The American Journal of Pathology, Vol. 172, No. 5, May 2008

DOI: 10.2353/ajpath.2008.070851

Tumorigenesis and Neoplastic Progression

Cross-Species Comparison of Human and Mouse Intestinal Polyps Reveals Conserved Mechanisms in Adenomatous Polyposis Coli (*APC*)-Driven Tumorigenesis

Claudia Gaspar,* Joana Cardoso,*‡ Patrick Franken,* Lia Molenaar,* Hans Morreau,§ Gabriela Möslein,¶ Julian Sampson,‖ Judith M. Boer,‡ Renée X. de Menezes,†‡ and Riccardo Fodde*

From the Department of Pathology, Josephine Nefkens Institute, and Department of Pediatric Oncology,† Erasmus Medical Center, Rotterdam, The Netherlands; Center for Human & Clinical Genetics,‡ and Department of Pathology,§ Leiden University Medical Center, Leiden, The Netherlands; St. Josefs-Hospital Bochum-Linden,¶ Bochum, Germany; and Institute of Medical Genetics,‖ Cardiff University, Cardiff, United Kingdom*

Expression profiling is a well established tool for the genome-wide analysis of human cancers. However, the high sensitivity of this approach combined with the well known cellular and molecular heterogeneity of cancer often result in extremely complex expression signatures that are difficult to interpret functionally. The majority of sporadic colorectal cancers are triggered by mutations in the adenomatous polyposis coli (*APC*) tumor suppressor gene, leading to the constitutive activation of the Wnt/β-catenin signaling pathway and formation of adenomas. Despite this common genetic basis, colorectal cancers are very heterogeneous in their degree of differentiation, growth rate, and malignancy potential. Here, we applied a cross-species comparison of expression profiles of intestinal polyps derived from hereditary colorectal cancer patients carrying *APC* germline mutations and from mice carrying a targeted inactivating mutation in the mouse homologue *Apc*. This comparative approach resulted in the establishment of a conserved signature of 166 genes that were differentially expressed between adenomas and normal intestinal mucosa in both species. Functional analyses of the conserved genes revealed a general increase in cell proliferation and the activation of the Wnt/β-catenin signaling pathway. Moreover, the conserved signature was able to resolve expression profiles from hereditary polyposis patients carrying *APC* germline mutations from those with bi-allelic inactivation of the *MYH* gene, supporting the usefulness of such comparisons to discriminate among patients with distinct genetic defects. ***(Am J Pathol 2008, 172:1363–1380; DOI: 10.2353/ajpath.2008.070851)***

Colorectal cancer (CRC) is one of the major causes of morbidity and mortality among Western societies. Although the vast majority of CRC cases are sporadic, a considerable fraction has been attributed to hereditary and familial factors.[1] Hereditary CRC syndromes have served as unique models to elucidate the molecular and cellular mechanisms underlying colorectal tumor initiation and progression to malignancy, as the same genes mutated in the germline of affected individuals are also known to play rate-limiting roles in the majority of the sporadic cases.[2] This is certainly the case of the adenomatous polyposis coli (*APC*) tumor suppressor gene, known to be mutated in the germline of individuals affected by familial adenomatous polyposis (FAP)[3–6] and in the majority of the sporadic CRC cases.[7–9] In fact, loss of *APC* function appears to play a rate-limiting and initiating role in the adenoma-carcinoma sequence.[10] Among the multiple functional domains characterized in its coding

These studies were supported by grants from the Dutch Cancer Society (EMCR 2001-2482), The Netherlands Organisation for Scientific Research (NWO/Vici 016.036.636), the BSIK program of the Dutch Government grant 03038, the EU FP6 (MCSCs), "Deutsche Krebshilfe, Verbundprojekt familiarer Darmkrebs", and the Centre for Medical Systems Biology (CMSB).

C.G. and J.C. equally contributed to the study.

Accepted for publication January 17, 2008.

Supplemental material for this article can be found on *http://ajp.amjpathol.org*.

Address reprint requests to Riccardo Fodde, Ph.D., Dept. of Pathology, Erasmus MC, PO Box 2040, 3000CA Rotterdam, The Netherlands. E-mail: r.fodde@erasmusmc.nl.

sequence, APC's ability to bind and down-regulate β-catenin, the main signaling molecule in the canonical Wnt pathway, is generally regarded as its main tumor-suppressing activity.[10] Loss of APC function or oncogenic activation of β-catenin, as observed in the vast majority of the sporadic CRC cases, leads to intracellular accumulation of β-catenin and its translocation to the nucleus, where it binds to transcription factors of the T cell factor/lymphoid enhancer-binding factor (TCF/LEF) family and modulates transcription of a broad spectrum of downstream target genes.[11,12] The vast majority (70 to 90%) of FAP patients have been shown to carry germline *APC* mutations.[13] More recently, biallelic mutations in the base excision repair gene *MYH* were found in a subset of polyposis families with attenuated clinical presentation and an autosomal recessive inheritance pattern, often referred to as MAP (*MYH*-associated polyposis).[14]

In view of its initiating role in intestinal cancer, several preclinical models carrying germline mutations in the endogenous mouse *Apc* tumor supressor gene have been generated, and their phenotype has been characterized.[15] The predisposition of these mouse models to multiple intestinal adenomas closely resembles the FAP phenotype at the molecular, cellular, and phenotypic level.[15] One exception to the latter is represented by the proximal localization and distribution of adenomas along the gastrointestinal (GI) tract of *Apc*-mutant mouse models when compared with the colorectal clustering of polyps among FAP patients.[15]

Expression profiling by cDNA and oligonucleotide microarrays represents a powerful tool for genome-wide transcriptional analysis. Several studies in the literature have reported on the comparison of expression profiles from colorectal tumors and normal intestinal mucosa in an attempt to identify differentially expressed genes, predict, whenever feasible, clinical outcome, and elucidate the molecular and cellular mechanisms underlying colorectal tumorigenesis.[16] However, the different lists of genes differentially expressed in CRC are often very extensive and only partially overlapping among independent studies, possibly reflecting differences in the methodologies and in cohorts used.[16] To pinpoint conserved and functionally relevant genes differentially expressed between normal and malignant tissues, cross-species comparison have been successfully applied by comparing expression signatures of hepatocellular carcinoma, and prostate and lung cancer derived from human patients and mouse models.[17–20]

Here, we report the cross-species comparison of expression profiles of intestinal adenomas from FAP patients with established *APC* germline mutations and from $Apc^{+/1638N}$, a mouse model for familial polyposis previously developed in our laboratory and characterized by the development of an average of five tumors in the upper GI tract, together with other extra-intestinal manifestations characteristic of FAP patients, such as epidermal cysts and desmoids.[21,22] A total of 166 genes were found to be highly conserved between the two species and are likely to play important roles in the cellular and molecular mechanisms underlying adenoma formation in the gastrointestinal tract. Among these, several Wnt downstream target genes are included, as also expected from the selection of *APC*-mutant mouse and human adenomas. Notably, the conserved *APC* signature also made it possible to distinguish FAP tumors from MAP ones in an unsupervised fashion.

Materials and Methods

Patients and Tumor Samples

Colorectal adenomatous polyps were obtained from a total of 13 patients from the Department of Surgery, Heinrich Heine University (Dusseldorf, Germany); the Institute of Medical Genetics, Cardiff University (Cardiff, UK); the Department of Pathology, Leiden University Medical Center (Leiden, The Netherlands); and the Department of Surgery, Erasmus Medical Center (Rotterdam, The Netherlands). Of the 13 polyposis patients, 8 carried truncating *APC* mutations, whereas 5 were found to carry biallelic *MYH* mutations. Detailed characterization of the polyposis patients carrying biallelic *MYH* or monoallelic *APC* germline mutations and of the corresponding tumor samples used in the present study has been reported elsewhere.[23] Normal epithelial mucosa from 3 healthy individuals was also collected (control samples NC1 to NC3). All tissue samples were snap-frozen, embedded in OCT medium, and stored at −80°C. Detailed sample processing procedures were as previously described.[24] All of the analyzed adenomatous lesions were matched for histology (low-grade dysplasia) and anatomical location (left-sided colon or rectum). Two to six polyps were analyzed for each individual patient.

Mouse Strains and Material

All wild-type and $Apc^{+/1638N}$ mice used in this study were inbred C57BL6/J, maintained under SPF conditions and fed *ad libitum*. Duodenal adenomas and normal mucosa samples were collected from 9-month-old males, briefly washed in PBS, and snap-frozen in OCT medium. Hematoxylin and eosin (H&E) staining of all tissues was performed to determine histology and tumor content.

Laser Capture Microdissection and RNA Isolation

Sample preparation and laser capture microdissection (LCM) were performed as previously described using a PALM MicroBeam microscope system (P.A.L.M. Microlaser Technologies, Bernereid, Germany).[23] In short, 10-μm cryosections were mounted on microscope slides with a polyethylene naphthalate membrane and stained by H&E to allow histological identification of the desired normal and tumor cells. Approximately 1000 to 2000 cells, corresponding to 600,000 (human samples)- and 1,200,000 (mouse specimens)-μm^2 areas, were microdissected.

RNA was isolated from the LCM samples by Mini RNeasy columns (QIAGEN, Valencia, CA), according to

the manufacturer's instructions, including a DNase step on the column. Quality of the isolated RNA was checked with RNA 6000 Pico LabChip kit (Agilent Technologies, Palo Alto, CA).

Expression Profiling and Data Analysis

Human Adenoma Samples

Each RNA sample that passed the quality controls was linearly amplified with two rounds of amplification using the MessageAmp kit (Ambion, Huntingdon, UK), according to the manufacturer's protocol. Quality and quantity of each amplified RNA sample was again evaluated by Nano Lab-on-Chip (Agilent Technologies). One-μg aliquots of target and reference amplified RNA were labeled with Cy5-dUTP and Cy3-dUTP, respectively (Amersham Biosciences, Amersham, UK) by reverse transcription using the CyScribe First Strand cDNA Labeling kit (Amersham Biosciences). Each labeling reaction was further purified with the CyScribe GFX purification kit (Amersham Biosciences). Subsequently, both labeled cDNAs were hybridized on a human 18K cDNA microarray encompassing 19,200 spots (representing 18,432 independent cDNAs) produced and obtained from The Netherlands Cancer Institute Microarray Facility (Amsterdam, The Netherlands). The cDNAs spotted in this array platform were PCR-amplified from a clone set purchased from Research Genetics (Huntsville, AL). Hybridization and washing procedures were performed according to the manufacturer's (The Netherlands Cancer Institute Microarray Facility) protocol. Sixteen-bit fluorescent images from the expression arrays were acquired with an Agilent DNA Microarray Scanner (Agilent Technologies), and the resulting TIFF images were analyzed with the software GenePix Pro 4.0 (Axon Instruments, Union City, CA). For each array, a GenePix results file (.gpr) with the extracted Cy3 and Cy5 spot and background raw intensities was generated.

Expression data analysis was performed with a set of functions implemented in R (*http://www.R-project.org/*[25]). Briefly, .gpr files from the array platforms were directly loaded into the R environment using the marray package to extract the background-corrected Cy3 and Cy5 median raw intensity per spot. Intensity data from both platforms were normalized with the variance stabilization and normalization function implemented in the vsn package.[26]

To find genes that could better characterize the histological and mutational status of the analyzed samples, we have used "mixed-effects" regression models. Briefly, we fitted the following model: $Y(i) = \alpha + \beta * histology + \gamma * mutation + \delta * patient + error(i)$, where Y represents the log-ratio of the expression value for clone i; *histology* and *mutation* are categorical variables represented as stages (normal or adenoma) and (APC or MYH), respectively, whereas α represents the baseline expression level of clone i. A patient effect must be included in the model to correctly handle multiple samples derived from the same patient. Although both histology and mutation are assumed to have fixed effects, the patient effect is assumed to be random, as the patients included in this study represent the heterogeneous (outbred) population of familial CRC patients. This model was fitted to the data using the MAANOVA package.[27] Moderated *F*-test statistics, which take advantage of the large number of genes being analyzed simultaneously to yield more reliable variance estimates, were extracted corresponding to histology and mutation effects. Their *P* values were subsequently corrected for multiple testing using Benjamini and Hochberg's false discovery rate (FDR) method.[28]

Mouse Adenoma Samples

RNA samples were submitted to a double round of amplification according to the Small Sample Labeling Protocol vII (Affymetrix, Inc, Santa Clara, CA). Quality of synthesized cRNA was checked using the RNA 6000 Nano LabChip kit (Agilent Technologies). Labeled cRNA products were hybridized to mouse arrays MOE430A (Affymetrix, Inc) according to the manufacturer's instructions. Data analysis was performed using R Statistical Computing software v2.4.1[25] complemented with BioConductor Packages affy,[29] limma,[30,31] and vsn.[26] Cel files were uploaded and summarized using the affy package and normalized with vsn at the probe level. Using an empirical-Bayesian linear model (implemented in the package limma), we identified genes that were differentially expressed, with multiple testing correction performed using Benjamini and Hochberg's FDR step-up method.[28] Hierarchical clustering (Euclidean similarity measure) was performed on vsn-normalized data for all probe identifications (IDs) using Spotfire DecisionSite 9.0. (*http:www.spotfire.com*).

To test and/or confirm that a set of genes yielded a differential expression signature between groups of samples, the globaltest[32] of vsn-normalized data was performed using R Statistical Computing software v2.4.1[25] complemented with the BioConductor Package globaltest.[32]

Functional Annotation Analysis

GenBank accession numbers or Affymetrix probe set IDs were assigned to biological process using the GO (Gene Ontology) chart feature offered by the Database for Annotation Visualization and Integrated Discovery (DAVID) 2006 (*http://david.abcc. ncifcrf.gov/home.jsp*). Individual genes from selected expression signatures were also placed in the context of their molecular and functional interactions by using Ingenuity Pathways Analysis (IPA) tools according to the manufacturer's instructions (Ingenuity Systems, Redwood City, CA).

Data Integration

The cross-species comparison was performed using the Sequence Retrieval System (SRS).[33] Both sets were first annotated using Unigene; next, the SRS system was used to pair the two species' Unigene entries based on Homologene. The MOE430A array includes a total of 22,626 probes, whereas the human cDNA array encom-

passes 19,200 probes. Of the 22,626 mouse probes in the mouse Affymetrix platform, the SRS system retrieved a total of 12,083 homologous genes encompassed by the human array. Due to probe multiplicity in both platforms, the total overlap between the platforms consists of 18,369 entries.

From the two original data sets (mouse adenomas versus normal mucosa and human adenomas versus normal mucosa), probe sets with the same direction of differential expression were selected, adding to a total of 9495 probes. Further selection was applied based on the statistical significance according to the following thresholds: mouse data FDR <5% and human data FDR <0.5% (n = 234 probes). To exclude the possibility that the observed overlap resulted by sheer chance, we have performed a χ^2 test with the selected probes and reached a highly significant P value (P = 0.007).

Immunohistochemistry

Immunohistochemistry of mouse and human normal and adenomatous intestinal sections was performed according to standard protocols. The following antibodies were used and optimized for human and mouse tissues: MacMarcks (Calbiochem cat.442708; EMD Biosciences, San Diego, CA), CyclinA (cat.GTX27956; Genetex, San Antonio, TX), AnnexinI (cat.71-3400; Zymed, South San Francisco, CA). Signal detection of these antibodies was obtained by the Rabbit Envision+ System-HRP (cat.K4011; DakoCytomation, Carpinteria, CA). CD44 (cat.553131; BD Pharmingen, San Diego, CA) was detected using a secondary antibody goat anti-rat, HRP labeled.

Results

The overall rationale and strategy of the present study was to attempt the comparative expression profiling of intestinal adenomas derived from FAP patients carrying *APC* germline mutations[23] and from the *Apc*$^{+/1638N}$ mouse model.[21,22] In both cases, we aimed at using tumor samples in which the initiating and rate-limiting mutation event is represented by loss of *APC* tumor-suppressing function to gain insight into conserved molecular and cellular pathways relevant for intestinal tumor initiation. Furthermore, we explored the ability of the conserved gene signature to discriminate between adenomas from polyposis patients carrying germline *APC* mutations from those with different genetic defects.

Expression Profiling Analysis of Familial Adenomatous Polyps

Colorectal adenomatous polyps (n = 42) have been obtained from a total of eight unrelated FAP patients carrying previously identified germline *APC* mutations.[23] To obtain expression signatures exclusively derived from parenchymal cells and avoid the confounding effects of infiltrating and adjacent stromal cells, dysplastic tumor cells were collected by laser capture microdissection (LCM). Control LCM samples were obtained from the intestinal mucosa of three individuals with no CRC history. RNA was extracted from the microdissected tumor and normal specimens and subsequently used for expression profiling by hybridization to human 18K cDNA arrays generated at The Netherlands Cancer Institute Microarray Facility (see Materials and Methods). The expression profiling data have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (*http://www.ncbi.nlm.nih.gov/geo*) and is accessible through GEO Series accession number GSE9689.

Two-dimensional hierarchical clustering was first applied to all 19,200 array probes to generate an overview, without any prior filtering, of the global gene expression differences among all samples (Figure 1a). Notably, the three normal colon mucosa specimens (NC1-3) do not cluster separately from the adenoma samples. Also, several samples from the same individual are often observed to cluster together, indicative of a patient effect. In view of the latter, a statistical approach using a mixed-effects regression model[34] was applied to all samples to determine whether specific patterns of gene expression could be associated with the adenomatous polyps. The linear mixed-effects model was fitted considering histology (adenoma versus normal mucosa) as having a fixed effect and patient as having a random effect. This procedure not only allowed us to calculate P values for all genes, but also to control the variance component associated with random patient-specific genetic variation, ie, the variability in gene expression related to the genetic background of each individual patient. With an FDR set to 0.5%, a total of 1859 probes appeared to be differentially expressed between normal colonic epithelium and adenomas (see Supplemental Tables S1 and S2 at *http://ajp.amjpathol.org*). The relatively high number of differentially expressed genes even under highly stringent conditions clearly illustrates the strong effect of histology on global gene expression. This is also further illustrated by the empirical cumulative P values distribution function, which is clearly distinct from what would be expected if no effect was detectable (Figure 1b). Of the 1859 probes, 839 (45%) and 1020 (55%) were, respectively, up- and down-regulated in tumor cells when compared with normal mucosa (see Supplemental Tables S1 and S2 at *http://ajp.amjpathol.org*).

To gain insight into the biological relevance of the newly generated list of genes differentially expressed in adenomatous polyps, we used the GO-based bioinformatics tool DAVID[35] (see Materials and Methods). An overview of the functional categories represented by the genes differentially expressed among *APC*-mutant adenomatous polyps when compared with normal colonic mucosa reveals a very broad spectrum of biological processes, ranging from different aspects of cellular metabolism to apoptosis, cell migration, and immune response (data not shown). The broadness of the transcriptional profiles of the colorectal polyps when compared with normal mucosa is likely to reflect the heterogeneity of these benign tumors even at this very early stage of the adenoma-carcinoma sequence.

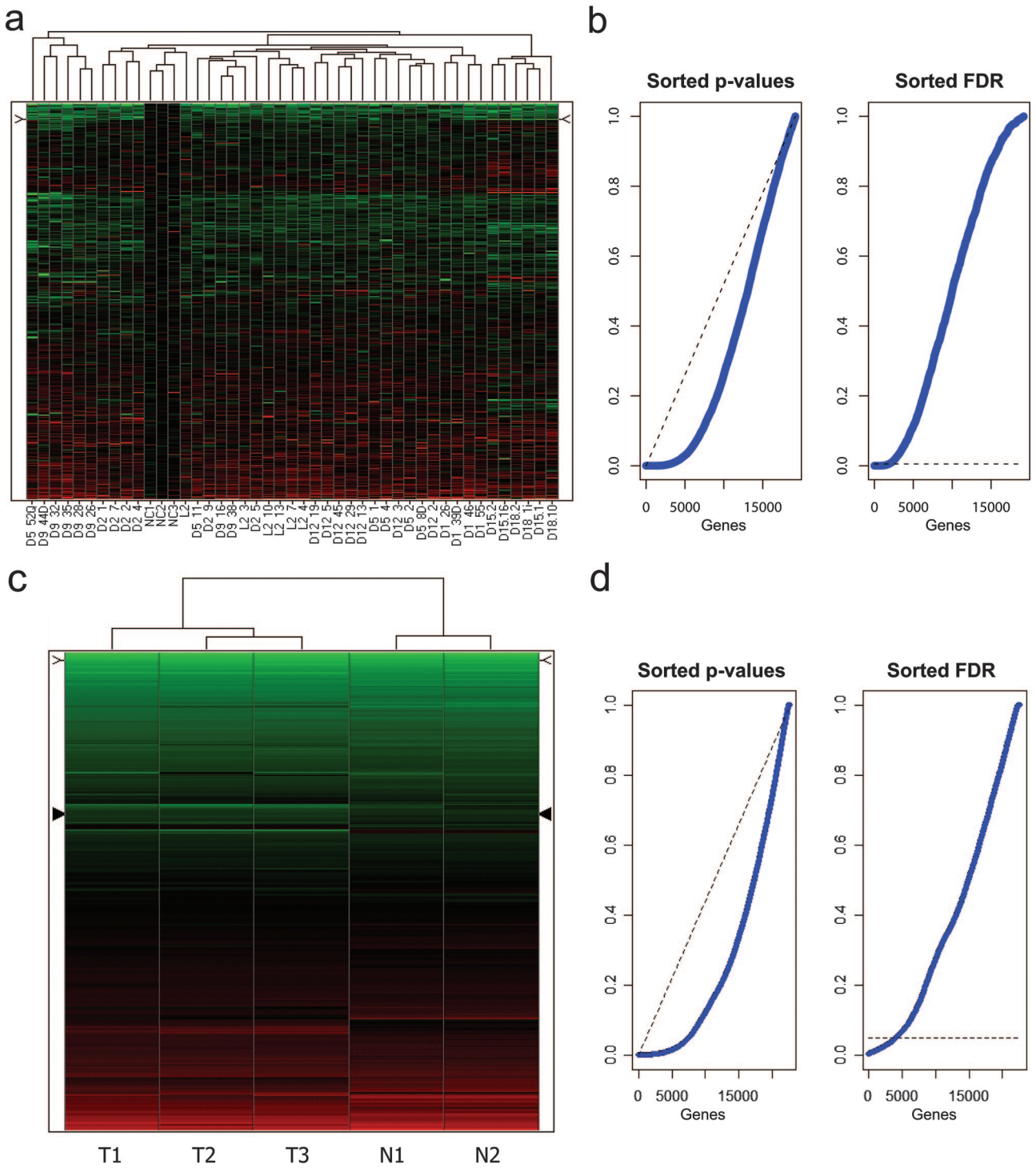

Figure 1. Unsupervised hierarchical cluster analysis of expression profiles from human and mouse intestinal polyps, without preliminary gene selection. Up- and down-regulated probes are depicted in red and green, respectively. **a:** Unsupervised hierarchical clustering analysis of 42 colorectal adenomatous polyps (obtained from eight unrelated FAP patients with previously identified germline *APC* mutations[23]) and 3 control normal mucosa samples (labeled as **NC1**-**3**) obtained from individuals with no history of CRC. **b:** Distribution of *P* values (**left plot**) relative to the comparison of patient-derived colorectal adenomas versus normal mucosa samples, sorted in ascending order (blue **line**). The **dashed** (black) **line** represents what would be expected if no effect was detectable. In the **right plot**, FDR-adjusted sorted *P* values are shown. The **dashed line** represents the FDR threshold used in our study to select the differentially expressed genes in the human set that led to the selection of 1859 probes. **c:** Unsupervised hierarchical clustering analysis of duodenal adenomas (n = 3, labeled **T1**–**T3**) and normal mucosa (n = 2, **N1**–**N2**) samples obtained from inbred C57BL6/J $Apc^{+/1638N}$ and $Apc^{+/+}$ mice, respectively. **d:** Distribution of *P* values (**left plot**) relative to the comparison of mouse duodenal adenomas versus normal tissue samples, sorted in ascending order (blue **line**). The **dashed** (black) **line** represents what would be expected if no effect was detectable. In the **right plot**, FDR-adjusted sorted *P* values are shown. The **dashed line** represents the FDR threshold used in our study to select the differentially expressed genes in the mouse set that led to the selection of 4137 probes.

Table 1. The Cross-Species Conserved 166-Gene Signature: Up-Regulated Genes

Mouse			Human		
Probe ID	Unigene*	Gene symbol	Unigene†	Gene symbol	Gene description
1448213_at	Mm.248360	*Anxa1*	Hs.494173	*ANXA1*	Annexin A1
1424460_s_at	Mm.284649	*Aytl2*	Hs.368853	*AYTL2*	Acyltransferase like 2
1424278_a_at	Mm.8552	*Birc5*	Hs.514527	*BIRC5*	Baculoviral IAP repeat-containing 5 (survivin)
1416815_s_at	Mm.927	*Bub3*	Hs.418533	*BUB3*	BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)
1455356_at	Mm.36834	*Camsap1*	Hs.522493	*CAMSAP1*	Calmodulin-regulated spectrin-associated protein 1
1416884_at	Mm.280968	*Cbx3*	Hs.381189	*CBX3*	Chromobox homolog 3 (HP1 gamma homolog, *Drosophila*)
1425616_a_at	Mm.36697	*Ccdc23*	Hs.113919	*CCDC23*	Coiled-coil domain containing 23
1427031_s_at	Mm.24035	*Ccdc52*	Hs.477144	*CCDC52*	Coiled-coil domain containing 52
1417911_at	Mm.4189	*Ccna2*	Hs.58974	*CCNA2*	Cyclin A2
1417419_at	Mm.273049	*Ccnd1*	Hs.523852	*CCND1*	Cyclin D1
1438560_x_at	Mm.296985	*Cct4*	Hs.421509	*CCT4*	Chaperonin containing TCP1, subunit 4 (delta)
1417258_at	Mm.282158	*Cct5*	Hs.1600	*CCT5*	Chaperonin containing TCP1, subunit 5 (epsilon)
1423760_at	Mm.423621	*Cd44*	Hs.502328	*CD44*	CD44 molecule (Indian blood group)
1452242_at	Mm.9916	*Cep55*	Hs.14559	*CEP55*	Centrosomal protein 55 kDa
1417457_at	Mm.222228	*Cks2*	Hs.83758	*CKS2*	CDC28 protein kinase regulatory subunit 2
1449300_at	Mm.200327	*Cttnbp2 nl*	Hs.485899	*CTTNBP2NL*	CTTNBP2 N-terminal like
1454268_a_at	Mm.271671	*Cyba*	Hs.513803	*CYBA*	Cytochrome b-245, alpha polypeptide
1419275_at	Mm.148693	*Dazap1*	Hs.222510	*DAZAP1*	DAZ associated protein 1
1424198_at	Mm.68971	*Dlg5*	Hs.500245	*DLG5*	Discs, large homolog 5 (*Drosophila*)
1435122_x_at	Mm.128580	*Dnmt1*	Hs.202672	*DNMT1*	DNA (cytosine-5-)-methyltransferase 1
1452052_s_at	Mm.27695	*Eif3s1*	Hs.404056	*EIF3S1*	Eukaryotic translation initiation factor 3, subunit 1 alpha
1426674_at	Mm.21671	*Eif3s9*	Hs.371001	*EIF3S9*	Eukaryotic translation initiation factor 3, subunit 9 eta
1448797_at	Mm.4454	*Elk3*	Hs.591015	*ELK3*	ETS-domain protein (SRF accessory protein 2)
1437211_x_at	Mm.427018	*Elovl5*	Hs.520189	*ELOVL5*	ELOVL family member 5, elongation of long chain fatty acids (FEN1/Elo2, SUR4/Elo3-like, yeast)
1420965_a_at	Mm.241073	*Enc1*	Hs.104925	*ENC1*	Ectodermal-neural cortex (with BTB-like domain)
1451550_at	Mm.6972	*Ephb3*	Hs.2913	*EPHB3*	EPH receptor B3
1417301_at	Mm.4769	*Fzd6*	Hs.591863	*FZD6*	Frizzled homolog 6 (*Drosophila*)
1419595_a_at	Mm.20461	*Ggh*	Hs.78619	*GGH*	Gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)
1419205_x_at	Mm.46029	*Gpatc4*	Hs.193832	*GPATC4*	G patch domain containing 4
1433736_at	Mm.248353	*Hcfc1*	Hs.83634	*HCFC1*	Host cell factor C1 (VP16-accessory protein)
1423051_at	Mm.426956	*Hnrpu*	Hs.166463	*HNRPU*	Heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)
1426705_s_at	Mm.21118	*Iars*	Hs.445403	*IARS*	Isoleucine-tRNA synthetase
1422546_at	Mm.440026	*Ilf3*	Hs.465885	*ILF3*	Interleukin enhancer binding factor 3
1456097_a_at	Mm.257094	*Itgb3bp*	Hs.166539	*ITGB3BP*	Integrin beta 3 binding protein (beta3-endonexin)
1421344_a_at	Mm.100253	*Jub*	Hs.655832	*JUB*	ajuba homolog (*Xenopus laevis*)
1452118_at	Mm.102761	*2600005C20Rik*	Hs.129621	*KIAA0179*	KIAA0179
1427080_at	Mm.29068	*2610036D13Rik*	Hs.370118	*KIAA0406*	KIAA0406
1448169_at	Mm.22479	*Krt18*	Hs.406013	*KRT18*	Keratin 18
1416621_at	Mm.285453	*Llgl1*	Hs.513983	*LLGL1*	Lethal giant larvae homolog 1 (*Drosophila*)
1434210_s_at	Mm.245210	*Lrig1*	Hs.518055	*LRIG1*	Leucine-rich repeats and immunoglobulin-like domains 1
1417511_at	Mm.28560	*Lyar*	Hs.425427	*LYAR*	Hypothetical protein FLJ20425
1439426_x_at	Mm.177539	*Lzp-s*	Hs.524579	*LYZ*	Lysozyme (renal amyloidosis)
1455941_s_at	Mm.325746	*Map2k5*	Hs.114198	*MAP2K5*	Mitogen-activated protein kinase kinase 5
1437226_x_at	Mm.424974	*Marcksl1*	Hs.75061	*MARCKSL1*	MARCKS-like 1
1439081_at	Mm.122725	*Mgea5*	Hs.500842	*MGEA5*	Meningioma expressed antigen 5 (hyaluronidase)
1424001_at	Mm.280311	*Mki67ip*	Hs.367842	*MKI67IP*	MKI67 (FHA domain) interacting nucleolar phosphoprotein
1449478_at	Mm.4825	*Mmp7*	Hs.2256	*MMP7*	Matrix metallopeptidase 7 (matrilysin, uterine)
1455129_at	Mm.130883	*Mtdh*	Hs.377155	*MTDH*	Metadherin
1419254_at	Mm.443	*Mthfd2*	Hs.469030	*MTHFD2*	Methylenetetrahydrofolate dehydrogenase ($NADP^+$ dependent) 2, methenyltetrahydrofolate cyclohydrolase
1452778_x_at	Mm.290407	*Nap1l1*	Hs.524599	*NAP1L1*	Nucleosome assembly protein 1-like 1
1423046_s_at	Mm.290027	*Ncbp2*	Hs.591671	*NCBP2*	Nuclear cap binding protein subunit 2
1455035_s_at	Mm.29363	*Nol5a*	Hs.376064	*NOL5A*	Nucleolar protein 5A (56 kDa with KKE/D repeat)
1416606_s_at	Mm.28203	*Nola2*	Hs.27222	*NOLA2*	Nucleolar protein family A, member 2 (H/ACA small nucleolar RNPs)
1449140_at	Mm.276504	*Nudcd2*	Hs.140443	*NUDCD2*	NudC domain containing 2
1428277_at	Mm.387724	*Otud6b*	Hs.30532	*OTUD6B*	OTU domain containing 6B

(table continues)

Table 1. *Continued*

Mouse			Human		
Probe ID	Unigene*	Gene symbol	Unigene†	Gene symbol	Gene description
1435368_a_at	Mm.277779	*Parp1*	Hs.177766	*PARP1*	poly-(ADP-ribose) polymerase family, member 1
1452620_at	Mm.29856	*Pck2*	Hs.75812	*PCK2*	Phosphoenolpyruvate carboxykinase 2 (mitochondrial)
1426838_at	Mm.37562	*Pold3*	Hs.82502	*POLD3*	Polymerase (DNA-directed), delta 3, accessory subunit
1427094_at	Mm.9199	*Pole2*	Hs.162777	*POLE2*	Polymerase (DNA directed), epsilon 2 (p59 subunit)
1449648_s_at	Mm.3458	*Rpo1–1*	Hs.584839	*POLR1C*	polymerase (RNA) I polypeptide C
1433552_a_at	Mm.273217	*Polr2b*	Hs.602757	*POLR2B*	polymerase (RNA) II (DNA directed) polypeptide B
1436505_at	Mm.11815	*Ppig*	Hs.470544	*PPIG*	Peptidylprolyl isomerase G (cyclophilin G)
1428265_at	Mm.7726	*Ppp2r1b*	Hs.546276	*PPP2R1B*	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform
1423775_s_at	Mm.227274	*Prc1*	Hs.567385	*PRC1*	protein regulator of cytokinesis 1
1452032_at	Mm.30039	*Prkar1a*	Hs.280342	*PRKAR1A*	Protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue-specific extinguisher 1)
1451576_at	Mm.71	*Prkdc*	Hs.491682	*PRKDC*	Protein kinase, DNA-activated, catalytic polypeptide
1435859_x_at	Mm.2462	*Psmc2*	Hs.437366	*PSMC2*	Proteasome (prosome, macropain) 26S subunit, ATPase, 2
1426631_at	Mm.58660	*Pus7*	Hs.520619	*PUS7*	Pseudouridylate synthase 7 homolog (*S. cerevisiae*)
1448899_s_at	Mm.204634	*Rad51ap1*	Hs.591046	*RAD51AP1*	RAD51-associated protein 1
1423700_at	Mm.12553	*Rfc3*	Hs.115474	*RFC3*	Replication factor C (activator 1) 3
1456375_x_at	Mm.314056	*Trim27*	Hs.440382	*RFP*	Tripartite motif-containing 27
1439403_x_at	Mm.435574	*Rnf12*	Hs.653288	*RNF12*	Ring finger protein 12
1437309_a_at	Mm.180734	*Rpa1*	Hs.461925	*RPA1*	Replication protein A1
1453362_x_at	Mm.16775	*Rps24*	Hs.356794	*RPS24*	Ribosomal protein S24
1416276_a_at	Mm.66	*Rps4x*	Hs.446628	*RPS4X*	Ribosomal protein S4, X-linked
1416120_at	Mm.99	*Rrm2*	Hs.226390	*RRM2*	Ribonucleotide reductase M2 polypeptide
1422864_at	Mm.4081	*Runx1*	Hs.149261	*RUNX1*	Runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)
1420824_at	Mm.33903	*Sema4 d*	Hs.655281	*SEMA4D*	Sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4D
1434972_x_at	Mm.391719	*Sfrs1*	Hs.68714	*SFRS1*	Splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)
1417623_at	Mm.399997	*Slc12a2*	Hs.162585	*SLC12A2*	Solute carrier family 12 (sodium/potassium/chloride transporters), member 2
1418326_at	Mm.27943	*Slc7a5*	Hs.513797	*SLC7A5*	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 5
1422771_at	Mm.325757	*Smad6*	Hs.153863	*SMAD6*	SMAD family member 6
1424206_at	Mm.246803	*Smarca5*	Hs.589489	*SMARCA5*	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 5
1452422_a_at	Mm.1323	*Snrpb2*	Hs.280378	*SNRPB2*	Small nuclear ribonucleoprotein polypeptide B
1419156_at	Mm.240627	*Sox4*	Hs.643910	*SOX4*	SRY (sex determining region Y)-box 4
1433502_s_at	Mm.220843	*Tsr1*	Hs.388170	*SRR*	TSR1, 20S rRNA accumulation, homolog (*S. cerevisiae*)
1415849_s_at	Mm.378957	*Stmn1*	Hs.693592	*STMN1*	Stathmin 1/oncoprotein 18
1450743_s_at	Mm.260545	*Syncrip*	Hs.571177	*SYNCRIP*	Synaptotagmin binding, cytoplasmic RNA interacting protein
1423601_s_at	Mm.2215	*Tcof1*	Hs.519672	*TCOF1*	Treacher Collins-Franceschetti syndrome 1
1416358_at	Mm.34564	*0610009O03Rik*	Hs.632581	*TETRAN*	Tetracycline transporter-like protein (TETRAN)
1434317_s_at	Mm.272025	*Tex10*	Hs.494648	*TEX10*	Testis expressed sequence 10
1426397_at	Mm.172346	*Tgfbr2*	Hs.82028	*TGFBR2*	Transforming growth factor, beta receptor II
1424641_a_at	Mm.219648	*Thoc1*	Hs.654460	*THOC1*	THO complex 1
1427318_s_at	Mm.34674	*Fer1l3*	Hs.500572	*FER1L3*	Fer-1-like 3, myoferlin (*C. elegans*)
1449041_a_at	Mm.27063	*Trip6*	Hs.534360	*TRIP6*	Thyroid hormone receptor interactor 6
1437906_x_at	Mm.19169	*Txnl1*	Hs.114412	*TXNL1*	Thioredoxin-like 1
1422842_at	Mm.3065	*Xrn2*	Hs.255932	*XRN2*	5′-3′ exoribonuclease 2
1448363_at	Mm.221992	*Yap1*	Hs.503692	*YAP1*	Yes-associated protein 1
1427208_at	Mm.289103	*Zfp451*	Hs.485628	*ZNF451*	Zinc finger protein 451
1416757_at	Mm.335237	*Zwilch*	Hs.21331	*ZWILCH*	Zwilch, kinetochore associated, homolog (*Drosophila*)

For simplicity, the gene description is only given for the human entry. *Unigene build 163; †Unigene build 202.

Expression Profiling Analysis of Apc$^{+/1638N}$ *Mouse Intestinal Adenomas*

Duodenal adenomas and normal mucosal samples from age- and sex-matched C57BL6/J *Apc*$^{+/1638N}$ mice (n = 3) and *Apc*$^{+/+}$ controls (n = 2) were collected and snap-frozen as for the above human polyps. Histological analysis of these lesions confirmed their benign adenomatous nature (not shown). Also in these cases, dysplastic epithelial cells were collected by LCM. Control expression signatures were obtained from wild-type C57BL6/J epithelial cells microdissected from the same anatomical location. Total RNA was extracted from normal and tumor samples and hybridized to Affymetrix MOE430A arrays. The corresponding data have been deposited in National Center for Biotechnology Information's Gene Expression Omnibus and are accessible through GEO Series accession number GSE9580. Unsupervised two-dimensional hierarchical clustering was able to discriminate and correctly cluster tumor from normal samples (Figure 1c). Empirical-Bayesian linear regression analysis allowed the identification of statistically significant differences between normal and tumor samples.[30,31] Notwithstanding the admittedly limited sample size, a strong gene expression signature of the tumor samples is detected as illustrated by the empirical cumulative distribution function of the *P* values, which is clearly distinct from what would be expected if no effect was detectable (Figure 1d). Such a strong histology-specific gene expression signature, despite the small sample size, may result from the use of inbred animals in which genetic background is identical among unrelated tumor-bearing mice. An FDR threshold of 5% resulted in the identification of as many as 4137 probes differentially regulated between normal and tumor tissue, and 2163 (52%) and 1974 (48%) up- and down-regulated, respectively (see Supplemental Tables S3 and S4 at *http://ajp.amjpathol.org*). As for the human gene list, annotation analysis of the mouse genes by DAVID revealed a very broad and partially overlapping spectrum of cellular functions (data not shown).

Cross-Species Comparison

As shown above, expression profiling analysis of human and mouse intestinal adenomas and their normal tissue counterparts resulted in very extensive lists of differentially expressed genes even when stringent parameters were used. We postulated that the cross-species comparison of the genes differentially expressed between the two independent data sets would limit bystander and adaptation effects and help in narrowing down the list to conserved transcripts more likely to play relevant roles in the tumorigenic process. As the two profiling data sets were generated with different microarray platforms (cDNA and oligonucleotide arrays for the human and mouse tumors, respectively), the comparative analysis was performed exclusively on the probes present in both platforms (see Materials and Methods).

Conserved probes were selected by applying the following inclusion criteria: FDR <5% for the mouse set (n = 4137 probes) and FDR <0.5% for the human set (n = 1859 probes). Further filtering was performed to eliminate probes with discordant transcriptional directions (eg, up- vs. down-regulated probes) between the two species. Following this procedure, a total of 234 probes representative of 166 genes were selected, 100 and 66 of which were up- and down-regulated, respectively (Tables 1 and 2). In those cases in which a gene is represented by more than one probe in the array platform, the probe ID associated with the lowest FDR value was selected.

Bioinformatic analysis of the 166 genes was performed by assigning them to functional groups based on their GO classification in addition to other information from the scientific literature (Table 3). Overall, up-regulation of genes with functions related to cell division was observed: DNA replication and repair, cell cycle regulation, and the maintenance of genomic integrity. Also, the transcriptional and translational machinery was up-regulated when compared with normal tissues.

To map the conserved genes to existing signaling and cellular pathways, we used the web-based software application IPA (Ingenuity Systems). As expected from our selection of human and mouse intestinal tumors arising from *APC* mutations, IPA revealed several differentially expressed genes encoding for members of the Wnt signal transduction pathway (Figure 2). Of the other pathways included in the IPA database, only the extracellular signal-regulated kinase/mitogen-activated protein kinase signaling network encompassed more than two differentially expressed genes in the conserved signature, namely *PKA*, *PKC*, *PP1*/*PP2A*, and *ELK3* (not shown).

Immunohistochemical Validation of Conserved Targets

To validate the results of our comparative cross-species expression analysis, we have performed immunohistochemistry (IHC) on mouse and human intestinal tissues with antibodies directed against proteins encoded by differentially expressed genes from the list reported in Table 1. Enhanced expression of the cell surface glycoprotein CD44 is an early event in the adenoma-carcinoma sequence both in mouse and human,[36,37] and is thought to result from direct *CD44* transcriptional up-regulation by Wnt/β-catenin signaling.[37] Accordingly, *CD44* was found to be conserved in our cross-species analysis and was used as an internal positive control for the IHC analysis (Figure 3, m–p).

The annexin A1 (*ANXA1*) gene, up-regulated in both human and mouse *APC*/*Apc*-mutant adenomas (Table 1), encodes for annexin A1, an anti-inflammatory protein induced by glucocorticoids and overexpressed in colitis in both human and rat.[38–40] Annexin A1 IHC analysis reveals a distinct perinuclear localization in normal intestinal mucosa, possibly in association with the endoplasmic reticulum (Figure 3, a and b). In *Apc*$^{+/1638N}$ and FAP intestinal tumors, cytoplasmic accumulation of annexin A1 is observed concomitantly with loss of the perinuclear localization (Figure 3, c and d). In distinct tumor areas, nuclear localization was also observed, possibly sugges-

Table 2. The Cross-Species Conserved 166-Gene Signature: Down-Regulated Genes

Mouse			Human		
Probe ID	Unigene*	Gene	Unigene†	Gene	Description
1424600_at	Mm.213898	*Abp1*	Hs.647097	*ABP1*	Amiloride binding protein 1 [amine oxidase (copper-containing)]
1427034_at	Mm.754	*Ace*	Hs.298469	*ACE*	Angiotensin I-converting enzyme (peptidyldipeptidase A) 1
1418553_at	Mm.170461	*Arhgef18*	Hs.465761	*ARHGEF18*	Rho/rac guanine nucleotide exchange factor (GEF) 18
1459924_at	Mm.340818	*Atp6v0a1*	Hs.463074	*ATP6V0A1*	ATPase, H^+ transporting, lysosomal V0 subunit a1
1416582_a_at	Mm.4387	*Bad*	Hs.370254	*BAD*	BCL2-antagonist of cell death
1423635_at	Mm.103205	*Bmp2*	Hs.73853	*BMP2*	Bone morphogenetic protein 2
1456616_a_at	Mm.726	*Bsg*	Hs.591382	*BSG*	BSG: Basigin (Ok blood group)
1424226_at	Mm.218590	*9030617O03Rik*	Hs.309849	*C14orf159*	Chromosome 14 open reading frame 159
1427944_at	Mm.150568	*C1qdc1*	Hs.234355	*C1QDC1*	C1q domain containing 1
1449248_at	Mm.177761	*Clcn2*	Hs.436847	*CLCN2*	Chloride channel 2
1416565_at	Mm.400	*Cox6b1*	Hs.431668	*COX6B1*	Cytochrome *c* oxidase subunit Vib polypeptide 1 (ubiquitous)
1420617_at	Mm.339792	*Cpeb4*	Hs.127126	*CPEB4*	Cytoplasmic polyadenylation element binding protein 4
1415677_at	Mm.21623	*Dhrs1*	Hs.348350	*DHRS1*	Dehydrogenase/reductase (SDR family) member 1
1416697_at	Mm.1151	*Dpp4*	Hs.368912	*DPP4*	Dipeptidylpeptidase 4 (CD26, adenosine deaminase complexing protein 2)
1450314_at	Mm.140332	*Dqx1*	Hs.191705	*DQX1*	DEAQ box polypeptide 1 (RNA-dependent ATPase)
1421136_at	Mm.9478	*Edn3*	Hs.1408	*EDN3*	Endothelin 3
1423005_a_at	Mm.264215	*Espn*	Hs.147953	*ESPN*	Espin
1421969_a_at	Mm.256025	*Faah*	Hs.528334	*FAAH*	Fatty acid amide hydrolase
1452117_a_at	Mm.170905	*Fyb*	Hs.370503	*FYB*	FYN binding protein (FYB-120/130)
1436889_at	Mm.338713	*Gabra1*	Hs.175934	*GABRA1*	γ -Aminobutyric acid (GABA) A receptor, alpha 1
1423236_at	Mm.30249	*Galnt1*	Hs.514806	*GALNT1*	UDP-*N*-acetyl-α -D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase 1 (GalNAc-T1)
1418863_at	Mm.247669	*Gata4*	Hs.243987	*GATA4*	GATA binding protein 4
1429076_a_at	Mm.283495	*Gdpd2*	Hs.438712	*GDPD2*	Glycerophosphodiester phosphodiesterase domain containing 2
1449144_at	Mm.260925	*Gna11*	Hs.654784	*GNA11*	Guanine nucleotide binding protein (G protein), alpha 11 (Gq class)
1419371_s_at	Mm.195451	*Gosr2*	Hs.463278	*GOSR2*	Golgi SNAP receptor complex member 2
1416416_x_at	Mm.37199	*Gstm1*	Hs.75652	*GSTM5*	Glutathione S-transferase M5
1425343_at	Mm.41506	*Hdhd3*	Hs.7739	*HDHD3*	Haloacid dehalogenase-like hydrolase domain containing 3
1422527_at	Mm.16373	*H2-DMa*	Hs.351279	*HLA-DMA*	HLA-DMA: Major histocompatibility complex, class II, DM alpha
1419455_at	Mm.4154	*Il10rb*	Hs.418291	*IL10RB*	Interleukin 10 receptor, beta
1418265_s_at	Mm.1149	*Irf2*	Hs.374097	*IRF2*	Interferon regulatory factor 2
1433775_at	Mm.331907	*C77080*	Hs.591502	*KIAA1522*	KIAA1522
1425547_a_at	Mm.279599	*Klc4*	Hs.408062	*KLC4*	Kinesin light chain 4
1451322_at	Mm.28108	*Cmbl*	Hs.192586	*CMBL*	Carboxymethylenebutenolidase homolog (*Pseudomonas*)
1425780_a_at	Mm.241387	*Tmem167*	Hs.355606	*TMEM167*	Transmembrane protein 167
1425704_at	Mm.192213	*BC022224*	Hs.462859	*MGC4172*	Short-chain dehydrogenase/reductase
1425930_a_at	Mm.628	*Mlx*	Hs.383019	*MLX*	MAX-like protein X
1450376_at	Mm.2154	*Mxi1*	Hs.501023	*MXI1*	MAX interactor 1
1425230_at	Mm.31686	*Nags*	Hs.8876	*NAGS*	*N*-Acetylglutamate synthase
1448331_at	Mm.29683	*Ndufb7*	Hs.532853	*NDUFB7*	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7
1415821_at	Mm.15125	*Nptn*	Hs.187866	*NPTN*	Neuroplastin
1451274_at	Mm.276348	*Ogdh*	Hs.488181	*OGDH*	Oxoglutarate (α -ketoglutarate) dehydrogenase (lipoamide)
1417677_at	Mm.32744	*Opn3*	Hs.534399	*OPN3*	Opsin 3
1449330_at	Mm.29872	*Pdzd3*	Hs.374726	*PDZD3*	PDZ domain containing 3
1435872_at	Mm.328931	*Pim1*	Hs.81170	*PIM1*	Pim-1 oncogene
1425542_a_at	Mm.240396	*Ppp2r5c*	Hs.368264	*PPP2R5C*	Protein phosphatase 2, regulatory subunit B′, gamma isoform
1422847_a_at	Mm.2314	*Prkcd*	Hs.155342	*PRKCD*	Protein kinase C, delta
1424456_at	Mm.4341	*Pvrl2*	Hs.655455	*PVRL2*	Poliovirus receptor-related 2 (herpesvirus entry mediator B)
1430527_a_at	Mm.261818	*Rnf167*	Hs.7158	*RNF167*	Ring finger protein 167
1448704_s_at	Mm.22362	*H47*	Hs.32148	*SELS*	Selenoprotein S

(table continues)

Table 2. *Continued*

Mouse			Human		
Probe ID	Unigene*	Gene	Unigene†	Gene	Description
1448299_at	Mm.246670	*Slc1a1*	Hs.444915	*SLC1A1*	Solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1
1424441_at	Mm.330113	*Slc27a4*	Hs.656699	*SLC27A4*	Solute carrier family 27 (fatty acid transporter), member 4
1433595_at	Mm.281800	*Slc35d1*	Hs.213642	*SLC35D1*	Solute carrier family 35 (UDP-glucuronic acid/UDP-*N*-acetylgalactosamine dual transporter), member D1
1421225_a_at	Mm.41044	*Slc4a4*	Hs.5462	*SLC4A4*	Solute carrier family 4, sodium bicarbonate cotransporter, member 4
1448783_at	Mm.45874	*Slc7a9*	Hs.408567	*SLC7A9*	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 9
1436797_a_at	Mm.300594	*Surf4*	Hs.512465	*SURF4*	Surfeit 4
1428095_a_at	Mm.33869	*Tmem24*	Hs.587176	*TMEM24*	Transmembrane protein 24
1417895_a_at	Mm.25295	*Tmem54*	Hs.534521	*TMEM54*	Transmembrane protein 54
1434553_at	Mm.26088	*Tmem56*	Hs.483512	*TMEM56*	Transmembrane protein 56
1420412_at	Mm.1062	*Tnfsf10*	Hs.478275	*TNFSF10*	Tumor necrosis factor (ligand) superfamily, member 10
1428327_at	Mm.305318	*Trak1*	Hs.535711	*TRAK1*	Trafficking protein, kinesin binding 1
1448737_at	Mm.18590	*Tspan7*	Hs.441664	*TSPAN7*	Tetraspanin 7
1448782_at	Mm.291015	*Txndc11*	Hs.313847	*TXNDC11*	Thioredoxin domain containing 11
1435110_at	Mm.290433	*Unc5b*	Hs.585457	*UNC5B*	Unc-5 homolog B
1426399_at	Mm.26515	*Vwa1*	Hs.449009	*VWA1*	Von Willebrand factor A domain containing 1
1436953_at	Mm.223504	*Wipf1*	Hs.654521	*WASPIP*	WAS/WASL interacting protein family, member 1
1416545_at	Mm.240076	*Zdhhc7*	Hs.592065	*ZDHHC7*	Zinc finger, DHHC-type containing 7

For simplicity, the gene description is only given for the human entry. *Unigene build 163; †Unigene build 202.

Table 3. GO Annotations of the Cross-Species Conserved Genes

Cellular function	Direction	Genes
Cell cycle	Up	*CCND1, CCNA2, CKS2*, CEP55*
DNA replication	Up	*RRM2, RFC3, PUS7*, POLE2, RPA1, NAP1L1*
DNA repair	Up	*XRN2*, PUS7*, POLD3, PARP1, RAD51AP1*
Apoptosis	Up	*DLG5, BIRC5*
	Down	*BAD, IRF2*, TNFSF10*, UNC5B, PIM1*
RNA and protein biogenesis, processing and transport	Up	*RPS4X, NOLA2, CCT5, ILF3, NCBP2, TCOF1, MKI67IP, THOC1, EIF3S9, EIF3S1, POLR2B, SYNCRIP, IARS, SNRPB2, RPS24, NOLA5A, HNRPU, XRN2*, POLR1C, SFRS1, PPIG, CCT4*
	Down	*ZDHHC7, GOSR2, GALNT1, TRAK1*
TGFβ	Up	*TGFBR2, SMAD6*
	Down	*BMP2*
Chromatin remodelization	Up	*CBX3, SMARCA5, DNMT1*
Cytoskeleton organization	Up	*LLGL1, KRT18, CTTNBP2NL*
	Down	*ARHGEF18, KLC4, WASPIP*, ESPN**
Genome integrity (mitotic checkpoint and telomere maintenance)	Up	*STMN1, ZWILCH, BUB3, CKS2*, PRC1, BIRC5*, PARP1*, PRKDC*
	Down	*ESPN**
Cell adhesion and migration	Up	*SEMA4D, JUB, CD44, TRIP6, DLG5, MAP2K5*
	Down	*NPTN, PVRL2, TSPAN7, WASPIP**
Transcription factors	Up	*SOX4, RUNX1, YAP1, MTDH, ITGB3BP, ZFP451, RFP, RNF12*
	Down	*MLX, MXI1, GATA4, IRF2*
G protein signaling	Up	*PRKAR1A*
	Down	*OPN3, PRKCD, GNA11*
Immune response	Up	*ANXA1*
	Down	*IRF2*, TNFSF10*, IL10RB, HLA-DMA, BSG, SELS*
Chromatin remodelization	Up	*CBX3, SMARCA5, DNMT1*
Metabolism	Up	*SLC7A5, MTHFD2, GGH, AYTL2, MGEA5, PCK2*
	Down	*DHRS1, GSTM1, DGAT1, SLC27A4, NAGS, HDHD3, GDPD2, SLC7A9, OGDH*
Proteolysis	Up	*ENC1, PSMC2, MMP7*
	Down	*DPP4, ACE, RNF167*

Functional annotations were retrieved from GO and the scientific literature. Genes belonging to more than one functional category are marked with*. Gene symbols refer to the human annotation.

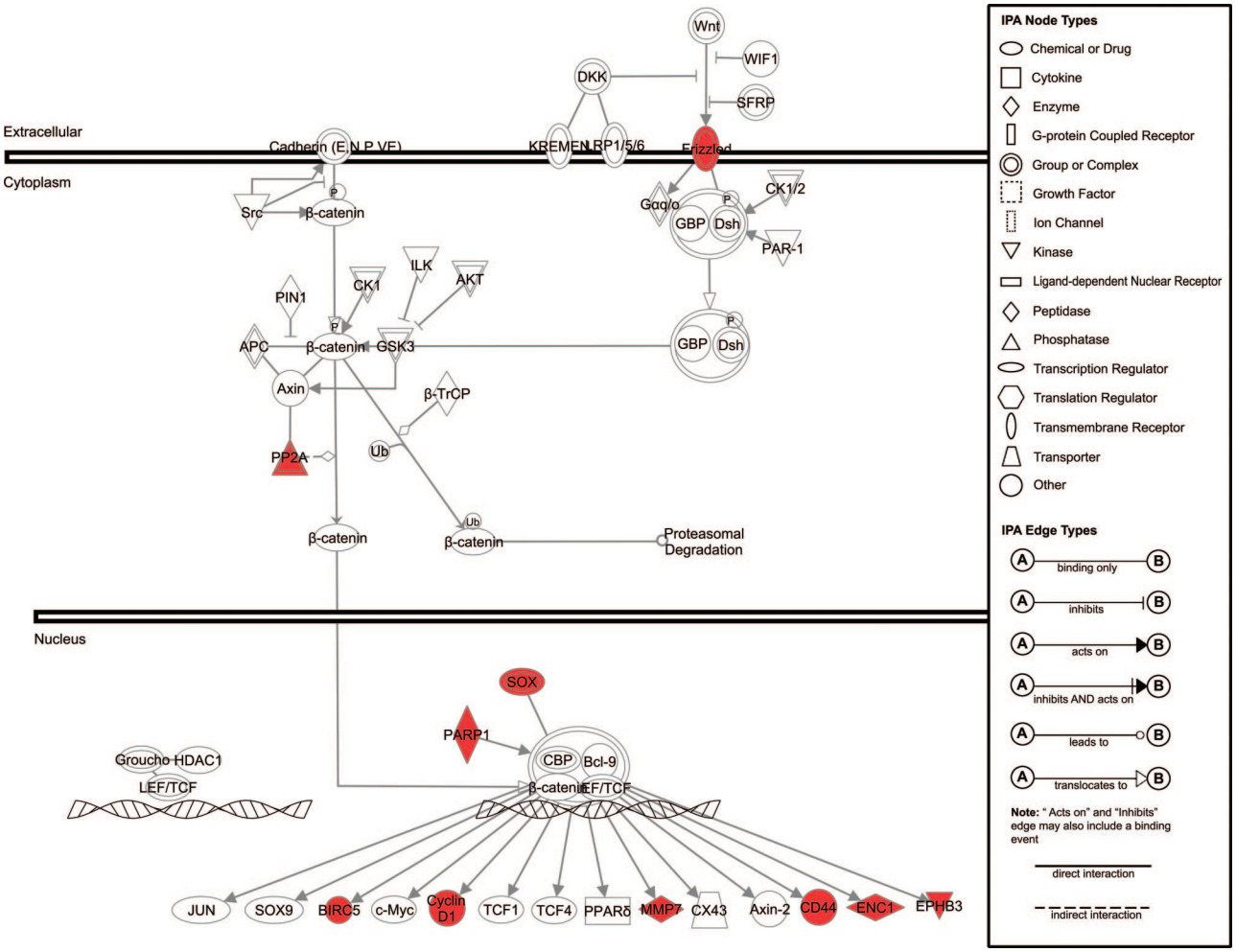

Figure 2. IPA of the genes encompassed by the conserved 166 signatures and belonging to the Wnt signal transduction pathway. The canonical Wnt pathway from the IPA database was slightly modified to accommodate additional Wnt target genes.[11,12] The signaling network is represented graphically as nodes (symbols representing genes) and **lines**/**arrows** (biological relationship between the genes according to the legend). Red and green gene symbols denote up- and down-regulated genes, respectively. White symbols denote genes not differentially expressed in the conserved signature.

tive of mitogenic stimulation as previously reported.[41] Also, annexin A1 expression appears not to be confined to the intestinal epithelium, but also to extend to the stromal compartment[42] (Figure 3, a–d).

Cyclin A2 (*CCNA2*) is a ubiquitously expressed regulator of cell cycle progression known to promote G_1/S and G_2/M transitions.[43] In normal mouse (upper GI) and human (colon) intestinal mucosa, CCNA2 IHC analysis shows nuclear expression mainly restricted to the crypts (Figure 3, e–h). *CCNA2* up-regulation in both mouse and human intestinal adenomas (Table 1) is reflected by an increase in the relative number of cells with nuclear CCNA2 staining spread throughout the tumors. The latter is indicative of enhanced cell proliferation of *APC*-mutant tumor cells, as was also confirmed by the GO analysis of the conserved gene list (Table 3).

To date, the function of Marks-like protein 1 (*MARCKSL1*) is not fully elucidated, although evidence in the literature indicates that it might be involved in the regulation of intracellular Ca^{2+} levels under the control of protein kinase C.[44] Up-regulation of this protein has been previously reported in prostate cancer,[45] and is here found in both FAP and $Apc^{+/1638N}$ adenomas (Table 1). Similar to what was observed for CCCNA2, MARCKSL1 expression is limited to a specific subset of cells within the normal human (colon) and mouse (upper GI) intestinal crypts. Likewise, a pronounced increase in cytoplasmic expression is observed in the vast majority of tumor cells (Figure 3, i–l). Overall, the above IHC results validate the cross-species expression profiling data for genes belonging to distinct GO and functional categories.

The Conserved Cross-Species Signature As a Tool to Differentiate Hereditary Polyposis Syndromes

Apart from its implications for the understanding of the molecular and cellular mechanisms underlying *APC*-driven colorectal tumorigenesis, the conserved cross-species signature may represent a useful tool to discriminate among adenomas from hereditary patients with different genetic syndromes, namely *APC*- and *MYH*-as-

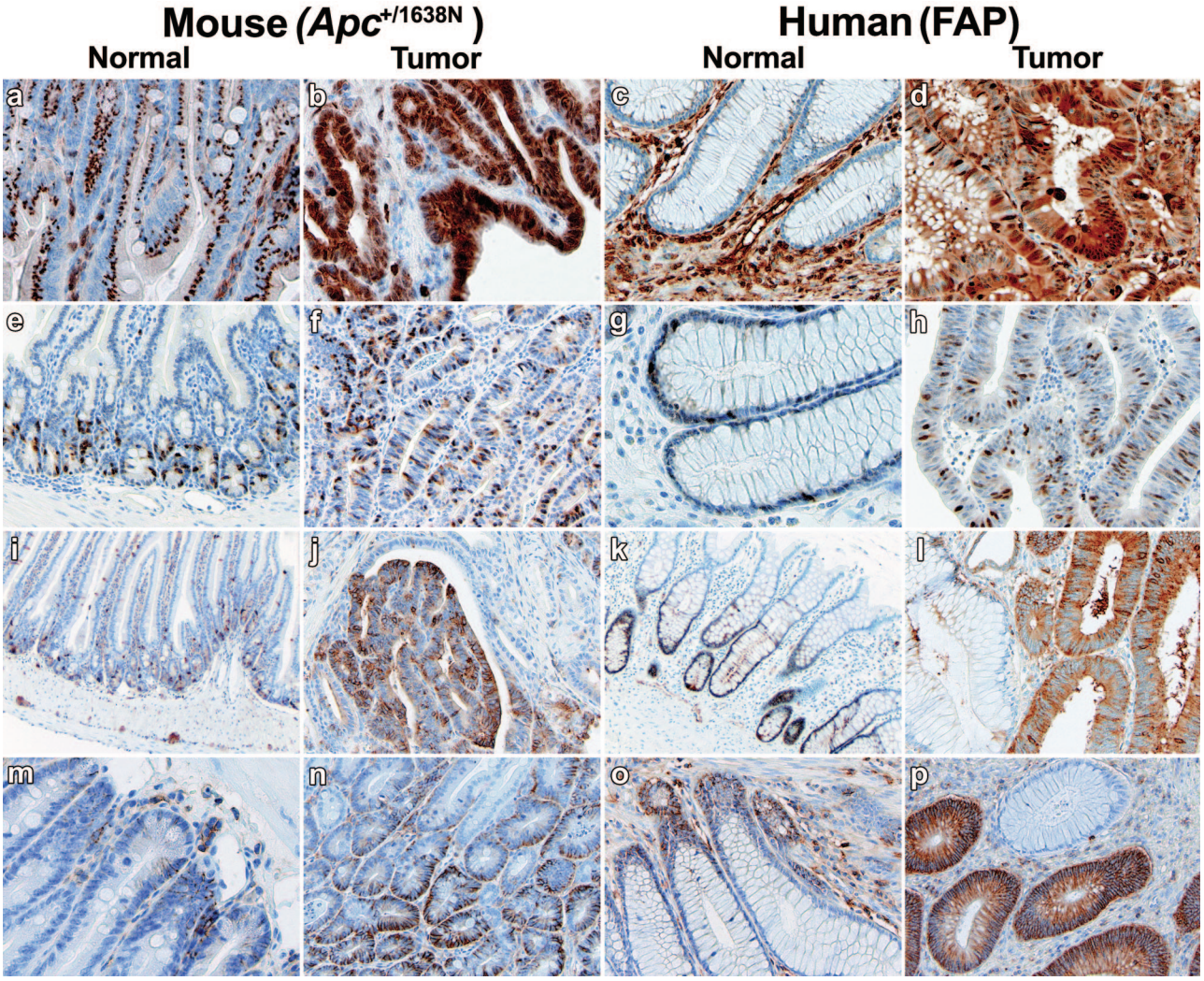

Figure 3. Immunohistochemistry validation analysis of cross-species conserved genes. Human (colorectal polyps and normal mucosa from FAP patients carrying germline *APC* mutations) and mouse (duodenal adenomas and normal mucosa from inbred C57BL6/J $Apc^{+/1638N}$ and $Apc^{+/+}$ mice) tissue sections were analyzed with specific antibodies (see Materials and Methods) for expression of the following proteins: ANXA1 (**a–d**), CCNA2 (**e–h**), MARCKSL1 (**i–l**), and CD44 (**m–p**).

sociated polyposis. To this aim, an additional 14 adenomas have been obtained from five unrelated patients with pathologically confirmed polyposis of the colon and carrying biallelic *MYH* germline mutations.[23] As for the *APC*-mutant polyps, RNA was extracted from the microdissected MAP adenomas and subsequently used for expression profiling. First, unsupervised hierarchical clustering was applied to all 56 profiles (from both the *APC*- and *MYH*-mutant patients) without any prior filtering to generate an overview of the global gene expression differences among all samples. Overall, we could not observe any clear association with mutation status (data not shown). The mixed-effects regression model[34] was again applied, this time fitted to consider mutation (*APC* and *MYH* germline mutation carriers) as having a fixed effect and patient as having a random effect. With an FDR set to 0.5%, we were able to select 49 genes differentially expressed between FAP and MAP adenomas (Table 4). To investigate further whether the 49-gene signature as a whole can predict the underlying *APC* or *MYH* gene defect, we applied the previously described globaltest for the analysis of microarray data.[32] This test assesses whether the global expression pattern of a group of genes is significantly related to any given parameter. It should be noted that, when applying the globaltest, the patient effect cannot be regarded as random and therefore be controlled by its inclusion in the model as a confounder, as it would also represent the genotype effect (all samples from a patient belong to the same genotype). We circumvented this problem by first selecting at random one sample at a time from each patient and then applying the globaltest. After repeating this process for 1000 random combinations of patients' samples, 57% of the computed *P* values were found to be below the 0.05 threshold, whereas the maximum value is close to 0.5 (Figure 4a). Next, we repeated this procedure with the conserved 166-gene signature. In sharp contrast with the previous result, 99.4% of the computed *P* values are below the 0.05 threshold with a maximum of 0.06 (Figure 4b). Accordingly, two-dimensional hierarchical clustering analysis of the expression profiles obtained from all of the FAP and MAP polyps with the

Table 4. The 49-Gene Signature Based on Statistically Significant Differences (FDR = 0.5%) Between Expression Profiles of FAP- and MAP-Derived Adenomatous Polyps, After Implementation of the Mixed-Effect Regression Model[34] Fitted Considering Mutation (*APC* vs. *MYH*) as Having Fixed Effect and Patient as Having a Random Effect

GenBank	Gene symbol	Gene description
H41285	*GDPD2*	Glycerophosphodiester phosphodiesterase domain containing 2
T46878	*EIF3S1*	Eukaryotic translation initiation factor 3, subunit 1 alpha
AA479795	*ISG20*	Interferon-stimulated exonuclease gene
AA151214	*G3BP2*	Ras-GTPase activating protein SH3 domain-binding protein 2
H28091	*PMP22*	Peripheral myelin protein 22
N59330	*NUP35*	Nucleoporin
H19333	*LOC285550*	Hypothetical protein LOC285550
AA449688	*FLJ32065*	Hypothetical protein FLJ32065
N/A	*N/A*	—
AA977417	*AA977417*	—
N50636	*RAP1GDS1*	RAP1, GTP-GDP dissociation stimulator 1
T61866	*IPO7*	Importin 7
AA453435	*LTV1*	LTV1 homolog (*S. cerevisiae*)
N91962	*EEF1E1*	Eukaryotic translation elongation factor 1 epsilon 1
H77636	*CD68*	CD68 antigen
AI308916	*PRSS3*	Protease, serine, 3 (mesotrypsin)
AA478589	*APOE*	Apolipoprotein E
AA459401	*KLK10*	Kallikrein 10
AA625765	*DDA1*	DDA1
AA205665	*SET*	SET translocation (myeloid leukemia-associated)
AA707453	*FLJ43855*	Similar to sodium- and chloride-dependent creatine transporter
AA464147	*CARS*	Cysteinyl-tRNA synthetase
AA456630	*ARHGEF18*	Rho/rac guanine nucleotide exchange factor (GEF) 18
N20475	*CTSD*	Similar to RIKEN cDNA 6330512M04 gene (mouse)
N31935	*ANGPTL1*	Angiopoietin-like 1
R77512	*PCDH1*	Protocadherin 1 (cadherin-like 1)
N31492	*FMO4*	Topoisomerase (DNA) I pseudogene 1
N45236	*KIAA0114*	KIAA0114 gene product
H60549	*CD59*	CD59 antigen, complement regulatory protein
AA907626	*KIF26B*	Kinesin family member 26B
AA917374	*TIMP2*	TIMP metallopeptidase inhibitor 2
AA983530	*VNN1*	Vanin 1
N90109	*NCL*	U23 small nucleolar RNA
H15431	*POLR2D*	Polymerase (RNA) II (DNA directed) polypeptide D
H52673	*BAK1*	BCL2-antagonist/killer 1
T72259	*CYP2A6*	Cytochrome P450, family 2, subfamily A, polypeptide 6
AA455910	*F2R*	Coagulation factor II (thrombin) receptor
AA775840	*C9orf123*	Chromosome 9 open reading frame 123
AA464566	*LRP1*	Low density lipoprotein-related protein 1
AA489640	*IFIT1*	Interferon-induced protein with tetratricopeptide repeats 1
AA962541	*LOC286167*	Hypothetical protein LOC286167
AA464421	*PCGF2*	Polycomb group ring finger 2
H25761 AI668603	—	—
AA447748	*DLD*	Dihydrolipoamide dehydrogenase
AA278755	*CEP27*	Centrosomal protein
AI261833	*SLC7A9*	Solute carrier family 7 (cationic amino acid transporter, y+ system) member 9
H72802	*ESPN*	Espin
AA608713	*C1QDC1*	C1q domain containing 1
AA047465	*SLC6A8*	Solute carrier family 6 (neurotransmitter transporter, creatine), member 8

49- and 166-gene signatures confirms that the latter is considerably more discriminative than the former in resolving tumors from carriers of *APC* germline mutations from those derived from MAP patients (Figure 4, c and d).

Discussion

Expression profiling by oligonucleotide and cDNA microarray platforms has rapidly become a commonly used tool for the qualitative and quantitative evaluation of the genome-wide transcriptional activity of human cancers. However, the outcomes of expression profiling of cancers are often very complex as they reflect the heterogeneity of cell types and biological activities present within the neoplastic mass, thus making their functional interpretation a difficult task. This is certainly the case for the expression (and genomic) profiles obtained to date from colorectal cancers. Although several studies have been published in the scientific literature, the degree of overlap between independent data sets is limited, possibly also as a consequence of differences in patient cohorts and methodologies used.[16] Cross-species comparison of cancer profiling data represents a valuable approach to i) decrease the complexity of omics signatures, ii) pinpoint

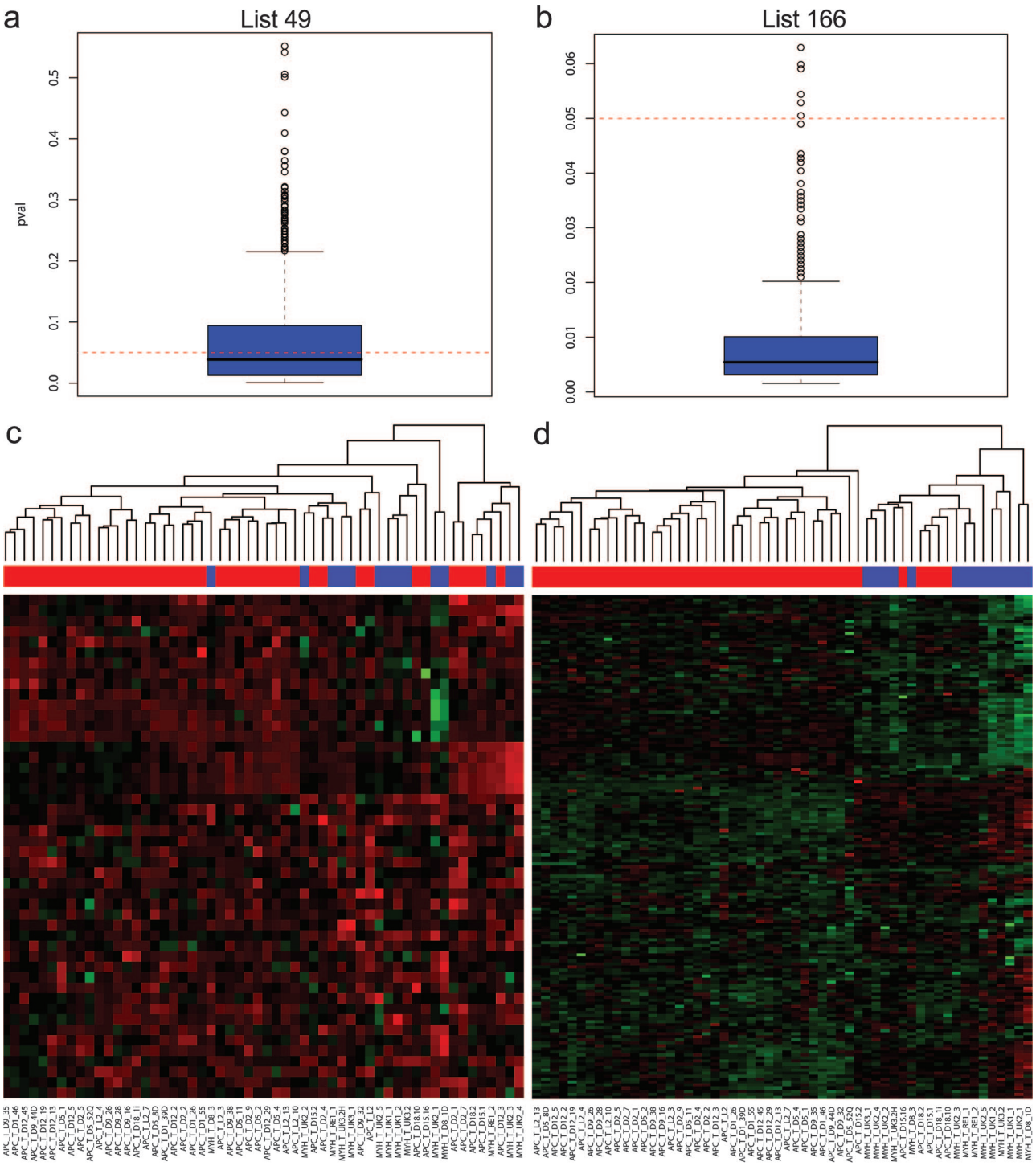

Figure 4. Analysis of the cross-species conserved signature as a tool to separate hereditary polyposis syndromes due to *APC* (FAP) or *MYH* (MAP) germline mutations. The globaltest[32] was performed with the 49 (**a**)- and 166 (**b**)-gene signature and graphically represented by box plots of the *P* values generated after 1000 iterations in which only one random sample from each patient was used at a time. Box plots were generated using the standard settings present in R2.4.1. The **filled** blue **boxes** encompass the range of *P* values representative of 50% of the data points, whereas the **central line** represents the median. two-dimensional hierarchical clustering analysis was performed with the 49 (**c**)- and 166 (**d**)-gene signature, respectively, on the expression profiles obtained from all 56 colorectal adenomas (42 from *APC* and 14 from *MYH* gene mutation carriers). The colored bar above the heat map represents the mutation status of the corresponding polyp samples: red, polyps from patients carrying germline *APC* mutations; blue, polyps from patients carrying bi-allelic germline *MYH* mutations.

conserved target genes more likely to play rate-limiting functional roles in tumor initiation and progression to malignancy, and iii) accelerate the development of tailor-made anticancer therapies.[46,47]

Notwithstanding the above-mentioned heterogeneity, colorectal cancer represents, at least from a genetic perspective, a relatively homogeneous disease as the vast majority of the sporadic cases is known to be trig-

gered by somatic mutations at the *APC* or *CTNNB1* (β-catenin) genes, leading to the constitutive activation of the canonical Wnt signaling pathway.[10] These mutations are known to initiate the formation of aberrant crypt foci and adenomatous polyps, the earliest benign precursors of the adenoma-carcinoma sequence. Also, germline *APC* mutations underlie FAP, an autosomal dominant predisposition to the development of multiple adenomatous polyps throughout the colon-rectum.[48] The availability of a unique collection of adenomatous polyps obtained from FAP patients carrying germline *APC* mutations and from a mouse model, $Apc^{+/1638N}$, carrying a targeted mutation in the endogenous *Apc* gene allowed us to perform the cross-species computational comparison of their gene expression profiles and derive a conserved 166-gene signature. It should be noted that whereas FAP patients mainly develop polyps in the colon-rectum, *Apc* mouse models are characterized by adenomas clustering in the upper gastrointestinal tract, mainly in the duodenum. This anatomical difference between the mouse and human adenomas used for the cross-species comparison may exert a confounding effect in our computational analysis as duodenum and colon-rectum represent distinct organs. However, it may also confer an additional advantage to our approach as tissue-specific differences between the two GI tracts are likely to be filtered out, thus retaining only those conserved differentially expressed genes more likely to play functional roles in intestinal tumor formation, regardless of anatomical sub-location. The same holds true for our methodological approach: different microarray platforms were used to derive the human (cDNA arrays) and mouse (oligonucleotide arrays) gene profiles. In both cases, laser-guided microdissection was used to enrich in tumor cells without the confounding effects of contaminating stromal cells. Overall, our IHC analysis of a subset of proteins encoded by the conserved genes confirmed their differential expression between normal tissue and adenomas in both species (Figure 3), thus validating our methodological strategy.

The significance thresholds used to generate the differentially expressed lists of genes for the human (approximately 10% of the represented genes, with FDR = 0.5%) and mouse (approximately 18% of the represented genes, with FDR = 5%) studies are admittedly arbitrary. They were chosen using two generic criteria: i) the gene lists would be representative of the differential signature without encompassing an excessive percentage of false positives, and ii) the resulting conserved list of differentially expressed genes would be sufficiently large to enable pathway analysis. The presence in our cross-species signature of several genes known to be differentially expressed in sporadic colorectal cancers[16] also represents an indirect confirmation of the general validity of our computational approach.

GO-based functional analysis of the 166 conserved genes reveals a general increase in cell division as shown by the up-regulation of genes related to DNA replication and repair, cell cycle regulation, and the maintenance of genomic integrity (Table 3). Notably, exclusively up-regulated genes were encompassed within these categories, indicative of the increased proliferation rate of tumor cells when compared with normal ones. Genes belonging to the transcriptional and translational machinery were also up-regulated when compared with normal tissues. These included genes involved in ribosome biogenesis, mRNA synthesis and maturation, and protein synthesis and folding (Table 3).

As expected from our selection of adenomas from *APC*-mutant patients and mouse models, several members of the Wnt signal transduction pathway are included among the conserved 166 differentially expressed genes (Figure 2). These included the Frizzled receptor homolog *FZD6*, the protein phosphatase type 2A (*PP2A*), the HMG box transcription factor *SOX4*,[49] and several Wnt downstream transcriptional targets, namely, the matrix metallopeptidase matrilysin (*MMP7*),[36,37,50] *CD44*,[36] *ENC1* (*ectodermal-neural cortex 1)*[51], ephrin receptor B3 (*EPHB3*),[52] cyclin D1 (*CCND1*),[53,54] and the apoptosis inhibitor survivin (*BIRC5*)[55,56] (Figure 2). However, *AXIN2*, a well known downstream Wnt target gene, is not included in this list simply because its probe is not encompassed by the human cDNA array. Other known Wnt target genes such as *EPHB2*, *SOX9*, and *MYC* were excluded because of the high stringency of the statistical thresholds used or to their absence in one of the platforms. Recently, an "intestinal Wnt/TCF4 signature" was obtained by integrating expression profiling data from CRC cell lines engineered with an inducible block of Wnt signaling and from sporadic human adenomas and carcinomas.[11] Comparison of this 208-gene Wnt/TCF gene signature with our cross-species conserved list revealed 10 common entries, 4 of which belong to the Wnt/β-catenin signaling pathway (*CD44*, *ENC1*, *EPHB3*, and *SOX4*). The latter is not surprising in view of the different computational approaches and tumor cohorts used. Moreover, the use of CRC cell lines with dominant negative *TCF4* constructs does not necessarily mimic the initial and rate-limiting loss of *APC* function characteristic of the mouse and human adenomas used in our cross-species analysis. Of more interest is the comparison with the study by Kaiser et al[57] in which a cross-species comparison was performed among human and mouse intestinal tumors together with mouse embryonic stages of intestinal development. As depicted in Supplemental Table S5 (see *http://ajp.amjpathol.org*), the overlap between the two studies is high, with 46 of 166 differentially expressed genes shared between the data sets. Notably, the overlap is considerably higher with genes showing similar behavior in intestinal tumorigenesis and embryonic development.

Apoptosis inhibition in the adenomas, as suggested by *BIRC5* up-regulation, is also strengthened by the conserved down-regulation of the *BAD* gene, encoding for a potent pro-apoptotic protein. BAD forms heterodimers with BCL2 and BCL-XL, thus repressing their anti-apoptotic function.[58,59]

Two members of the TGF-β signaling pathways are up-regulated among the cross-species conserved genes, namely *SMAD6* and *TGFBR2*. The TGF-β ligand mediates its effects through the transmembrane type I (TGFBR1) and type II receptor subunits (TGFBR2), and in

the cytoplasm through stimulatory and inhibitory SMADs. The up-regulation of the *TGFBR2* gene encoding for the type II receptor is remarkable in view of its frequent mutational inactivation in a substantial proportion of sporadic colon cancers.[60] The *SMAD6* gene encodes for an inhibitory SMAD protein that becomes up-regulated as the result of a negative feedback loop. SMAD6 is thought to represent a key component in the integration of signals from different pathways and was shown to exert BMP inhibitory activity.[61] Down-regulation of the bone morphogenetic *BMP2* gene apparently confirms the inhibition of this TGF-β-related pathway. Although its role in tumorigenesis is yet unclear, SMAD6 up-regulation has been reported in other tumor types.[62] Overall, the conserved gene signature is indicative of the activation of TGF-β and inhibition of BMP signaling at early stages of intestinal tumorigenesis. However, this observation needs to be validated by additional expression and reporter assays.

Among the many genes encompassed by the cross-species conserved signature, the up-regulation of *ANXA1* is of interest in view of its phospholipase A2 (PPA2) inhibitory activity, an enzyme involved in the synthesis of prostaglandins during inflammation.[42] Antibodies against annexin A1 have been found in patients with inflammatory bowel disorders.[38] Also, its up-regulation was shown to occur in mitogenically stimulated cells in a PKC phosphorylation-dependent fashion, accompanied by its translocation from the cytoplasm to the nucleus.[41] Notably, changes in ANXA1 subcellular localization were also observed in our IHC validation analysis (Figure 3).

Apart from its implications for the understanding and elucidation of the molecular and cellular mechanisms underlying *APC*-driven intestinal tumor formation, the cross-species conserved gene signature may also represent a useful tool to discriminate among hereditary polyposis patients with distinct genetic defects. Expression profiling analysis of the additional set of 14 colorectal adenomas obtained from patients carrying bi-allelic mutations at the *MYH* gene showed a high degree of similarity with the *APC* profiles. This could be explained by previous observations, according to which the *APC* gene is a preferential target for somatic mutations in colorectal adenomas from carriers of bi-allelic *MYH* germline mutations.[63] The observed high degree of similarity between expression profiles from FAP and MAP polyps could then be explained provided that the somatic *APC* mutation does represent the initiating event in *MYH*-associated polyp formation. Alternatively, human adenoma profiles may be similar notwithstanding the initiating genetic defect, as indicated by our own most recent results with the expression analysis of three polyposis patients of unknown genetic basis (and no germline mutations found after sequencing of the *MYH* and *APC* genes). Also in these cases, the resulting profiles were virtually indistinguishable from those derived from *MYH*- and *APC*-mutant polyps (data not shown).

Nevertheless, by applying an FDR threshold of 0.5%, we could generate a 49-gene signature based on differences between *MYH*- and *APC*-mutant human polyps. Yet, both globaltest and two-dimensional hierarchical clustering analyses showed that the conserved 166-signature clusters more accurately the expression profile data from FAP and MAP patients than does the 49-gene signature (Figure 4).

In conclusion, cross-species comparison of expression profiles of intestinal adenomas obtained from hereditary polyposis patients and mouse models carrying germline *APC* mutations resulted in a signature of 166 differentially expressed genes. Functional annotation of the conserved genes indicates an overall increase in cell division and the up-regulation of the Wnt/β-catenin signaling pathway. These main cellular and molecular changes are accompanied by a plethora of gene-specific changes yet to be tested by functional assays to determine their relative contribution to intestinal tumor formation. Additional validation on independent polyp cohorts and further fine-tuning of the conserved gene signature are needed toward the development of an expression-based assay to classify hereditary polyposis syndromes.

Acknowledgments

We thank Dr. Guido Jenster and Dr. Don de Lange for granting access and helping with the Sequence Retrieval System (SRS), and for fruitful discussions; Dr. Bruce J. Aronow for providing the detailed list of genes from his group's microarray results used here for a comparison; and Mr. Frank van der Panne for his assistance with the artwork.

References

1. de la Chapelle A: Genetic predisposition to colorectal cancer. Nat Rev Cancer 2004, 4:769–780
2. Kinzler KW, Vogelstein B: Lessons from hereditary colorectal cancer. Cell 1996, 87:159–170
3. Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M, Sargeant L, Krapcho K, Wolff E, Burt R, Hughes J, Warrington J, McPherson J, Wasmuth J, LePaslier D, Abderrahim H, Cohen D, Leppert M, White R: Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991, 66:589–600
4. Joslyn G, Carlson M, Thliveris A, Albertsen H, Gelbert L, Samowitz W, Groden J, Stevens J, Spirio L, Robertson M, Sargeant L, Krapcho K, Wolff E, Burt R, Hughes JP, Warrington J, McPherson J, Wasmuth J, Le Paslier D, Abderrahim H, Cohen D, Leppert M, White R: Identification of deletion mutations and three new genes at the familial polyposis locus. Cell 1991, 66:601–613
5. Kinzler KW, Nilbert MC, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hamilton SR, Hedge P, Markham A, Carlson M, Joslyn G, Groden J, White R, Miki Y, Miyoshi Y, Nishisho I, Nakamura Y: Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. Science 1991, 251:1366–1370
6. Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, Hedge P: Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 1991, 253:665–669
7. Smith KJ, Johnson KA, Bryan TM, Hill DE, Markowitz S, Willson JK, Paraskeva C, Petersen GM, Hamilton SR, Vogelstein B, Kinzler KW: The APC gene product in normal and tumor cells. Proc Natl Acad Sci USA 1993, 90:2846–2850
8. Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN, Vogelstein B, Kinzler KW: APC mutations occur early during colorectal tumorigenesis. Nature 1992, 359:235–237
9. Miyoshi Y, Nagase H, Ando H, Horii A, Ichii S, Nakatsuru S, Aoki T, Miki Y, Mori T, Nakamura Y: Somatic mutations of the APC gene in

colorectal tumors: mutation cluster region in the APC gene. Hum Mol Genet 1992, 1:229–233
10. Fodde R, Smits R, Clevers H: APC, signal transduction and genetic instability in colorectal cancer. Nat Rev Cancer 2001, 1:55–67
11. Van der Flier LG, Sabates-Bellver J, Oving I, Haegebarth A, De Palo M, Anti M, Van Gijn ME, Suijkerbuijk S, Van de Wetering M, Marra G, Clevers H: The intestinal Wnt/TCF signature. Gastroenterology 2007, 132:628–632
12. Vlad A, Rohrs S, Klein-Hitpass L, Muller O: The first five years of the Wnt targetome, Cell Signal 2008, 20:795–802
13. Lipton L, Tomlinson I: The genetics of FAP and FAP-like syndromes. Fam Cancer 2006, 5:221–226
14. Cheadle JP, Sampson JR: Exposing the MYtH about base excision repair and human inherited disease. Hum Mol Genet 2003, 12 Spec No 2:R159–R165
15. Fodde R, Smits R: Disease model: familial adenomatous polyposis. Trends Mol Med 2001, 7:369–373
16. Cardoso J, Boer J, Morreau H, Fodde R: Expression and genomic profiling of colorectal cancer. Biochim Biophys Acta 2007, 1775:103–137
17. Sweet-Cordero A, Mukherjee S, Subramanian A, You H, Roix JJ, Ladd-Acosta C, Mesirov J, Golub TR, Jacks T: An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. Nat Genet 2005, 37:48–55
18. Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, Matusik R, Thomas GV, Sawyers CL: Myc-driven murine prostate cancer shares molecular features with human prostate tumors. Cancer Cell 2003, 4:223–238
19. Lee JS, Chu IS, Mikaelyan A, Calvisi DF, Heo J, Reddy JK, Thorgeirsson SS: Application of comparative functional genomics to identify best-fit mouse models to study human cancer. Nat Genet 2004, 36:1306–1311
20. Lee JS, Grisham JW, Thorgeirsson SS: Comparative functional genomics for identifying models of human cancer. Carcinogenesis 2005, 26:1013–1020
21. Fodde R, Edelmann W, Yang K, van Leeuwen C, Carlson C, Renault B, Breukel C, Alt E, Lipkin M, Khan PM, Kucherlapati R: A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. Proc Natl Acad Sci USA 1994, 91:8969–8973
22. Smits R, van der Houven van Oordt W, Luz A, Zurcher C, Jagmohan-Changur S, Breukel C, Khan PM, Fodde R: Apc1638N: a mouse model for familial adenomatous polyposis-associated desmoid tumors and cutaneous cysts. Gastroenterology 1998, 114:275–283
23. Cardoso J, Molenaar L, de Menezes RX, van Leerdam M, Rosenberg C, Moslein G, Sampson J, Morreau H, Boer JM, Fodde R: Chromosomal instability in MYH- and APC-mutant adenomatous polyps. Cancer Res 2006, 66:2514–2519
24. Cardoso J, Molenaar L, de Menezes RX, Rosenberg C, Morreau H, Moslein G, Fodde R, Boer JM: Genomic profiling by DNA amplification of laser capture microdissected tissues and array CGH. Nucleic Acids Res 2004, 32:e146
25. Ihaka R, Gentleman R: R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 1996, Vol 5, No 3, 299–314
26. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M: Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002, 18(Suppl 1):S96–S104
27. Wu H, Kerr MK, Cui X, Gary A: MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments. 2005, New York, Churchill, p 314–341
28. Benjamini Y, Hochberg Y: Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc B Met 1995, 289–300
29. Gautier L, Cope L, Bolstad BM, Irizarry RA: affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004, 20:307–315
30. Smyth GK: Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004, 3:Article 3
31. Smyth GK: Limma: linear models for microarray data. Bioinformatics and Computational Biology Solutions using R and Bioconductor. Edited by R Gentleman, V Cary, S Dudoit, R Irizarry, W Huber. New York, Springer, 2005, pp 397–420
32. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC: A global test for groups of genes: testing association with a clinical outcome. Bioinformatics 2004, 20:93–99
33. Veldhoven A, de Lange D, Smid M, de Jager V, Kors JA, Jenster G: Storing, linking, and mining microarray databases using SRS. BMC Bioinformatics 2005, 6:192
34. Kerr MK, Martin M, Churchill GA: Analysis of variance for gene expression microarray data. J Comput Biol 2000, 7:819–837
35. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003, 4:P3
36. Abbasi AM, Chester KA, Talbot IC, Macpherson AS, Boxer G, Forbes A, Malcolm AD, Begent RH: CD44 is associated with proliferation in normal and neoplastic human colorectal epithelial cells. Eur J Cancer 1993, 29A:1995–2002
37. Wielenga VJ, Smits R, Korinek V, Smit L, Kielman M, Fodde R, Clevers H, Pals ST: Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. Am J Pathol 1999, 154:515–523
38. Perretti M, Gavins FN: Annexin 1: an endogenous anti-inflammatory protein. News Physiol Sci 2003, 18:60–64
39. Vergnolle N, Pages P, Guimbaud R, Chaussade S, Bueno L, Escourrou J, Comera C: Annexin 1 is secreted in situ during ulcerative colitis in humans. Inflamm Bowel Dis 2004, 10:584–592
40. Vergnolle N, Comera C, Bueno L: Annexin 1 is overexpressed and specifically secreted during experimentally induced colitis in rats. Eur J Biochem 1995, 232:603–610
41. Kim YS, Ko J, Kim IS, Jang SW, Sung HJ, Lee HJ, Lee SY, Kim Y, Na DS: PKCdelta-dependent cleavage and nuclear translocation of annexin A1 by phorbol 12-myristate 13-acetate. Eur J Biochem 2003, 270:4089–4094
42. Pepinsky RB, Sinclair LK, Browning JL, Mattaliano RJ, Smart JE, Chow EP, Falbel T, Ribolini A, Garwin JL, Wallner BP: Purification and partial sequence analysis of a 37-kDa protein that inhibits phospholipase A2 activity from rat peritoneal exudates. J Biol Chem 1986, 261:4239–4246
43. Pagano M, Pepperkok R, Verde F, Ansorge W, Draetta G: Cyclin A is required at two points in the human cell cycle. EMBO J 1992, 11:961–971
44. Raufman JP, Malhotra R, Xie Q, Raffaniello RD: Expression and phosphorylation of a MARCKS-like protein in gastric chief cells: further evidence for modulation of pepsinogen secretion by interaction of Ca2+/calmodulin with protein kinase C. J Cell Biochem 1997, 64:514–523
45. Chaib H, Cockrell EK, Rubin MA, Macoska JA: Profiling and verification of gene expression patterns in normal and malignant human prostate tissues by cDNA microarray analysis. Neoplasia 2001, 3:43–52
46. Graeber TG, Sawyers CL: Cross-species comparisons of cancer signaling. Nat Genet 2005, 37:7–8
47. Peeper D, Berns A: Cross-species oncogenomics in cancer gene identification. Cell 2006, 125:1230–1233
48. Fodde R, Khan PM: Genotype-phenotype correlations at the adenomatous polyposis coli (APC) gene. Crit Rev Oncog 1995, 6:291–303
49. Reichling T, Goss KH, Carson DJ, Holdcraft RW, Ley-Ebert C, Witte D, Aronow BJ, Groden J: Transcriptional profiles of intestinal tumors in Apc(Min) mice are unique from those of embryonic intestine and identify novel gene targets dysregulated in human colorectal tumors. Cancer Res 2005, 65:166–176
50. Crawford HC, Fingleton BM, Rudolph-Owen LA, Goss KJ, Rubinfeld B, Polakis P, Matrisian LM: The metalloproteinase matrilysin is a target of beta-catenin transactivation in intestinal tumors. Oncogene 1999, 18:2883–2891
51. Fujita M, Furukawa Y, Tsunoda T, Tanaka T, Ogawa M, Nakamura Y: Up-regulation of the ectodermal-neural cortex 1 (ENC1) gene, a downstream target of the beta-catenin/T-cell factor complex, in colorectal carcinomas. Cancer Res 2001, 61:7722–7726
52. Batlle E, Henderson JT, Beghtel H, van den Born MM, Sancho E, Huls G, Meeldijk J, Robertson J, van de Wetering M, Pawson T, Clevers H: Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell 2002, 111:251–263
53. Tetsu O, McCormick F: Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. Nature 1999, 398:422–426
54. Shtutman M, Zhurinsky J, Simcha I, Albanese C, D'Amico M, Pestell

R, Ben-Ze'ev A: The cyclin D1 gene is a target of the beta-catenin/LEF-1 pathway, Proc Natl Acad Sci USA 1999, 96:5522–5527
55. Zhang T, Fields JZ, Ehrlich SM, Boman BM: The chemopreventive agent sulindac attenuates expression of the antiapoptotic protein survivin in colorectal carcinoma cells. J Pharmacol Exp Ther 2004, 308:434–437
56. Zhu HX, Zhang G, Wang YH, Zhou CQ, Bai JF, Xu NZ: [Indomethacin induces apoptosis through inhibition of survivin regulated by beta-catenin/TCF4 in human colorectal cancer cells]. Ai Zheng 2004, 23:737–741
57. Kaiser S, Park YK, Franklin JL, Halberg RB, Yu M, Jessen WJ, Freudenberg J, Chen X, Haigis K, Jegga AG, Kong S, Sakthivel B, Xu H, Reichling T, Azhar M, Boivin GP, Roberts RB, Bissahoyo AC, Gonzales F, Bloom GC, Eschrich S, Carter SL, Aronow JE, Kleimeyer J, Kleimeyer M, Ramaswamy V, Settle SH, Boone B, Levy S, Graff JM, Doetschman T, Groden J, Dove WF, Threadgill DW, Yeatman TJ, Coffey RJ Jr, Aronow BJ: Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. Genome Biol 2007, 8:R131
58. Kelekar A, Chang BS, Harlan JE, Fesik SW, Thompson CB: Bad is a BH3 domain-containing protein that forms an inactivating dimer with Bcl-XL. Mol Cell Biol 1997, 17:7040–7046
59. Yang E, Zha J, Jockel J, Boise LH, Thompson CB, Korsmeyer SJ: Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. Cell 1995, 80:285–291
60. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B, Brattain M, Willson JKV: Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 1995, 268:1336–1338
61. Hata A, Lagna G, Massague J, Hemmati-Brivanlou A: Smad6 inhibits BMP/Smad1 signaling by specifically competing with the Smad4 tumor suppressor. Genes Dev 1998, 12:186–197
62. Park SH: Fine tuning and cross-talking of TGF-beta signal by inhibitory Smads. J Biochem Mol Biol 2005, 38:9–16
63. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP: Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. Nat Genet 2002, 30:227–232