Preinvasive colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid versus nonpolypoid).

Cattaneo et al.

SUPPLEMENTARY METHODS

Analysis of microarray data

Statistical analysis of high-throughput data from microarray studies is commonly regarded as a challenge since the number of variables (i.e., expression intensity values) far exceeds the number of samples. The analytical tools we employed were chosen with two primary aims: 1) dimension reduction, i.e., a process by which the multitude of original variables is transformed into a smaller number of *derived* variables that accurately summarize the original information; and 2) to reveal patterns in the data, common to multiple samples, that could not be found if each variable were analyzed separately. One way to identify sample groups with shared patterns is to plot the sample values in a low-dimensional space, where each dimension represents a new derived variables. This process is also known as *ordination*. The samples are ordered along each axis, and the distance between 2 samples in the low-dimensional space reflects their biological similarity. PCA is an unsupervised method that reduces data dimensionality by identifying directions - called principal components (PC) - along which the variation in the data is maximal (*Ringner M., Nat Biotechnol 2008;26:303-4*; *Baty F. et al., BMC Bioinformatics 2008;9:289*). By using few components, each sample can be represented by relatively few numbers instead of by values for thousands of variables. The first 3 PCs of our 3-dimensional PCA score plot are shown in Figure 1A. The components are extracted so that PC1 explains most of the variation, and PC2, which is uncorrelated with PC1, explains most of the variation that was not accounted for by PC1. Therefore, each PC can be interpreted as the direction that maximizes the variance of the samples when projected onto the

component. The sample scores can then be plotted on a scatter plot whose axes represent the first 3 PCs. The interpretation of this plot is straightforward: the distance between 2 samples is inversely related to the similarity of their variable values (i.e., gene expression profiles), based on the components being a summary of the original variables (conversely for samples further apart).

Correspondence analysis (CoA) also reveals the principal axes of a high-dimensional space, allowing it to be projected into a subspace of low dimensionality (e.g., a plane identified by two axes, like the biplot in Figure 1B). CoA was used to identify samples associated with particular gene expression changes. In typical CoA, a matrix of *n* gene expression levels from *p* samples is treated as a two-way contingency table (genes by samples, or vice versa), with *n* and *p* specifications for the ‘‘factors’’ gene and sample, respectively. Each intensity value thus reflects the abundance of a given transcript in a given sample. The between-group analysis (BGA) in our study was based on CoA.

BGA is a supervised, multivariate analysis method for sample discrimination and class prediction (*Baty F. et al., BMC Bioinformatics 2008;9:289*; *Culhane AC. et al., Bioinformatics 2002;18:1600-8*). Each sample is preassigned to one of a small number of groups, and the axes are those that split/separate these groups as much as possible (see the 1-dimensional plots in Figures 2B and 3A and the biplots Figure 3B and C). The analysis consists in identifying those axes that maximize the between-group variance. To this end, the groups' centroids are regarded as the objects to be analyzed, and they are subjected to CoA. BGA is especially powerful when combined with CoA as it reveals correspondences between the grouped samples and identifies the genes whose altered expression changes are most responsible for the differentiation of these groups. We used BGA based on $\log_2$ ratio expression values to reveal the association between independent classifiers (such as *tissue type*, *histology*, *diameter* of the lesion, or *degree of dysplasia* in our study) and gene expression patterns.

Our third analysis, RDA, can be considered a constrained extension of PCA. It identifies trends in the scatter of gene expression data points that are maximally and linearly related to a second set of independent or explanatory variables, i.e., phenotypic features of the samples that are potentially useful for their classification. In our study, the clinical descriptors were considered as independent variables, and the gene expression as a dependent variable set that can be predicted from the clinical descriptors. The proportion of total variance in the dependent variables that can be explained by a linear combination of the independent variables is termed *redundancy*. RDA was used to determine which of the explanatory variables investigated in the study was capable of segregating the plotted samples and hence to discover gene expression changes associated with an observed phenotype. The result is a PCA model that is constrained to optimize the prediction of the gene expressions in the RDA model from the clinical variables. The relations between the independent and dependent variable sets are visualized by a correlation circle (e.g., those shown in Figure 2A and in Figure 3). In a correlation circle, the positions of the uniformly scaled variables on the surface of a multidimensional sphere are shown as projections on a plane defined by a pair of relevant ordination axes (RDA axes 1 and 2, or 1 and 3, in our figures). The proximity of the variables inside the unit circle reflects affinities and antagonisms between the variables.

KRAS and BRAF mutational analysis

Our analysis focused on the major hot spots for gain-of-function mutations in *KRAS* (exon 2, which includes codons 12 and 13) and *BRAF* (exon 15, which comprises codon 600). This approach allowed us to identify tumors harboring the most common *KRAS* and *BRAF* point mutations. *KRAS* exon 2 and exon 15 of *BRAF* were PCR-amplified from tumor DNA as fragments containing 173 and 103 bps, respectively. PCR products were purified at 37°C for 45 min and then at 80°C for 15 min (ExoSAP-IT® kit, USB Corporation, Cleveland, OH).

Mutations at the relevant nucleotides were then assessed with the ABI PRISM SNaPshot Multiplex Kit (Applied Biosystems, Foster City, CA, USA) using single base-extension SNaPshot primers similarly as in Di Fiore F. *et al*. (Br J Cancer 2007; 96:1166-9) and Sieben NL. *et al*. (J Pathol 2004; 202:336-40). The multiplex SNaPshot reaction was performed in a final volume of 10μl, containing 1 μl of each PCR product, 2 μl of the SNaPshot Multiplex Ready Reaction Mix, 1 μl of 5x sequencing buffer of Big Dye V3.1 Terminator Kit and SNaPshot primers at 0.01-0.06 μM concentration. Cycling conditions were those recommended by the kit manufacturer, and the products were purified by treatment (45 min, 37°C) with 2 U of shrimp alkaline phosphatase (USB Corporation, Cleveland, Ohio). The enzyme was deactivated (heating to 80°C for 15 min), and the fluorescently labeled products were separated with an ABI Prism 3130 DNA sequencer (7 min). The data were analyzed using GeneMapper Analysis Software (Applied Biosystems).

SUPPLEMENTARY FIGURE LEGENDS

Sup_2.jpg (Supplementary Figure 1)**. Between-group analysis of gene expression data for the colorectal tissue samples included in this study after exclusion of the 3 serrated nonpolypoid lesions that were not dysplastic (Table 1)**.
(See legend to Figure 1 for explanation).

Sup_4.jpg (Supplementary Figure 2)**. Venn diagrams illustrating the quantitative difference between the nonpolypoid and polypoid lesions in terms of the number of genes that were differentially expressed compared with the normal mucosa.** Circles represent gene lists generated by pairwise comparison of expression levels in nonpolypoid lesions (*blu*e) or polypoid lesions (*green*) vs normal mucosa. Differentially expressed genes were identified by 2-way ANOVA (patient vs tissue type, including only contrasts for the tissue types, i.e., nonpolypoid vs normal mucosa, polypoid vs normal mucosa, and nonpolypoid vs polypoid), and results are shown for 3 false discovery rate (FDR) levels (step-down method). For example (*top panel*), a total of 1749 genes were significantly dysregulated in preinvasive lesions ($p < 0.001$ vs. normal mucosa), 1003 in nonpolypoid lesions and 1429 in the polypoid group. The difference, as estimated from the FDR, was statistically significant. The 683 genes that were concomitantly dysregulated in both groups represented 68% of those altered in nonpolypoid lesions but only 48% of those that were dysregulated in polypoid lesions. Similar percentages were observed when the *p*-value cutoffs were relaxed (*middle and lower panels*).

Sup_5.jpg (Supplementary Figure 3). **Between-group analysis of gene expression data for proximal-colon vs. distal-colon tissue samples** (See *Results* section and legend to Figure 1 for explanation).

Sup_6.jpg (Supplementary Figure 4). **Beta-catenin expression patterns in polypoid and nonpolypoid lesions.** A large series of nonpolypoid (n=31) and polypoid (n=26) lesions, including most of those analyzed in this study, were incubated for 24 h at 4°C with mouse monoclonal antibody (1:20) against β-catenin (antibody M3539, DakoCytomation, Glostrup, DK). Staining patterns (membrane, cytoplasm, and nucleus) in different regions of the tumors and in the corresponding normal mucosa (evaluated by an experienced pathologist [B.H.]) were similar in the two types of lesions. Examples of increased cytoplasmic and nuclear staining of β-catenin (reflecting aberrantly activated Wnt signalling) are shown for a polypoid lesion (**A** and **B**, respectively) and a nonpolypoid lesion (**C** and **D**, respectively). *N.B.* Increased nuclear localization of β-catenin is associated with a high degree of cellular dysplasia.

Sup_7.jpg (Supplementary Figure 5). ***TMIGD1* mRNA expression in extracolonic tissues.** *TMIGD1* mRNA was present at relatively low levels (compared with those found in the normal colon) in placental, tracheal, renal, and testicular tissues. Total RNAs extracted from the different tissues were purchased from Ambion (Huntington, UK). Their quality was found to be optimal, but information is not available on the organ regions sampled. The coding regions of *TMIGD1* (target) and *GAPDH* (reference) were amplified. *y* axis: [target concentration/reference concentration] for sample / [target concentration/reference concentration] for normal mucosa.

Sup_8.jpg (Supplementary Figure 6). **Exon array expression patterns of 3 transcription factors putatively involved in the control of *TMIGD1* expression.** In silico analysis of the 10,000-bp genomic region upstream from the transcription start site of *TMIGD1* (see Results) revealed putative binding sites for 10 transcription factors (BPTF, EVI1, GATA1, HNF4A, MEIS1, NFE2L1, POU2F1, POU2F2, POU3F2, TLX2), three of which (MEIS1, HNF4A,

NFE2L1) were characterized by significantly downregulated transcription (p values < 0.001 vs. normal mucosal levels) in the 42 precancerous colorectal lesions. y axis: normalized $\log_2$ expression intensity values of mRNA detected in polypoid (green dots), nonpolypoid (blue dots) and in the matched normal mucosa samples (red dots). Box plots show the 25th, 50th (median), and 75th percentiles for each data set. Fold changes (FC) for each pairwise comparison are indicated.

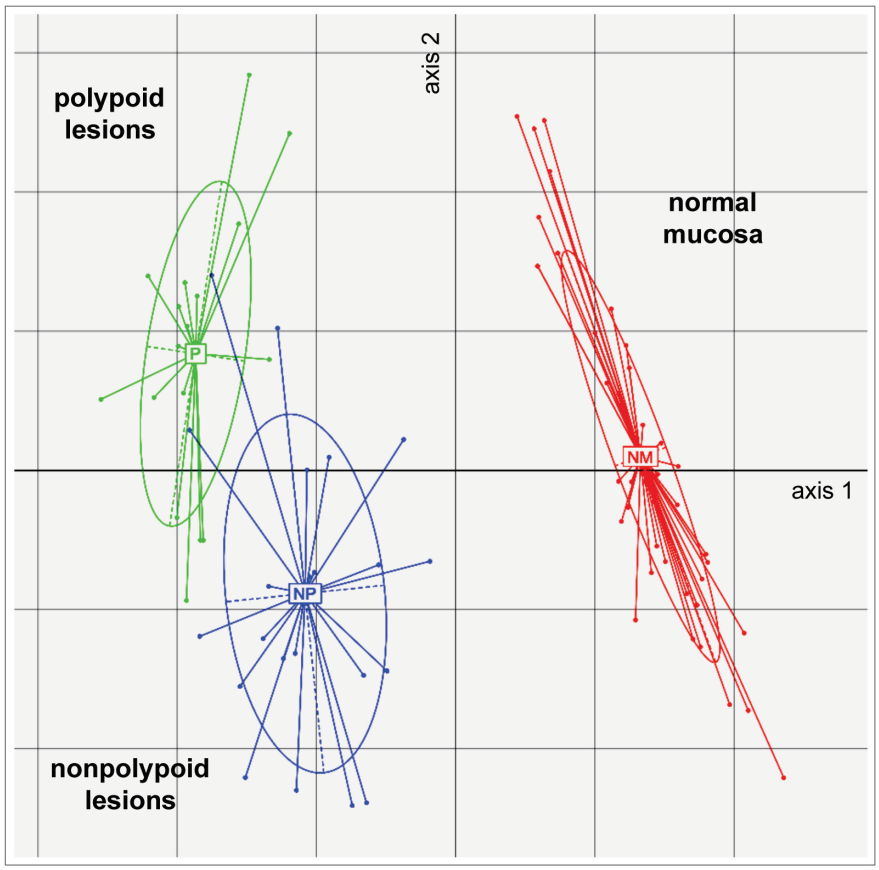
Sup_9.jpg (Supplementary Figure 7). **Immunohistochemical assessment of TMIGD1 expression in the terminal ileum and kidney.** TMIGD1 is expressed in the villi of the terminal ileum (A), particularly at the brush border (B). It is virtually absent in the proliferative compartment of ileal crypts (C). Of the normal extracolonic tissues of a tissue microarray subjected to immunostaining , renal tubules exhibited the highest expression of TMIGDI (D) and labeling was most intense at the brush border. TMIGD1 expression was preserved in the papillary renal carcinoma shown in (E), but it was very low in the clear cell renal carcinomas of panel (F).

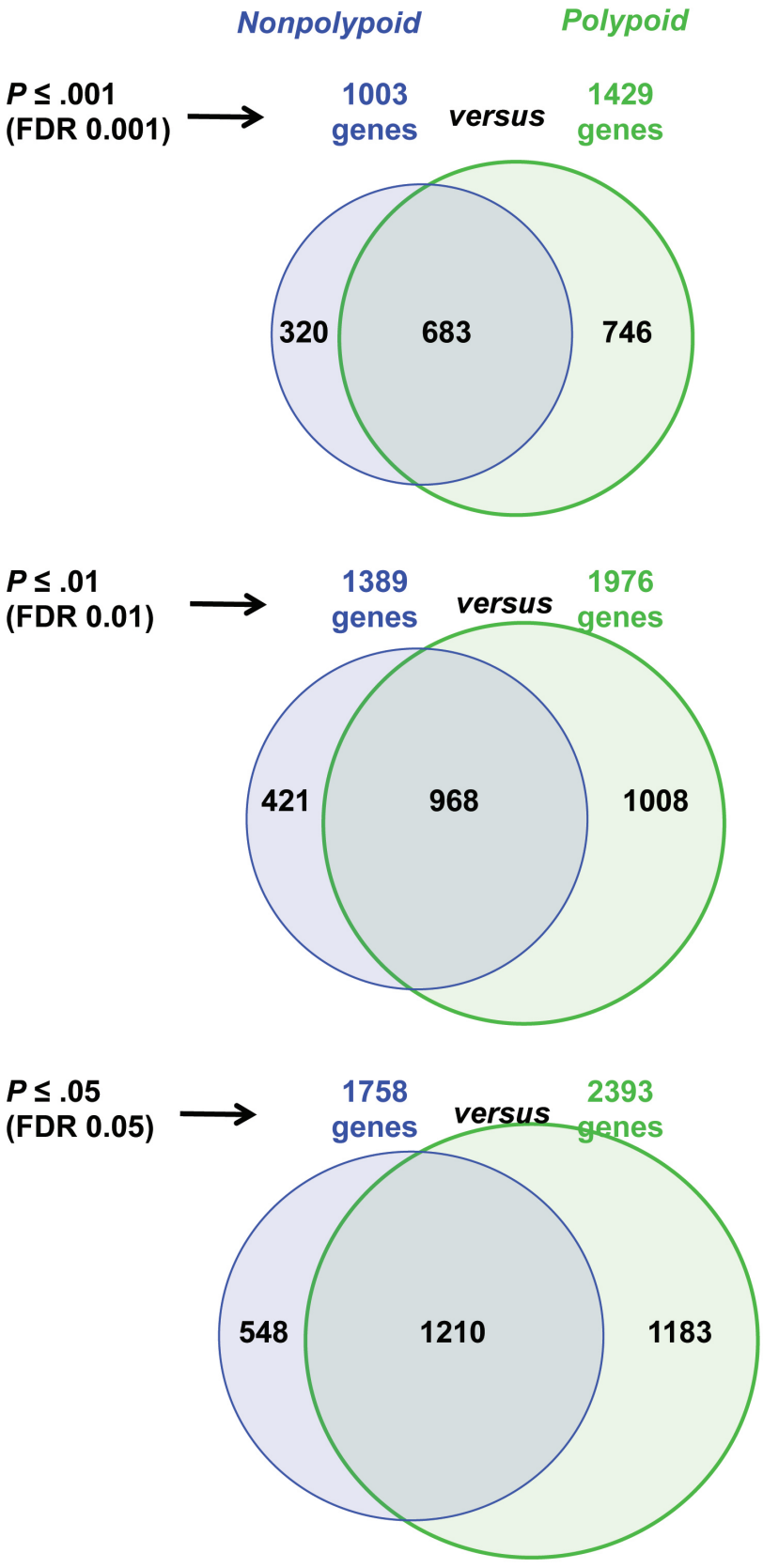
Sup_10.jpg (Supplementary Figure 8). **mRNA expression change trends observed in nonpolypoid and polypoid colorectal lesions.** Normalized $\log_2$ expression intensity values for the 100 genes displaying the most significant alterations with respect to normal mucosal expression in exon array analysis. *Left panel*: upregulations; *right panel*: downregulations. Alterations found in nonpolypoid and polypoid lesions were usually concordant in terms of direction, but the magnitude of the alteration (with respect to normal mucosa) was usually smaller in nonpolypoid lesions.

Sup_11.jpg (Supplementary Figure 9). **GeneGo Metacore map of the pathway *Cell cycle: Start of DNA replication in early S phase.*** DNA replication begins with the ordered assembly of a multiprotein *prereplicative complex*, which licenses DNA for replication during S phase. This complex consists of the origin recognition complex (ORC), the cell division cycle 6 homolog (CDC6), Cdt1, and the minichromosome maintenance protein complex (MCM). Detailed descriptions of the map symbols are available at www.genego.com. Red and blue thermometers indicate up- or downregulation of the gene they flank, respectively; the numbers on the thermometers indicate whether the gene is differentially regulated in nonpolypoid (1) or polypoid (2) lesions. The predominance of red thermometers with the number 2 in this map reflects the upregulation of numerous genes in polypoid lesions.

Sup_12.jpg (Supplementary Figure 10). **GeneGo Metacore map of the *Oxidative Phosphorylation* pathway.** (See legend to Sup_10.jpg for explanation of map symbols.) This map shows the inner mitochondrial membrane where electron transport and oxidative phosphorylation take place. The electron-transport proteins are grouped into 4 complexes. Complex I (NADH:CoQ oxidoreductase) is responsible for transferring electrons from the NADH to CoQ (also known as ubiquinone). Complex II (succinate dehydrogenase), together with complex I, mediates the transfer of a pair of electrons from NADH to oxygen, which generates 3 ATP equivalents. Complex III contains several heme prosthetic groups and accepts electrons from CoQ (therefore, indirectly from both complexes I and II). The electrons received by Complex III are donated, one at a time, to cytochrome c. Cytochrome c oxidase (Complex IV) accepts electrons from cytochrome c. Complex IV carries out the oxidative phosphorylation reaction that converts ADP to ATP. In this map, a large proportion of the genes are flanked by blue thermometers labeled with the number 2, indicating significant downregulation in nonpolypoid lesions.

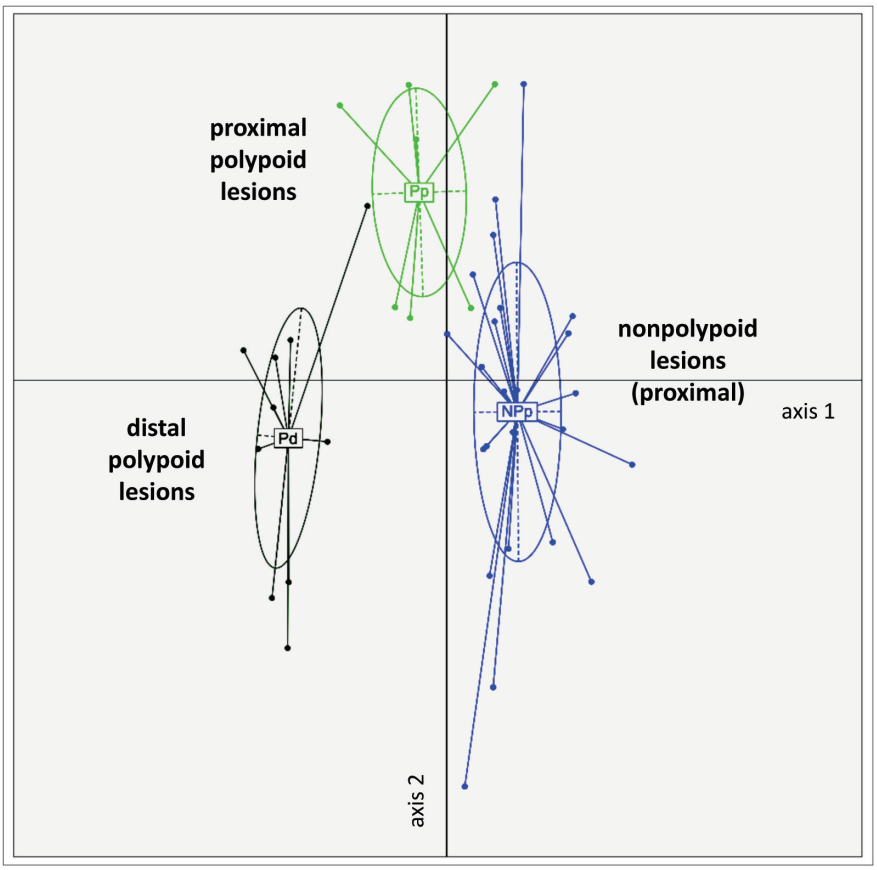
Between Group Analysis



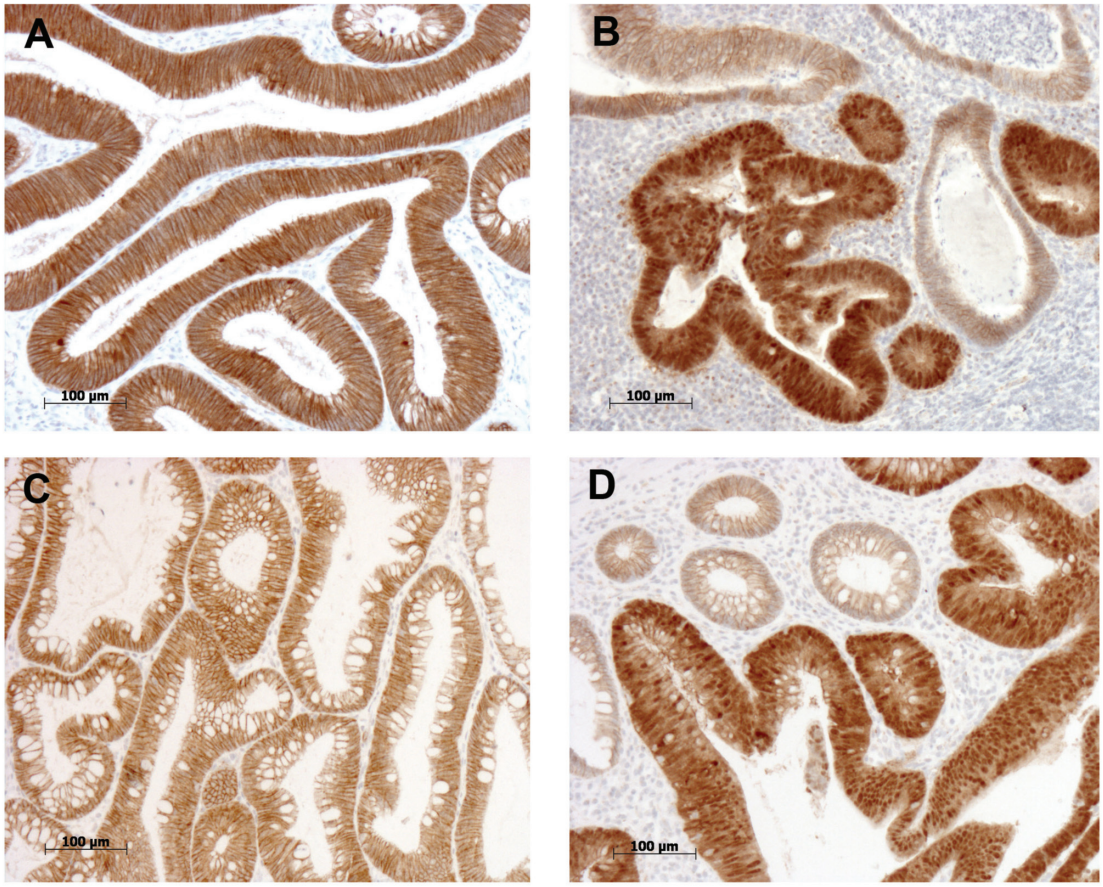

Nonpolypoid
Polypoid
P ≤ .001
(FDR 0.001)
1003 genes
versus
1429 genes
320
683
746
P ≤ .01
(FDR 0.01)
1389 genes
versus
1976 genes
421
968
1008
P ≤ .05
(FDR 0.05)
1758 genes
versus
2393 genes
548
1210
1183

Cattaneo et al., Suppl. Figure 3

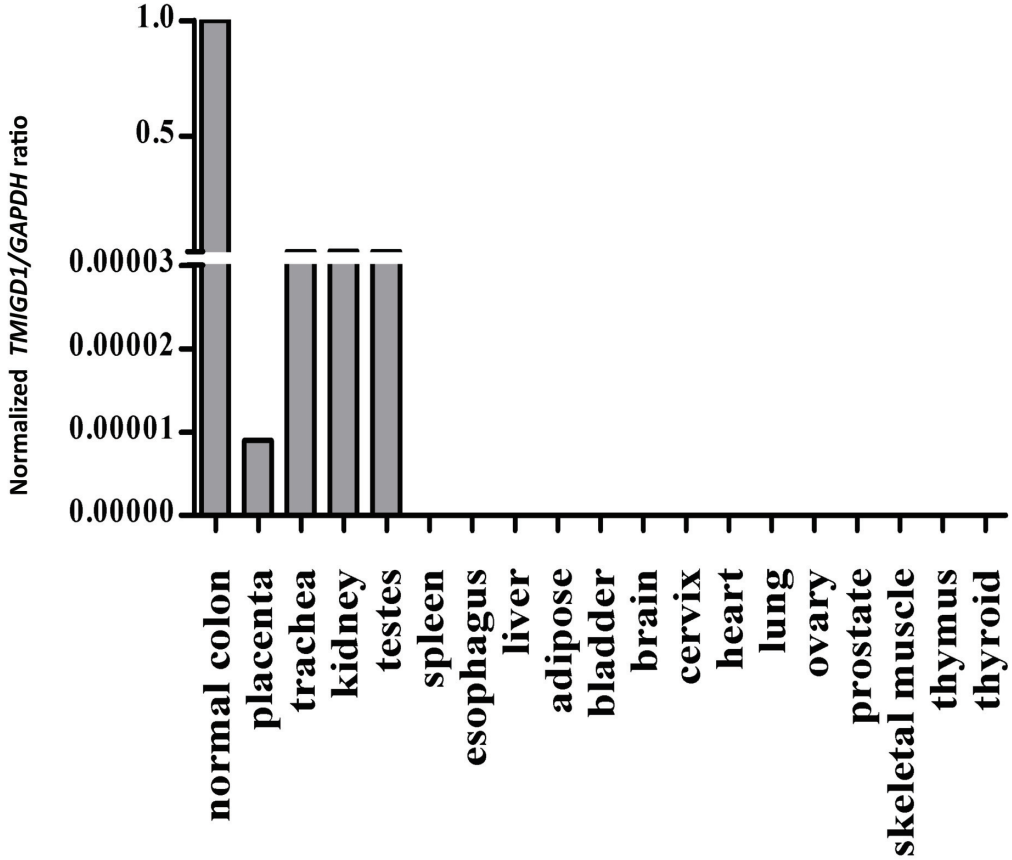
Between Group Analysis



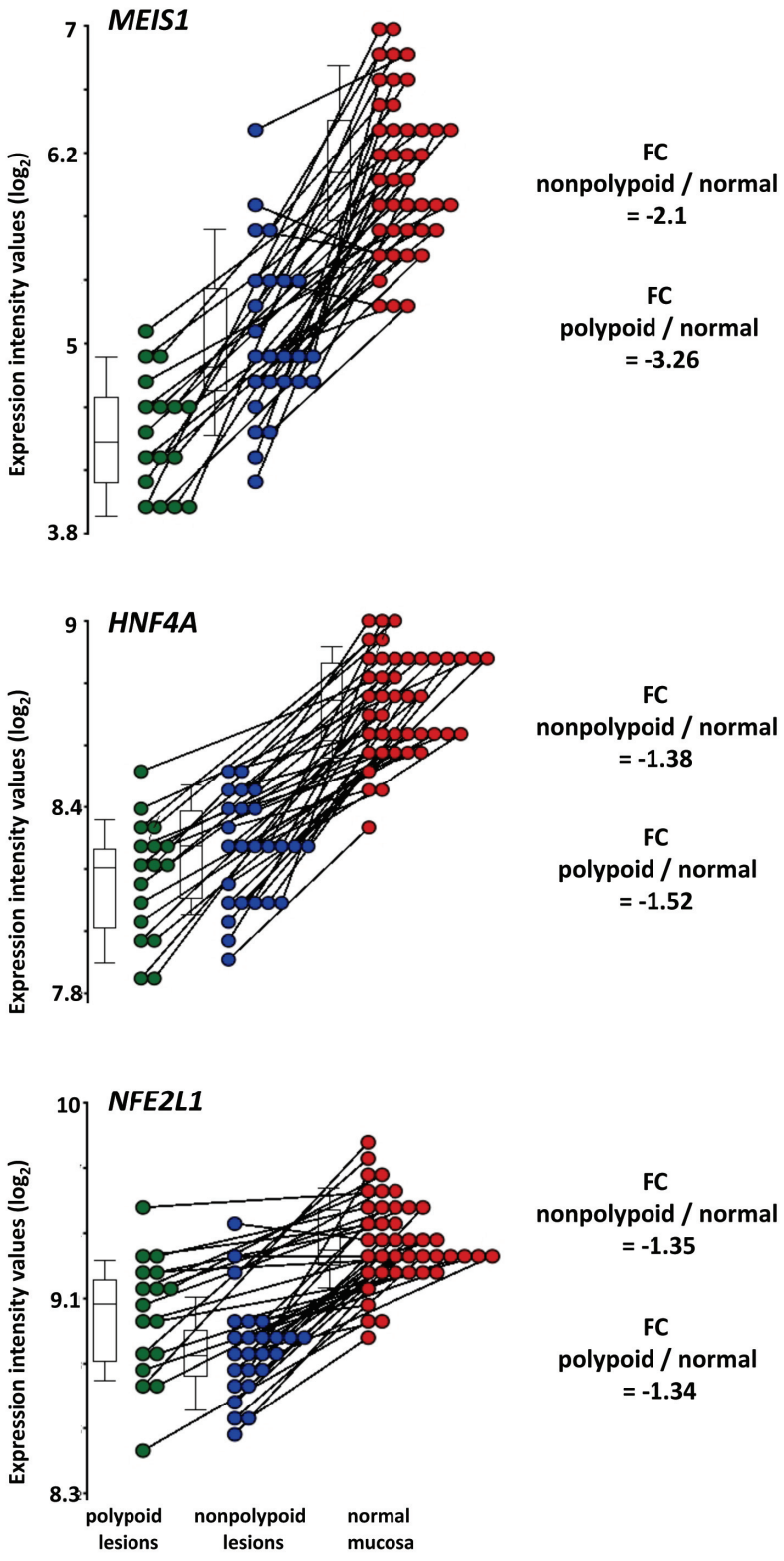
Cattaneo et al., Suppl. Figure 4

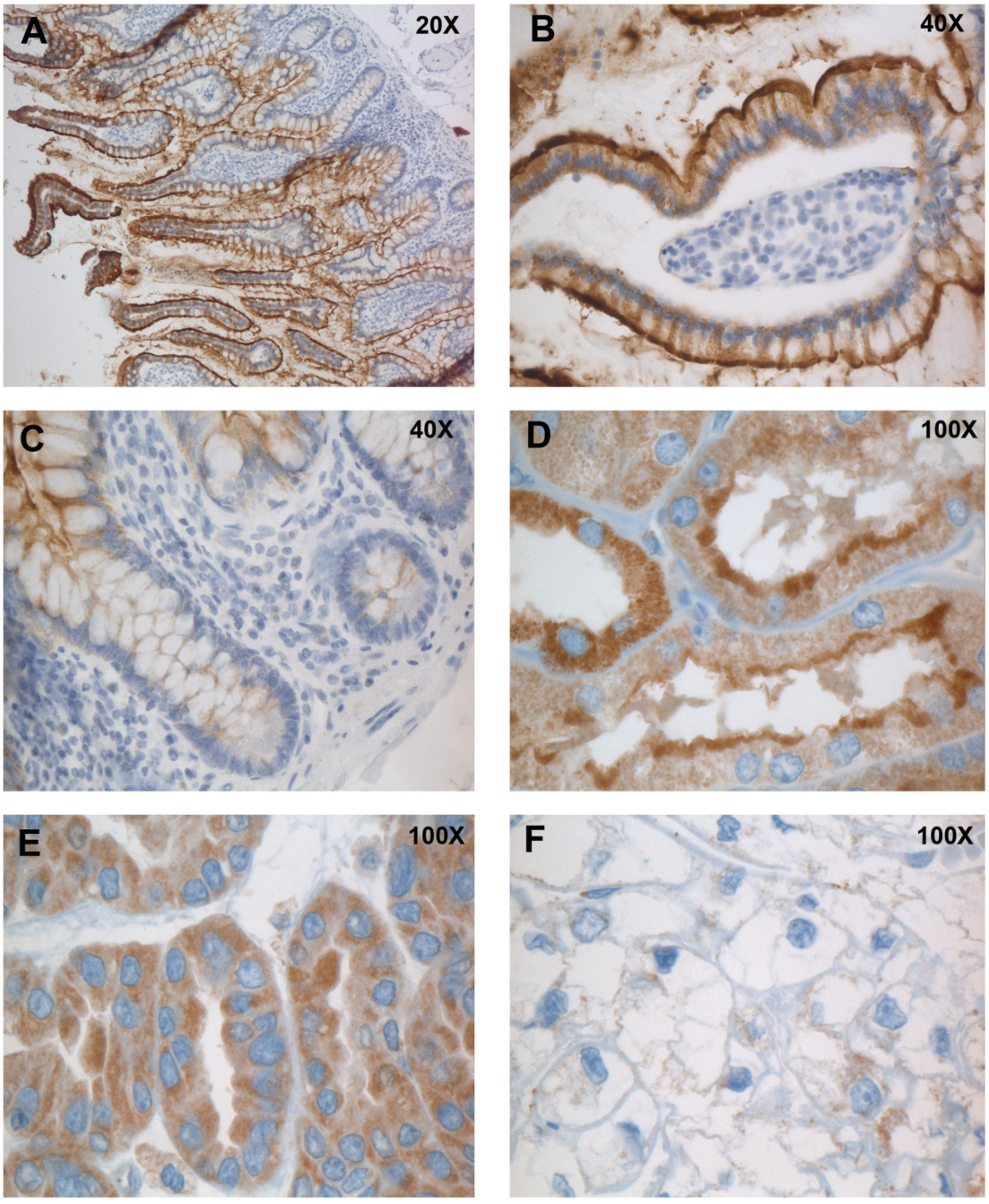


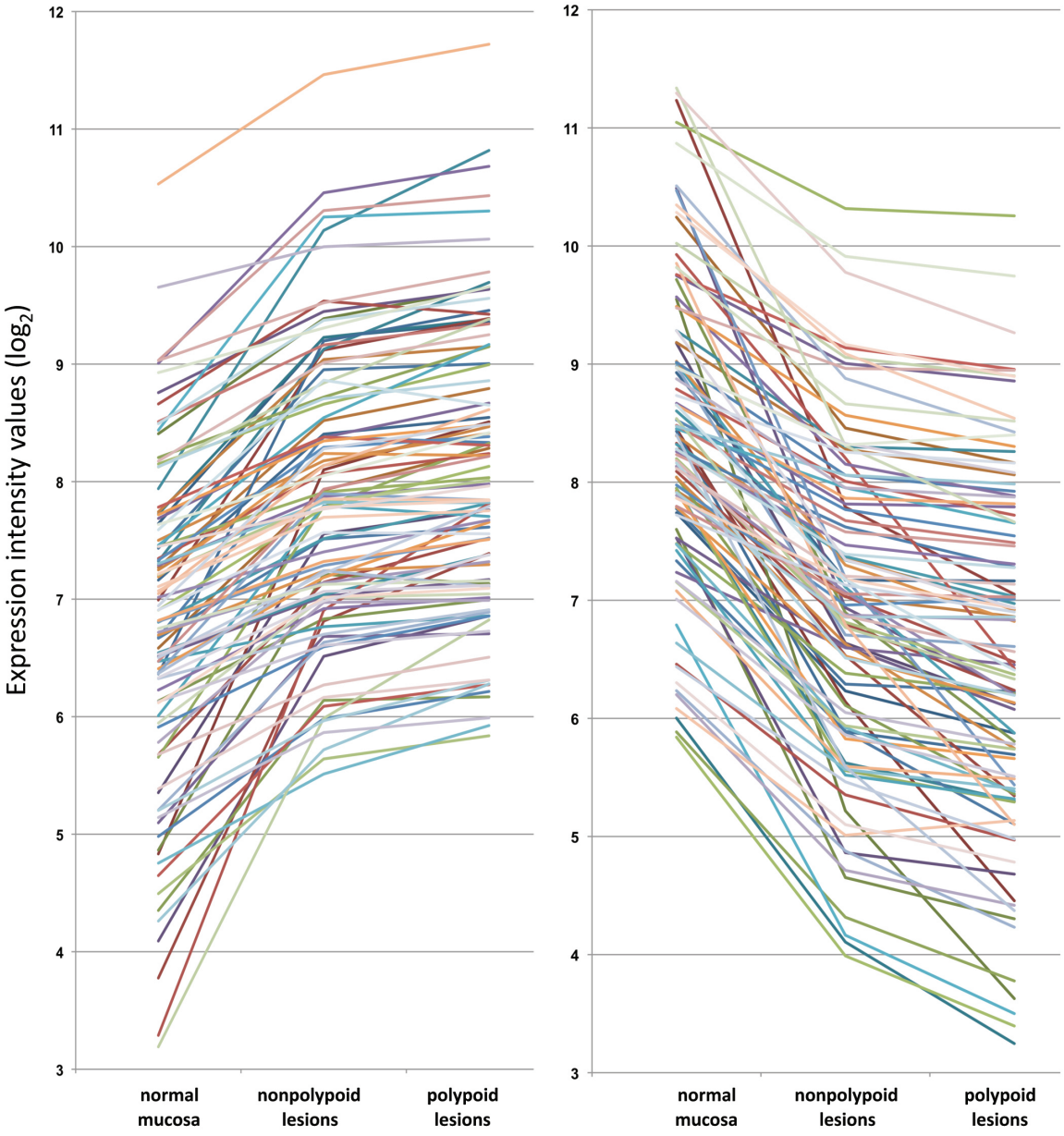
Cattaneo et al., Suppl. Figure 5



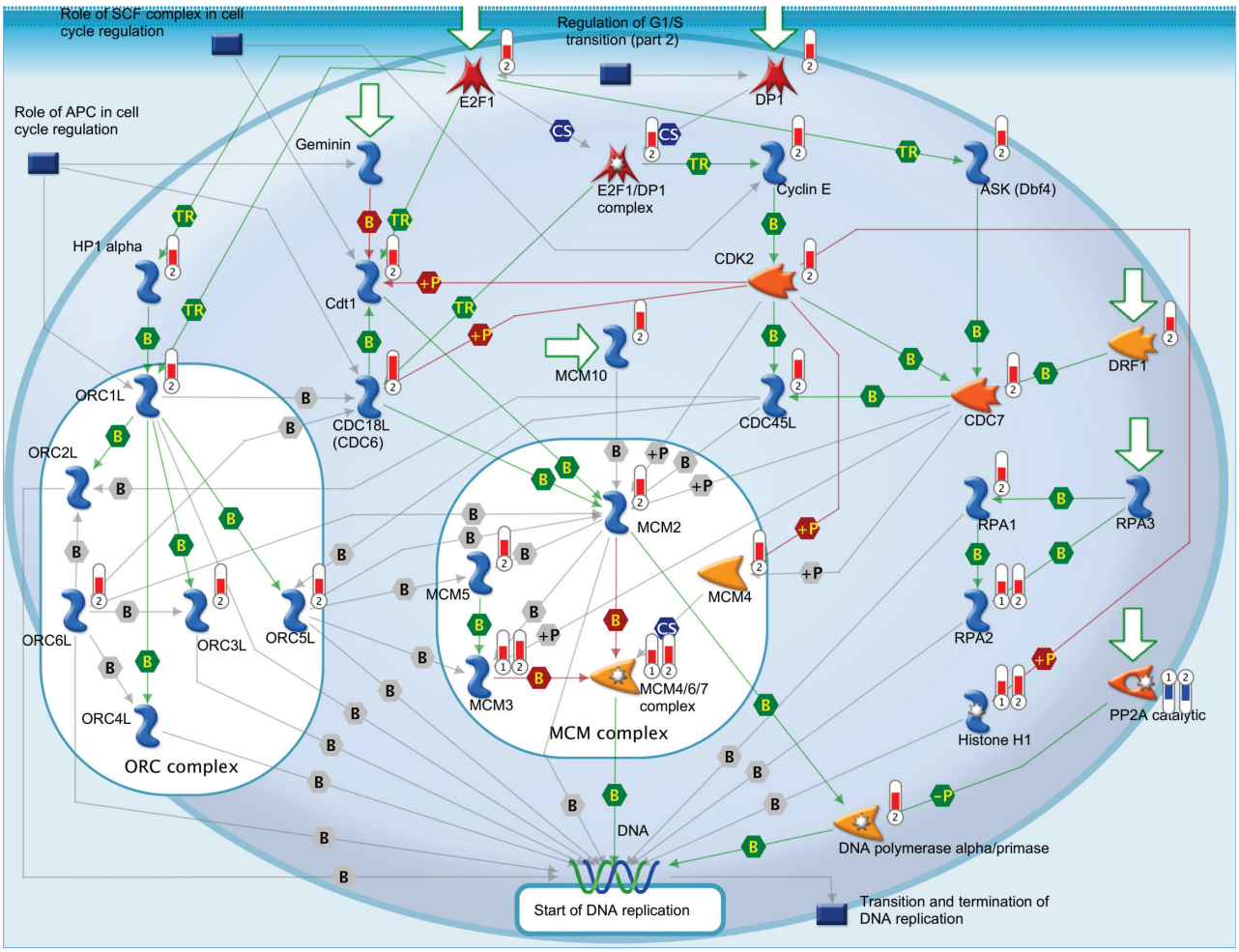
Cattaneo et al. Suppl. Figure 6

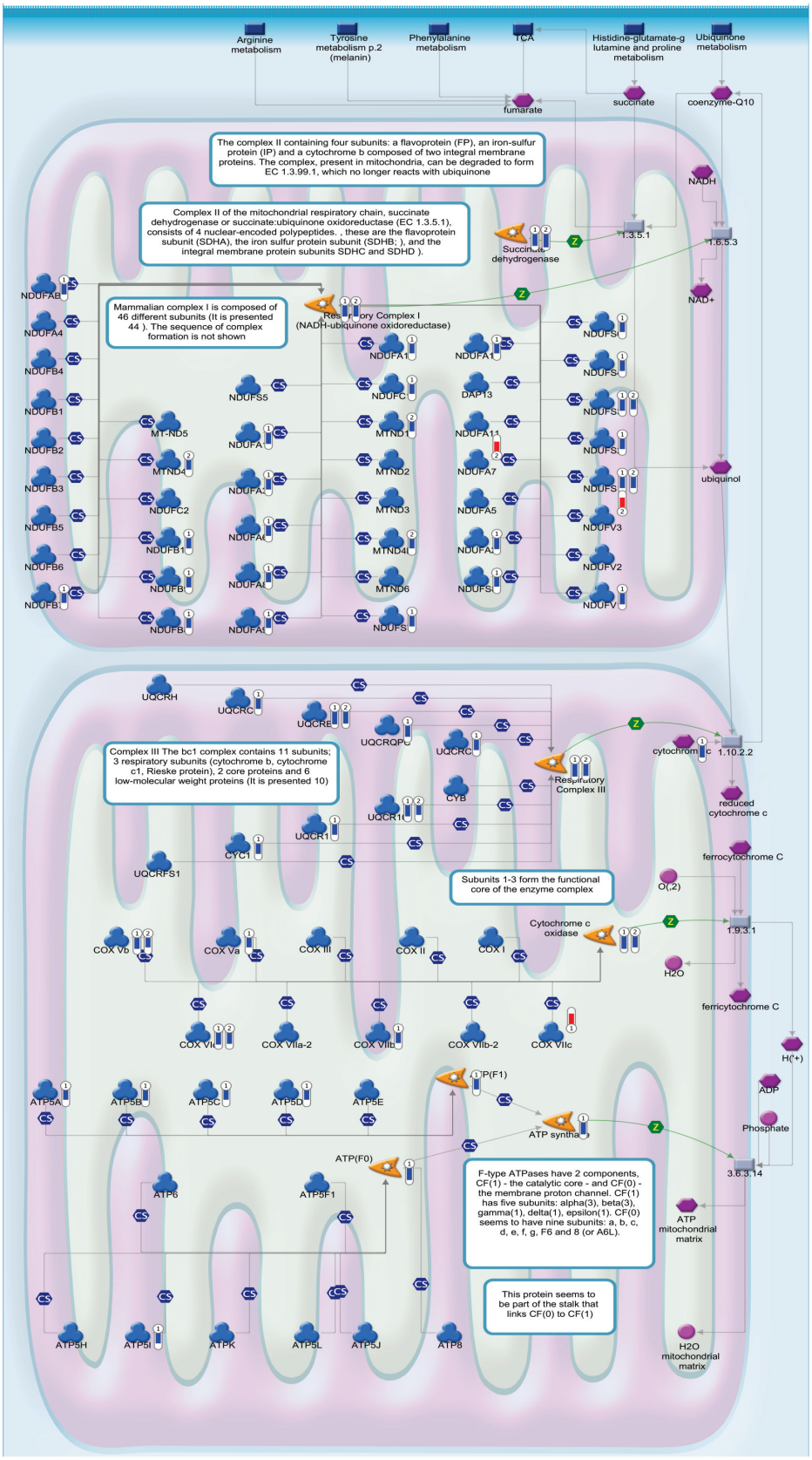



A
20X
B
40X
C
40X
D
100X
E
100X
F
100X


Expression intensity values (log2)
12
11
10
9
8
7
6
5
4
3
normal mucosa
nonpolypoid lesions
polypoid lesions
12
11
10
9
8
7
6
5
4
3
normal mucosa
nonpolypoid lesions
polypoid lesions

Cattaneo et al., Suppl. Figure 9




Arginine metabolism
Tyrosine metabolism p.2 (melanin)
Phenylalanine metabolism
TCA
Histidine-glutamate-glutamine and proline metabolism
Ubiquinone metabolism
fumarate
succinate
coenzyme-Q10
The complex II containing four subunits: a flavoprotein (FP), an iron-sulfur protein (IP) and a cytochrome b composed of two integral membrane proteins. The complex, present in mitochondria, can be degraded to form EC 1.3.99.1, which no longer reacts with ubiquinone
Complex II of the mitochondrial respiratory chain, succinate dehydrogenase or succinate:ubiquinone oxidoreductase (EC 1.3.5.1), consists of 4 nuclear-encoded polypeptides. , these are the flavoprotein subunit (SDHA), the iron sulfur protein subunit (SDHB;), and the integral membrane protein subunits SDHC and SDHD).
NADH
Succinate dehydrogenase
1.3.5.1
1.6.5.3
NAD+
Mammalian complex I is composed of 46 different subunits (It is presented 44). The sequence of complex formation is not shown
Respiratory Complex I (NADH-ubiquinone oxidoreductase)
NDUFAB
NDUFA4
NDUFB4
NDUFB1
NDUFB2
NDUFB3
NDUFB5
NDUFB6
NDUFB7
NDUFS5
MT-ND5
MTND4
NDUFC2
NDUFB1
NDUFB8
NDUFB9
NDUFA1
NDUFC1
MTND1
MTND2
MTND3
MTND4L
MTND6
NDUFS7
DAP13
NDUFA11
NDUFA7
NDUFA5
NDUFA2
NDUFS4
NDUFS6
NDUFS8
NDUFS1
NDUFS2
NDUFS3
NDUFV3
NDUFV2
NDUFV1
ubiquinol
UQCRH
UQCRC1
UQCRE1
UQCRQPC
UQCRC2
Complex III The bc1 complex contains 11 subunits; 3 respiratory subunits (cytochrome b, cytochrome c1, Rieske protein), 2 core proteins and 6 low-molecular weight proteins (It is presented 10)
Respiratory Complex III
CYB
UQCR10
UQCR1
CYC1
UQCRFS1
cytochrome c
1.10.2.2
reduced cytochrome c
ferrocytochrome C
Subunits 1-3 form the functional core of the enzyme complex
O(,2)
Cytochrome c oxidase
1.9.3.1
H2O
ferricytochrome C
COX Vb
COX Va
COX III
COX II
COX I
COX VIc
COX VIIa-2
COX VIIb
COX VIIb-2
COX VIIc
H(+)
ATP5A
ATP5B
ATP5C
ATP5D
ATP5E
P(F1)
ATP synthase
ADP
Phosphate
3.6.3.14
ATP(F0)
ATP6
ATP5F1
F-type ATPases have 2 components, CF(1) - the catalytic core - and CF(0) - the membrane proton channel. CF(1) has five subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0) seems to have nine subunits: a, b, c, d, e, f, g, F6 and 8 (or A6L).
ATP mitochondrial matrix
This protein seems to be part of the stalk that links CF(0) to CF(1)
ATP5H
ATP5I
ATPK
ATP5L
ATP5J
ATP8
H2O mitochondrial matrix

Supplementary Table I. Histological and molecular characteristics of the 25 nonpolypoid lesions [a].

Patient [b]	Microscopic appearance [c]	Dysplasia [d]	MLH1 immunostaining [e]	BRAF status [f]	KRAS status [f]
1a *	VA	high	positive	wt	G12S
1b *	VA	high	positive	wt	G12S
2a *	TA	low	positive	wt	G12V
2b *	TA	low	positive	wt	G13D
3	TA	high	positive	wt	wt
4	SA (mixed)	high	positive	V600E	wt
5	TA	low	positive	wt	wt
6	TA	low	positive	wt	G12V
7	TA	low	positive	wt	G12D
8	TA	low	positive	wt	G12D
9	TVA	low	positive	wt	G12D
10	TA	low	positive	wt	G13D
11	TA	high	positive	wt	wt
12	TVA	high	positive	wt	wt
13a *	TVA	low	positive	wt	G12S
13b *	TA	low	positive	wt	wt
14	TVA	low	positive	wt	G12A
15	SA (mixed)	low	positive	wt	wt
16	SSA	no dysplasia	positive	V600E	wt
17a *	TA	high	positive	wt	wt
17b *	TA	low	positive	wt	wt
18	TA	low	positive	wt	G12A
19a *	VA	low	positive	wt	wt
20	MVSP	no dysplasia	positive	V600E	wt
21	MVSP	no dysplasia	positive	wt	G12D

[a] Analyses performed in formalin fixed, paraffin embedded tissues.

[b] Two lesions were analyzed from patients marked with an asterisk.

[c] *Abbreviations*: TA, tubular adenoma; TVA, tubulovillous adenoma; VA, villous adenoma; MVSP, microvescicular serrated polyp; SA, serrated adenoma; SSA, sessile serrated adenoma.

[d] Highest degree of dysplasia in the lesion based on the WHO classification of tumors of the digestive system (Editorial and consensus conference in Lyon, France, November 6-9, 1999 [IARC]).

[e] Mouse monoclonal antibody BD5510 (BD Pharmingen) used at a 1:300 dilution, 2 hours incubation at 4C).

[f] Mutations at *BRAF* (exon 15) and *KRAS* (exon 2) were analyzed as described in Supplementary Methods. *Abbreviations*: wt, wild-type; V, valine; E, glutamate; G, glycine; S, serine; D, aspartate; A, alanine (e.g., V600E, valine to glutamate substitution at codon 600)

Supplementary Table II. Top 200 genes whose expression displayed significant association with the variable *tissue type* (nonpolypoid vs polypoid lesions). The genes are ranked based on their highest positive or negative score along BGA* axis 1.

Highest positive score along BGA axis 1

No.	Gene Symbol	Affymetrix transcript ID	*BGA score*	FC (NP*vs*NM)^	FC (P*vs*NM)^	FC (NP*vs*P)^	*p*-value #
1	*SLC38A4*	3452417	0.07725326	-2.00346	1.29258	-2.58963	2.10E-05
2	***MMP1***	3388807	0.06265226	1.39065	3.44344	-2.47614	8.80E-07
3	***PSAT1***	3175971	0.05996282	1.64946	4.01335	-2.43314	1.06E-10
4	*OSTalpha*	2659393	0.05902134	-4.1584	-2.22614	-1.86799	3.55E-14
5	***HBB***	3360401	0.05856116	-1.3559	1.58364	-2.14725	6.39E-04
6	***CDC6***	3720896	0.05830322	1.00722	2.24315	-2.22708	3.11E-07
7	***LOC729680***	3504369	0.05648983	-1.18504	1.79221	-2.12384	1.55E-02
8	*OLFM4*	3490892	0.05546722	2.22941	5.25604	-2.3576	3.77E-08
9	***SLC25A33***	2319340	0.05535631	-2.33467	-1.21231	-1.9258	2.13E-06
10	***LOC51581***	4036196	0.05299886	-1.22132	1.65253	-2.01828	2.39E-03
11	***FGFBP1***	2761829	0.05194929	-1.74237	1.09676	-1.91096	7.39E-05
12	***PBK***	3129149	0.05153867	1.15612	2.37889	-2.05765	6.41E-07
13	***LOC729680***	3504367	0.05148803	-1.12896	1.76733	-1.99525	1.25E-04
14	*NETO2*	3690154	0.04867329	-1.29645	1.45952	-1.89219	3.67E-02
15	***MMP3***	3388830	0.04855053	2.79735	6.06084	-2.16664	1.92E-11
16	***C12orf5***	3401786	0.04767323	-1.28506	1.45441	-1.86901	2.28E-02
17	*MEP1B*	3783788	0.04758012	-8.89031	-5.82649	-1.52584	6.52E-14
18	***KCTD16***	2833691	0.047384	1.05489	2.02708	-1.9216	8.07E-08
19	***NAG18***	3843677	0.04717329	-1.11899	1.68313	-1.88341	2.79E-03
20	***DNA2***	3292694	0.04633674	1.13845	2.17277	-1.90853	3.97E-05
21	*CYP2C18*	3258966	0.04621138	-2.99916	-1.78258	-1.68248	7.64E-10
22	***CENPN***	3670663	0.04613195	1.02083	1.92139	-1.88218	1.48E-05
23	***MELK***	3168508	0.04605585	1.06823	2.0177	-1.88883	4.47E-07
24	*ARG2*	3541383	0.04510095	-1.41886	1.2608	-1.78891	7.43E-03
25	***DNAJC8***	2327373	0.0448664	-1.22943	1.47122	-1.80876	8.17E-03
26	***ASF1B***	3852565	0.04467892	1.00466	1.85052	-1.84193	1.52E-07
27	***HGF***	3058944	0.04449618	-1.3685	1.30156	-1.78119	6.40E-03
28	***CDC20***	2333136	0.04385709	1.04497	1.91035	-1.82814	1.34E-04
29	***TUBA3E***	2575949	0.04343352	-1.32085	1.33446	-1.76262	3.05E-01
30	***SPINK4***	3166815	0.04300328	2.60301	5.12492	-1.96884	4.89E-09
31	***CDC2***	3248289	0.04291239	1.16621	2.12618	-1.82316	1.04E-06
32	***GINS1***	3880827	0.04290054	1.04399	1.88336	-1.804	1.20E-09
33	***PNMA2***	3128731	0.04244226	1.03423	1.85239	-1.79107	2.28E-04
34	***DTL***	2378937	0.04237348	1.05053	1.88252	-1.79197	3.84E-07
35	***RRM2***	2469252	0.0423537	1.39155	2.55879	-1.8388	1.34E-07
36	***STEAP1***	3011838	0.04232019	-1.36861	1.2651	-1.73142	1.08E-02
37	***RPSA***	3104758	0.04226581	-1.27459	1.36643	-1.74163	4.73E-02
38	***FANCI***	3607510	0.04203269	1.15204	2.07226	-1.79878	4.54E-09
39	*AKR1B10*	3249533	0.04196512	-3.07458	-1.92168	-1.59994	6.00E-12
40	***KIAA0101***	3629103	0.04155367	1.30136	2.35116	-1.80669	5.74E-07
41	*NP*	3527514	0.04122344	-1.19995	1.43949	-1.72732	3.05E-04
42	***SLC7A5***	3703885	0.04112539	1.46557	2.66004	-1.81502	2.36E-10
43	***MMP10***	3388785	0.04101146	1.64556	3.01298	-1.83097	4.29E-08
44	***L1TD1***	2339334	0.04043087	2.26442	4.22859	-1.86741	1.31E-06
45	***RFC3***	3485074	0.04039811	1.13956	2.00161	-1.75648	8.07E-10
46	*HMOX1*	3944129	0.0400356	-2.06767	-1.27543	-1.62116	9.84E-10
47	*B4GALNT2*	3725572	0.03982181	-2.69086	-1.70278	-1.58027	5.46E-10
48	***SLC39A2***	3527785	0.03939606	-1.07685	1.58039	-1.70185	7.15E-04
49	***FAM111B***	3331903	0.03928544	1.28604	2.24745	-1.74758	2.99E-08
50	***MAD2L1***	2783715	0.03922022	1.13961	1.96917	-1.72794	2.46E-05
51	*HSD3B2*	2354433	0.03878498	-5.95578	-4.08007	-1.45973	1.39E-12

52	***TUBG1***	3721926	0.03872455	1.01671	1.72796	-1.69956	2.86E-08
53	*IGFBP2*	2527253	0.03852405	1.27291	2.19876	-1.72735	2.30E-09
54	***KRT14***	3747399	0.03839115	-1.04038	1.61853	-1.68388	1.54E-01
55	***FIBIN***	3324447	0.03819729	-1.14795	1.45103	-1.66571	1.89E-03
56	***FLJ38028***	3797543	0.03808956	1.05556	1.78409	-1.69019	4.99E-07
57	***EXO1***	2388219	0.03795459	1.08548	1.83554	-1.69099	9.15E-08
58	***PHGDH***	2354634	0.03772093	1.06587	1.79388	-1.68303	3.49E-05
59	***NQO1***	3696666	0.03741638	1.46334	2.5171	-1.72011	2.02E-11
60	***TERF1***	3800245	0.03732576	-1.26616	1.29005	-1.63341	1.75E-02
61	***FLJ25694***	3516443	0.03722758	-1.16407	1.41103	-1.64254	2.78E-02
62	***ARHGAP11A***	3587457	0.03714087	1.14079	1.91532	-1.67895	3.10E-05
63	***CYP1B1***	2548699	0.03704653	-1.09296	1.50692	-1.64701	9.41E-03
64	***GINS2***	3703112	0.03690436	-1.03772	1.59078	-1.65079	1.79E-05
65	***CDC25A***	2673085	0.03689021	-1.08767	1.51169	-1.64421	8.20E-08
66	***TK1***	3772158	0.0368502	1.00481	1.66312	-1.65516	5.36E-08
67	***SPAG11B***	3122984	0.03682759	-1.03289	1.59718	-1.6497	3.80E-02
68	***RRM1***	3318052	0.03668197	1.01701	1.68111	-1.65299	1.16E-02
69	*SLC20A1*	2500919	0.03666186	-3.8984	-2.63495	-1.47949	4.00E-14
70	***GLB1L2***	3357397	0.0366095	-1.01318	1.62598	-1.6474	5.78E-12
71	***CDCA5***	3377423	0.03645837	1.08754	1.80186	-1.65682	3.73E-10
72	***CDC45L***	3936913	0.03631068	1.10364	1.82696	-1.6554	5.70E-11
73	**"na"**	3023803	0.03627047	-1.48209	1.07352	-1.59105	1.38E-02
74	***SHCBP1***	3689880	0.03621993	1.17049	1.94427	-1.66107	1.94E-11
75	***ODC1***	2540157	0.03597429	1.19816	1.9871	-1.65846	2.86E-07
76	***ANKRD22***	3299469	0.03578192	2.02768	3.49506	-1.72367	2.71E-17
77	***CTSL2***	3216671	0.03572055	1.19793	1.97968	-1.65259	2.27E-09
78	***ORM1***	3186123	0.03563511	1.0059	1.63787	-1.62826	8.39E-03
79	*OASL*	3474831	0.03561252	-2.2169	-1.44987	-1.52903	5.70E-10
80	***RAD51***	3590086	0.03545095	1.00124	1.62563	-1.62362	2.39E-09
81	***CKS2***	3178583	0.03533152	1.36334	2.26343	-1.66021	5.78E-12
82	***ANLN***	2997376	0.03532371	1.50767	2.52234	-1.673	4.20E-10
83	***HIST1H1B***	2947073	0.03531689	1.10932	1.81217	-1.63359	1.41E-03
84	**"na"**	3477610	0.03525377	-1.22256	1.30398	-1.5942	1.97E-01
85	***OBP2B***	3228545	0.03511535	-1.17015	1.36452	-1.59669	1.16E-02
86	***XRCC2***	3080280	0.03510734	1.03868	1.68336	-1.62067	2.15E-04
87	***SLC27A2***	3593575	0.03496043	-1.3531	1.16459	-1.5758	6.77E-03
88	***BRCA1***	3758317	0.03490084	1.09496	1.77675	-1.62266	6.02E-09
89	***DEFA6***	3122703	0.0348991	2.49917	4.31938	-1.72833	8.76E-05
90	***MYBL2***	3886223	0.03484146	1.16167	1.89198	-1.62867	2.86E-09
91	***CCNE2***	3145107	0.03482165	-1.14986	1.38508	-1.59265	6.29E-05
92	***LOC100128010***	3761608	0.03469886	-1.00986	1.59016	-1.60585	5.06E-02
93	*ADFP*	3200648	0.03462388	-1.12032	1.4207	-1.59164	5.53E-04
94	***FMO2***	2367015	0.03453234	-1.3109	1.19835	-1.57092	7.05E-02
95	***GPT2***	3658980	0.03447324	1.27399	2.07883	-1.63174	4.26E-12
96	***SNRPG***	3808678	0.03442209	-1.27906	1.22871	-1.57159	4.73E-02
97	***TREM1***	2953570	0.03435962	1.02695	1.64611	-1.60291	1.13E-04
98	***HIST1H2AL***	2900091	0.03434644	1.14609	1.85199	-1.61592	5.71E-07
99	***TMPRSS11E***	2729821	0.03433158	-1.25741	1.25001	-1.57177	8.15E-02
100	***TPX2***	3881443	0.03421436	1.29782	2.11297	-1.62809	2.49E-07

Highest negative score along BGA axis 1

N°	Gene Symbol	Affymetrix transcript ID	*BGA score*	FC (NP*vs*NM)	FC (P*vs*NM)	FC (NP*vs*P)	*p*-value
1	***CLCA4***	2345061	-0.22341567	-8.41057	-40.453	4.80979	1.50E-19
2	***TMIGD1***	3751859	-0.22194127	-18.959	-66.3636	3.50038	1.76E-25
3	***GCG***	2584113	-0.18146967	-5.92818	-24.62	4.15305	9.56E-18
4	***SLC26A3***	3067250	-0.17154663	-5.16429	-20.8777	4.04271	3.77E-17
5	***BMP3***	2733483	-0.1481741	-4.81337	-16.7064	1.52489	4.00E-26

6	***SI***	2703750	-0.14534472	-1.34463	-6.38041	4.7451	5.77E-08
7	***MS4A12***	3332424	-0.13771629	-8.90193	-24.4702	2.74887	1.74E-18
8	***ADAMDEC1***	3090294	-0.1321929	-3.33561	-11.1878	3.35406	1.74E-20
9	***CA1***	3142967	-0.12508547	-6.60221	-17.8465	2.70311	2.39E-18
10	***TPH1***	3365203	-0.11816102	-4.95187	-13.541	2.73452	1.26E-19
11	*CLDN8*	3928415	-0.11634628	-2.50043	-7.80063	3.11972	6.18E-10
12	*INSL5*	2417008	-0.11612724	-1.17591	-4.30081	3.65745	5.90E-06
13	*B3GNT7*	2531908	-0.11538896	-1.68415	-5.66427	3.36329	8.71E-13
14	*TRPM6*	3210013	-0.11345669	-2.79135	-8.29021	2.96996	1.46E-15
15	***PCK1***	3890640	-0.1104965	-3.09548	-8.76711	2.83223	9.13E-16
16	***CDKN2BAS***	3165025	-0.10690454	-3.15355	-8.62233	2.73417	3.75E-24
17	***FAM23A***	3237282	-0.10345595	-3.69581	-7.78262	2.75162	5.86E-15
18	***GUCA2A***	2408832	-0.10342575	-9.85327	-21.0042	2.1317	1.37E-23
19	***CYP3A4***	3063501	-0.09704554	1.26603	-2.54109	3.2171	3.96E-03
20	***SLC4A4***	2730746	-0.09704165	-3.67611	-8.95996	2.43735	1.52E-20
21	***CCL19***	3204285	-0.09662263	-1.60358	-4.52058	2.81906	1.63E-07
22	***CXCL13***	2732508	-0.09658658	-1.81405	-5.00003	2.75628	1.81E-08
23	***CHGA***	3549092	-0.09278322	-3.31257	-7.9226	2.39168	3.49E-21
24	*SST*	2709750	-0.09212062	-3.12595	-7.50679	2.40144	1.47E-19
25	***ADH1B***	2779231	-0.08977269	-1.77628	-4.59149	2.58489	1.47E-12
26	*CCL21*	3204301	-0.08964529	-1.75467	-4.53913	2.58689	2.20E-09
27	***SEPP1***	2855285	-0.08929915	-1.72855	-4.46719	2.58436	6.17E-14
28	***SLC17A4***	2898986	-0.08872451	-3.46903	-7.94106	2.28913	7.74E-22
29	*C21orf88*	3932397	-0.08751069	-1.30943	-3.47801	2.65612	1.81E-07
30	***ITLN1***	2440413	-0.08697576	1.09095	-2.56554	2.79888	8.52E-04
31	***SEMA3D***	3059667	-0.08646256	-1.63608	-4.14544	2.53376	4.26E-12
32	***SLC26A2***	2835300	-0.08385746	-3.94644	-8.47779	2.14821	1.84E-15
33	*B3GALT5*	3921490	-0.08366052	-1.66679	-4.09176	2.45487	3.90E-09
34	***MT1M***	3662150	-0.08348001	-2.99125	-6.68863	2.23606	6.51E-10
35	*PYY*	3758775	-0.08331981	-2.76827	-6.25616	2.25995	1.46E-18
36	***CEACAM7***	3863214	-0.08272676	-4.00819	-8.50563	2.12206	2.12E-13
37	***GPR174***	3982612	-0.08107902	-1.29605	-3.21942	2.48402	3.33E-09
38	***FAM23A***	3237280	-0.0804329	-3.69581	-7.78262	2.10579	5.86E-15
39	***CA2***	3105600	-0.08036145	-4.34241	-8.9188	2.05388	3.97E-16
40	***ANO5***	3323748	-0.07978082	-1.66409	-3.92653	2.35957	2.90E-11
41	*CMAH*	2945882	-0.07787393	-1.4437	-3.41036	2.36223	7.60E-15
42	***LOC727820***	2432336	-0.0740117	1.69682	-1.514	2.56898	1.83E-02
43	***FN1***	2598261	-0.07201708	-1.23156	-2.79316	2.26798	4.10E-06
44	***CP***	2700244	-0.07179646	-2.19528	-4.59291	2.09217	1.91E-17
45	*GLDN*	3593931	-0.06946346	-1.01232	-2.29048	2.26261	1.03E-04
46	***RARRES3***	3333899	-0.06911894	-1.27434	-2.78705	2.18705	1.15E-10
47	***LGALS2***	3960174	-0.06870247	-1.71903	-3.59992	2.09415	7.28E-09
48	***LIFR***	2854088	-0.06799249	-1.95536	-3.99792	2.04459	9.83E-15
49	***SLC15A1***	3522327	-0.06776007	-1.8654	-3.82796	2.05209	8.23E-08
50	***MT1H***	3662201	-0.06693898	-2.27855	-4.52056	1.98396	2.15E-07
51	*SCNN1B*	3652902	-0.06672708	-2.61999	-5.09646	1.94522	3.38E-18
52	***DPP4***	2584018	-0.06557954	-1.61015	-3.29026	2.04345	6.09E-08
53	***UGT2B17***	2772088	-0.06536633	-2.21135	-4.33515	1.96041	9.12E-09
54	***RGS2***	2372858	-0.06533515	-1.18328	-2.50529	2.11725	1.76E-06
55	***C3***	3848039	-0.0647728	1.31087	-1.69373	2.22027	2.15E-03
56	***CR2***	2377283	-0.06470257	-1.47484	-3.01813	2.04641	2.12E-07
57	***CD69***	3443868	-0.06467651	-1.23901	-2.58943	2.08993	8.87E-07
58	***F13A1***	2940202	-0.06456554	-2.61462	-4.98177	1.90535	2.07E-11
59	*PLA2G10*	3681488	-0.0643777	-1.22588	-2.55683	2.08571	2.67E-06
60	***SPON1***	3321361	-0.06436793	-1.64226	-3.30508	2.01252	6.61E-10
61	***NR5A2***	2374126	-0.06432769	-2.03239	-3.98378	1.96014	2.01E-19
62	***ADH1C***	2779271	-0.06390431	-1.9574	-3.83767	1.9606	6.82E-09
63	***CFH***	2373336	-0.06388455	-1.57428	-3.16825	2.01251	3.23E-11

64	*HOXB13*	3761538	-0.06345392	3.42979	1.39817	2.45306	4.80E-05
65	***PDZK1***	2356425	-0.0628472	-1.05828	-2.20829	2.08668	1.66E-04
66	***SLC28A2***	3592304	-0.06282745	2.02987	-1.12562	2.28485	1.19E-03
67	***GOLGA8B***	3617574	-0.06178931	1.40153	-1.54043	2.15896	1.76E-03
68	***RGS1***	2372781	-0.0607034	-1.15637	-2.33067	2.0155	1.06E-03
69	*CPNE8*	3450655	-0.06018552	-2.39744	-4.42049	1.84384	1.04E-18
70	*TMEM200A*	2925590	-0.05972803	-1.0019	-2.02989	2.02604	4.40E-07
71	***MS4A8B***	3332465	-0.0597143	-1.74274	-3.31589	1.90268	1.14E-11
72	***EDIL3***	2865390	-0.05934968	-2.17733	-4.02477	1.84849	5.88E-17
73	***RNF152***	3811000	-0.05920401	-2.68905	-4.84713	1.80254	2.36E-19
74	***PLA2G2A***	2400027	-0.05854392	1.34797	-1.53285	2.06624	1.74E-01
75	***VNN1***	2974592	-0.05853718	1.84808	-1.15781	2.13972	3.11E-03
76	***OSTbeta***	3598267	-0.05826325	-7.95725	-12.6067	1.5843	9.49E-22
77	***GIMAP7***	3031517	-0.0581458	-1.31369	-2.53632	1.93068	4.47E-09
78	***PAG1***	3142217	-0.05802421	-2.25257	-4.09289	1.81699	2.68E-21
79	***PTPRC***	2373842	-0.05800879	-1.29609	-2.50225	1.93061	1.01E-07
80	***ZG16***	3655621	-0.05774016	-8.14808	-12.8235	1.5738	2.57E-18
81	***ANKRD36B***	2565935	-0.05770494	1.18596	-1.70057	2.0168	2.25E-01
82	***PRKAA2***	2337740	-0.05743557	-1.70153	-3.16873	1.86228	6.43E-08
83	***IL7R***	2806468	-0.05743141	1.11314	-1.79349	1.9964	1.18E-03
84	***CD3G***	3351300	-0.05730689	-1.23559	-2.37921	1.92557	9.70E-11
85	***GREM1***	3587553	-0.05704747	-1.18474	-2.28504	1.92872	3.28E-04
86	***CA4***	3729419	-0.05696684	-5.5067	-8.9876	1.63212	2.89E-24
87	***EFEMP1***	2554018	-0.05664407	-1.20763	-2.31382	1.916	8.90E-10
88	***SAMD9***	3061438	-0.05662587	-1.26443	-2.41021	1.90617	3.92E-08
89	***CD96***	2635741	-0.05658624	-1.22715	-2.34565	1.91145	2.16E-12
90	***HEPACAM2***	3061484	-0.05647412	-1.12969	-2.17584	1.92605	7.02E-03
91	***RHOQ***	2480585	-0.05630797	-1.11665	-2.14931	1.92479	3.71E-04
92	***NEUROD1***	2590491	-0.05601349	-2.67774	-4.68067	1.74799	1.20E-15
93	***EPB41L4A***	2870964	-0.05588518	-1.23019	-2.33236	1.89593	3.17E-09
94	***SMOC2***	2937144	-0.05574625	1.24947	-1.58552	1.98106	0.0194252
95	***ABCG2***	2777276	-0.0556232	-7.08986	-11.1396	1.5712	2.10E-23
96	***ZHX2***	3114022	-0.05552447	1.16716	-1.68069	1.96162	2.84E-03
97	***CD3E***	3351280	-0.05497976	-1.11965	-2.12199	1.89522	1.37E-09
98	*CPB1*	2647073	-0.05483416	-1.03021	-1.9662	1.90854	2.81E-04
99	***DOCK10***	2601648	-0.05481578	-1.41883	-2.61861	1.84562	1.93E-13
100	***DUSP1***	2887309	-0.05466722	-1.70319	-3.07935	1.80799	4.43E-09

***** BGA, between-group analysis of $\log_2$ ratio expression intensity values with the variable *tissue type* (nonpolypoid vs polypoid) as grouping factor. The first 10 genes in each panel are also reported in Figure 2B. Abbreviations: FC, fold change; NP, nonpolypoid; P, polypoid; NM, normal mucosa. "na": gene symbol not assigned.

^ Fold changes based on mean log2 intensity values are given for NP vs NM, P vs NM, and NP vs P.

p values were determined by 2-way ANOVA for the factors patient x tissue type (*p* value are given here for the factor tissue type)

Expression profiles of the genes shown in boldface type were influenced exclusively by *tissue type*; those of the other genes were also related to the *location* of the lesion (proximal vs. distal colon), as demonstrated by 2-way ANOVA assuming tissue type and location as interacting factors.