinate (and thus pointing upward), whereas the reverse is true for crypt-enriched metabolites. To allow an estimation of the significance of the peaks in the plot, each peak is color-coded according to a scale from 0 to 1 representing the weight of the contribution of each resonance signal at a given chemical shift region to the O-PLS-DA model for the first PLS component. Thus peaks in yellow to red colors represent the metabolites that are most important for the discrimination between crypts and villi.

Gene expression and NMR data were combined in a single model by classical PLS regression (for a review see Ref. 1) using the software SIMCA-P 10.0 (Umetrics, Umeå, Sweden). The gene expression data were used as the independent variables defining the $X$-matrix, and the NMR data were used as the dependent variables defining the $Y$-matrix. A total of two components were calculated, and the model explained 83% of variances in the dataset with a predictability of 0.82. The PLS regression model predicts the dependent variables ($\delta$) from the set of independent variables (the gene expression measures). Each dependent variable (e.g., $\delta = 1.29$ ppm) is predicted by multiple regression: $\delta_{n\ ppm} = a_1 \times \text{expression}_{gene-1} + a_2 \times \text{expression}_{gene-2} + \ldots a_n \times \text{expression}_{gene-n}$, where $a_1$ to $a_n$ are regression coefficients that are calculated from the parameters derived from the PLS model. The genes with the highest positive regression coefficients have the highest positive influence on the dependent variable ($\delta$). To find genes that are positively correlated with the increased villus lipid resonances, the genes with the highest positive regression coefficient for the lipid resonance at 1.29 ppm were accordingly extracted (Supplementary Table 9) and used in a subsequent PRIMO analysis for promoter cis-element overrepresentation analysis (see above).

## RESULTS

*Experimental strategy.* The overall experimental strategy is depicted in Fig. 1. The starting point was mouse embryonic endoderm, adult crypt, and villus epithelium. Transcriptome and metabolome data were subsequently collected by high-throughput procedures and finally analyzed biostatistically and bioinformatically to yield information about the biological processes and metabolites that are upregulated during the differentiation of immature intestinal epithelial cells. Information about transcription factors that might be important in mediating these differences was also obtained. Validations were carried out at the single gene level by immunocytochemistry, chromatin immunoprecipitation, and transfection experiments to support the high-throughput studies. A model that integrates the findings was finally generated.

*Generation of quantitative genome-wide endoderm, crypt, and villus gene expression data.* Gene expression data were obtained by Affymetrix high-density oligonucleotide array analysis. To allow easy and meaningful mining of our expression data, we constructed a public resource in the form of two databases with web access: one database for the crypt-villus gene expression data (MouseCVDB: http://gastro.imbg.ku.dk/mousecv/) (Fig. 2) and one database (FETALINTDB: http://gastro.imbg.ku.dk/fetalint/) for the endoderm gene expression data that are presented together with gene expression data from its mesenchymal counterpart.

To evaluate the overall quality of the hybridization results, we took advantage of our published crypt-villus in situ hybridization database (29) that stores information of previously reported intestinal in situ hybridization experiments. The expression patterns of genes represented in both databases were compared. Probe sets representing 56 genes in the crypt-villus in situ hybridization database are present on the MOE 430 A 2.0 GeneChip array used. In summary, 47% of the probe sets representing transcripts considered to be crypt-specific by in situ hybridization and 69% of the probe sets representing transcripts considered to be villus-specific by in situ hybridization showed the expected tendency in the differences in their mean crypt and villus expression measures calculated from the GeneChip hybridizations. Moreover, the majority of the probe sets that did not show the expected difference in their mean crypt and villus expression measures had small expression measures that were not significantly different. The signals from these probe sets are presumably below the detection threshold for the GeneChip hybridization procedure.

*Overall functional interpretation of gene expression changes during endoderm-villus and crypt-villus enterocyte differentiation.* For an initial characterization of the gene expression data, genes that displayed a fourfold difference ($P < 0.01$) in expression levels, either between the endoderm and the villus epithelium or between the crypt and the villus epithelium, were identified and subjected to an analysis of gene ontology annotations for biological processes. Most differences were found between endoderm and villus. We found 1,122 probe sets to have a fourfold higher villus expression measure than endodermal expression measure; we found 1,715 probe sets to have a fourfold higher endodermal expression measure than villus expression measure. When the two lists of probe sets were analyzed for overrepresentation of specific gene ontology terms, we found that genes annotated with the gene ontology terms for the biological processes related to immune response, molecular transport, carbohydrate metabolism, and lipid metabolism were upregulated in the adult villus epithelium compared with the endoderm. In contrast, genes annotated with the gene ontology terms related to the biological processes DNA repair, organelle biogenesis, cell cycle regulation, protein, and DNA and RNA metabolism were downregulated in the adult villus epithelium compared with the endoderm (Fig. 3A). Many fewer probe sets displayed a fourfold difference in gene expression measures when we compared hybridization probes generated from either adult crypt or adult villus RNA (143 and 250, respectively). Of note, the gene ontology terms related to lipid metabolism were overrepresented in the villus-expressed genes, whereas the gene ontology terms related to the cell cycle and DNA metabolism were overrepresented in the annotations of the crypt-expressed genes (Fig. 3B).

*Combination of cis-element overrepresentation and gene expression analysis.* It has previously been demonstrated that an eukaryotic cell contains at least three classes of transcripts that differ in their abundance in the cell (8, 43), and we recently showed that this is also the case for the mouse small intestinal epithelium (38). For the bioinformatic analysis of overrepresentation of potential transcription factor binding sites in the promoters for differentially expressed genes, we chose to focus our analysis on promoters for genes encoding transcripts that change expression level from one abundance class to another

Fig. 2. The MouseCVDB web portal. The web page (http://gastro.imbg.ku.dk/mousecv/) provides query possibilities to the crypt-villus gene expression data reported in the present work. The web page is divided into 3 frames. An upper search frame, a result frame (*bottom left*), and the crypt-villus (CV) navigator frame (*right*). A search parameter (descriptive text or a gene symbol) is entered into the "search genes" field. When the button is hit, probe sets matching the search criteria are displayed below in the result frame. The output can be sorted according to gene title, probe ID, mean crypt, mean villus, the crypt-to-villus fold difference, and the *P* value of the unpaired *t*-test. For probe sets where the *t*-test is considered significant ($P < 0.05$), the descriptive text is displayed in purple. Clicking on a desired probe set changes the display of the result frame, which now gives relevant gene information with links to external databases. Once such gene-specific information is displayed, the colors of the CV navigator change to reflect the expression changes along the crypt-villus axis for the specific probe set. The CV navigator can subsequently be used for a search for genes matching the same expression pattern by clicking the "search pattern" button. The CV navigator can, in addition, be used to search a group of genes with a specified crypt-villus expression pattern: each time the crypt or villus segment in the CV navigator is clicked, it changes its level and color. Once a desired crypt-villus expression pattern is set, a list of genes fulfilling the expression pattern can be retrieved by clicking the "search pattern" button. For the sake of simplicity, the CV navigator only displays three levels, low, medium, and high expression, based on the expression measures. The original logarithmic (with the base of 2) expression measures are displayed in 2 columns; 1 for villi and 1 for crypts, together with a "fold change" column that displays the calculated fold change between the mean crypt and villus expression measures. The upper search frame also contains a field entitled "search functional terms": this allows the user to sort out the genes linked to a particular gene function, biological process, or structural component according to the principles defined by the Gene Ontology (GO) Consortium (5). Filling in a functional term as search text and hitting the associated button results in the display of a list of GO terms in the result frame. Only GO terms with associated probe sets are displayed. The appropriate term can then be followed down to the gene level by first clicking the "go to genes" link. This displays a list of probe sets annotated with the chosen GO term in the result frame. The information for the probe sets can then be followed further down to the gene level by clicking the probe set link.

during development from endoderm to villus or during crypt to villus differentiation. Three different abundance classes, corresponding to low expression, medium expression, and hi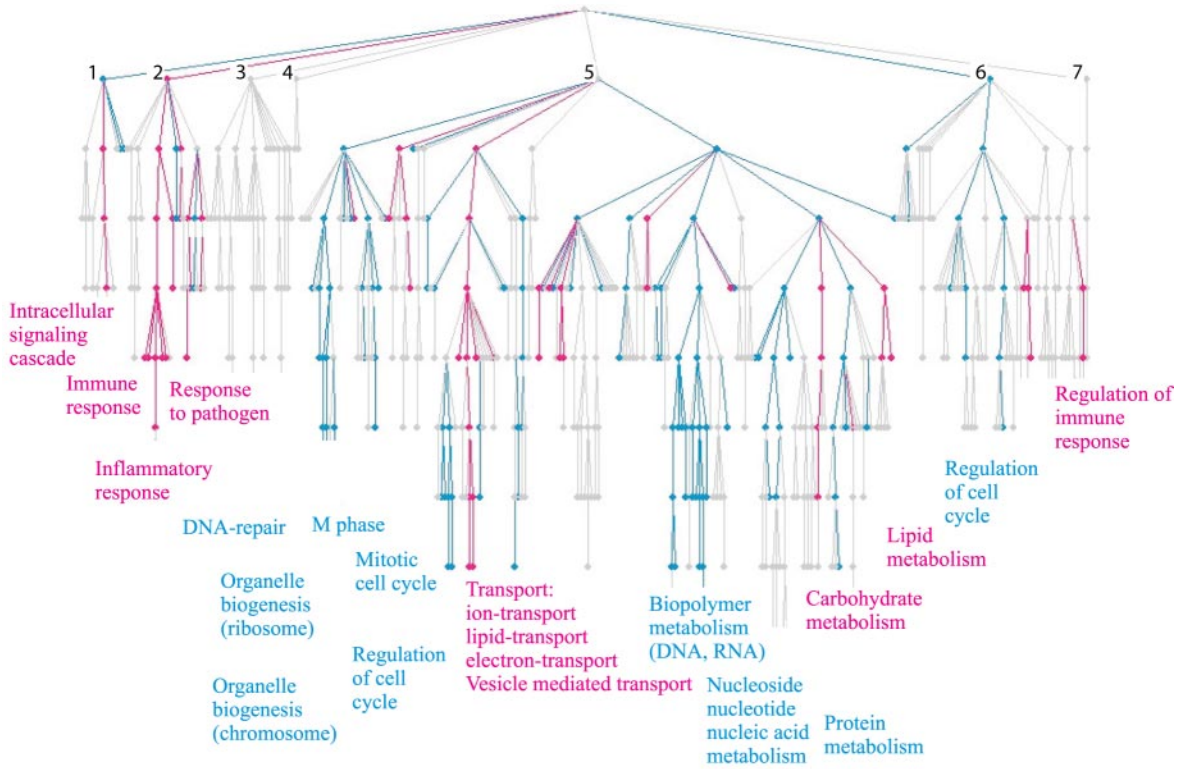gh expression, were defined on the basis of gene expression measures (see METHODS for details). We therefore concentrated on the corresponding 12 relevant gene expression patterns, and we constructed six lists of promoters controlling genes that change expression from one abundance class to another during

Fig. 3. *A*: functional interpretation of gene expression differences between embryonic *day 13* endoderm and adult villus epithelium. Lists of probe sets from the Affymetrix MOE430 A 2.0 GeneChip, which showed a 4-fold ($P < 0.01$) difference in expression measure following hybridization with probes generated with either endoderm or adult mouse villus epithelium, were generated. The lists with probe sets were compared with respect to functional annotation of biological processes defined by the GO Consortium (5). Only branches having >10 genes represented and with significant overrepresentation ($P < 0.01$) are displayed. The tree structure begins with 7 different types of processes (1, cellular process; 2, response to stimulus; 3, development; 4, growth; 5, physiological process; 6, regulation of biological process; 7, reproduction). Each branch becomes more and more detailed as one progress downward into the tree structure. Purple branches represent biological processes that are upregulated in the adult villus epithelium compared with the endoderm, and blue branches represent downregulated processes. Gray branches represent processes that are not significantly differently, distributed in crypts and villi. Representative GO terms (5) for significant branches are written below with the same color. The majority of the written terms are taken from the 4th node in each branch, and they are positioned to match the positions of the branches they represent. Some terms are found in several branches and are then only written once. The analysis and the display of the tree structure was generated using the GoSurfer software (51). *B*: functional interpretation of gene expression differences between adult crypt and villus epithelium. The same analysis was performed as described in *A*, but in this case the crypt-villus expression patterns are compared. The same number and color codes as in *A* are used. Fewer genes satisfied the criteria for differential expression (4-fold, $P < 0.01$) than for the villus and endoderm comparisons. Therefore fewer branches had 10 or more genes represented and thus fewer branches are displayed.

**A**

**Biological process
Villus versus endoderm**



Intracellular
signaling
cascade

Immune
response

Response
to pathogen

Inflammatory
response

DNA-repair

M phase

Mitotic
cell cycle

Organelle
biogenesis
(ribosome)

Regulation
of cell
cycle

Organelle
biogenesis
(chromosome)

Transport:
ion-transport
lipid-transport
electron-transport

Vesicle mediated transport

Biopolymer
metabolism
(DNA, RNA)

Nucleoside
nucleotide
nucleic acid
metabolism

Protein
metabolism

Carbohydrate
metabolism

Lipid
metabolism

Regulation
of cell
cycle

Regulation
of immune
response

**B**

**Biologicalprocess
Villus versus crypt**

Significant in villus
Significant in endoderm
Not significant



Cell cycle

Cell
division

Cellular
lipid
metabolism

DNA
metabolism

Lipid
metabolism

Generation
of precursor
metabolites and
energy

Amine
amino acid
metabolism
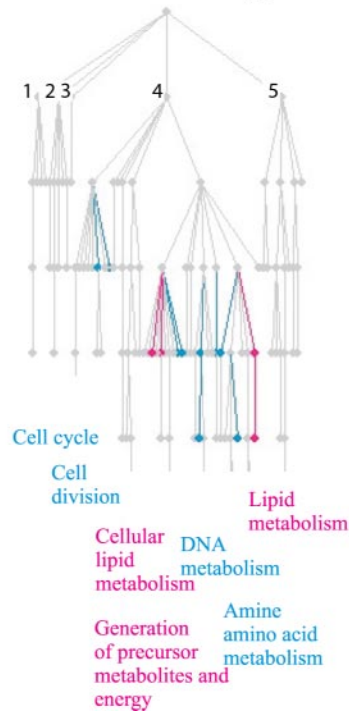
endoderm to villus development and six lists of genes that change expression pattern from one abundance class to another during crypt to villus differentiation. The genes chosen for the lists should show a shift in mRNA abundance, and the difference in expression should also be significant using an unpaired *t*-test ($P < 0.05$). We subsequently analyzed these promoter lists for overrepresentation of potential transcription factor binding sites using a search algorithm based on PWMs for transcription factor binding sites. The list of PWMs contained 65 PWMs (Supplementary Table 8) for vertebrate transcription factors and was derived from the Transfac database. Although the search algorithm was similar to a previously published algorithm (14), we used an in-house implementation that was slightly different and considerably faster.

The most significant finding was the overrepresentation of potential HNF-4 binding sites in the promoters of genes that were upregulated to a high expression level in the villi compared with the endoderms or to the crypts from adult mice (Figs. 4 and 5, Supplementary Tables 2–4). Some interesting features also arose from the analysis of the genes with lower expression in the villi compared with the endoderms or to the crypts (Figs. 6 and 7, Supplementary Tables 5–7). First, the PWM with the accession number M0050 (describing potential binding sites for the E2F transcription factor) had overrepresentation of hits in four of the six expression patterns for downregulated genes during differentiation. Second, the PWMs describing potential binding sites for the Myc transcription factor had overrepresentation of hits in the promoters of the genes that changed expression from a medium level of expression in the crypts or in the endoderm to a low level of expression in the villi. Third, PWMs describing potential binding sites for the transcription factors nuclear factor (NF)-Y, cAMP responsive element binding (CREB), and YY1 had an overrepresentation of hits in the promoters of genes that decreased expression from a medium or high level of expression in the endoderm to a lower level of expression in the villi. Finally, PWMs describing potential binding sites for STAT, ELK, and ETS transcription factors also had overrepresentation of hits in the comparisons between downregulated genes during endoderm to villus development.

*HNF-4 binds to target genes in the villus epithelium.* Immunocytochemical analysis (Fig. 8) with an HNF-4 antibody showed that the HNF-4 protein is absent from the epithelial cells located in the lower third of the crypts but expressed in the nuclei of cells located from the upper two-thirds of the crypt to the tips of the villi. Villus epithelial cells were subsequently isolated and macromolecules cross-linked with formalin. After sonication, DNA cross-linked to HNF-4 was precipitated using the same HNF-4 antibody that was used for the immunocytochemical analysis; the precipitations of specific promoter regions were analyzed by real-time quantitative PCR. Four promoters (*Apoa4*, apolipoprotein A4; *Anpep*, aminopeptidase N; *Numb*, numb gene homolog; *Mep1a*, meprin 1a) were selected from the list of genes (Supplementary Table 10) that both are upregulated during crypt to villus differentiation and contain potential HNF-4 binding sites as predicted by our search algorithm. The *Cd24a* gene, which is downregulated during crypt-villus differentiation and which does not contain a predicted potential HNF-4 site in its promoter region, was selected as a negative control promoter. The *Apoa4* and *Mep1a* promoter fragments were enriched in the
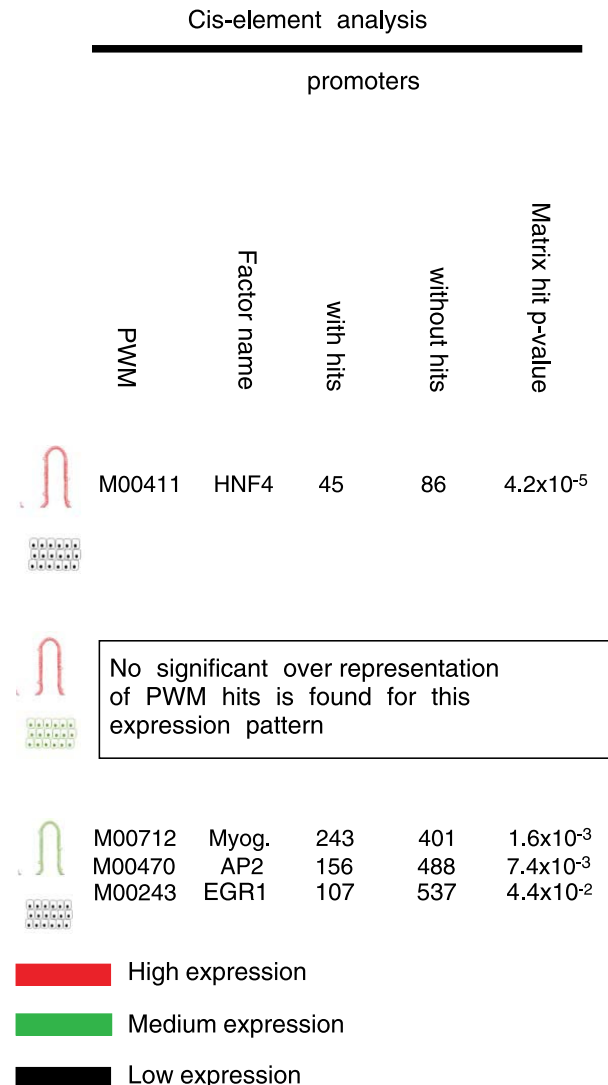


Fig. 4. Analysis of cis-element overrepresentation for genes displaying higher expression levels in the villus epithelium than in the endoderm. Lists of promoters (Supplementary Tables 2–4) for genes with the depicted expression patterns in fetal endoderm and adult villus epithelium were generated. The promoters were analyzed for overrepresentation of potential transcription factor binding sites defined by 65 selected position weight matrices (PWMs) (Supplementary Table 1) using the PRIMO program. The *P* value has been corrected for 65 tests by the Bonferroni procedure. HNF, hepatocyte nuclear factor; Myog., myogenin; AP, activating enhancer binding protein; EGR, early growth response.

HNF-4 immunoprecipitated cross-linked chromatin, both compared with the negative control *Cd24a* promoter and compared with the amounts precipitated without the primary HNF-4 antibody (Fig. 9). The *Anpep* and *Numb* promoters were not significantly enriched compared either with the negative *Cd24a* control promoter or when the primary HNF-4 antibody was omitted. The Cd24a negative control promoter itself was also not enriched in the HNF-4-immunoprecipitated chromatin compared with the control situation without the primary HNF-4 antibody. The *Apoa4* promoter is already known to be regulated by HNF-4 (3), whereas the *Mep1a* promoter has not previously been reported as an HNF-4 target promoter. We therefore also tested whether the *Mep1a* promoter was responsive to cotransfection with an expression vector for HNF-4. We

## Cis-element analysis

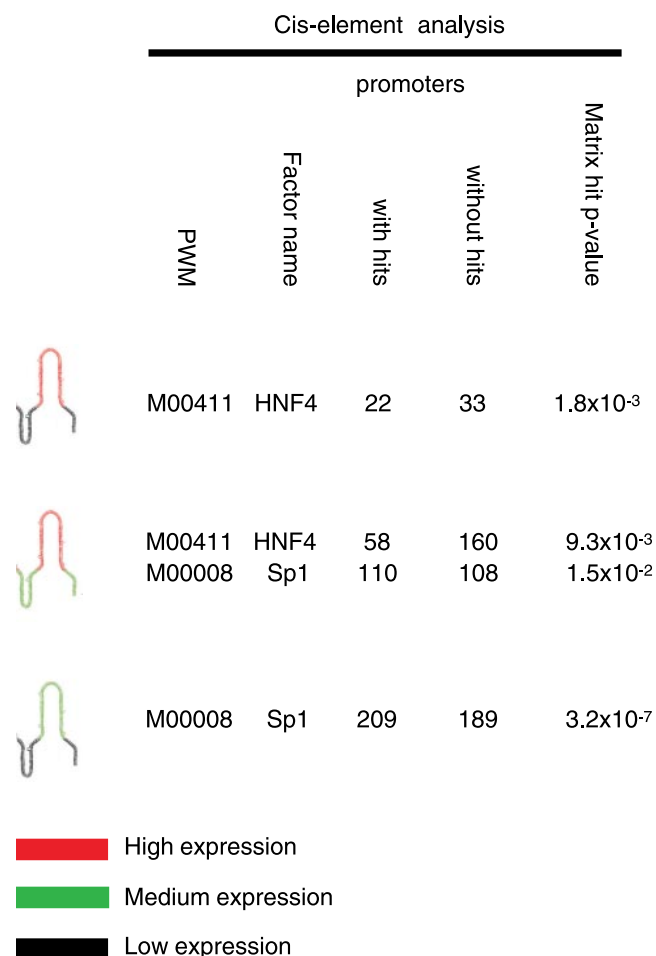| PWM | Factor name | promoters | | Matrix hit p-value |
|-----|-------------|-----------|--|--------------------|
| | | with hits | without hits | |
| M00411 | HNF4 | 22 | 33 | $1.8 \times 10^{-3}$ |
| M00411 | HNF4 | 58 | 160 | $9.3 \times 10^{-3}$ |
| M00008 | Sp1 | 110 | 108 | $1.5 \times 10^{-2}$ |
| M00008 | Sp1 | 209 | 189 | $3.2 \times 10^{-7}$ |

High expression
Medium expression
Low expression

Fig. 5. Analysis of cis-element overrepresentation for genes displaying up-regulated crypt-villus expression patterns. Lists of promoters (Supplementary Tables 2–4) for genes with the depicted expression pattern along the crypt-villus axis were generated. The promoters were analyzed for overrepresentation of potential transcription factor binding sites defined by 65 selected PWMs (Supplementary Table 1) using the PRIMO program. The *P* value has been corrected for 65 tests by the Bonferroni procedure.

used cotransfection in HeLa cells, and we have previously shown that in this system that cotransfection of an expression vector for HNF-4 activates the human intestinal alkaline phosphatase promoter (*ALPI*) 1.5- to 2-fold and that this activation depends on the presence of an HNF-4 binding site in the ALPI promoter (28). As shown in Fig. 10, the Mep1a promoter is stimulated significantly (1.8-fold) by HNF-4 cotransfection in HeLa cells; furthermore, the activation is comparable to the activation of the positive control *ALPI* promoter.

*Villus and crypt epithelial cells differ in their content of lipid metabolites.* Our analysis thus far implicated HNF-4 as a villus gene regulatory node with many connected genes in the villus enterocyte. In the liver, HNF-4 is involved in lipid metabolism (47); we therefore investigated crypts and villi for their content of lipid metabolites. Eight crypt and seven villus preparations were prepared for magic angle ¹H NMR spectroscopic analysis. Protons in a magnetic field will at the correct resonance frequency absorb energy from electromagnetic radiation. This absorption of energy can be measured in an NMR spectrometer, and the signal strength is proportional to the concentration of resonating protons in the sample. The resonance frequency

depends on the chemical environment the protons are situated in. The shift in resonance frequency for protons in a specific molecular environment compared with protons in the environment of a reference compound is referred to as the chemical shift ($\delta$) and is measured in ppm. Figure 11*A* shows a ¹H NMR spectrum generated with crypt and villus samples, respectively. For illustration purposes, two peaks are pointed out. One signal, at 1.29 ppm, is higher in the villus sample compared with the crypt sample, whereas another peak, at 3.21 ppm, is higher in the crypt sample compared with the villus sample. The signal at 1.29 ppm comes from protons in the chemical environment —($\mathbf{CH_2}$)n— (the protons giving the signal are indicated in bold) and is a signal typically obtained from lipid carbon chains such as fatty acid chains. The signal obtained at 3.21 ppm was generated by protons in the three methyl groups

## Cis-element analysis

| PWM | Factor name | promoters | | Matrix hit p-value |
|-----|-------------|-----------|--|--------------------|
| | | with hits | without hits | |
| M00050 | E2F | 36 | 137 | $5.1 \times 10^{-13}$ |
| M00287 | NFY | 67 | 106 | $7.0 \times 10^{-8}$ |
| M00179 | CREBP | 27 | 146 | $1.3 \times 10^{-2}$ |
| M00025 | ELK | 65 | 203 | $1.6 \times 10^{-4}$ |
| M00179 | CREBP | 41 | 227 | $5.5 \times 10^{-4}$ |
| M00287 | NFY | 79 | 189 | $1.8 \times 10^{-3}$ |
| M00287 | ETS | 66 | 202 | $1.2 \times 10^{-2}$ |
| M00050 | E2F | 101 | 831 | $4.6 \times 10^{-15}$ |
| M00025 | ELK | 201 | 731 | $7.5 \times 10^{-9}$ |
| M00496 | STAT1 | 144 | 788 | $8.4 \times 10^{-7}$ |
| M00287 | NFY | 241 | 691 | $6.9 \times 10^{-6}$ |
| M00793 | YY1 | 228 | 704 | $7.8 \times 10^{-5}$ |
| M00032 | ETS | 196 | 736 | $1.3 \times 10^{-3}$ |
| M00322 | MYC | 126 | 806 | $1.9 \times 10^{-3}$ |
| M00497 | STAT3 | 177 | 755 | $1.0 \times 10^{-2}$ |

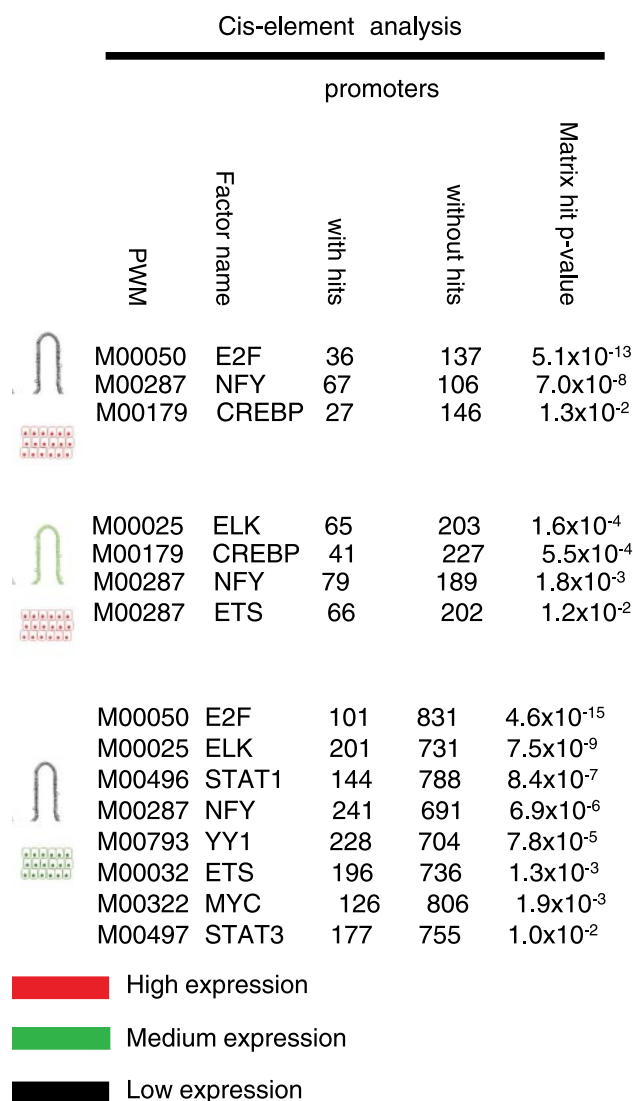High expression
Medium expression
Low expression

Fig. 6. Analysis of cis-element overrepresentation for genes displaying higher endodermal than villus expression levels. Lists of promoters (Supplementary Tables 5–7) for genes with the depicted expression pattern in fetal endoderm and adult villus epithelium were generated. The promoters were analyzed for overrepresentation of potential transcription factor binding sites defined by 65 selected PWMs (Supplementary Table 1) using the PRIMO program. The *P* value has been corrected for 65 tests by the Bonferroni procedure. NF-Y, nuclear factor Y; CREBP, cAMP responsive element binding protein.
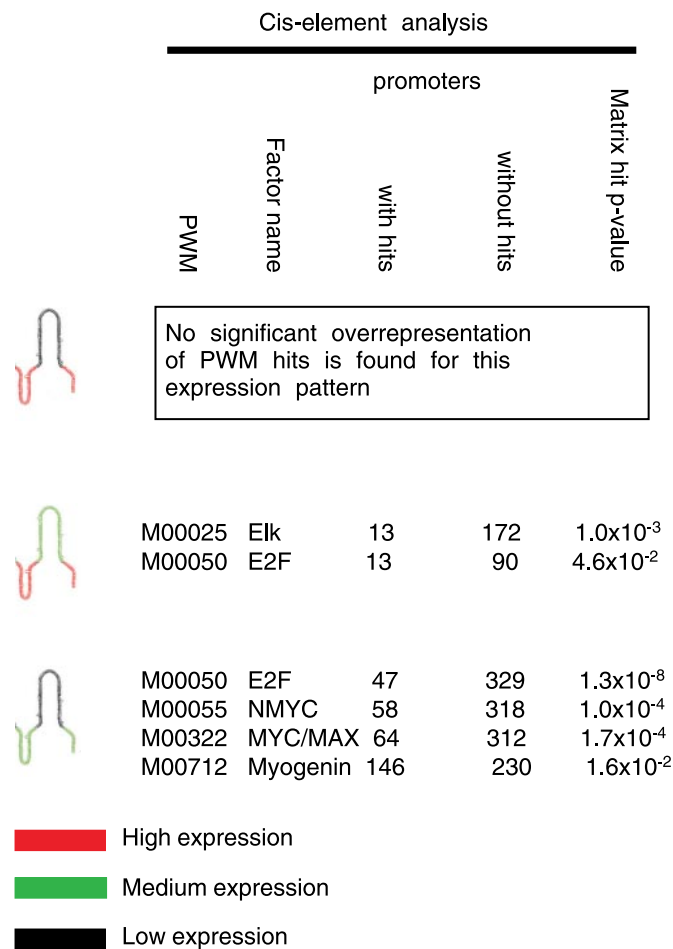
## Cis-element analysis

| PWM | Factor name | promoters | | Matrix hit p-value |
|---|---|---|---|---|
| | | with hits | without hits | |
| | | No significant overrepresentation of PWM hits is found for this expression pattern | | |
| M00025 | Elk | 13 | 172 | $1.0 \times 10^{-3}$ |
| M00050 | E2F | 13 | 90 | $4.6 \times 10^{-2}$ |
| M00050 | E2F | 47 | 329 | $1.3 \times 10^{-8}$ |
| M00055 | NMYC | 58 | 318 | $1.0 \times 10^{-4}$ |
| M00322 | MYC/MAX | 64 | 312 | $1.7 \times 10^{-4}$ |
| M00712 | Myogenin | 146 | 230 | $1.6 \times 10^{-2}$ |

■ High expression (red)
■ Medium expression (green)
■ Low expression (black)

Fig. 7. Analysis of cis-element overrepresentation for genes displaying down-regulated crypt-villus expression patterns. Lists of promoters (Supplementary Tables 5–7) for genes with the depicted expression pattern along the crypt-villus axis were generated. The promoters were analyzed for overrepresentation of potential transcription factor binding sites defined by 65 selected PWMs (Supplementary Table 1) using the PRIMO program. The *P* value has been corrected for 65 tests by the Bonferroni procedure.

respectively. The most valid resonance signals that characterize villi (positive, yellow to red peaks) are almost all related to lipid carbon chains. Thus the molecular structures —CH=CH—, =CH—$CH_2$—CH=, —CO—$CH_2$—$CH_2$—, —$CH_2$—CH=, —$(CH_2)n$—$CH_3$ can all be found in either saturated or in unsaturated fatty acids present in membrane lipids, triglycerides, or lipoproteins. Apart from lipids, more lactate is present in the villi compared with the crypts. The metabolites that characterize crypts are glucose, glycogen, and choline-containing compounds. In conclusion, the results suggest that lipids related to saturated and unsaturated fatty acid chains are present in higher concentrations in villi compared with crypts.

*Bioinformatic support for a connection between genes having potential HNF-4 binding sites in their promoters and lipid metabolites in the villus enterocyte.* The PLS multivariate analysis procedure can also be used to model metabolite data as a function of the gene expression data and thereby uncover functionality of genes. For such an analysis, the gene expression data were used as *X*-variables and the NMR data as *Y*-variables in ordinary PLS regression (for a review see Ref. 1). To find genes that are positively correlated with the increased lipid resonances, the genes with the highest positive regression coefficients with respect to the lipid resonance at 1.29 ppm (see Fig. 11) were extracted. We selected 235 probe sets in this way. Of the corresponding genes (Supplementary Table 9), 113 had a promoter represented in our database and were selected for a cis-element overrepresentation analysis. The PWM M00411, representing binding sites for HNF-4, was the only one of the 65 matrices that had a significant overrepresentation of hits in

of choline, and choline-containing metabolites are responsible for generating this peak, which is higher in the crypt samples compared with the villi samples. In Fig. 11*B*, the spectra from all samples are integrated into a single figure that shows the average difference in the resonance signal strength between the villus and crypt samples as a function of the chemical shift. Peaks representing NMR signals from metabolites with highest concentration in villi point upward (peaks with positive values), whereas NMR signals from crypt enriched metabolites point downward (peaks with negative values). O-PLS regression was used to construct a multivariate model for classification of crypts and villi samples based on the NMR spectra. Figure 11*C* displays a score plot of this O-PLS model. The model separates crypt and villus in the first dimension because the crypt samples are plotted to the left on the *x*-axis, whereas the villus samples are plotted to the right. The regression weights from the O-PLS model were used to give an estimation of the validity of each resonance peak displayed in Fig. 11*B*. Thus the peaks with yellow to red colors contribute the most to discriminate villi from crypts in the O-PLS model; they therefore reflect the most villus- or crypt-enriched metabolites,
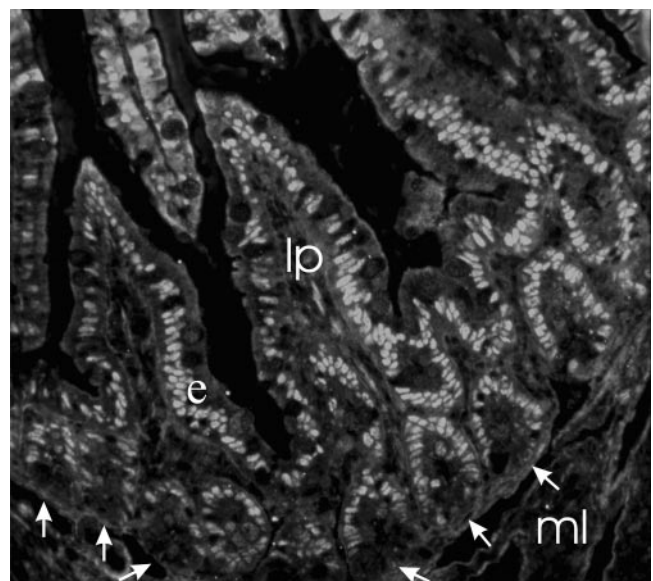


Fig. 8. Immunocytochemical localization of HNF-4 in the mouse ileal epithelium. A section of paraffin-embedded mouse ileum was stained with an anti-HNF-4 antibody, which reacts with all forms of HNF-4. Bound antibody was detected using an Alexa-488-coupled secondary goat anti-rabbit IgG antibody. Staining is absent in the lower ⅓ of the crypts (marked with arrows), whereas clear staining is seen in the epithelial nuclei from the upper ⅔ of the crypts to the tips of the villi. Abbreviations: ml, muscular layer; e, epithelium; lp, lamina propria.
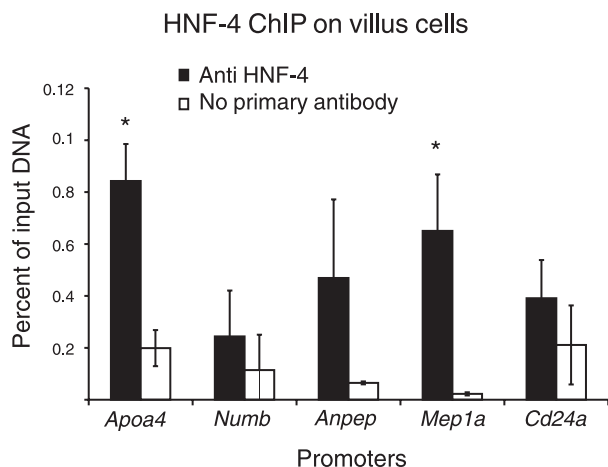
## HNF-4 ChIP on villus cells



Fig. 9. In vivo binding of HNF-4 to target genes in the villus ileal epithelium. Villus epithelial cells were isolated from 5 mice. The pooled cells were treated with formaldehyde to cross-link protein and DNA. After sonication, fragmented chromatin was incubated with (or without) an HNF-4 polyclonal antibody and immunoprecipitated using protein G-Sepharose. Enrichment of promoter DNA for specific genes was measured by quantitative PCR. The apolipoprotein A4 (*Apoa4*), numb gene homolog (*Numb*), aminopeptidase N (*Anpep*), and meprin 1a (*Mep1a*) promoters were selected for analysis from a list of genes that are upregulated during crypt-villus differentiation and that have potential HNF-4 binding sites in their promoter regions, as determined by our PWM search algorithm. The *Apoa4* promoter is also a positive control promoter because it has previously been shown to bind HNF-4 (3). The *Cd24a* promoter is a negative control promoter without a potential HNF-4 binding site. The percentage of recovered promoter DNA in the immunoprecipitated chromatin is expressed relative to the amounts of promoter fragment present in the input DNA before the addition of the HNF-4 antibody. The *Apoa4* and *Mep1a* promoters (marked by an asterisk) are significantly ($P < 0.01$) enriched in the HNF-4 immunoprecipitated chromatin when compared with the Cd24a negative control promoter and compared with the amounts precipitated without the addition of the primary HNF-4 antibody. ChIP, chromatin immunoprecipitation.

the promoters (36 promoters with hits and 77 without; $P = 0.004$ after Bonferroni correction). Thus, in villus cells, there is a correlation between the presence of villus-enriched lipids and the expression of genes that have potential HNF-4 binding sites in their promoters. Clearly some of these genes might display a correlation in their expression pattern with the concentration of lipids simply by chance, even without being involved in the metabolism of lipids. Thus an independent approach was taken to obtain additional support for a connection between potential HNF-4 binding sites and lipid metabolism in the villus enterocyte. Two lists of genes that were upregulated in the villi compared with either the crypts or the endoderm and which are annotated with the gene ontology term "lipid metabolism" were generated by the GoSurfer program (see Fig. 3). The promoters for these genes were subsequently analyzed for overrepresentation of potential HNF-4 binding sites. In both cases, a significant overrepresentation of HNF-4 binding sites was detected ($P = 2 \times 10^{-3}$ for the crypt-villus gene list and $P = 1 \times 10^{-5}$ for the endoderm-villus gene list). Thus genes that are upregulated in the villi during enterocyte differentiation and that are annotated with the term "lipid metabolism" have an overrepresentation of potential HNF-4 binding sites in their promoters.

## DISCUSSION

*Significance of a genome-wide approach to enterocyte transcriptional gene regulation.* In the present work we took a systems biology approach to enterocyte differentiation and physiology, an approach based on metabolome and quantitative gene expression data from endoderm, crypt, and villus epithelium. Transcriptional regulation of gene expression during enterocyte differentiation has previously been approached by studying single genes. Such studies formulated the hypothesis that HNF-1 and CDX2 were transcription factors that might be of a more general importance for gene expression in the differentiated enterocyte (39). Much focus has subsequently been placed on the CDX2 transcription factor, which was found to regulate the small intestine-specific disaccharidases sucrase-isomaltase (37) and lactase-phlorizin hydrolase (40). Surprisingly, potential CDX2 binding sites were not found to be overrepresented in the genes with higher villus than crypt or endodermal expression in our work. Our search algorithm did detect CDX2 and HNF-1 binding sites in the correct positions of the human sucrase-isomaltase promoter. The lack of overrepresentation of CDX2 and HNF-1 binding sites in our analyses is therefore not due to a poor PWM; rather, it is due to the low number of potential target promoters found. The sucrase-isomaltase and lactase-phlorizin hydrolase genes were not represented on the Affymetrix MOE430 A 2.0 GeneChip used in our work, and they were therefore not included in our analysis. The lack of a few genes, however, did not disturb the overrepresentation analysis. The genes upregulated from a medium level in the crypts to a high level of expression in the villi had, for example, only 16 promoters with hits for the CDX2 PWM (M00729), whereas 202 promoters were without hits. HNF-4 in contrast had 58 promoters with hits and 160 promoters without hits for its PWM. Thus a substantial increase in the number of promoters with hits would
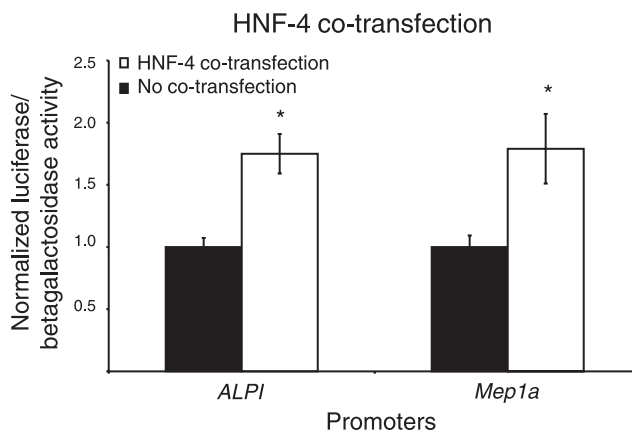
## HNF-4 co-transfection



Fig. 10. Activation of the *Mep1a* promoter by HNF-4. The mouse *Mep1a* promoter was cloned in front of the firefly luciferase gene and transfected into HeLa cells, with or without cotransfection with an expression vector for rat HNF-4a. The human alkaline phosphatase promoter (*ALPI*) was used as a positive control because it has previously been shown to be stimulated by HNF-4a cotransfection in HeLa cells and also that this stimulation depends on the presence of an HNF-4 binding site in the *ALPI* promoter (28). Cotransfection with the HNF-4a expression vector results in a significant ($P < 0.01$) and comparable stimulation of both promoters. The luciferase activity was normalized to β-galactosidase expression driven from the internal CMV-LacZ control plasmid and expressed relative to the activity obtained without HNF-4a cotransfection.

be needed to yield significant overrepresentation of CDX2 binding sites in the upregulated promoters. CDX2 is undoubtedly very important for intestine-specific gene expression, and it might also target some important regulators of enterocyte differentiation; yet it seems to be less important than HNF-4 when it comes to activating differentiation-induced genes, which carry out the physiological functions of the differentiated enterocyte. The same arguments are true for HNF-1. The PWM (M00790) used for HNF-1 can find the correctly positioned hits in known target promoters (e.g., *Anpep*), but it is again the low number of potential HNF-1 binding sites in the promoters for upregulated genes along the crypt villus axis that

explain why HNF-1 is not found to be overrepresented (21 with hits vs. 197 without hits for same list of promoters as mentioned above for CDX2).

Therefore, an important lesson from our work is that a genome-wide approach is likely to generate different conclusions concerning transcription factor importance than an approach based on a few selected model promoters. Moreover, our conclusions from the genome-wide approach are supported by our findings that c-Myc and E2F transcription factors are important for crypt cell functions (see below), as is expected from the known functions of these transcription factors.

*The c-Myc and E2F crypt and endoderm gene regulatory nodes are likely to reflect differences in cell proliferation and stem cell maintenance.* An important role of the family of E2F proteins is to regulate the cell cycle during G1/S transition and DNA synthesis (for reviews see Refs. 11, 13). The activities of the E2F transcription factors are under the regulation of pocket proteins of the retinoblastoma protein (Rb) family during these processes. The Rb and E2F interaction is in turn regulated by cyclin-dependent kinase complexes (for reviews see Refs. 2, 10). Elkon and colleagues (14) convincingly demonstrated this connection between E2F binding sites and genes involved in cell cycle control by coupling the gene ontology annotation with promoter cis-element analysis. In addition, potential binding sites for NF-Y, Sp1, and nuclear respiratory factor-1 were
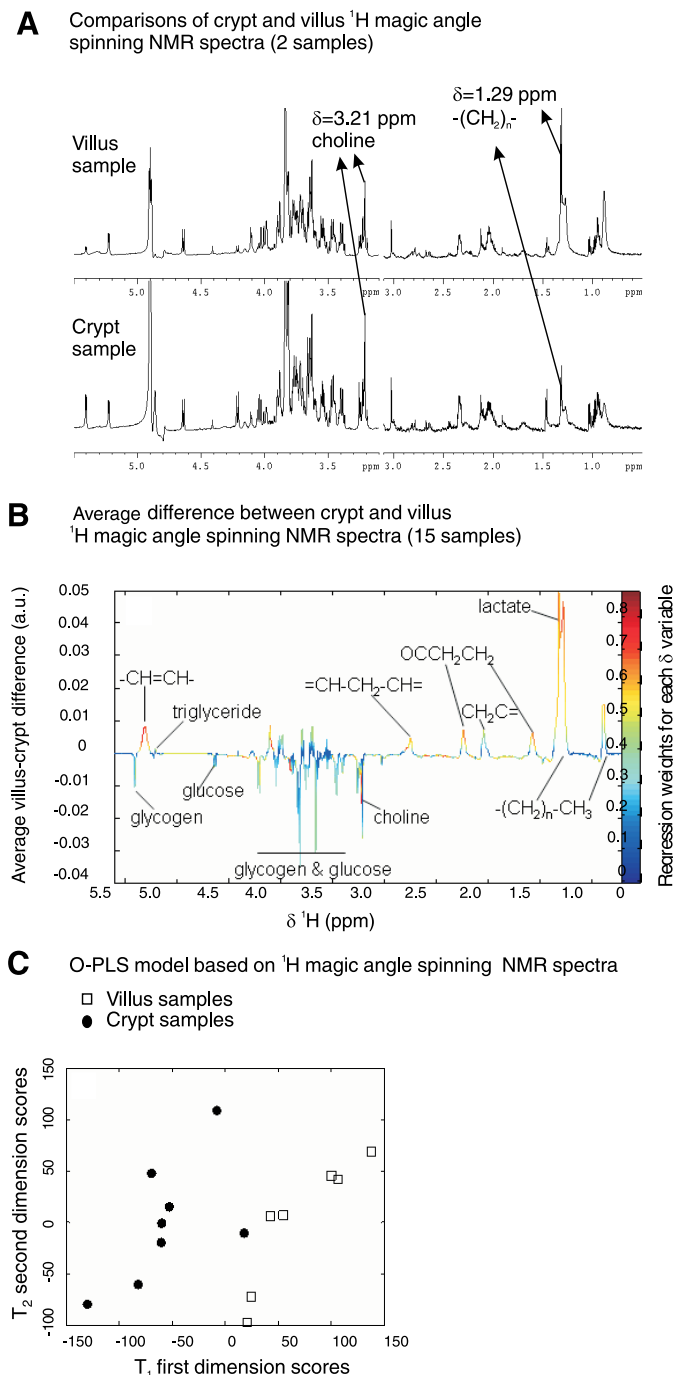


Fig. 11. *A*: $^1$H magic-angle spinning NMR spectra from crypt and villus preparations. Signal intensities are given as a function of the chemical shift ($\delta$). The 2 spectra are normalized to the total sum of each spectrum, and the peak heights reflect the concentration of the compounds containing the resonating protons. The peak pointed out at $\delta = 1.29$ ppm comes from alkane protons typically found in lipids (e.g., fatty acid chains). This signal is stronger with the villus sample compared with the crypt sample, indicating that higher concentrations of alkane protons are present in the villus sample. The peak pointed out at $\delta = 3.21$ ppm comes from the 9 protons in the 3 methyl groups of choline. This signal is stronger with the crypt sample, which indicates that higher concentrations of choline containing compounds are present in the crypt sample. *B*: metabolites discriminating between crypts and villi. The average difference between the villus and the crypt $^1$H NMR signal intensities (centered to the mean and unit variance for each $\delta$ variable) from the 15 samples used in *A* are plotted as a function of $\delta$. Positive peak values point upward and represent resonances from metabolites with higher concentration in villi than in crypts. Negative peak values point downward and represent resonances from metabolites with higher concentrations in crypts than in villi. The peaks are color coded with a scale from 0 to 1 according the validity of the crypt-villus difference in resonance peak signal intensity (see legend for *C* for details). The molecular environments for the resonating protons are indicated. Except for lactate, the most important (yellow to red peaks) resonances with positive values are related to lipids enriched in villi. Glycogen, glucose, and choline are enriched in crypts. *C*: discrimination between crypt and villus samples based on magic angle spinning $^1$H NMR spectroscopy. We sampled 30,000 $^1$H NMR signal intensity values (from the range $\delta$ 0.5–8 ) from each of 15 individual $^1$H NMR spectra obtained from 8 crypt preparations and 7 villus preparations. The extracted signal intensities were treated as *X*-variables in multivariate modeling using orthogonal partial least-squares regression (O-PLS). The *Y*-variables in the model were the 2 classes, crypt or villus. The figure shows the score plot of the individual samples on the first 2 extracted components. Thus the original projection of the samples in a space with 30,000 dimensions is transformed into a projection of only 2 dimensions. Despite the reduction in dimensions, crypt and villus samples are clearly separated. In fact, it can be seen that even plotting the samples onto the 1-dimensional *x*-axis would place all the circles representing the crypt samples to the *left* of the squares representing the villus samples. Thus the O-PLS model is able to distinguish between crypts and villi based on the $^1$H NMR spectrum. The regression weights from the O-PLS model of the resonance signals onto the first axis were used for the color code used in *B* and are an indication of the validity of the crypt or villus enrichment for the corresponding metabolites.

reported to be overrepresented in the promoters of genes annotated with the terms "cell cycle control," "mitotic cell cycle," and "DNA metabolism." Here we report the overrepresentation of potential binding sites for E2F in the promoters of genes that are downregulated during crypt to villus differentiation. In addition, overrepresentation of potential binding sites for E2F and NF-Y are found in the promoters of genes with higher expression in the endoderm than in the villus. Considering the fact that the crypts and endoderm harbor proliferative epithelial cells, the E2F and NF-Y gene regulatory nodes most likely reflect activity in the cell cycle process in the crypts and in the endoderm. This is also supported by the functional annotation analysis that showed overrepresentation of genes annotated with functions related to the cell cycle and DNA metabolism among the downregulated genes.

In the promoters of genes with higher expression in endoderms and crypts than in villi, we also find overrepresentation of potential c-Myc binding sites. The c-Myc oncoprotein is known to be a downstream nuclear target in the Wingless signaling pathway (18). The secreted Wnt proteins bind to seven transmembrane receptors (from the Frizzled family) and mediate β-catenin stabilization. The stabilization process allows β-catenin to associate with the T-cell factor (Tcf)-4 transcription factor in the small intestinal epithelial cells. The Tcf-4/β-catenin complex subsequently translocates to the nucleus and activates target genes including *Myc* (for a review see Ref. 7). Our GeneChip analysis also shows that the *Myc* probe set expression measure is 2.3-fold higher in the crypts than in the villi and 4.2-fold higher in the endoderm than in the villi (use NM_010849 as search criteria at the MouseCVDB and FETALINTDB web pages). A preferential expression of *c-Myc* mRNA and c-Myc protein in crypts was recently also reported by Mariadason and colleagues (25), further suggesting that the c-Myc gene regulatory node is indeed important in the intestinal crypt cells. An intact Wingless signaling pathway has been shown to be crucial for survival of small intestinal stem cells, since the inactivation of both the mouse *Tcf4* alleles leads to stem cell depletion and small intestinal dysfunction in early postnatal life (23). In our opinion, the c-Myc crypt gene regulatory node therefore most likely reflects Wingless signaling in stem cells and their immediate progeny.

*The villus HNF-4 gene regulatory node integrates enterocyte physiology.* Lipid carbon chain metabolites distinguished the villi from the crypts in the NMR metabolite spectra. Fat absorption takes place in the differentiated villus enterocytes and is one major metabolic difference existing between crypt and villus cells (35). The synthesis of specialized lipoproteins, the chylomicrons, is essential for fat absorption. The NMR chemical shift reported for lipoprotein (17) coincides with the villus lipid peaks reported here (Fig. 11). It is therefore likely that a higher concentration of lipoproteins was observed in villi compared with crypts. To support the hypothesis that the differences in the lipid metabolite profiles between the villi and the crypt enterocytes are related to lipoprotein synthesis, we inspected the list of genes that both have increased expression in the villi (to a medium or a high level of expression) and contain potential HNF-4 binding sites in their promoters (Supplementary Table 10). Three genes with clear relevance to chylomicron synthesis were found on this list; these are the apolipoprotein C-III (*Apoc3*) gene, the *Apoa4* gene, and the microsomal triglyceride transfer protein gene. The formation of the chylomicron precursor occurs in the endoplasmic reticulum and is followed by the transfer of triglycerides into the chylomicron precursor, a process catalyzed by the microsomal triglyceride transfer protein (for reviews see Refs. 19, 49). Apoc3 is an apolipoprotein found in mature chylomicrons, whereas Apoa4 stimulates chylomicron formation by an unknown mechanism (for a review see Ref. 42). In addition, we directly demonstrated by chromatin immunoprecipitation that the Apoa4 promoter binds HNF-4 in the villus enterocytes. These findings can therefore explain how HNF-4 in the villus enterocytes directs the villus-specific expression of genes that can establish marked differences in the lipid profile between villus and crypt cells.

The list of genes with higher villus than crypt expression and with potential HNF-4 binding sites in their promoters contained other interesting genes that might play important functional roles. Of particular note is the *Mep1a* gene, which encodes a brush border metalloproteinase (for a review see Ref. 44). In the present work, we directly demonstrate the binding of HNF-4 to the *Mep1a* promoter in villus epithelial cells and the activation of the promoter in HeLa cells by
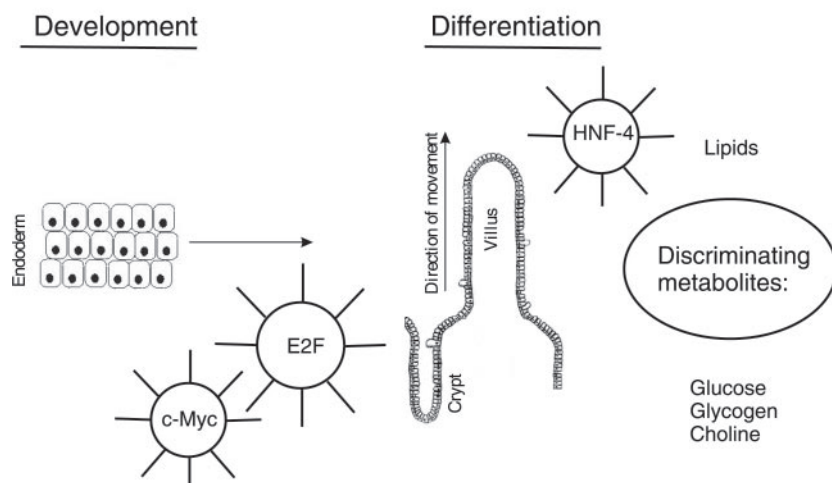


Fig. 12. Integrated model for the relationship between transcription factor nodes, epithelial cell differentiation, and metabolites in the small intestine. The c-Myc transcription factor constitutes a gene regulatory node involved in stem cell maintenance. This c-Myc gene regulatory node control genes with higher expression in endodermal and crypt cells than in villus epithelial cells. The E2F gene regulatory node also regulates genes with higher expression in the endoderm and crypts than in villi. This gene regulatory node is involved in regulating cell proliferation in the embryonic endoderm and in the proliferative crypt zone. The HNF-4 gene regulatory node regulates genes that are expressed at higher levels in villus cells than in crypt and endodermal cells. The HNF-4 villus gene regulatory node is involved in generating the physiological properties (e.g., lipid absorption and brush border hydrolytic activity) typical of the differentiated villus epithelial cell.

HNF-4 overexpression. The human lactase phlorizin hydrolase gene was recently demonstrated to be regulated by an upstream enhancer that also contained and HNF-4 binding site (24). Thus HNF-4 also affects the expression of genes involved in the extracellular hydrolysis of carbohydrates and proteins.

*Integrated model.* Figure 12 depicts our integrated model, which is the outcome of our experimental strategy depicted in Fig. 1. Three significant gene regulatory nodes with clear functions are detected by our bioinformatic cis-element over-representation analysis. The E2F and c-Myc transcription factors form the two crypt gene regulatory nodes, and these transcription factors are involved in regulating cell proliferation and stem cell maintenance. The involvement of E2F and c-Myc in these processes has already been demonstrated experimentally by others; in our work, we tie the two transcription factors to the function of the undifferentiated intestinal epithelial cells using a completely different approach. The HNF-4 transcription factor forms a villus gene regulatory node. The crypt-villus expression gradient of potentially HNF-4-regulated genes correlates with crypt-villus concentration gradient of metabolites with lipid carbon chains. Together with the functional annotation analysis, this supports the hypothesis that one consequence of HNF-4-mediated transcription in the villus enterocyte is to increase the villus content of lipids by stimulating the expression of genes involved in lipid metabolism.

### REFERENCES

1. **Abdi H.** Partial least squares regression (PLS-regression). In: *Encyclopedia for Research Methods for the Social Sciences,* edited by Lewis-Beck M, Bryman A, and Futing T. Thousands Oaks CA: Sage, p. 792–795, 2003.
2. **Adams PD.** Regulation of the retinoblastoma tumor suppressor protein by cyclin/cdks. *Biochim Biophys Acta* 1471: M123–M133, 2001.
3. **Archer A, Sauvaget D, Chauffeton V, Bouchet PE, Chambaz J, Pincon-Raymond M, Cardot P, Ribeiro A, and Lacasa M.** Intestinal apolipoprotein A-IV gene transcription is controlled by two hormone-responsive elements: a role for hepatic nuclear factor-4 isoforms. *Mol Endocrinol* 19: 2320–2334, 2005.
4. **Armitage P, Berry G, and Matthews JNS.** *Statistical Methods in Medical Research.* Oxford: Blackwell Science, 2002.
5. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, and Sherlock G.** Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29, 2000.
6. **Babu MM, Luscombe NM, Aravind L, Gerstein M, and Teichmann SA.** Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* 14: 283–291, 2004.
7. **Barker N and Clevers H.** Catenins, Wnt signaling and cancer. *Bioessays* 22: 961–965, 2000.
8. **Bishop JO, Morton JG, Rosbash M, and Richardson M.** Three abundance classes in HeLa cell messenger RNA. *Nature* 250: 199–204, 1974.
9. **Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, Lindon JC, Nicholson JK, and Holmes E.** Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in H-1 NMR spectroscopic metabonomic studies. *Anal Chem* 77: 517–526, 2005.
10. **Cobrink D.** Pocket proteins and cell cycle control. *Oncogene* 24: 2796–2809, 2005.
11. **Dimova DK and Dyson NJ.** The E2F transcriptional network: old acquaintances with new faces. *Oncogene* 24: 2810–2826, 2005.
12. **Duluc I, Freund JN, Leberquier C, and Kedinger M.** Fetal endoderm primarily holds the temporal and positional information required for mammalian intestinal development. *J Cell Biol* 126: 211–221, 1994.
13. **Dyson N.** The regulation of E2F by pRB-family proteins. *Genes Dev* 12: 2245–2262, 1998.
14. **Elkon R, Linhart C, Sharan R, Shamir R, and Shiloh Y.** Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res* 13: 773–780, 2003.
15. **Flint N, Cove FL, and Evans GS.** A low-temperature method for the isolation of small-intestinal epithelium along the crypt-villus axis. *Biochem J* 280: 331–334, 1991.
16. **Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, and Zhang J.** Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80, 2004.
17. **Hamilton JA, Small DM, and Parks JS.** 1H NMR studies of lymph chylomicra and very low density lipoproteins from nonhuman primates. *J Biol Chem* 258: 1172–1179, 1983.
18. **He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, and Kinzler KW.** Identification of c-MYC as a target of the APC pathway. *Science* 281: 1509–1512, 1998.
19. **Hussain MM, Fatma S, Pan X, and Iqbal J.** Intestinal lipoprotein assembly. *Curr Opin Lipidol* 16: 281–285, 2005.
20. **Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, and Speed TP.** Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31: e15, 2003.
21. **Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, and Kent WJ.** The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32: D493–D496, 2004.
22. **Kedinger M, Simon PM, Grenier JF, and Haffen K.** Role of epithelial–mesenchymal interactions in the ontogenesis of intestinal brush-border enzymes. *Dev Biol* 86: 339–347, 1981.
23. **Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters PJ, and Clevers H.** Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet* 19: 379–383, 1998.
24. **Lewinsky RH, Jensen TG, Moller J, Stensballe A, Olsen J, and Troelsen JT.** T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Hum Mol Genet* 14: 3945–3953, 2005.
25. **Mariadason JM, Nicholas C, L'Italien KE, Zhuang M, Smartt HJ, Heerdt BG, Yang W, Corner GA, Wilson AJ, Klampfer L, Arango D, and Augenlicht LH.** Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. *Gastroenterology* 128: 1081–1088, 2005.
26. **Nicholson JK, Foxall PJ, Spraul M, Farrant RD, and Lindon JC.** 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Anal Chem* 67: 793–811, 1995.
27. **Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, Fraenkel E, Bell GI, and Young RA.** Control of pancreas and liver gene expression by HNF transcription factors. *Science* 303: 1378–1381, 2004.
28. **Olsen L, Bressendorff S, Troelsen JT, and Olsen J.** Differentiation-dependent activation of the human intestinal alkaline phosphatase promoter by HNF-4 in intestinal cells. *Am J Physiol Gastrointest Liver Physiol* 289: G220–G226, 2005.
29. **Olsen L, Hansen M, Ekström CT, Troelsen JT, and Olsen J.** CVD: The intestinal crypt/villus in situ hybridization database. *Bioinformatics* 20: 1327–1328, 2004.
30. **Potten CS.** Stem cells in gastrointestinal epithelium: numbers, characteristics and death. *Philos Trans R Soc Lond B Biol Sci* 353: 821–830, 1998.

31. **Potten CS, Booth C, and Pritchard DM.** The intestinal epithelial stem cell: the mucosal governor. *Int J Exp Pathol* 78: 219–243, 1997.

32. **Pruitt KD, Tatusova T, and Maglott DR.** NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33: D501-D504, 2005.

33. **Roberts DJ.** Embryology of the gastrointestinal tract. In: *Development of the Gastrointestinal Tract,* edited by Sanderson IR and Walker WA. Hamilton: B. C. Decker, 1999, p. 1–12.

34. **Schlitt T and Brazma A.** Modelling gene networks at different organisational levels. *FEBS Lett* 579: 1859–1866, 2005.

35. **Shiau YF.** Lipid digestion and absorption. In: *Physiology of the Gastrointestinal Tract,* edited by Johnson LR. New York: Raven, 1987, p. 1527–1556.

36. **Stappenbeck TS, Wong MH, Saam JR, Mysorekar IU, and Gordon JI.** Notes from some crypt watchers: regulation of renewal in the mouse intestinal epithelium. *Curr Opin Cell Biol* 10: 702–709, 1998.

37. **Suh E, Chen L, Taylor J, and Traber PG.** A homeodomain protein related to caudal regulates intestine- specific gene transcription. *Mol Cell Biol* 14: 7340–7351, 1994.

38. **Tadjali M, Seidelin J, Olsen J, and Troelsen JT.** Transcriptome changes during intestinal cell differentiation. *Biochim Biophys Acta* 1589: 160–167, 2002.

39. **Traber PG.** Control of gene expression in intestinal epithelial cells. *Philos Trans R Soc Lond B Biol Sci* 353: 911–914, 1998.

40. **Troelsen JT, Mitchelmore C, Spodsberg N, Jensen AM, Noren O, and Sjostrom H.** Regulation of lactase-phlorizin hydrolase gene expression by the caudal-related homoeodomain protein Cdx-2. *Biochem J* 322: 833–838, 1997.

41. **Trygg J.** $O_2$-PLS for qualitative and quantitative analysis in multivariate calibration. *J Chemometr* 16: 283–293, 2002.

42. **Tso P, Liu M, Kalogeris TJ, and Thomson AB.** The role of apolipoprotein A-IV in the regulation of food intake. *Annu Rev Nutr* 21: 231–254, 2001.

43. **Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, and Kinzler KW.** Characterization of the yeast transcriptome. *Cell* 88: 243–251, 1997.

44. **Villa JP, Bertenshaw GP, Bylander JE, and Bond JS.** Meprin proteolytic complexes at the cell surface and in extracellular spaces. *Biochem Soc Symp* 53–63, 2003.

45. **Wang Y, Bollard ME, Keun H, Antti H, Beckonert O, Ebbels TM, Lindon JC, Holmes E, Tang H, and Nicholson JK.** Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning 1H nuclear magnetic resonance spectroscopy of liver tissues. *Anal Biochem* 323: 26–32, 2003.

46. **Wang Y, Tang H, Holmes E, Lindon JC, Turini ME, Sprenger N, Bergonzelli G, Fay LB, Kochhar S, and Nicholson JK.** Biochemical characterization of rat intestine development using high-resolution magic-angle-spinning (1)H NMR spectroscopy and multivariate data analysis. *J Proteome Res* 4: 1324–1329, 2005.

47. **Watt AJ, Garrison WD, and Duncan SA.** HNF4: a central regulator of hepatocyte differentiation and function. *Hepatology* 37: 1249–1253, 2003.

48. **Wiginton DA.** Gene regulation: the key to intestinal development. In: *Development of the Gastrointestinal Tract*, edited by Sanderson IR and Walker WA. London: B. C. Decker, 2000, p. 13–36.

49. **Williams CM, Bateman PA, Jackson KG, and Yaqoob P.** Dietary fatty acids and chylomicron synthesis and secretion. *Biochem Soc Trans* 32: 55–58, 2004.

50. **Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, and Schacherer F.** TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* 28: 316–319, 2000.

51. **Zhong S, Storch KF, Lipan O, Kao MC, Weitz CJ, and Wong WH.** GoSurfer: a graphical interactive tool for comparative analysis of large gene sets in Gene Ontology space. *Appl Bioinformatics* 3: 261–264, 2004.