



<http://www.aimspress.com/journal/MBE>

Research article

Identification of potential biomarkers with colorectal cancer based on bioinformatics analysis and machine learning

Ahmed Hammad^{1,2}, Mohamed Elshaer^{1,3} and Xiuwen Tang^{1,*}

¹ Department of Biochemistry and Department of Thoracic Surgery of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China

² Radiation Biology Department, National Center for Radiation Research and Technology, Egyptian Atomic Energy Authority, Cairo 13759, Egypt

³ Labeled Compounds Department, Hot Labs Center, Egyptian Atomic Energy Authority, Cairo 13759, Egypt

* **Correspondence:** E-mail: xiuwentang@zju.edu.cn; Tel: +0086(0)57188981270;
Fax: +0086(0)57188208266

Supplementary

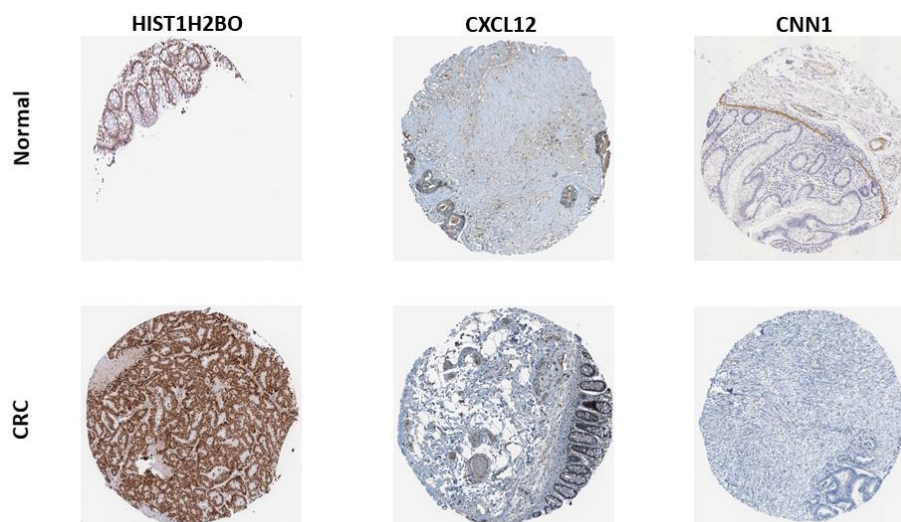


Figure S1. HPA Immunohistochemistry staining. Figures are showing the protein expression of some discovered biomarkers in cancer and normal cases. IHC staining indicated the HIST1H2BO expression was overexpressed CRC tissue while, the staining indicated that CXCL12 and CNN1 expression were downregulated cancer tissues.

Table S1. A full list of DEGs with P-value and Log2FC.

Gene name	P-value	Log2FC
ACTG2	7.82E-05	-2.5132474
ADAMDEC1	0.000700638	-1.7876829
ADAMTS1	2.24E-05	-1.69847431
ADH1C	0.007243681	-1.59266988
AG2	4.13E-05	-1.59385306
ANXA3	1.18E-06	1.559581619
AOC3	5.31E-05	-2.13942318
C3	3.17E-06	-2.01134417
C7	2.50E-05	-2.3168779
CA2	0.005057096	-1.5902095
CAV1	0.000197723	-1.52214255
CCDC80	3.54E-07	-2.21482029
CCL14	/// 4.43E-08	-2.30839102
CCL19	2.66E-05	-1.74360619
CEACAM6	0.001006531	1.866345637
CFH /// CFHR1	1.98E-07	-1.67902202
CHRD2	7.89E-07	-1.68762697
CILP	0.001324465	-1.99451604
CLDN2	2.16E-05	1.900529826
CLU	7.80E-06	-1.60790214
CNN1	9.30E-05	-2.6517931
COL14A1	1.34E-05	-1.78737828
CRISPLD2	5.25E-05	-1.85896882
CRYAB	2.05E-05	-2.02142998
CSRP1	0.000563726	-1.62848457
CXCL12	2.18E-07	-2.22811673
DES	7.43E-05	-2.80974563
DES	/// 0.000218387	-1.81633264
DPT	0.000447806	-1.77503881
F13A1	0.000294533	-1.7919965
FAM129A	0.000112728	-1.98612695
FHL1	1.14E-06	-2.50794128

FOLR1	0.001581522	-1.7568562
FOS	1.19E-06	-1.68413731
FOSB	3.15E-05	-2.36991221
FOXQ1	0.000265983	1.7089554
GEM	5.56E-05	-1.88016995
GLYR1	/// 0.000105894	-1.57959686
GPX3	4.51E-07	-2.37781122
GREM1	8.54E-06	-1.54259729
HIST1H2BO	2.49E-06	1.605592532
HIST1H3B	1.05E-07	1.72855819
HIST1H3F	7.18E-07	1.529398961
HIST1H3G	1.22E-08	1.60169528
HSPB6	2.06E-07	-4.03663425
IGF1	1.39E-05	-1.90069424
IGH@	/// 0.000184056	-1.68879247
IGH@	/// 2.81E-05	-1.53576796
IGH@	/// 6.79E-05	-1.806699
IGH@	/// 0.000454008	-1.67147553
IGH@	/// 0.000411719	-1.58001289
IGHA1 /// IGHD	7.56E-05	-1.50066412
IGHA1 /// IGHD	8.62E-05	-1.51511136
IGHA1	/// 5.46E-05	-1.55388263
IGHD	/// 0.00017173	-1.62939335
IGHG1	/// 0.000143445	-1.5220659
IGHM	3.06E-05	-1.7549489
IGJ	9.50E-05	-2.54055758
IGK@ /// IGKC	0.000310753	-1.62683781
IGKV1D-13	0.000617092	-1.56910815
IGLV2-23	0.000133503	-1.61559941
IL8	8.47E-05	1.555366609
KCNMB1	0.000165677	-1.55947491
KCTD12	7.70E-06	-1.58605327
KRT18	0.000117868	1.579756135
LMOD1	8.29E-05	-2.11378074
LOC100293440	4.56E-05	-1.69117357

LOC728264	3.83E-06	-1.92291917
LYVE1	0.000204774	-1.80907003
MAB21L2	0.00016604	-2.51409302
MAP1B	4.05E-05	-1.73308303
MATN2	2.53E-07	-1.67901827
MATR3	/// 8.12E-07	1.508025975
MFAP4	2.07E-06	-1.70011094
MGP	4.94E-06	-1.59730189
MS4A12	0.010878535	-1.77444662
MT1M	1.77E-05	-2.56041282
MUSTN1	1.14E-05	-1.50105751
MYH11	9.77E-06	-2.27408858
MYL9	7.44E-05	-1.80945377
OGN	0.000177035	-1.83477577
PLA2G2A	2.50E-05	-1.92006561
PPAP2A	3.43E-09	-1.6450938
PPP1R12B	6.71E-06	-1.58890461
PRIMA1	1.31E-05	-1.6770041
PTGDS	3.46E-08	-2.13910887
PTGIS	6.17E-05	-1.77190901
RBPM5	0.000176269	-1.51459038
RGS2	9.07E-06	-1.94948183
RNF43	1.62E-08	1.520676126
S100P	3.58E-06	1.588066163
SERPINF1	2.50E-06	-1.62262725
SFRP2	6.42E-06	-2.76057518
SLC26A2	0.017995797	-1.52474327
SPARCL1	2.09E-06	-2.17011969
SRPX	5.15E-05	-1.62030404
SYNM	0.000370914	-1.5937141
SYNPO2	4.73E-05	-2.63197896
TGFB1I1	0.000300593	-1.64357972
TMEM97	6.96E-10	1.507875507
TNS1	7.94E-05	-1.69788772
TPSAB1	/// 1.41E-06	-1.63270559

TPX2	3.95E-08	2.173175143
UBE2C	1.73E-08	2.087131174
UCA1	5.47E-05	2.080060879

Table S2. Table showing the Degrees of Connectivity for discovered biomarkers.

Gene name	Degree Centrality	Betweenness Centrality	Closeness Centrality
MYH11	9	0.23629686	0.36641221
IGF1	9	0.47695563	0.45714286
CLU	8	0.3471399	0.42477876
FOS	7	0.29827761	0.4
MYL9	7	0.02581898	0.2962963
CXCL12	6	0.19401701	0.36090226
LMOD1	6	0.04532569	0.29090909
CNN1	6	0.00709748	0.29090909
HIST1H2BO	5	0.11968085	0.3
C3	5	0.09802326	0.33566434

Table S3. Table showing the AUC values for discovered biomarkers.

Marker	AUC
IGF1	0.9845
MYH11	0.96608
FOS	0.95409
CLU	0.97047
MYL9	0.92515
CXCL12	0.99357
LMOD1	0.95702
C3	0.95585
CNN1	0.9383
HIST1H2BO	0.97076