Insights into developmental mechanisms and cancers in the mammalian intestine derived from serial analysis of gene expression and study of the hepatoma-derived growth factor (HDGF)

Maina Lepourcelet[1,2,*], Liqiang Tou[1,2,*], Li Cai[1], Jun-ichi Sawada[1,3], Alexander J. F. Lazar[4], Jonathan N. Glickman[4], Jessica A. Williamson[1], Allen D. Everett[5], Mark Redston[4], Edward A. Fox[1,6], Yoshihiro Nakatani[1,3] and Ramesh A. Shivdasani[1,2,†]

[1]Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[4]Department of Pathology, Brigham & Women's Hospital, 75 Francis Street, Boston, MA 02115, USA
[5]Department of Pediatric Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[6]Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
*These authors contributed equally to this work
†Author for correspondence (e-mail: ramesh_shivdasani@dfci.harvard.edu)

Accepted 5 November 2004

Development 132, 415-427
Published by The Company of Biologists 2005
doi:10.1242/dev.01579

Summary

The vertebrate intestine is a model for investigating inductive cellular interactions and the roles of epithelial stem cells in tissue regeneration, and for understanding parallels between development and cancer. We have used serial analysis of gene expression to measure transcript levels across stages in mouse intestine development. The data (http://genome.dfci.harvard.edu/GutSAGE) identify novel differentiation products, potential effectors of epithelial-mesenchymal interactions, and candidate markers and regulators of intestinal epithelium. Transcripts that decline significantly during intestine development frequently are absent from the adult gut. We show that a significant proportion of such genes may be reactivated in human colon cancers. As an example, hepatoma-derived growth factor (HDGF) mRNA is expressed prominently in early gut tissue, with substantially reduced levels after villous epithelial differentiation. HDGF expression is dramatically increased in human colorectal cancers, especially in tumors proficient in DNA mismatch repair, and thus represents a novel marker for a distinctive tumor subtype. HDGF overexpression in fetal intestine explants inhibits maturation, suggesting a role in epithelial differentiation. To investigate the molecular basis for HDGF functions, we isolated components of a nuclear HDGF complex, including heterogeneous nuclear ribonucleoproteins implicated in processing RNA. These genes are regulated in tandem with HDGF during intestine development and one factor, TLS/Fus, is commonly overexpressed in colon cancers. Tumor expression of fetal genes may underlie similarities between developing and malignant tissues, such as self-renewal, invasion and angiogenesis. Our findings also advance understanding of HDGF functions and implicate this developmentally regulated gene in RNA metabolic pathways that may influence malignant behaviors in colorectal cancer.

Key words: Developmental gene expression, Epithelial-mesenchymal interaction, Oncofetal genes, Hepatoma-derived growth factor, Heterogeneous nuclear ribonucleoprotein, Mouse

Introduction

The mammalian intestinal epithelium is arranged in villi (small bowel) or flat cuffs (colon) associated with mucosal invaginations known as the crypts of Lieberkühn. Each crypt houses a monoclonal population of stem and progenitor cells that proliferate near its base; differentiated progeny reside higher along small intestinal villi (Cheng and Leblond, 1974; Potten, 1998; Simon and Gordon, 1995). Differentiated epithelial cells arise in intimate contact with the underlying mesenchyme, and intestinal organogenesis relies on epithelial-mesenchymal interactions (Haffen et al., 1987; Kedinger et al., 1998; Rubin et al., 1992). Directed investigation of selected signaling pathways has uncovered roles for hedgehog, TGF-β family and Wnt proteins in this process (Batlle et al., 2002; Korinek et al., 1998; Ramalho-Santos et al., 2000; Roberts et al., 1995), whereas mouse gene knockouts have implicated other regulators, often serendipitously (Hentsch et al., 1996; Kaestner et al., 1997; Karlsson et al., 2000; Katz et al., 2002; Yang et al., 2001). However, the mediators of many essential cellular interactions remain unknown and understanding of intestine organogenesis is incomplete.

The fetal mouse gut endoderm, which appears as a

pseudostratified epithelium between embryonic days (E) 12 and 13, undergoes a pivotal cellular transition over the ensuing 2 days to form the first rudimentary intestinal villi (Fig. 1A) and to initiate expression of lineage-specific genes (Maunoury et al., 1992; Simon et al., 1993). We reasoned that accurate transcriptional profiles over this developmental interval could provide fundamental information about underlying mechanisms, and characterized intestinal gene expression using serial analysis of gene expression (SAGE) (Velculescu et al., 1995). Coupled with localization of gene products, the resulting mRNA profiles (http://genome.dfci.harvard.edu/GutSAGE) identify candidate regulators of cell interactions and mucosal differentiation. We also applied this resource to address longstanding questions about reactivation of fetal genes in cancer.

Cancer cells display immature features and dysregulated gene expression, with attenuation of tumor suppressors and aberrant expression of genes that are inactive in normal adult tissue. Some proteins expressed exclusively in tumors and in developing embryos, such as carcinoembryonic antigen, are onco-fetal markers that assist in clinical cancer management (Carr et al., 1997; Uriel, 1975). Besides some recent advances (Kho et al., 2004), the extent to which tumor cells recapitulate embryonic gene expression and the significance of this phenomenon are unknown. We find that transcripts whose abundance declines significantly during gut development frequently are absent from the adult intestine, and many such genes are reactivated in human tumors of the same fetal origin. Among developmentally regulated genes, we report here frequent overexpression in colorectal tumors of the hepatoma-derived growth factor (*Hdgf*) and the results of studies that investigate its cellular and differentiation functions.

HDGF was first isolated from conditioned fibroblast media as a factor that stimulated proliferation of heterologous cultured cells (Nakamura et al., 1994; Oliver and Al-Awqati, 1998). Although it is hence regarded as a growth factor, the amino acid sequence lacks a signal peptide but includes a region homologous to the high-mobility group (HMG) box of nuclear proteins (Nakamura et al., 1994) and a nuclear localization signal that is essential for its activities (Everett et al., 2001; Kishima et al., 2002). Two groups have characterized a related protein LEDGF (lens epithelium-derived growth factor) as a nuclear protein and apparent transcriptional regulator (Fatma et al., 2001). To date, insight into HDGF functions derives largely from study of cell lines, and its physiological roles in development or cell differentiation are obscure. Prompted by our demonstration of its significant regulation in the developing intestine and frequent re-expression in DNA mismatch repair (MMR)-proficient colon cancers, we identified HDGF-associated nuclear proteins and investigated HDGF functions in gut epithelial differentiation.

Materials and methods

Serial analysis of gene expression

Libraries were constructed as described (Velculescu et al., 1995) with RNA extracted from pooled fetal intestines, using *Nla*III and *Bsm*FI as the anchoring and tagging restriction enzymes, respectively. SAGE tags were extracted, filtered and tabulated using SAGE2000 software, and gene annotations were updated using the mouse gene version released on 18 August 2004. Statistical analysis was performed using Monte Carlo simulation, as described (Velculescu et al., 1995). Listed values represent the probability (P-chance) that observed differences in SAGE tag frequency might result from chance alone. Because assignment of genes to SAGE tags can be equivocal, the analytic program assigns multiple genes to some tags and scores each assignment according to the degree of confidence: transcripts with well characterized cDNA clones receive the highest score, whereas assignments based exclusively on poorly characterized or EST clones receive low scores. Our database lists the highest scoring gene assignments; in the few cases where the top two or three possible assignments received similar scores, each of these possibilities is listed. We noted ~91% concordance between computational (highest score) and manual gene assignments for ~400 tested tags. Functional classification of genes was according to Gene Ontology Consortium criteria, and genes were clustered according to patterns of temporal variation using published methods (Cai et al., 2004).

RT-PCR

Pooled RNA samples, different from those used to construct SAGE libraries, were extracted with Trizol (Invitrogen), treated with RNase-free DNaseI (Ambion) and reverse transcribed with oligo-(dT) primer. First-strand cDNA was used as the template for PCR amplification in the presence of 0.01 mCi/ml [^{32}P] dCTP with T_m of 62°C. Cycle numbers were adjusted to ensure linear amplification, and PCR products were resolved by non-denaturing polyacrylamide gel electrophoresis and detected by autoradiography. PCR primers are listed in Tables S2A,B in the supplementary material, unless reported previously (Tou et al., 2004). Glyceraldehyde 3-phosphate dehydrogenase (GAPD), human β-amyloid and mouse hypoxanthine phosphoribosyl transferase (HPRT) mRNAs verified equal sample loading.

In situ hybridization

RNA in situ hybridization was performed on 6-10 μm paraffin sections of paraformaldehyde-fixed whole embryos or isolated fetal or adult intestine, using digoxigenin-labeled (Roche) riboprobes. Rehydrated tissue sections were treated successively with 50 μg/ml proteinase K, 0.2 M HCl, 0.1 M triethanolamine in 0.25% acetic anhydride and 2×SSPE. Slides were incubated in hybridization buffer containing 50% formamide for 4 hours at 60°C, then hybridized overnight with the probe at 60°C, washed in 0.2×SSC at 60°C, blocked for 2 hours at room temperature with 10% goat serum and 2 mg/ml bovine serum albumin diluted in PBS containing 0.1% Triton-X100 (PBT), and incubated overnight at 4°C with alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche). After washing, slides were treated with 5 mM levamisole (Sigma), rinsed in alkaline buffer, treated with NBT/BCIP solution (Roche), post-fixed in 4% paraformaldehyde, counterstained with Methyl Green, and examined by light microscopy. Images were captured on a CCD camera (QCapture) with Photoshop7.0 software (Adobe).

Immunostaining

Trp53 (DO-1, Santa Cruz), Ki67 (NovoCastra), HDGF (Everett et al., 2001), hnRNPK (gift of K. Bomsztyk, University of Washington), TLS/Fus (gift of D. Goodman, Vollum Institute, Oregon) and other antibodies were used to stain 7 μm paraffin wax embedded sections of human colorectal tumors, mouse fetal intestine or cultured HeLa cells. Sections were rehydrated, fixed further in 4% paraformaldehyde in PBS, and treated successively with warm 0.01 M citrate (pH 6), 40% methanol in 0.3% hydrogen peroxide, and PBS containing 10% goat and 10% fetal bovine sera or 5% milk. Slides were incubated sequentially with antibodies diluted in PBS containing 5% goat serum, horseradish peroxidase-conjugated anti-mouse (Amersham Pharmacia) or anti-rabbit (Santa Cruz Biotechnology) IgG and diaminobenzidine peroxidase substrate (Vector Laboratories). Mounted sections were examined for image capture as described above. Cultured cells were fixed in 4% paraformaldehyde for 15

minutes, permeabilized in 0.1% NP-40 for 5 minutes, and blocked with 3% bovine serum albumin (BSA) in PBS containing 0.1% Triton (PBST). Antibody incubation was in PBST with 3% BSA and fluorophore-conjugated secondary antibody (Jackson ImmunoResarch Laboratories) incubation in PBST for 30 minutes. Cells were washed in PBST, mounted in aqueous medium containing DAPI, and confocal images were captured on the LSM510 imaging system (Zeiss).

Organ explant and cell culture

Intact mouse fetal intestines were dissected and cultured as described (Tou et al., 2004), over filter disks in Fitton-Jackson-modified BGJb medium (Life Technologies) containing 0.1 mg/ml ascorbic acid. Explants were maintained in 5% CO_2 atmosphere, with medium changed every 2 days. pK7 plasmid DNA (2 mg/ml) encoding green fluorescent protein (GFP)-tagged full-length or truncated HDGF (Everett et al., 2001) was flushed into the gut lumen by capillary injection, followed by placement between platinum electrodes and delivery of three 10 ms pulses of 80 V each using a BTX830 square-wave electroporator (BTX, San Diego, CA), washing and culture.

HeLa cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 50 μg/ml streptomycin. Apoptosis was induced by treating the cells for 16 hours with 2 ng/ml human TNFα and 35 μM cycloheximide (Sigma). Necrosis was induced by treatment for 16 hours with ionomycin (Sigma) in serum-free DMEM. For permeabilization, cells was washed with ice-cold buffer containing 20 mM HEPES (pH 7.3), 110 mm K acetate, 5 mM Na acetate, 2 mM Mg acetate, 1 mM EGTA, 2 mM DTT and protease inhibitors, incubated on ice in the same buffer with 0.1% NP-40, washed twice with PBS and continued for immunostaining.

Isolation of HDGF-associated proteins

The detailed purification procedure has been described previously (Nakatani and Ogryzko, 2003). Briefly, HeLa cells stably expressing Flag and HA epitope-tagged HDGF were propagated in suspension. Nuclear extracts were isolated from 12 L cultures and HDGF complexes were purified successively with anti-Flag and anti-HA antibodies (Sigma). Purified complexes were separated by 4-20% gradient SDS-PAGE and stained with silver stain. Individual protein bands were excised, digested with trypsin and analyzed by tandem mass spectrometry (Taplin Biological Mass Spectrometry Facility, Harvard Medical School). Tryptic peptides were matched to proteins by the Sequest database-searching program. For glycerol gradient sedimentation, concentrated HeLa or mouse E13.5 intestine nuclear extracts were loaded over 5 ml of a 10-40% glycerol gradient and centrifuged for 10 hours at 236,000 ***g*** (Beckman SW 55 Ti rotor). Individual 100 μl fractions were resolved by SDS-PAGE and analyzed by immunoblotting.

Results

SAGE analysis of mouse gut development

SAGE sequence tags (14 bp) identify individual transcripts with high accuracy and the frequency of individual SAGE tags reliably reflects abundance of the corresponding mRNA in vivo. SAGE also reveals novel transcripts that may escape gene prediction or isolation (Chen et al., 2002; Moreno et al., 2001), and the results are readily disseminated to assist hypothesis-driven investigation. We applied SAGE to identify transcriptional changes that accompany development of the mouse small intestine at post-coital days (E) 12, 13 and 15. Intestinal morphology between E12 and E13 is similar (Fig. 1A), and their expression profiles may be enriched for transcripts that regulate the subsequent villous transformation. We identified over 65,000 SAGE tags representing over 10,000 unique transcripts from each tested developmental stage. Analysis of the E15 data revealed nearly every tissue-enriched mRNA reported in the intestine (data not shown), indicating an adequate sequencing depth. Because there can be ambiguity in assigning transcripts to individual SAGE tags, we checked manually the annotation of transcripts corresponding to ~400 SAGE tags and found 91% concordance between the manual and highest-scoring computational assignments (see Materials and methods). The data thus provide good representation of the relative tissue abundance of currently annotated transcripts.

Thirty-eight out of 40 transcripts tested independently by reverse transcription (RT)-PCR showed expression changes concordant with the SAGE data (represented in Fig. 1B), as did eight out of nine independent genes by northern analysis (represented in Fig. 1E), attesting to the fidelity of SAGE results. As an example, *Gli1* mRNA, a target and effector of hedgehog signaling (Marigo et al., 1996), is significantly downregulated after E13 (Fig. 1B), consistent with reported expression dynamics (Ramalho-Santos et al., 2000) and with prior evidence for obligate attenuation of hedgehog signaling in vertebrate gut differentiation (Zhang et al., 2000). Similarly, the levels of *Fabpi* (*Fabp2* – Mouse Genome Informatics), *Fabpl* (*Fabp1* – Mouse Genome Informatics), *Cdx2* and other mRNAs vary in agreement with published results (see Table S1 in the supplementary material; and data not shown). Our complete dataset (http://genome.dfci.harvard.edu/GutSAGE), annotated with comprehensive functional and genetic information, thus provides a unique resource to outline the composition of molecular pathways active in mammalian intestine organogenesis.

Insights into gut development derived from SAGE expression profiles coupled with transcript localization

We record 703 significant (P<0.015) temporal changes in transcript abundance, revealing notable modulation of gene activity as the epithelium differentiates (http://genome.dfci.harvard.edu/GutSAGE). A representative amount of this analysis (see Table S1 in the supplementary material) outlines prominent transcriptional alterations that accompany the villous transition in intestine development. Expression of selected adult epithelial markers is known to occur by E15 (Maunoury et al., 1992; Simon et al., 1993). Numerous transcripts associated with mucosal architecture (e.g. cadherin 17 and transgelin), metabolic or secretory functions (e.g. lipid-binding proteins and cryptdin peptides), and surface markers of cell maturation (e.g. galactose-binding lectins) are detected concomitantly. To capture changes occurring in all major cellular compartments, we prepared SAGE libraries from unfractionated tissue. Moreover, the intestinal long axis undergoes regional patterning during development, whereas our profiling considered the small bowel as a single organ and excluded the cecum and colon. Consequently, the expression data alone do not identify sites of gene expression within the developing gut. For genes of particular interest, this limitation is readily overcome by complementary approaches, including RNA in situ hybridization, to localize transcripts. Fig. 1C illustrates the value of these methods to map spatiotemporal patterns of fetal gene expression.

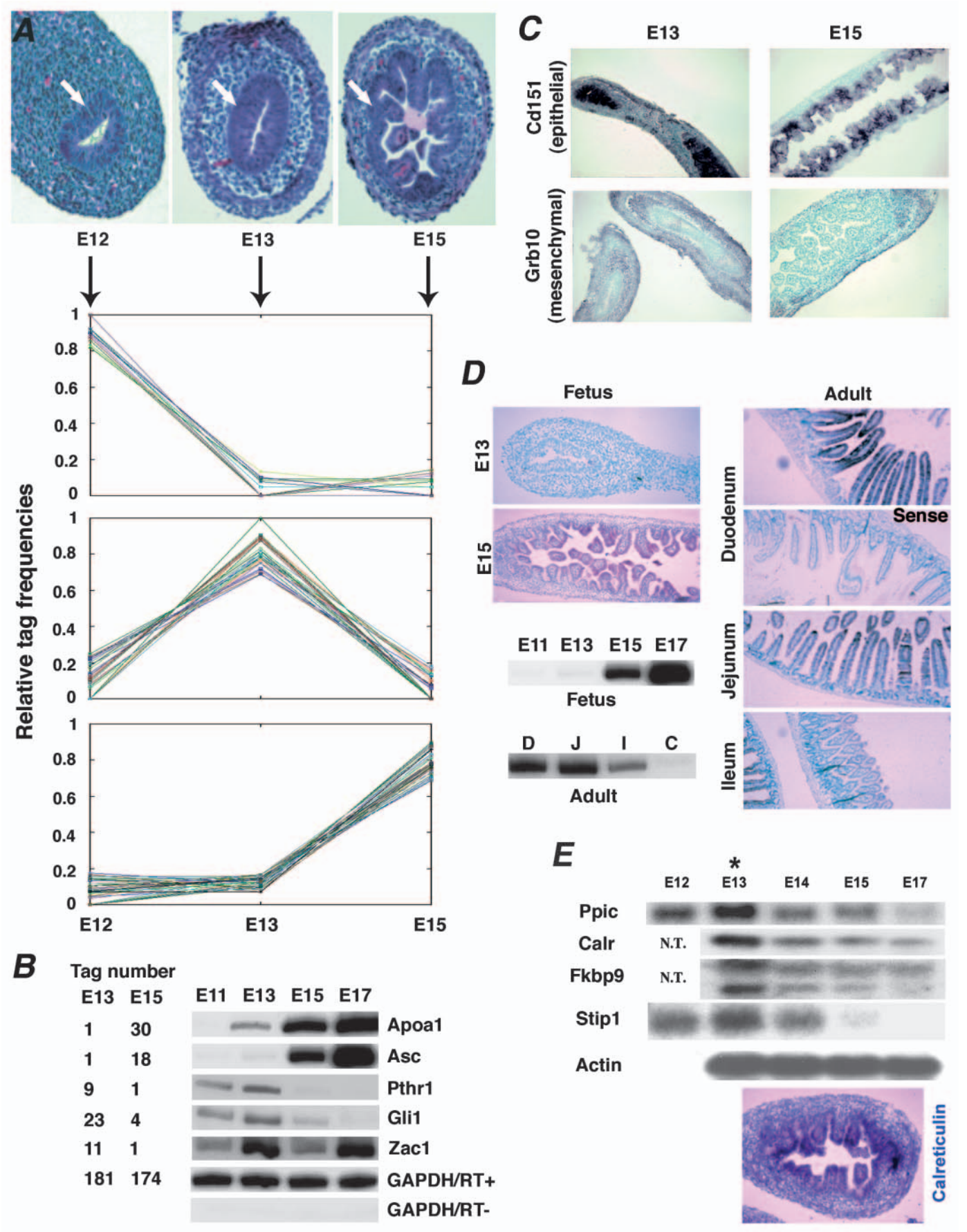
A
E12
E13
E15
Relative tag frequencies
1
0.8
0.6
0.4
0.2
0
E12
E13
E15
B
Tag number
E13 E15
1 30
1 18
9 1
23 4
11 1
181 174
E11 E13 E15 E17
Apoa1
Asc
Pthr1
Gli1
Zac1
GAPDH/RT+
GAPDH/RT-
C
E13
E15
Cd151 (epithelial)
Grb10 (mesenchymal)
D
Fetus
Adult
E13
E15
E11 E13 E15 E17
Fetus
D J I C
Adult
Duodenum
Jejunum
Ileum
Sense
E
*
E12 E13 E14 E15 E17
Ppic
Calr
Fkbp9
Stip1
Actin
N.T.
N.T.
Calreticulin

Fig. 1. Tissue and molecular anatomy of mouse intestine development. (A) Histology of the mid-gestation small intestine revealed in Hematoxylin- and Eosin-stained transverse sections. Gut endoderm (arrows) transforms into a simple villous epithelium between E13 and E15. Also shown are temporal clusters of a subset of SAGE tags representing 703 genes that vary significantly (P<0.015) over the E12-E15 interval. Each colored line corresponds to one SAGE tag and denotes its relative frequency among the three SAGE libraries. (B) Validation and extension of SAGE results for representative transcripts from three classes of temporally regulated genes [increasing: apolipoprotein 1A (*Apoa1*) and apoptosis-specific protein with CARD domain (*Asc*; *Pycard* – Mouse Genome Informatics); decreasing: parathyroid hormone receptor (*Pthr1*) and *Gli1*; and dynamically regulated: zinc-finger regulator of apoptosis and cell-cycle arrest (*Zac1*; *Plagl1* – Mouse Genome Informatics)]. RT-PCR analysis of gut tissue isolated at four fetal stages is compared with E13 and E15 SAGE data (tag numbers). Glyceraldehyde 3-phosphate dehydrogenase (*Gapd*) serves as a loading control and the lack of PCR products in samples not treated with reverse transcriptase (RT–) indicates absence of DNA in the RNA preparations. (C) RNA in situ hybridization analysis to distinguish predominantly epithelial (e.g. *CD151*) from stromal (e.g. growth receptor-binding protein 10, *Grb10*) transcripts. Sense probes (not shown) gave negligible staining. (D) Characterization of one gene product (proline-rich acidic protein 1, *Prap1*) suggested in SAGE to represent a marker of intestine differentiation. RT-PCR and in situ hybridization analysis confirm onset of epithelial expression after E13. In adults, *Prap1* expression is concentrated in rostral segments, duodenum (D) and jejunum (J), compared with the ileum (I) or colon (C). (E) Northern analysis confirms SAGE data that a group of genes traditionally associated with the cellular stress response peaks in expression around E13, coincident with the villous epithelial transition. Sample results are shown for peptidylprolyl isomerase C (*Ppic*), calreticulin (*Calr*), FK506-binding protein 9 (*Fkbp9*) and stress-induced phosphoprotein 1 (*Stip1*). N.T., not tested. In situ hybridization localized expression of these transcripts to the epithelial compartment, as shown for *Calr*.

Transcripts that increase significantly at E15 are usually present at high levels in the adult organ (three examples shown in Fig. 2, 'Increasing') and some of them can serve as molecular markers of tissue differentiation. Examples include the pyrin-CARD domain caspase-recruiting protein ASC and proline-rich acidic protein 1 (Prap1). The latter was originally described as a selective marker of the pregnant uterus (Kasik and Rice, 1997) and later found to be expressed at high levels in the adult small bowel (Bates et al., 2002). Detailed examination of *Prap1* mRNA expression confirms its activation after E13 in a regionally restricted manner that anticipates its distribution in the adult organ (Fig. 1D). We have also shown *Prap1* to be a useful in vitro marker of intestinal epithelial cytodifferentiation (Tou et al., 2004).

The transition from a squamous into a columnar epithelium is associated with establishment of new intercellular junctions and cell polarity. Accordingly, the gene expression profiles reveal substantial changes in the concentrations of transcripts encoding structural or regulatory cytoskeletal proteins, including keratins 18 and 19, cadherins, a number of actin-binding factors and genes such as *Lgl1*, which determines cell polarity in other species and tissues (Klezovitch et al., 2004). Our results further reveal temporal regulation of the components of several signaling pathways (Table 1), including secreted factors that can bind extracellular matrix (ECM) and may function in short-range intercellular communication. Expression of these factors in the gut has not been revealed through other experimental approaches and they are attractive candidate mediators of tissue interactions in gut organogenesis. Examples include ECM-binding factors such as midkine; cell surface receptor-associated proteins such as Grb10 and the integrin-linked kinase; suppressors of cytokine signaling such as WSB-1; and several phosphatases. Consistent with possible functions in influencing tissue morphogenesis, many of these transcripts peak in expression at E12 or E13. One surprise is the abundant and regulated expression of the parathyroid hormone receptor gene *Pthr1* (Fig. 1B, Table 1), which prompted us to examine fetal gut histology prior to the gestational lethality of $Pthr1^{-/-}$ mouse embryos (Lanske et al., 1996). Although this study failed to reveal gross defects in $Pthr1^{-/-}$ intestines (data not shown), the expression dynamics could reflect a developmental role with compensated or subtle manifestations.

Among transcripts classified according to presumptive cellular functions, the E13 mouse intestine shows peak expression of a number of genes associated with protein folding or cellular stress responses, a finding verified by tissue northern analysis (Fig. 1E). Expression of these genes is confined to the epithelial compartment (shown for *Calr* in Fig. 1E). These observations raise the possibility that certain developmental transitions share molecular features with cellular stress; alternatively, the regulated expression of heat- and stress-response genes may reflect metabolic or biosynthetic functions associated with emergence of an absorptive and secretory epithelium. Transcriptional profiles, coupled with mRNA localization studies, can thus lead to specific hypotheses about the functions of groups of co-expressed genes. Here, we focus on the significance of specific trends in gene expression.

Significance of transcripts that decline in abundance with progressive epithelial differentiation

Genes with reduced expression past E13 potentially serve functions that are restricted to the period of organogenesis. Indeed, the proportion of the genome that is dedicated to developmental functions is unclear. We assessed expression of 32 transcripts, selected arbitrarily from a total of 254 genes that showed higher SAGE representation at E12 or E13 than at E15. Seven of these mRNAs were undetectable past E15 (Fig. 2, 'Decreasing', four examples shown), whereas 25 transcripts could be detected again at E17 ('Dynamic', seven examples). Half of the Dynamic, and all the Decreasing, transcripts are weakly or not detected in any segment of the adult intestine with a sensitive RT-PCR assay (Fig. 2). Because analysis of RNA isolated from the whole organ may underestimate expression that is confined to minor cell populations, we confirmed absence of most of these transcripts in the adult mouse gut by in situ hybridization (data not shown). Transient expression of many genes during organogenesis suggests that a substantial fraction of mammalian genes may serve dedicated functions in development, although some mRNAs in this class (e.g. *Pthr1*) may have separate roles in other sites (Urena et al., 1993). Moreover, groups of fetally expressed genes may be subject to common molecular mechanisms of gene silencing concomitant with tissue differentiation.

Silencing of genes during normal development may be

especially relevant to cancer, a disease in which cells display immature morphology and gene expression is broadly dysregulated. Some proteins expressed exclusively in tumors and in developing embryos, such as carcinoembryonic antigen and α-fetoprotein, are oncofetal markers of clinical utility (Carr et al., 1997; Uriel, 1975); however, the extent to which tumor cells recapitulate embryonic gene expression and the significance of this phenomenon are unknown. We therefore

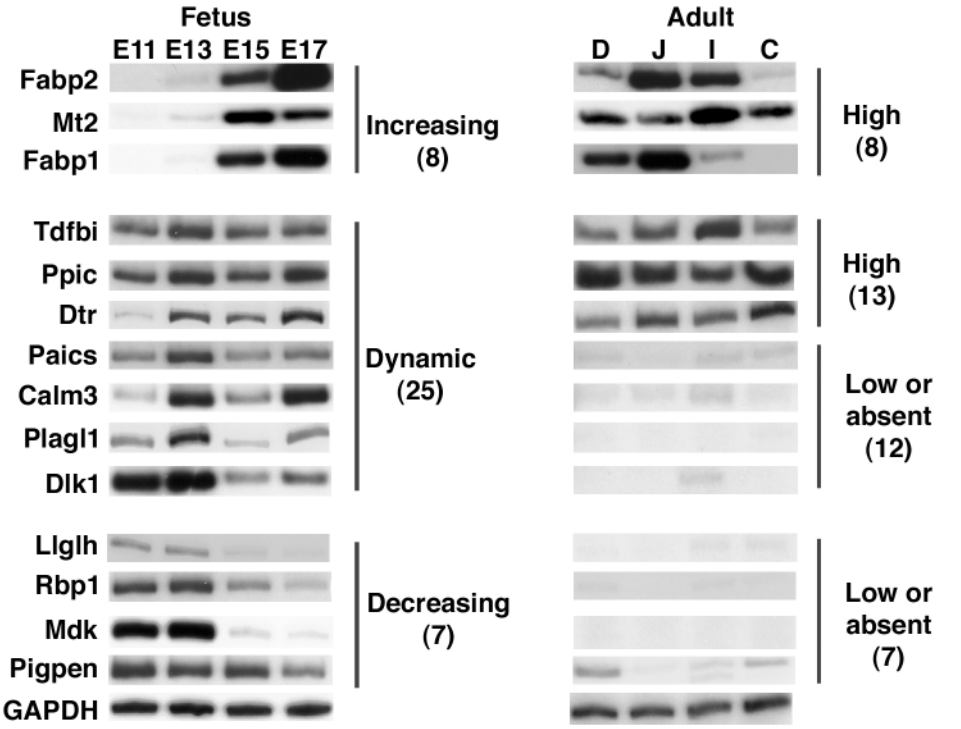

Fig. 2. Expression of developmentally regulated genes in the adult mouse intestine. E11-E17 fetal gut and adult (12-week) duodenum (D), jejunum (J), ileum (I) and colon (C) were analyzed by RT-PCR for selected transcripts that display developmentally regulated expression in SAGE analysis. Representative examples are shown for three categories, defined according to the relationship among expression levels at E13, E15 and E17, i.e. Increasing, Dynamically regulated or Decreasing; numbers in parentheses refer to the number of transcripts investigated in each class.

Table 1. Signaling genes that show significant temporal regulation in mouse intestine development

SAGE tag	Tag number E12	E13	E15	P_12 v 13	P_13 v 15	Unigene Mm.	Gene symbol (and name or description)
GGCTTCCGCG	24	62	13	4.3E-05	0	906	*Mdk* (midkine)
TAGTTCCGAC	11	29	11	0.00571	0.003967	30195	*Gps1* (G protein pathway suppressor 1)
CATTTTCTGG	101	60	30	0.00026	0.001459	3862	*Igf2* (insulin-like growth factor 2)
TAACCGTTTT	32	25	7	0.1594	0.001269	273117	*Grb10* (growth factor receptor bound protein 10)
GAAGGTTCTA	12	11	2	0.42377	0.012708	289681	*Dtr* (diphtheria toxin receptor)
GGGACCGAAA	32	32	9	0.6218	0.000313	302504	*Rbp1* (retinol binding protein 1, cellular)
GAAAGCAATG	28	30	13	0.58273	0.009037	90787	*Ngfrap1* [nerve growth factor receptor (TNFRSF16) associated protein 1]
GGCAGATGGA	5	9	1	0.23288	0.010862	3542	*Pthr1* (parathyroid hormone receptor 1)
TCCCTGACTT	12	23	3	0.0601	8.33E-05	2769	*Mlp* (MARCKS-like protein)
GGTTTTGTTT	4	12	1	0.04861	0.001818	307022	*Wsb1* (WD-40-repeat-containing protein with a SOCS box 1)
CCCCTCTTCC	10	21	3	0.05677	0.000193	292208	*Hdgf* (hepatoma-derived growth factor)
GGGAACAACT	3	9	1	0.08066	0.010862	8687	*Cap1* [adenylyl cyclase-associated CAP protein homolog 1 (*S. cerevisiae*, *S. pombe*)]
ATAATGAATA	31	43	15	0.14153	0.00018	157069	*Dlk1* [delta-like 1 homolog (*Drosophila*)]
TGGGGACCTG	4	11	2	0.07143	0.012708	2402	*Rhoip3* (Rho interacting protein 3; *AA536749* – Mouse Genome Informatics)
GAGGACCTGG	2	8	0	0.07374	0.004647	258771	*Ptprs* (protein tyrosine phosphatase, receptor type, S)
GTGTAGCTGC	2	8	0	0.07374	0.004647	3128	*Pitpn* (phosphatidylinositol transfer protein)
TAGCTCAGGC	2	10	0	0.0253	0.001082	153891	*Ptp4a3* (protein tyrosine phosphatase 4a3)
TGGGGGCCCC	5	13	3	0.05821	0.012192	276367	*Efemp2* (epidermal growth factor-containing fibulin-like extracellular matrix protein 2)
GGTCCAAGAG	27	41	17	0.08045	0.001603	298728	*Nisch* (nischarin)
ATTTGAAATA	12	21	8	0.10539	0.014242	196464	*Gnai2* (guanine nucleotide binding protein, alpha inhibiting 2)
GGAGGGATCA	15	29	13	0.03479	0.013535	274846	*Ilk* (integrin linked kinase)
ACGTTTGTGG	7	20	10	0.01383	0.057766	271249	*Ppp2r3a* [protein phosphatase 2 (formerly 2A), regulatory subunit B″, α]
ATCCGCACCC	2	11	3	0.01433	0.030314	288630	*Calm3* (calmodulin 3)
GCTTCTCCCT	1	9	2	0.01419	0.033196	216227	*Csnk1d* (casein kinase 1, δ)
TAACGCCCTT	5	17	13	0.01209	0.330915	8385	*Mapk3* (mitogen activated protein kinase 3)
CCGGACGAGG	1	10	3	0.00705	0.05332	2343	*Ppp1r14a* [protein phosphatase 1, regulatory (inhibitor) subunit 14A]
GACTGACGAG	0	10	6	0.00118	0.240611	19080	*Gtpbp1* (GTP binding protein 1)
TATCCAGCCA	7	0	2	0.00676	0.23157	295226	*Dgkz* (diacylglycerol kinase ζ)
CATTAAATGA	9	1	8	0.00834	0.018837	41933	*Pitrm1* (pitrilysin metalloprotease 1)
GGAAACCTGA	0	0	8	na	0.003638	175989	*1810015H18Rik* (RIKEN cDNA 1810015H18 gene)
AGAGACAAGG	1	8	24	0.02696	0.002955	30837	*Ndr1* (N-myc downstream regulated 1; *Ndrg1* – Mouse Genome Informatics)

attempted to examine expression of all 32 Dynamic or Decreasing transcripts in seven samples of malignant and adjacent normal human colon tissue. For this purpose, we regard the small intestine and colon to have highly similar biology and gene expression profiles, as shown or suggested in previous studies (Bates et al., 2002; Potten, 1998); colon cancer is the second leading cause of cancer death in many developed nations, whereas primary tumors of the small bowel are rare (http://seer.cancer.gov). First, we identified the known or presumptive orthologous human gene and designed suitable primers for RT-PCR; 27 primer pairs for the corresponding human genes amplified the predicted fragment reliably. Second, we rigorously established two independent standards (GAPD and β-amyloid) to ensure equal loading of RNA from each matched pair of tumor and normal tissue (Fig. 3A,B) and we amplified *Tgfbi*, a TGFβ-inducible gene whose expression is known to be elevated in colon cancers (Zhang et al., 1997), as a positive control.

Like *Tgfbi*, seven out of the 27 reaction sets showed overexpression of the corresponding gene in at least two out of the seven tumors (Fig. 3B), a result we quantified by densitometry (e.g. Fig. 3C). The remaining transcripts show variable levels in normal and tumor samples, without consistent patterns (e.g. *Calm3*). Histological inspection of the tumors excluded a higher epithelial:stromal ratio as a trivial explanation for elevated levels of epithelial transcripts (data not shown), and for those markers we could test (e.g. *Trp53* and *Hdgf*; Fig. 3D and Fig. 4D), the protein is present in tumor cells but not in the stroma. The tumors show heterogeneity in expression of fetal genes, as expected, although two transcripts, *Tls*/*Fus* and *Trp53*, are increased in nearly every tumor. *Trp53* is commonly mutated in cancers and elevated protein levels are believed to result from post-translational stabilization (Kubbutat et al., 1997); our study reveals increased mRNA levels of this developmentally regulated gene. *Tls*/*Fus* [also called pigpen in the mouse (Alliegro and Alliegro, 1996)] is a recurring target of chromosomal translocation in human myxoid liposarcoma (Rabbitts et al., 1993), where its pathogenic mechanisms are unclear.

Because our results reveal that many developmentally downregulated genes are re-expressed in tumors, understanding their respective cellular roles can inform both developmental and cancer biology. Here, we focus on one such product. Hepatoma-derived growth factor (HDGF) was isolated from conditioned cell culture media as a protein that weakly stimulated proliferation of heterologous cells (Nakamura et al., 1994). However, HDGF harbors a nuclear localization signal (NLS) that is essential for its mitogenic

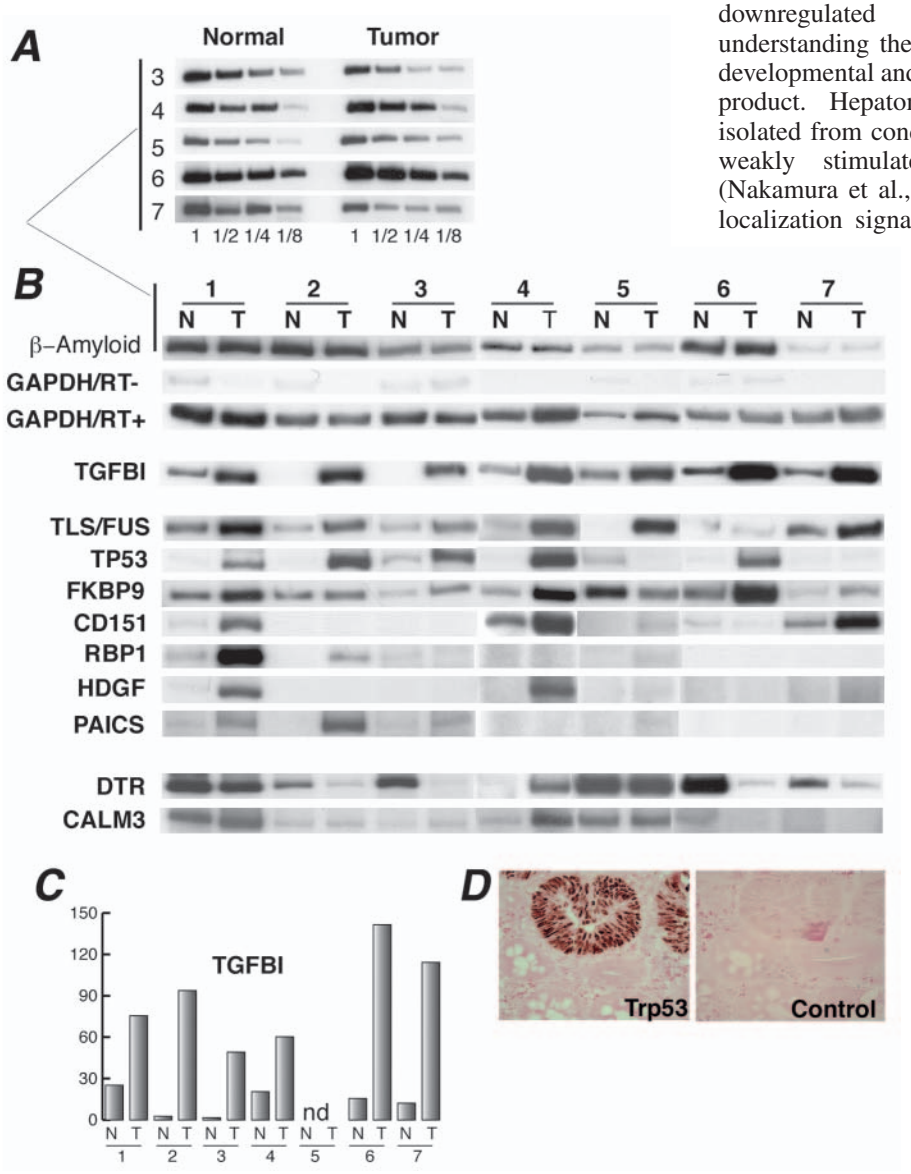

Fig. 3. Re-expression of developmentally attenuated genes in human colorectal cancers. (A) Normalization of cDNAs prepared from five of seven independent pairs of tumors (T) and matched normal (N) colonic tissue by RT-PCR analysis of β-amyloid mRNA, which is known to be expressed to similar levels in cancerous and normal human colon. Labels refer to serial dilution of cDNA before PCR. (B) Expression (RT-PCR) analysis for the seven out of 27 transcripts that show increased expression in two or more of seven T-N pairs, and illustrative results for two transcripts that show minimal change or inconsistent differences between T and N. (C) Results were quantified by gel densitometry, as shown here for one transcript, which encodes a 68 kDa TGF-β-induced protein. (D) Immunohistochemistry reveals the representative oncofetal marker *Trp53* to be expressed in tumor cells and not in the surrounding stroma.

activity (Everett et al., 2001; Kishima et al., 2002). Conclusions about its proposed functions (Everett et al., 2000; Oliver and Al-Awqati, 1998; Zhou et al., 2004) rely mainly on studies in cultured cell lines; the physiological roles and molecular mechanisms are poorly understood.

HDGF expression in mammalian fetal gut development and colorectal cancer

HDGF mRNA is downregulated concomitantly with transition of the gut endoderm into a villous epithelium and there is a corresponding, albeit delayed, decline in protein levels (Fig. 4A). In the fetal gut, HDGF mRNA and protein localize in the developing mucosa and are excluded from the mesenchyme (Fig. 4B); HDGF protein is restricted to cell nuclei. In the adult intestine, weak residual HDGF expression is confined to nuclei in the lower half of the villous projections (Fig. 4C), the site of temporary residence of undifferentiated epithelial cells. HDGF expression thus correlates with undifferentiated states in the developing and adult gut mucosa; there is no variation along the rostrocaudal axis of the digestive tract (data not shown).

In Fig. 3B, we show elevated *HDGF* mRNA in 2 of 7 human colon cancers. HDGF protein was overexpressed in 14 of 28 colorectal cancers we tested independently by immunohistochemistry (Fig. 4D). Again, HDGF was present only in nuclei and overexpressed in the tumor compartment; cells at the invasive front, particularly isolated invasive foci and large cells with undifferentiated morphology, were often the most strongly stained (Fig. 4E). This finding may be pertinent to the role of HDGF-containing complexes discussed below, and staining throughout the cancerous tissue argues against HDGF simply marking rapidly dividing cells. Colon cancers with deficient or intact DNA mismatch repair (MMR) express distinctive genetic, biological, and prognostic features but there is limited understanding about the molecular basis for these differences (Peltomaki, 2003). Notably, HDGF was overexpressed in 11 out of the 16 tumors proficient for MMR, compared with three out of 12 MMR-deficient colon cancers (Fig. 4F; P=0.027 by the Fisher exact test). Thus, although HDGF is expressed weakly in normal adult gut mucosa, our findings reveal oncofetal properties and especially elevated expression in MMR-proficient colon tumors.

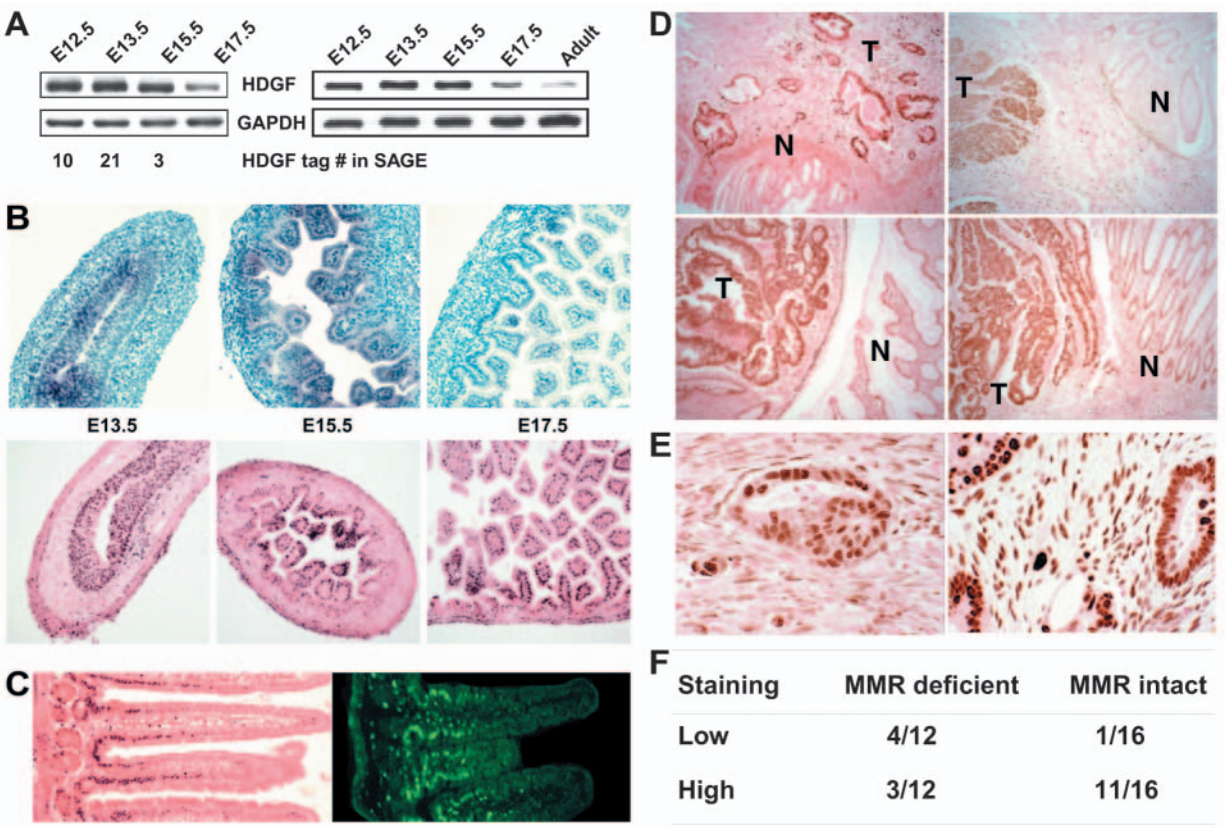

Fig. 4. Expression of the hepatoma-derived growth factor (HDGF) in development and neoplasia. (A) Relative expression levels of HDGF mRNA (left) and protein (right) in the mouse small intestine at the indicated fetal stages, assessed by semi-quantitative RT-PCR and immunoblot analysis, respectively. (B) Localization of HDGF mRNA (top) and protein (bottom) in the fetal mouse small intestine at the indicated developmental stages, as judged by in situ hybridization and immunohistochemistry, respectively. (C) Localization of HDGF in the adult (12-week-old) mouse small intestine, assessed by immunohistochemistry (left) and immunofluorescence (right). Exposure times were increased over fetal samples to permit localization of weaker signals. (D,E) Strong expression of HDGF in human colon tumors (T) relative to adjacent normal (N) mucosa, shown here in representative low- (D, 100×) and high- (E, 600×) magnification photomicrographs. (F) Relative HDGF staining signals in human colon cancer specimens according to the status of DNA mismatch repair (MMR) in the tumor, which was determined separately by the presence or absence of MMR gene (*Mlh1* and *Msh2*) expression.

Isolation and characterization of an HDGF-containing nuclear protein complex

HDGF has few sequence motifs that suggest possible cellular functions, except for partial homology with a subfamily of high-mobility group (HMG)-box proteins exemplified by HMGB1. The latter is also a nuclear protein that may be found in culture media, owing to its weak association with chromatin, which is disrupted in necrotic cells but enhanced when cells die by apoptosis (Scaffidi et al., 2002). Indeed, as noted for HMGB1, native HGDF could be extracted from HeLa cell nuclei within 1 minute of mild NP-40 detergent treatment and completely within 5 minutes, whereas the chromatin-associated histone 2Az protein resisted such extraction (Fig. 5A). In HeLa cells treated with ionomycin to induce necrosis, HDGF was lost from the cell nucleus, whereas after treatment with tumor necrosis factor α and cycloheximide, which induces apoptosis, HDFG remained in the nucleus (data not shown). These findings affirm nuclear expression of HDGF and might explain its appearance in conditioned cell media.

To better characterize HDGF function, we sought to isolate associated proteins with established cellular roles. We expressed double epitope-tagged HDGF by retroviral infection in HeLa cells, followed by immunoprecipitation from nuclear extracts using antibodies against the two synthetic epitopes (Nakatani and Ogryzko, 2003). We identified several HDGF-associated proteins using mass spectrometric analysis and computer-assisted sequence searching. In two experiments (Fig. 5B; data not shown) we identified tryptic peptides that cover nearly 22% of the amino acid sequence of two novel and highly conserved proteins with predicted molecular mass of 80 kDa and 57 kDa (see Fig. S1 in the supplementary material). We also recovered three peptides each corresponding to heterogeneous nuclear ribonucleoproteins (hnRNPs) K and I, and the sarcoma oncoprotein TLS/Fus. hnRNPI has been shown independently to interact with each of the other proteins (Kim et al., 2000; Meissner et al., 2003), and all three factors have well-characterized RNA-binding motifs; like HDGF, TLS/Fus mRNA is commonly increased in colorectal tumors (Fig. 3B). These factors have not been isolated previously in HeLa cell co-precipitation with diverse protein baits.

Available antibodies against the putative HDGF-associated proteins lack immunoprecipitating activity, which precludes independent assessment of protein interactions. Instead, we

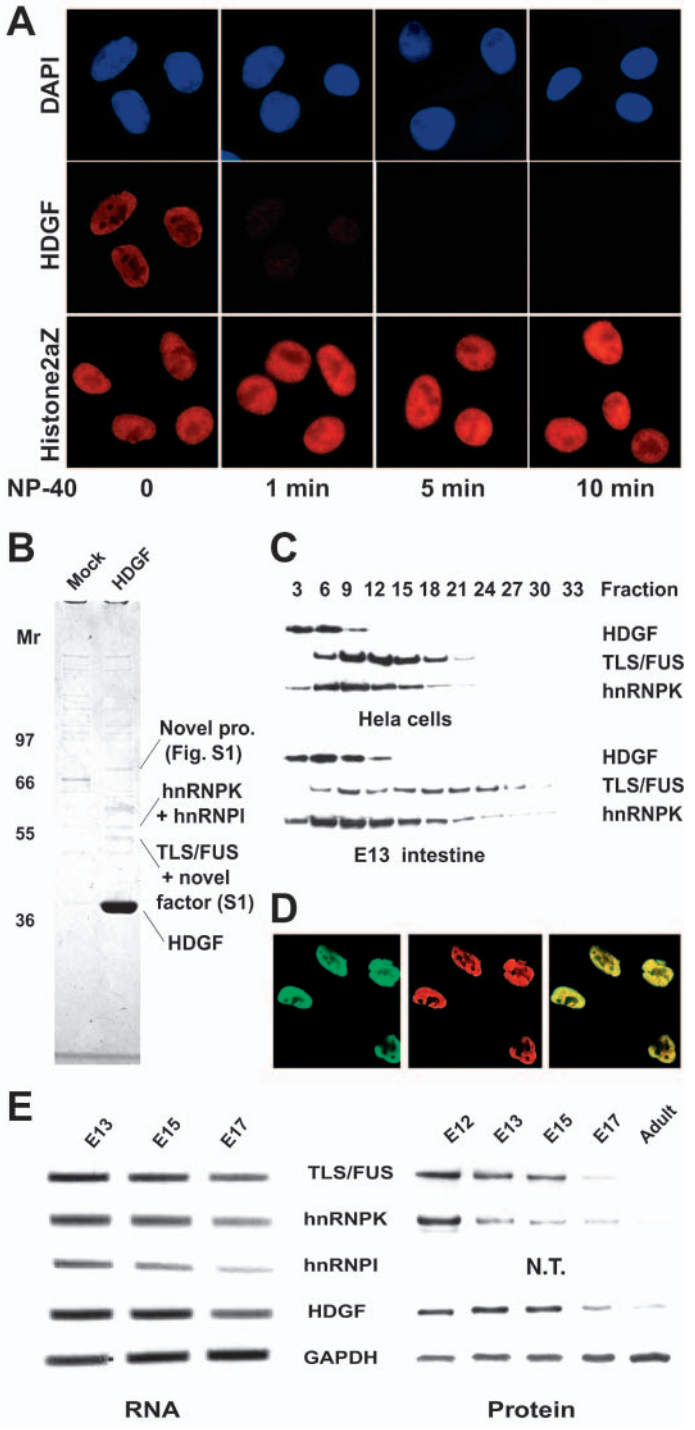

Fig. 5. HDGF properties and associations revealed through sequence similarities and isolation of a nuclear multi-protein complex. (A) Rapid release of nuclear HDGF but not histone 2Az protein following treatment of cultured HeLa cells with low concentrations of NP-40 detergent. This behavior parallels that of the related protein HMGB1 (Scaffidi et al., 2002) and may explain why HDGF previously was found intact in conditioned culture media. (B) Silver-stained gel of HDGF immunoprecipitates from mock-transfected or epitope-tagged HDGF-overexpressing HeLa cells resolved by SDS-PAGE. The marked specific bands were extracted and identified by mass spectrometric analysis to contain the indicated proteins: two novel factors (Fig. S1), hnRNPs K and I, and TLS/Fus, also known as mouse pigpen. (C) Immunoblot analysis of HeLa cell (top) and E13 mouse intestine (bottom) nuclear fractions resolved over a glycerol gradient and probed for the presence of HDGF, hnRNPK and TLS/Fus proteins. Resolution of protein complexes is similar in the two cell sources and indicates that a significant proportion of nuclear HDGF may be complexed with hnRNPK (red) and TLS/Fus (green), which in turn may associate independently with each other and with other factors in bulkier protein complexes. (D) Immunofluorescence analysis of subcellular localization of HDGF and TLS/Fus (similar results for hnRNPK are not shown) reveals the abundance of each in tiny nuclear dots, although the resolution cannot unambiguously define protein association in this context. (E) mRNA (RT-PCR, left) and protein (immunoblot, right) levels of the putative HDFG-associated factors TLS/Fus and hnRNPs K and I are downregulated in tandem with HDGF during mouse intestine development. These results reinforce the possibility of the four proteins functioning within a common cellular pathway that regulates epithelial differentiation and cancer. N.T., not tested.

resolved nuclear extracts from HeLa cells and from E13 mouse intestine by glycerol gradient sedimentation. Specific antibodies against hnRNPK and TLS/Fus (hnRNPI antibody is not readily available) revealed partial co-sedimentation with HDGF, mainly in fractions 6-9 (Fig. 5C). These results support the idea that nuclear HDGF associates, probably sub-stoichiometrically, with these factors. Immunostaining for HDGF, TLS/Fus and hnRNPK confirmed diffuse nuclear localization of each within innumerable discrete foci (Fig. 5D), although the subnuclear resolution is insufficient to verify co-localization of these abundant proteins. However, expression dynamics for the HDGF-associated proteins provide independent support for their participation in a common developmental pathway: mRNA and protein levels for both hnRNPK and TLS/Fus parallel that of HDGF in the fetal mouse gut (Fig. 5E). The three HDGF-associated factors we studied are RNA-binding proteins. Our observations thus combine to place previously obscure HDGF mechanisms within a RNA metabolic pathway that is developmentally controlled and pertinent to cancer. Although the two novel HDGF-associated proteins (Fig. S1) lack discernible functional motifs, their identification also should enable further study of HDFG function.

Differentiation and cellular functions of HDGF

The strong correlation between HDGF expression and undifferentiated epithelial states (Fig. 4) led us to hypothesize that HDGF regulates mucosal cell differentiation negatively. We reasoned that if declining HDGF levels during gut development enable expression of maturation-associated genes, then forced HDGF expression may prevent or delay their activation. We introduced DNA by luminal injection and electroporation in intestines explanted from E14 mouse fetuses, when endogenous mRNA levels are dropping and prior to native expression of differentiation genes (Tou et al., 2004). Epithelial expression of GFP-tagged HDGF fusion protein (Fig. 6A,B) reduced mRNA levels of several intestine-specific differentiation markers: *Apo1a*, liver (*Fabpl*)- and intestine (*Fabpi*)-specific fatty acid-binding proteins, metallothionein 2 (*Mt2*), and *Prap1* (Fig. 6C). Villin RNA levels were unchanged, which implies that HDGF influences differentiation genes selectively, and expression of the inhibited markers recovered partially after additional culture, indicating that HDGF overexpression did not compromise explant viability. An inactive HDGF form lacking the NLS (Everett et al., 2001) did not affect differentiation-related mRNAs (Fig. 6C); levels of the intact and mutant proteins were comparable (Fig. 6B). We have previously shown good correlation between molecular and histological maturation in fetal gut explants (Tou et al., 2004). However, the conditions for DNA electroporation affect tissue morphology adversely and preclude evaluation of HDGF effects on epithelial cytodifferentiation.

The logical conclusion from these experiments, that high HDGF levels retard epithelial cell differentiation, ideally should be complemented by depletion of early HDGF expression. Although we could apply HDFG-specific short hairpin (sh) shRNAs to reduce transcript levels in cultured cells and organ explants, the protein resisted our depletion efforts (data not shown). Moreover, proteins related to HDGF (Dietz et al., 2002) might substitute for the loss of a single family member. Nevertheless, the overexpression studies indicate that excessive HDGF in the fetal intestine impairs expression of gut epithelial differentiation markers. Forced expression of TLS/Fus or hnRNPK did not abrogate or delay fetal gut epithelial differentiation, which suggests that HDGF may be a limiting factor.

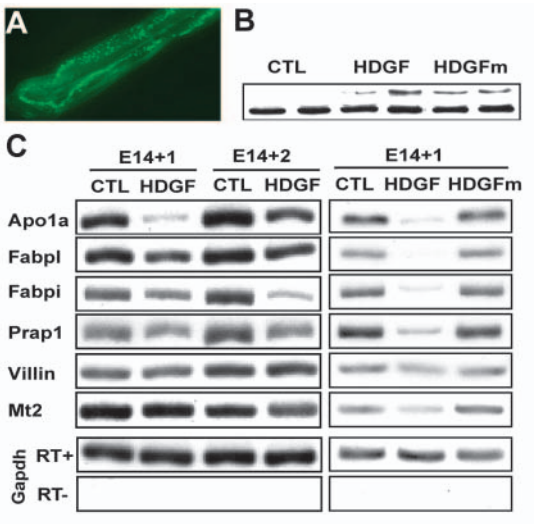

Fig. 6. Overexpression of HDGF in mouse fetal gut explants retards epithelial differentiation. (A) UV micrograph showing expression of GFP-tagged HDGF in a sample E14 explant 18 hours after plasmid electroporation. (B) Immunoblot confirming that full-length and mutant (m) HDGF were expressed to similar levels. Duplicate samples are shown for each; the prominent faster-migrating protein band represents a non-specific immunoblot signal and surrogate loading control. (C) RT-PCR analysis of cultured intestinal explants for molecular markers of gut epithelial differentiation after forced expression of GFP (CTL) or GFP-tagged wild-type or mutant (m) HDGF. Explant RNA was isolated 1 and 2 days after transfection and analyzed by RT-PCR for transcript levels of differentiation markers: apolipoprotein A1 (Apo1a), liver (Fabpl) and intestinal (Fabpi) fatty acid-binding proteins, proline-rich acidic protein 1 (Prap1), villin and metallothionein 2 (Mt2). The results represent five independent experiments with intact and two separate studies with the mutant form of HDGF. Gapd, loading control. RT+ and RT– refer to mRNA samples treated with and without reverse transcriptase, respectively.

Discussion

Insight into how individual genes influence development and differentiation derives largely from genetic and developmental studies. Broader understanding is often constrained by difficulties in tracking the many pathways that act simultaneously in organogenesis. Expression profiling can map active pathways and uncover processes not previously known to participate in development. The goals of expression profiling in developing organs include: (1) recording a reliable and searchable dataset that is validated and annotated with relevant and current gene information; (2) defining the repertoire of genes whose expression is modulated in space and time; (3) discovering novel molecular markers for distinct developmental stages; and (4) generating hypotheses about the functions of co-expressed genes. Here, we report the creation and use of a SAGE resource to advance these goals.

Experimental tests reproduce sample results from our analysis with more than 90% accuracy. Building on the quantitative aspects of SAGE, our web resource (http://genome.dfci.harvard.edu/GutSAGE) provides primary data on numbers of sequence tags and enumerates the statistical significance of differences measured across developmental stages. A versatile search function permits users to parse genes according to expression or functional criteria, with immediate access to the primary data. Tag and gene annotations are continually updated, as each SAGE entry is linked to three public resources: LocusLink, GeneOntology and SAGEmap; LocusLink will soon be replaced by its versatile successor Entrez Gene (http://www.ncbi.nlm.nih.gov/entrez/query/static/help/genehelp.html). These features assure the currency of the data and will enable investigators to apply a developmental dataset to address many outstanding problems. In particular, the basis for important interactions between epithelial and mesenchymal cells may be revealed through signaling components that vary over the E12-E15 interval. Although the data presented in Table 1 do not imply a vital role for every listed product, they represent an enriched gene set for targeted investigation. Such studies could be facilitated by localizing gene expression using in situ hybridization and an array of stage-specific molecular markers identified in this study. These advances enhance the collective ability to address issues such as the relationship between fetal and adult mucosal stem cells and how or why certain classes of genes, such as the stress-response group, are co-regulated during intestinal differentiation.

Genetic pathways that initiate intestinal tumors are well characterized (Batlle et al., 2002; Kinzler and Vogelstein, 1996), whereas most alterations that accompany disease progression (Stoler et al., 1999) are not. Yet, it is the latter changes that probably underlie especially ominous features of colonic and other epithelial tumors, including tissue invasion, viability at distant sites and drug resistance. Each of these features is strongly correlated with the degree of tumor cell differentiation (Deans et al., 1994), but the basis for this correlation is unknown. Although it is commonly assumed that poor differentiation of tumor cells reflects reversion to a primitive state, the definition of such states is vague. By contrast, in some leukemias and lymphomas, where relationships between tumor cells and steps in normal development or cell differentiation are defined precisely, the resulting insights guide both understanding and treatment of the disease (Tenen, 2003). We have therefore begun to ask to what extent a common carcinoma might recapitulate the gene expression program associated with its developmental origins. Current cancer treatments, even when directed against molecular carcinogenic mechanisms, frequently are confounded by toxicities that limit the dose or discourage drug development. Such toxicity typically reflects expression of the drug target in non-tumor tissues. Oncofetal proteins that are essential for malignant behaviors thus represent attractive targets for cancer treatment because their absence or reduced levels in normal adult tissues may confer especially wide therapeutic windows. An approach to identify oncofetal proteins could thus find useful therapeutic applications.

Our findings implicate the oncofetal protein HDGF in regulation of normal and pathologic gut mucosal differentiation. In the mammalian gut, HDGF localizes in epithelial cells, peak expression coincides with the developmental villous transition and the protein is overexpressed in human colorectal cancers, especially those with intact DNA mismatch repair. HDGF is a nuclear protein rather than a classical, secreted growth factor and forced expression in the fetal mouse intestine retards epithelial development. HDGF shares discernible sequence similarity and cellular properties with the ubiquitous and abundant nuclear protein HMGB1 (Nakamura et al., 1994), which may regulate gene transcription but also has differentiation and pathological roles when released from necrotic cells (Melloni et al., 1995; Muller et al., 2001; Scaffidi et al., 2002). Presumably, both HDGF and HMGB1 are released passively from necrotic cells following lysis. More important from a mechanistic standpoint is our isolation of nuclear complexes wherein HDGF associates with well-characterized RNA-binding proteins.

Not only does HDGF associate with two hnRNPs and TLS/Fus but the four mRNAs are regulated in tandem in intestine organogenesis and both HDGF and TLS/Fus are commonly overexpressed in human colon cancers. TLS/Fus, which can interact with DNA and RNA, regulates transcription and, in the context of a pathogenic fusion oncoprotein, confers a transcriptional activation domain (Rabbitts et al., 1993). Besides having a role in preserving genomic integrity (Baechtold et al., 1999; Hicks et al., 2000; Kuroda et al., 2000), TLS/Fus binds pre- and processed mRNAs, and engages in nucleocytoplasmic RNA transport (Iko et al., 2004; Lerga et al., 2001; Zinszner et al., 1997). It contains both RNA-binding domains and functions shared with hnRNPs and is known to interact with hnRNPI (Meissner et al., 2003). hnRNPI binds intronic polypyrimidine tracts, and both hnRNPI and hnRNPK associate in a protein complex (Kim et al., 2000) and interact with Ro ribonucleoprotein-associated Y RNAs (Fabini et al., 2001); both factors are implicated in gene transcription and pre-RNA splicing, nuclear export of mature mRNAs, translation and regulation of mRNA stability (Dreyfuss et al., 2002; Reed and Hurt, 2002). Their interaction and developmental co-regulation with HDGF thus strongly suggests related cellular functions for the latter and represents a novel insight into putative HDGF mechanisms. Recently, cytosolic TLS/Fus, hnRNPs K and I, other RNA-binding proteins, and RNA were found within novel focal adhesion-related structures called spreading initiation centers (de Hoog et al., 2004). Our studies do not address the possibility of similar roles for HDGF, although notably, HDGF expression in colon tumors is especially prominent in isolated foci of large, invasive cells (Fig. 5E).

In summary, *HDGF* is one of several genes we identify whose peak expression coincides with fetal gut epithelial morphogenesis and in carcinomas of the same developmental origin. HDGF acts to limit epithelial differentiation, which is significant in light of its expression pattern, and associates with a group of nuclear proteins that share biochemical properties and functions in RNA metabolism. Our approach can be extended to identify other oncofetal proteins, which could present safer therapeutic targets, and to better understand molecular mechanisms common to development and cancer.

Supported by grants from the Robert Black Charitable Foundation and National Institutes of Health grant R01DK61139. R.A.S. is a

Scholar of the Leukemia and Lymphoma Society. We thank K. Kinzler for permission to use SAGE methods and software; K. Polyak and M. Loda for technical advice; P. Latham for effort in preparing SAGE libraries; Q. Liu and F. David for assistance with experiments; M. Alliegro, K. Bomsztyk and D. Goodman for antisera; U. Chung and H. Kronenberg for *Pthr* mutant mice; and M. Bianchi, D. Hill, M. Hu and S. Parkinson for helpful discussions.

Supplementary material

Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/132/2/415/DC1

References

Alliegro, M. C. and Alliegro, M. A. (1996). A nuclear protein regulated during the transition from active to quiescent phenotype in cultured endothelial cells. *Dev. Biol.* **174**, 288-297.

Baechtold, H., Kuroda, M., Sok, J., Ron, D., Lopez, B. S. and Akhmedov, A. T. (1999). Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation. *J. Biol. Chem.* **274**, 34337-34342.

Bates, M. D., Erwin, C. R., Sanford, L. P., Wiginton, D., Bezerra, J. A., Schatzman, L. C., Jegga, A. G., Ley-Ebert, C., Williams, S. S., Steinbrecher, K. A. et al. (2002). Novel genes and functional relationships in the adult mouse gastrointestinal tract identified by microarray analysis. *Gastroenterology* **122**, 1467-1482.

Batlle, E., Henderson, J. T., Beghtel, H., van den Born, M. M., Sancho, E., Huls, G., Meeldijk, J., Robertson, J., van de Wetering, M., Pawson, T. et al. (2002). Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251-263.

Cai, L., Huang, H., Blackshaw, S., Liu, J. S., Cepko, C. and Wong, W. H. (2004). Clustering analysis of SAGE data using a Poisson approach. *Genome Biol.* **5**, R51.

Carr, B. I., Flickinger, J. C. and Lotze, M. T. (1997). Hepatobiliary cancers. In *Cancer: Principles & Practice of Oncology* (ed. V. T. DeVita, S. Hellman and S. A. Rosenberg). Philadelphia, PA: Lippincott-Raven.

Chen, J., Sun, M., Lee, S., Zhou, G., Rowley, J. D. and Wang, S. M. (2002). Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc. Natl. Acad. Sci. USA* **99**, 12257-12262.

Cheng, H. and Leblond, C. P. (1974). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian theory of the origin of the four epithelial cell types. *Am. J. Anat.* **141**, 537-561.

de Hoog, C. L., Foster, L. J. and Mann, M. (2004). RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. *Cell* **117**, 649-662.

Deans, G. T., Patterson, C. C., Parks, T. G., Spence, R. A., Heatley, M., Moorehead, R. J. and Rowlands, B. J. (1994). Colorectal carcinoma: importance of clinical and pathological factors in survival. *Ann. R. Coll. Surg. Engl.* **76**, 59-64.

Dietz, F., Franken, S., Yoshida, K., Nakamura, H., Kappler, J. and Gieselmann, V. (2002). The family of hepatoma-derived growth factor proteins: characterization of a new member HRP-4 and classification of its subfamilies. *Biochem. J.* **366**, 491-500.

Dreyfuss, G., Kim, V. N. and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3**, 195-205.

Everett, A. D., Lobe, D. R., Matsumura, M. E., Nakamura, H. and McNamara, C. A. (2000). Hepatoma-derived growth factor stimulates smooth muscle cell growth and is expressed in vascular development. *J. Clin. Invest.* **105**, 567-575.

Everett, A. D., Stoops, T. and McNamara, C. A. (2001). Nuclear targeting is required for hepatoma-derived growth factor- stimulated mitogenesis in vascular smooth muscle cells. *J. Biol. Chem.* **276**, 37564-37568.

Fabini, G., Raijmakers, R., Hayer, S., Fouraux, M. A., Pruijn, G. J. and Steiner, G. (2001). The heterogeneous nuclear ribonucleoproteins I and K interact with a subset of the ro ribonucleoprotein-associated Y RNAs in vitro and in vivo. *J. Biol. Chem.* **276**, 20711-20718.

Fatma, N., Singh, D. P., Shinohara, T. and Chylack, L. T., Jr (2001). Transcriptional regulation of the antioxidant protein 2 gene, a thiol- specific antioxidant, by lens epithelium-derived growth factor to protect cells from oxidative stress. *J. Biol. Chem.* **276**, 48899-48907.

Haffen, K., Kedinger, M. and Simon-Assmann, P. (1987). Mesenchyme-dependent differentiation of epithelial progenitor cells in the gut. *J. Pediat. Gastroenterol. Nutr.* **6**, 14-23.

Hentsch, B., Lyons, I., Li, R., Hartley, L., Lints, T. J., Adams, J. M. and Harvey, R. P. (1996). Hlx homeo box gene is essential for an inductive tissue interaction that drives expansion of embryonic liver and gut. *Genes Dev.* **10**, 70-79.

Hicks, G. G., Singh, N., Nashabi, A., Mai, S., Bozek, G., Klewes, L., Arapovic, D., White, E. K., Koury, M. J., Oltz, E. M. et al. (2000). Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. *Nat. Genet.* **24**, 175-179.

Iko, Y., Kodama, T. S., Kasai, N., Oyama, T., Morita, E. H., Muto, T., Okumura, M., Fujii, R., Takumi, T., Tate, S. I. et al. (2004). Domain architectures and characterization of an RNA binding protein, TLS/FUS. *J. Biol. Chem.* **279**, 44834-44840.

Kaestner, K. H., Silberg, D. G., Traber, P. G. and Schutz, G. (1997). The mesenchymal winged helix transcription factor Fkh6 is required for the control of gastrointestinal proliferation and differentiation. *Genes Dev.* **11**, 1583-1595.

Karlsson, L., Lindahl, P., Heath, J. K. and Betsholtz, C. (2000). Abnormal gastrointestinal development in PDGF-A and PDGFR-(alpha) deficient mice implicates a novel mesenchymal structure with putative instructive properties in villus morphogenesis. *Development* **127**, 3457-3466.

Kasik, J. and Rice, E. (1997). A novel complementary deoxyribonucleic acid is abundantly and specifically expressed in the uterus during pregnancy. *Am. J. Obstet. Gynecol.* **176**, 452-456.

Katz, J. P., Perreault, N., Goldstein, B. G., Lee, C. S., Labosky, P. A., Yang, V. W. and Kaestner, K. H. (2002). The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development* **129**, 2619-2628.

Kedinger, M., Duluc, I., Fritsch, C., Lorentz, O., Plateroti, M. and Freund, J. N. (1998). Intestinal epithelial-mesenchymal cell interactions. *Ann. N. Y. Acad. Sci.* **859**, 1-17.

Kho, A. T., Zhao, Q., Cai, Z., Butte, A. J., Kim, J. Y., Pomeroy, S. L., Rowitch, D. H. and Kohane, I. S. (2004). Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers. *Genes Dev.* **18**, 629-640.

Kim, J. H., Hahm, B., Kim, Y. K., Choi, M. and Jang, S. K. (2000). Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* **298**, 395-405.

Kinzler, K. W. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell* **87**, 159-170.

Kishima, Y., Yamamoto, H., Izumoto, Y., Yoshida, K., Enomoto, H., Yamamoto, M., Kuroda, T., Ito, H., Yoshizaki, K. and Nakamura, H. (2002). Hepatoma-derived growth factor stimulates cell growth after translocation to the nucleus by nuclear localization signals. *J. Biol. Chem.* **277**, 10315-10322.

Klezovitch, O., Fernandez, T. E., Tapscott, S. J. and Vasioukhin, V. (2004). Loss of cell polarity causes severe brain dysplasia in Lgl1 knockout mice. *Genes Dev.* **18**, 559-571.

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P. J. and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat. Genet.* **19**, 379-383.

Kubbutat, M. H., Jones, S. N. and Vousden, K. H. (1997). Regulation of p53 stability by Mdm2. *Nature* **387**, 299-303.

Kuroda, M., Sok, J., Webb, L., Baechtold, H., Urano, F., Yin, Y., Chung, P., de Rooij, D. G., Akhmedov, A., Ashley, T. et al. (2000). Male sterility and enhanced radiation sensitivity in TLS(–/–) mice. *EMBO J.* **19**, 453-462.

Lanske, B., Karaplis, A. C., Lee, K., Luz, A., Vortkamp, A., Pirro, A., Karperien, M., Defize, L. H., Ho, C., Mulligan, R. C. et al. (1996). PTH/PTHrP receptor in early development and Indian hedgehog-regulated bone growth. *Science* **273**, 663-666.

Lerga, A., Hallier, M., Delva, L., Orvain, C., Gallais, I., Marie, J. and Moreau-Gachelin, F. (2001). Identification of an RNA binding specificity for the potential splicing factor TLS. *J. Biol. Chem.* **276**, 6807-6816.

Marigo, V., Johnson, R. L., Vortkamp, A. and Tabin, C. J. (1996). Sonic hedgehog differentially regulates expression of GLI and GLI3 during limb development. *Dev. Biol.* **180**, 273-283.

Maunoury, R., Robine, S., Pringault, E., Leonard, N., Gaillard, J. A. and Louvard, D. (1992). Developmental regulation of villin gene expression in the epithelial cell lineages of mouse digestive and urogenital tracts. *Development* **115**, 717-728.

Meissner, M., Lopato, S., Gotzmann, J., Sauermann, G. and Barta, A.

(2003). Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins. *Exp. Cell Res.* **283**, 184-195.

Melloni, E., Sparatore, B., Patrone, M., Pessino, A., Passalacqua, M. and Pontremoli, S. (1995). Extracellular release of the 'differentiation enhancing factor', a HMG1 protein type, is an early step in murine erythroleukemia cell differentiation. *FEBS Lett.* **368**, 466-470.

Moreno, J. C., Pauws, E., van Kampen, A. H., Jedlickova, M., de Vijlder, J. J. and Ris-Stalpers, C. (2001). Cloning of tissue-specific genes using serial analysis of gene expression and a novel computational substraction approach. *Genomics* **75**, 70-76.

Muller, S., Scaffidi, P., Degryse, B., Bonaldi, T., Ronfani, L., Agresti, A., Beltrame, M. and Bianchi, M. E. (2001). New EMBO members' review: the double life of HMGB1 chromatin protein: architectural factor and extracellular signal. *EMBO J.* **20**, 4337-4340.

Nakamura, H., Izumoto, Y., Kambe, H., Kuroda, T., Mori, T., Kawamura, K., Yamamoto, H. and Kishimoto, T. (1994). Molecular cloning of complementary DNA for a novel human hepatoma-derived growth factor. Its homology with high mobility group-1 protein. *J. Biol. Chem.* **269**, 25143-25149.

Nakatani, Y. and Ogryzko, V. (2003). Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* **370**, 430-444.

Oliver, J. A. and Al-Awqati, Q. (1998). An endothelial growth factor involved in rat renal development. *J. Clin. Invest.* **102**, 1208-1219.

Peltomaki, P. (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J. Clin. Oncol.* **21**, 1174-1179.

Potten, C. S. (1998). Stem cells in gastrointestinal epithelium: numbers, characteristics and death. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **353**, 821-830.

Rabbitts, T. H., Forster, A., Larson, R. and Nathan, P. (1993). Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. *Nat. Genet.* **4**, 175-180.

Ramalho-Santos, M., Melton, D. A. and McMahon, A. P. (2000). Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763-2772.

Reed, R. and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523-531.

Roberts, D. J., Johnson, R. L., Burke, A. C., Nelson, C. E., Morgan, B. A. and Tabin, C. (1995). Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development* **121**, 3163-3174.

Rubin, D. C., Swietlicki, E., Roth, K. A. and Gordon, J. I. (1992). Use of fetal intestinal isografts from normal and transgenic mice to study the programming of positional information along the duodenal-to- colonic axis. *J. Biol. Chem.* **267**, 15122-15133.

Scaffidi, P., Misteli, T. and Bianchi, M. E. (2002). Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* **418**, 191-195.

Simon, T. C. and Gordon, J. I. (1995). Intestinal epithelial cell differentiation: new insights from mice, flies and nematodes. *Curr. Opin. Genet. Dev.* **5**, 577-586.

Simon, T. C., Roth, K. A. and Gordon, J. I. (1993). Use of transgenic mice to map cis-acting elements in the liver fatty acid-binding protein gene (Fabpl) that regulate its cell lineage-specific, differentiation-dependent, and spatial patterns of expression in the gut epithelium and in the liver acinus. *J. Biol. Chem.* **268**, 18345-18358.

Stoler, D. L., Chen, N., Basik, M., Kahlenberg, M. S., Rodriguez-Bigas, M. A., Petrelli, N. J. and Anderson, G. R. (1999). The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proc. Natl. Acad. Sci. USA* **96**, 15121-15126.

Tenen, D. G. (2003). Disruption of differentiation in human cancer: AML shows the way. *Nat. Rev. Cancer* **3**, 89-101.

Tou, L., Liu, Q. and Shivdasani, R. A. (2004). Regulation of mammalian epithelial differentiation and intestine development by class I histone deacetylases. *Mol. Cell. Biol.* **24**, 3132-3139.

Urena, P., Kong, X. F., Abou-Samra, A. B., Juppner, H., Kronenberg, H. M., Potts, J. T., Jr and Segre, G. V. (1993). Parathyroid hormone (PTH)/PTH-related peptide receptor messenger ribonucleic acids are widely distributed in rat tissues. *Endocrinology* **133**, 617-623.

Uriel, J. (1975). Fetal characteristics of cancer. In *Cancer, a comprehensive treatise* (ed. F. F. Becker). New York, NY: Plenum Press.

Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* **270**, 484-487.

Yang, Q., Bermingham, N. A., Finegold, M. J. and Zoghbi, H. Y. (2001). Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* **294**, 2155-2158.

Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. and Kinzler, K. W. (1997). Gene expression profiles in normal and cancer cells. *Science* **276**, 1268-1272.

Zhang, J., Rosenthal, A., de Sauvage, F. J. and Shivdasani, R. A. (2000). Downregulation of hedgehog signaling is required for organogenesis of the small intestine in Xenopus. *Dev. Biol.* **229**, 188-202.

Zhou, Z., Yamamoto, Y., Sugai, F., Yoshida, K., Kishima, Y., Sumi, H., Nakamura, H. and Sakoda, S. (2004). Hepatoma-derived growth factor is a neurotrophic factor harbored in the nucleus. *J. Biol. Chem.* **279**, 27320-27326.

Zinszner, H., Sok, J., Immanuel, D., Yin, Y. and Ron, D. (1997). TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling. *J. Cell Sci.* **110**, 1741-1750.