

Supplementary Tables

Supplementary Table 1. Number of differentially expressed, upregulated, and downregulated genes in UC and CRC datasets

Ulcerative colitis			
Accession ID	DEG (adj. $p \leq 0.05$ ve $ \log FC \geq 1$)	Up-Regulated mRNAs	Down-Regulated mRNAs
GSE83687	6633	3391	3242
GSE92415	3853	2224	1629
GSE87473	3707	2141	1566

Colorectal cancer			
Accession ID	DEG (adj. $p \leq 0.05$ ve $ \log FC \geq 1$)	Up-Regulated mRNAs	Down-Regulated mRNAs
GSE110224	2216	1171	1045
GSE22598	5622	3053	2569
GSE21815	7010	3794	3216

Supplementary Table 2. Top ten up-regulated and down-regulated genes in CRC across datasets

GSE110224			GSE21815			GSE22598		
Up-regulated			Up-regulated			Up-regulated		
Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC
MMP3	0.002	4.12	KRT23	0.0000	6.87	FOXQ1	0.000	5.58
INHBA	0.002	4.00	MMP7	0.0000	6.06	KRT23	0.000	4.28
REG1A	0.007	3.85	DPEP1	0.0000	5.87	DPEP1	0.000	4.13
FOXQ1	0.002	3.57	KRT80	0.0000	5.19	CEMIP	0.000	4.05
MMP7	0.002	3.37	FABP6	0.0000	4.95	LY6G6F	0.000	3.64
REG1B	0.033	3.33	LINC00659	0.0000	4.93	PPBP	0.001	3.54
CEMIP	0.002	3.31	H19	0.0001	4.79	TACSTD2	0.000	3.32
MMP1	0.003	3.24	ACSL6	0.0000	4.71	SLC35D3	0.000	3.25
TCN1	0.004	3.23	NKD2	0.0000	4.70	SLC6A6	0.000	3.25
CXCL11	0.003	3.20	INHBA	0.0000	4.66	CTHRC1	0.000	3.15
Down-regulated			Down-regulated			Down-regulated		
Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC
CLCA4	0.006	-4.46	ZG16	0.000000	-9.42	CLCA4	0.000	-4.91
MS4A12	0.005	-4.45	CA1	0.000000	-8.75	ZG16	0.000	-4.44
AQP8	0.002	-4.32	GCG	0.000000	-7.65	GCG	0.000	-4.14
CA1	0.004	-4.23	CLCA1	0.000000	-7.50	ABCG2	0.000	-3.95
CA4	0.003	-4.03	SLC4A4	0.000000	-7.21	PKIB	0.000	-3.94
GCG	0.012	-3.61	ITLN1	0.000001	-7.09	CA2	0.000	-3.77
SLC26A3	0.012	-3.56	CLCA4	0.000000	-6.75	GUCA2A	0.000	-3.73
GUCA2A	0.005	-3.42	CA4	0.000000	-6.67	ABI3BP	0.000	-3.72
CLDN8	0.015	-3.33	CA2	0.000000	-6.61	BEST4	0.000	-3.68
ABCA8	0.003	-3.22	GUCA2B	0.000000	-6.59	CLDN8	0.000	-3.66

Supplementary Table 3. Top ten up-regulated and down-regulated genes in UC across datasets

GSE83687			GSE92415			GSE87473		
Up-regulated			Up-regulated			Up-regulated		
Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC
MMP3	0.0000	6.72	CHI3L1	0.00000	5.29	DUOX2	0.0000	5.00
CHI3L1	0.0000	6.25	MMP3	0.00000	5.07	CHI3L1	0.0000	4.74
CXCL5	0.0000	5.57	SLC6A14	0.00000	5.06	MMP3	0.0000	4.73
S100A12	0.0000	5.38	DUOX2	0.00000	4.97	SLC6A14	0.0000	4.71
MMP1	0.0000	5.38	DEFB4A	0.00000	4.62	DEFB4A	0.0000	4.37
MCEMP1	0.0000	5.33	S100A8	0.00000	4.07	S100A8	0.0000	4.01
WTAPP1	0.0000	5.13	SAA2///SAA1	0.00000	4.01	TNIP3	0.0000	4.00
CXCL11	0.0000	4.84	TNIP3	0.00000	3.85	CXCL8	0.0000	3.90
S100A8	0.0000	4.84	REG3A	0.00000	3.84	SAA2///SAA1	0.0000	3.88
ELK2AP	0.0000	4.84	CXCL1	0.00000	3.82	CXCL1	0.0000	3.74
Down-regulated			Down-regulated			Down-regulated		
Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC	Gene symbol	adj. P Val	log2FC
AQP8	0.0000	-6.70	AQP8	-4.80	0.0000	AQP8	0.0000	-4.62
DAO	0.0000	-4.45	SLC51A	-3.71	0.0000	SLC51A	0.0000	-3.83
GUCA2B	0.0000	-4.27	SLC26A2	-3.52	0.0000	CLDN8	0.0000	-3.60
OTOP2	0.0000	-4.27	CLDN8	-3.52	0.0000	HMGCS2	0.0000	-3.59
GUCA2A	0.0000	-4.26	DPP10-AS1	-3.50	0.0000	DPP10-AS1	0.0000	-3.45
CLDN8	0.0000	-4.24	PCK1	-3.49	0.0000	PCK1	0.0000	-3.45
CA1	0.0000	-4.09	HMGCS2	-3.28	0.0000	ABCG2	0.0000	-3.27
INSL5	0.0000	-3.98	ABCG2	-3.19	0.0000	MEP1B	0.0000	-2.97
TMIGD1	0.0000	-3.76	SLC38A4	-3.04	0.0000	GBA3	0.0000	-2.94
VWA5B1	0.0000	-3.64	CYP2B7P///CYP2B6	-3.03	0.0000	SLC38A4	0.0000	-2.94

Supplementary Table 4. Publicly available GEO datasets used in this study

Accession	Disease Context	Platform	Disease Samples (n)	Control Samples (n)	Purpose in Study
GSE110224	Colorectal cancer	GPL570	17	17	Differential expression analysis
GSE22598	Colorectal cancer	GPL570	17	17	Differential expression analysis
GSE21815	Colorectal cancer	GPL6480	132	9	Differential expression analysis
GSE83687	Ulcerative colitis	GPL16791	32	60	Differential expression analysis
GSE92415	Ulcerative colitis	GPL13158	162	21	Differential expression analysis
GSE87473	Ulcerative colitis	GPL13158	106	21	Differential expression analysis
GSE5206	Colorectal cancer	GPL96	100	5	Validation of upregulation of 24 EMT-related genes in colorectal cancer
GSE32323	Colorectal cancer	GPL570	17	17	Validation of upregulation of 24 EMT-related genes in colorectal cancer
GSE47908	Ulcerative colitis	GPL570	45	15	Validation of upregulation of 24 EMT-related genes in UC, ssGSEA analysis of z-score-based risk model
GSE59071	Ulcerative colitis	GPL6244	97	11	Validation of upregulation of 24 EMT-related genes in UC
GSE37283	Ulcerative colitis	GPL13158	11(UC-neoplasia) 4 (UC quiescent)	5	Gene selection, Random Forest modeling, ROC analysis, co-expression analysis
GSE17536	Colorectal cancer	GPL570	177	–	Overall survival and recurrence analyses of EMT-related genes, z-score-based model validation, co-expression analysis, and ssGSEA analysis
GSE75214	Active ulcerative colitis	GPL6244	74 (UC active) 23 (UC inactive)	11	Validation of upregulation trends of 14 EMT-related genes in active ulcerative colitis
GSE179285	Active ulcerative colitis	GPL6480	32 (UC active) 23 (UC inactive)	23	Validation of upregulation trends of 14 EMT-related genes in active ulcerative colitis
GSE15781	Colorectal cancer	GPL2986	22	20	Validation of the z-score-based risk model in colorectal cancer
GSE39582	Colorectal cancer	GPL570	566	19	Validation of the z-score-based risk model and co-expression analysis in colorectal cancer
GSE37364	Colorectal neoplasia	GPL570	29 (adenoma) 27(CRC)	38	Evaluation of Z-score across adenoma–carcinoma sequence