Sporadic colon cancer murine models demonstrate the value of autoantibody detection for preclinical cancer diagnosis

Rodrigo Barderas[1+†], Roi Villar-Vazquez[1+], María Jesús Fernández-Aceñero[2], Ingrid Babel[1], Alberto Peláez-García[1], Sofía Torres[1] & J. Ignacio Casal[1*]

([+]. Equal contribution)

[1]Department of Cellular and Molecular Medicine. Centro de Investigaciones Biológicas (CSIC). 28040 Madrid. Spain.

[2]Pathology Department. Fundación Jiménez Díaz. Madrid. Spain

†. Current address: Biochemistry and Molecular Biology I Department. Facultad de Ciencias Químicas. Universidad Complutense de Madrid. 28040 Madrid

*Corresponding author:
J. Ignacio Casal
Department of Cellular and Molecular Medicine
Centro de Investigaciones Biológicas (CIB-CSIC)
Ramiro de Maeztu, 9
28040 Madrid. Spain
Phone: +34 918373112
Fax: +34 91 5360432
e-mail: icasal@cib.csic.es

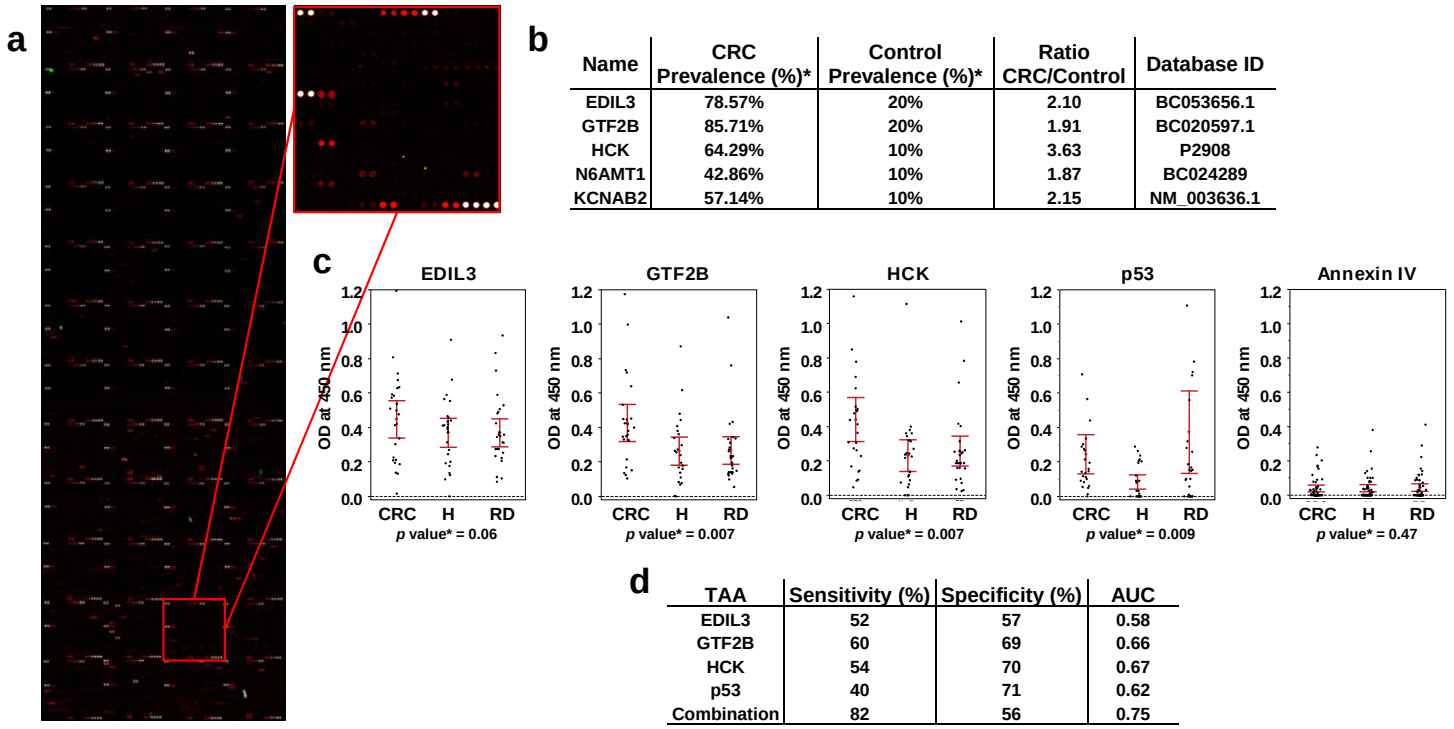
Supplementary Note 1

We investigated the profiling of colorectal cancer specific autoantibodies on a modified version of ProtoArrays with different surface properties; we probed 20 sera from CRC and control patients (Supplementary Fig. S1a and Supplementary Table S1). After performing a high-stringent bioinformatic analysis we selected only those proteins that were coincident in three bioinformatic analyses and showed the highest prevalence in CRC patients. By using the Prospector Analyzer tool, we identified 24 proteins showing immunoreactivity with cancer sera, having a p value<0.04 and a prevalence >50% in the CRC group (Supplementary Table S2). In addition, normalized data were processed using T-Rex from the GEPAS web-tool database and Pomelo II. From 24 initially-identified candidates, we found 5 CRC-associated antigens coincident in all three analyses (Supplementary Fig. S1b). From these five, we selected three antigens (EDIL3, GTF2B and HCK) that showed prevalence >60% in the CRC group and <20% in the control group (Supplementary Fig. S1b).

Although, we have not found any association of these proteins to colorectal carcinoma, EDIL3, GTF2B and HCK might play an important role in colon tumorigenesis and as CRC diagnostic markers since they have been reported to develop important cellular functions. Indeed, EDIL3 binds αvβ3 integrin and plays a role in angiogenesis and vascular morphogenesis[1]. High expression of EDIL3 has been associated to poor prognosis in hepatocellular carcinoma[2]. GTF2B, general transcription factor IIB, is required for transcription initiation by RNA polymerase II. It plays a major role in the activation of eukaryotic genes transcribed by RNA polymerase II[3]. HCK (hematopoietic cell kinase) belongs to the Src family of tyrosine kinases and has been reported to be involved in phagocytosis, adhesion and migration, to regulate formation

of membrane protrusions, lysosome exocytosis, podosome formation, actin polymerization, and has been associated to cancer[4, 5, 6].

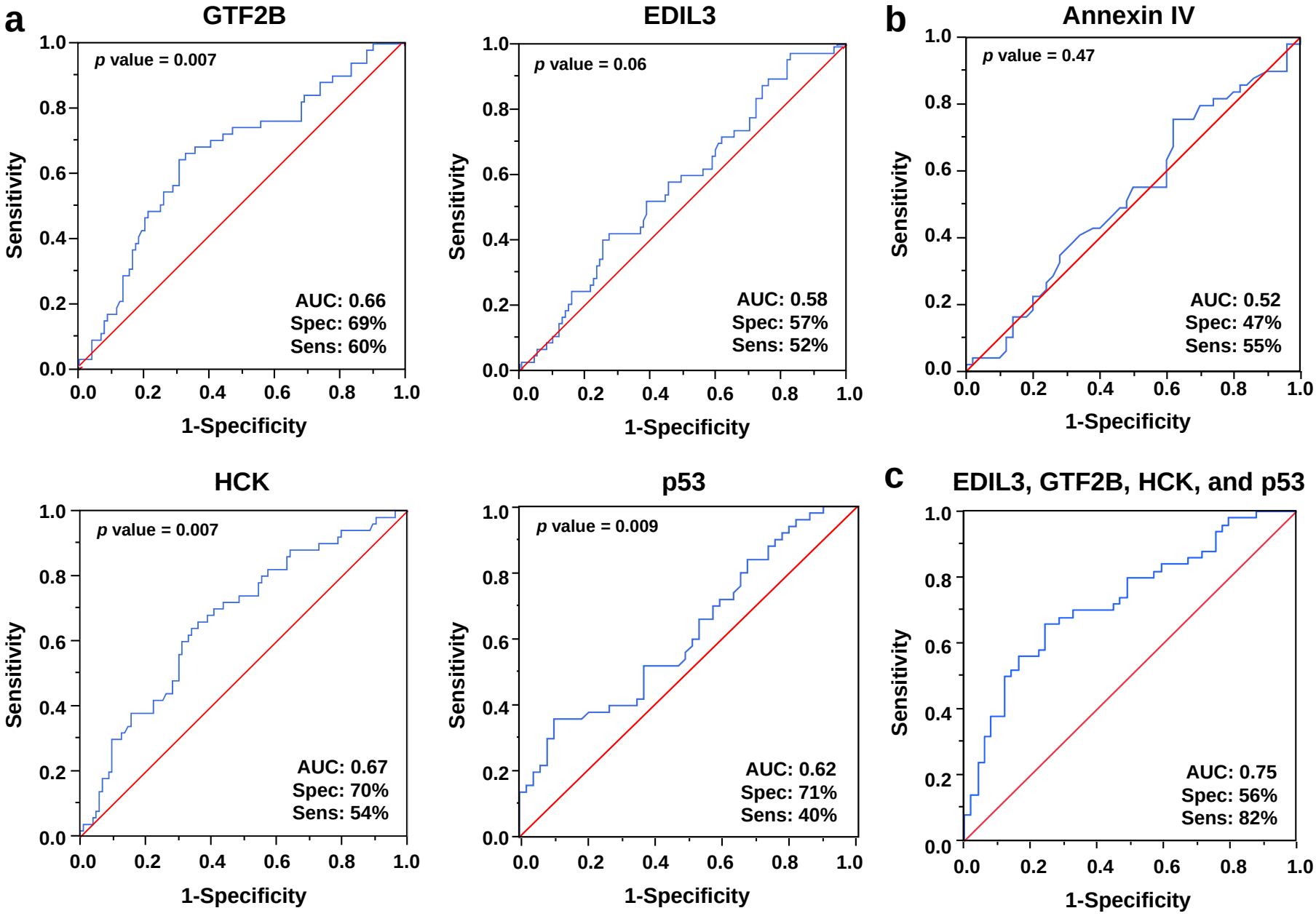
Then, to confirm the suitability of these new identified TAAs as potential CRC diagnostic markers, we used them, in combination with p53, for the screening of an independent set of 153 serum samples (50 CRC, 49 healthy samples, and 54 sera from other related diseases) (Supplementary Table S1). Although all the CRC TAAs were able to discriminate between CRC and control samples, best differential values were obtained for GTF2B and HCK, with p values <0.01 (Supplementary Fig. S1c). In both cases, the comparison of healthy samples *versus* other related cancers did not give significant values. However, p53 reactivity was similar for CRC and samples from other cancers, indicating a low specificity of p53 antibodies for colon cancer. No reactivity was observed against control Annexin IV (Supplementary Fig. S1c). To determine specificity and sensitivity, receiver operating characteristics (ROC) curves were generated for each of these TAAs to discriminate CRC. The areas under the curve (AUC) values were relatively low for individual antigens, ranging between 0.58 to 0.67 for EDIL3, GTF2B and HCK and 0.62 for p53. However, the combination of markers gave an AUC of 0.75 (Supplementary Fig. S1d and Supplementary Fig. S2). Collectively, these data support the capacity of EDIL3, GTF2B and HCK TAAs as candidate biomarkers for the diagnosis of CRC. Therefore, we incorporated these TAAs in subsequent studies for autoantibody responses in murine models of sporadic colon cancer.



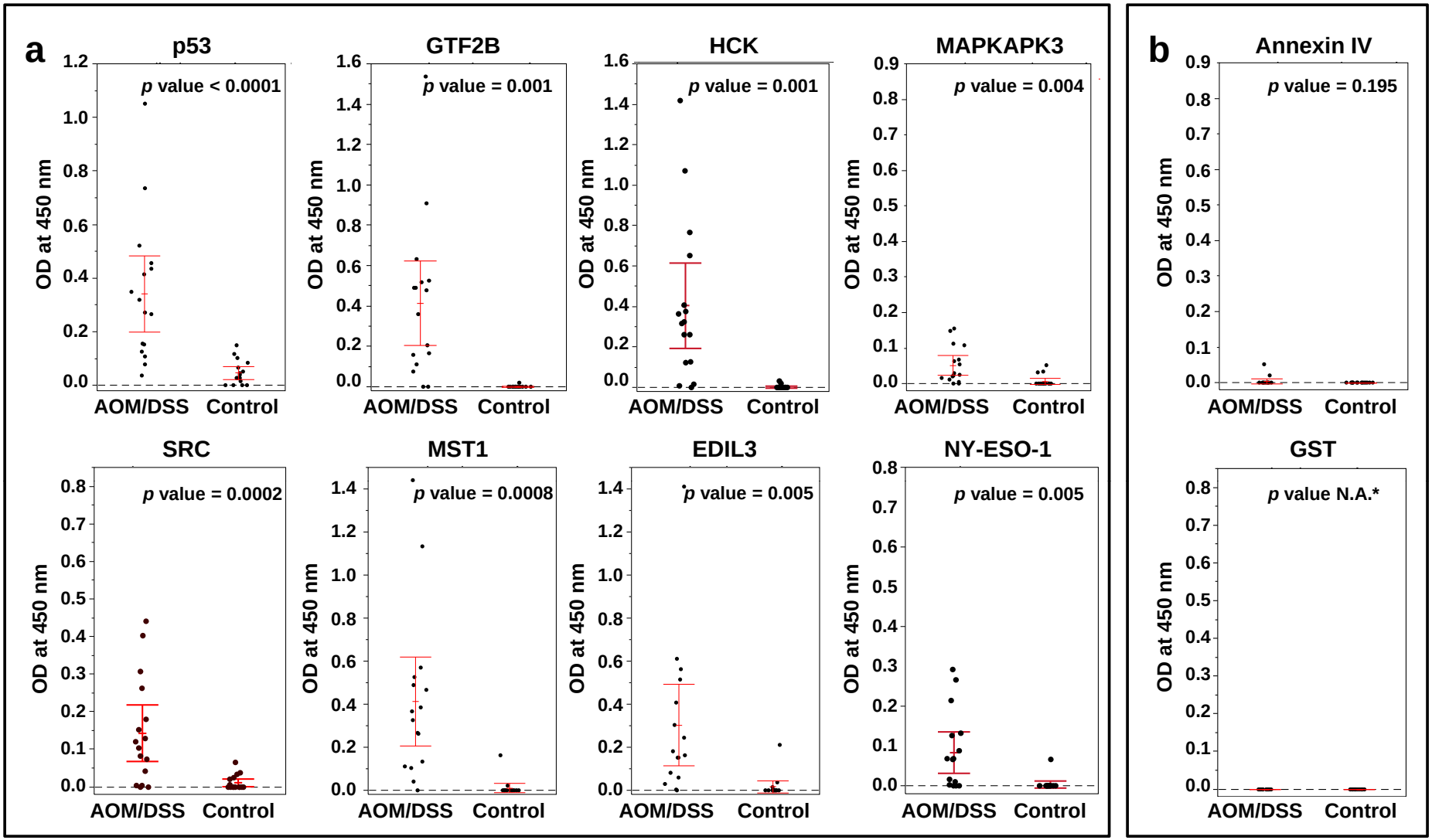
Name	CRC Prevalence (%)*	Control Prevalence (%)*	Ratio CRC/Control	Database ID
EDIL3	78.57%	20%	2.10	BC053656.1
GTF2B	85.71%	20%	1.91	BC020597.1
HCK	64.29%	10%	3.63	P2908
N6AMT1	42.86%	10%	1.87	BC024289
KCNAB2	57.14%	10%	2.15	NM_003636.1

TAA	Sensitivity (%)	Specificity (%)	AUC
EDIL3	52	57	0.58
GTF2B	60	69	0.66
HCK	54	70	0.67
p53	40	71	0.62
Combination	82	56	0.75

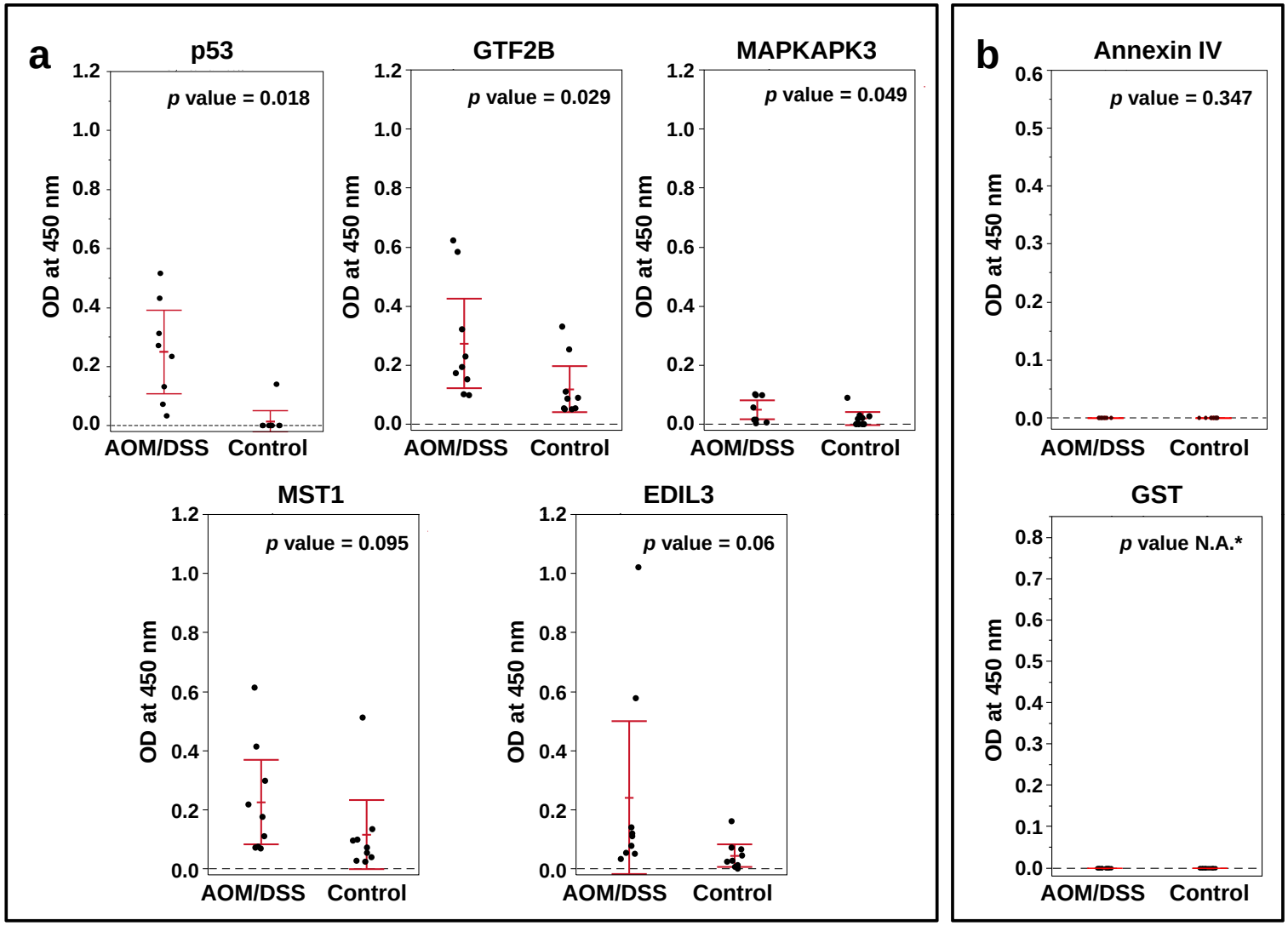
Supplementary Figure S1 | Identification and validation of proteins reactive to autoantibodies as CRC diagnostic markers. (**a**) Twenty ProtoArrays v4.1 were tested with 20 sera from CRC patients and controls (Supplementary Table S1). A representative microarray incubated with sera from a CRC patient is shown, together with a highlighted sub-array to observe the presence of reactive proteins recognized by autoantibodies present in the sera of CRC patients. Control spots were located in both corners of each subarray. In red, proteins reacting with CRC serum. In white, saturated control spots. (**b**) The five TAAs identified with all three bioinformatic approaches: i) Pomelo II, ii) T-Rex, and iii) ProtoArray Prospector Analyzer. Prevalence values correspond to that obtained from the Prospector Analyzer. Ratio was calculated with the fluorescence intensity for each TAA from the median of CRC and Control group. Database ID, accession number of the TAA. (**c**) ELISA with serum samples from CRC patients, healthy individuals (H) and related diseases patients (RD) was performed using human recombinant EDIL3, GTF2B, HCK, and p53. Annexin IV was used as negative control (Supplementary Table S1). *, *p* values correspond to the comparison CRC vs healthy controls. (**d**) ROC curves were built comparing CRC samples *vs* healthy samples for individual proteins and the combination of TAAs.



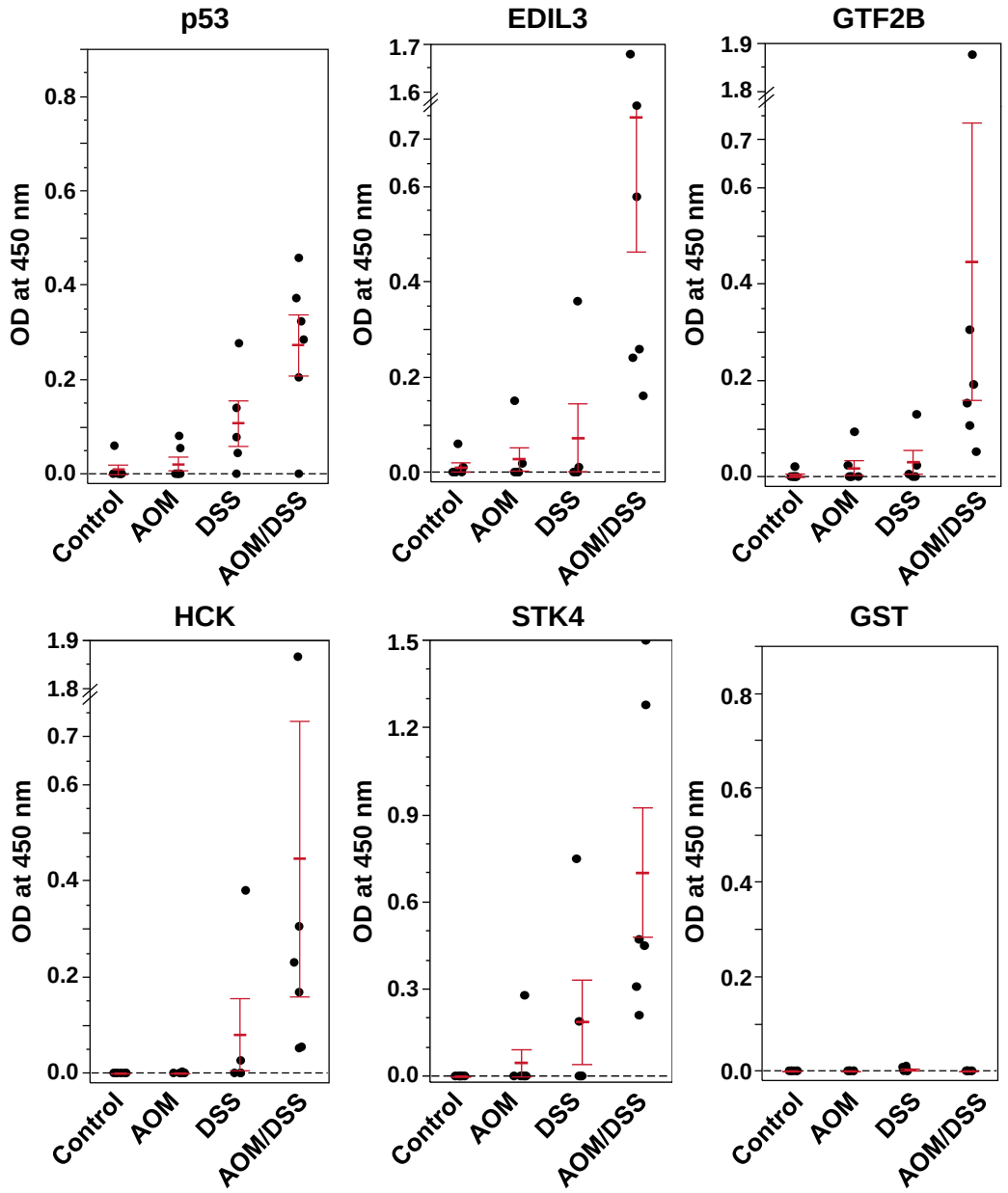
Supplementary Figure S2 | Validation of the TAAs for the diagnosis of CRC. Individual performance and best combination performance of the proteins in the validation set. Receiver-operating-characteristic curves were constructed with the ELISA values and based on multiplex analyses of (**a**) the four TAAs, (**b**) Annexin IV as control, and (**c**) the best combination using 50 CRC and 49 healthy serum samples.



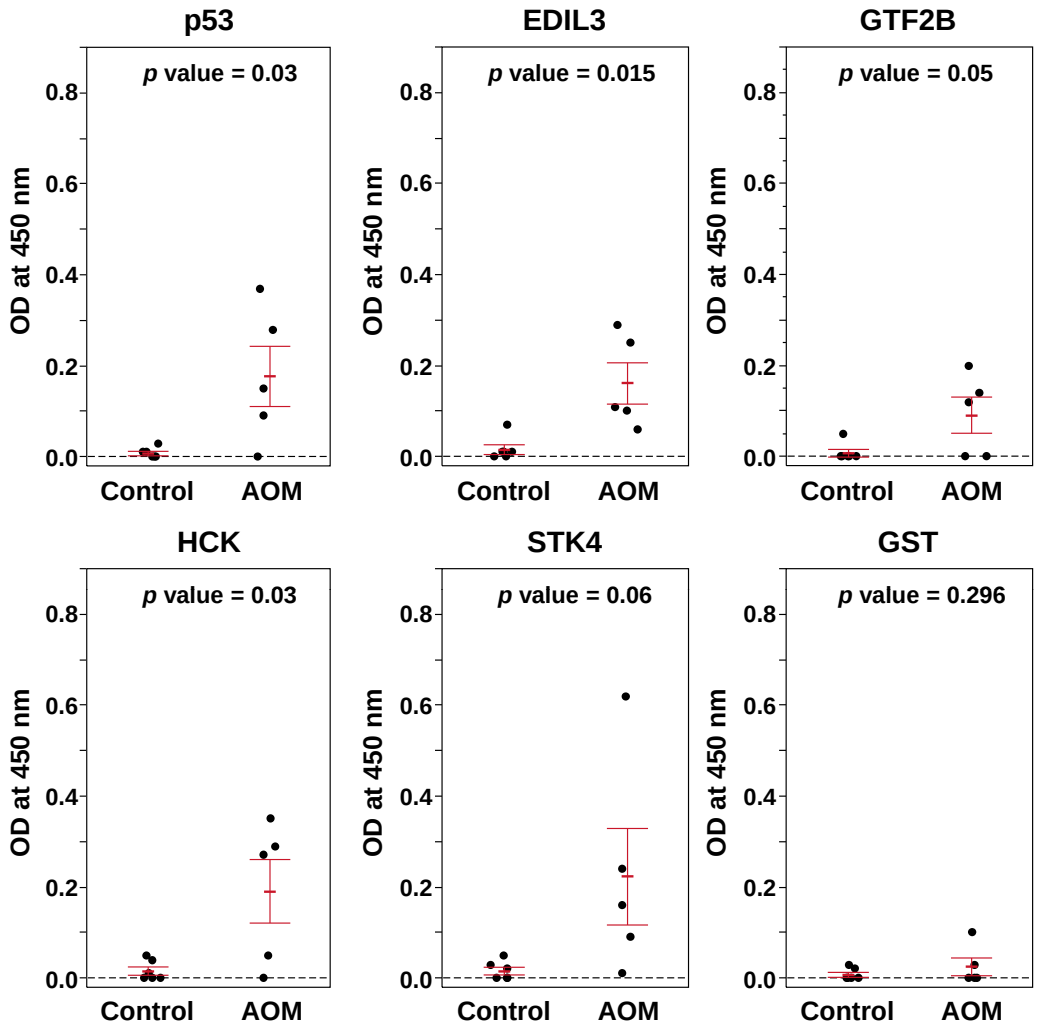
Supplementary Figure S3 | ELISA-based analysis of murine serum samples using selected human colorectal TAAs and controls. (**a**) Absorbance values obtained with 16 AOM/DSS-treated and 16 vehicle-treated control mice by indirect ELISA coating ELISA plates with 0.3 μg of individual human CRC TAAs protein per well. Dots represent the individual binding for each serum. Mean of the assay is indicated by a red horizontal dash. Errors bars represent the CI95 estimation of the mean. (**b**) Annexin IV and GST were used as control of the assay. N.A.* , the *p* value for GST was immeasurable since the OD for both groups was background.



Supplementary Figure S4 | Analysis of autoantibodies to p53, GTF2B, MAPKAPK3, MST1/STK4 and EDIL3 as early colorectal cancer markers. Optical density (OD) data at 450 nm were obtained with the 9 AOM/DSS-treated and 9 vehicle-treated control mice using murine sera collected just after the first AOM/DSS cycle where no presence of adenoma was observed (Fig. 1*B*). (**a**) Data were obtained by indirect ELISA coating ELISA plates with 0.3 μg of individual human colorectal cancer TAAs protein per well. Dots represent the individual binding for each serum. Mean of the assay is indicated by a red horizontal dash. Errors bars represent the CI95 estimation of the mean. (**b**) Annexin IV and GST were used as controls of the assay. N.A.*, the *p* value for GST was immeasurable since the OD for both groups was background.



Supplementary Figure S5 | ELISA-based analysis of murine serum samples collected at day 63 using selected human colorectal TAAs and controls. Absorbance values obtained with a new independent set of murine samples from AOM/DSS-, AOM-, DSS-, and vehicle-treated control mice by indirect ELISA coating ELISA plates with 0.3 μg of individual human colorectal cancer TAAs protein per well. Serum samples were collected at the end of the protocol of AOM/DSS-, and DSS-treated mice, whereas mice from AOM-treated group were maintained 187 extra days. Dots represent the individual binding for each serum. Mean of the assay is indicated by a red horizontal dash. Errors bars represent the CI95 estimation of the mean.



Supplementary Figure S6 | ELISA-based analysis of murine serum samples collected at day 250 using selected human colorectal TAAs and controls. Absorbance values obtained with the murine samples from AOM-, and vehicle-treated control mice by indirect ELISA coating ELISA plates with 0.3 μg of individual human colorectal cancer TAAs protein per well. Serum samples were collected at the end of the protocol of AOM-treated mice. Dots represent the individual binding for each serum. Mean of the assay is indicated by a red horizontal dash. Errors bars represent the CI95 estimation of the mean.

Supplementary Table S1 | Clinical information of the 20 human serum samples used for the screening of the ProtoArrays v4.1 to identify colorectal cancer-specific TAAs and the 153 human serum sample for the validation of EDIL3, GTF2B, and HCK in comparison to p53 and Annexin IV as control of immunoreactivity.

			Sex (%)		Age (years)		
		Total number	Female	Male	Mean	Minimum	Maximum
Screening	**Healthy**	8	25	75	65.3	38	81
	CRC Dukes' stage D	12	25	75	64.8	41	84
Validation	**Healthy**	49	36.7	63.3	59.9	34	89
	CRC cases	50	34	66	70.8	41	91
	CRC Dukes' stage A	6	50	50	59.7	51	82
	CRC Dukes' stage B	11	27.3	72.7	72.5	55	85
	CRC Dukes' stage C	16	50	50	76.1	55	85
	CRC Dukes' stage D	17	17.6	82.4	68.7	41	91
	Related Diseases	54	62.2	37.8	57.1	26	79.7
	Bladder cancer	9	36	64	67.7	58	78
	Breast cancer	8	100	0	52.5	30	66.2
	Lung cancer	8	25	75	62.9	41	76.9
	Pancreas cancer	8	37.5	62.5	64.3	36.9	74
	Stomach cancer	8	75	25	62.4	37	79.7
	Family Record of cancer	9	77.8	22.2	46.7	26	66
	Patient Record of cancer	1	0	100	73	73	73
	Hyperplastic polyp	1	100	0	61	61	61
	Ulcerative colitis	2	50	50	38.5	28	49

Supplementary Table S2 | Identification of proteins reactive to autoantibodies using the software provided by the manufacturer.

Protein Name	*p* value	CRC prevalence	Control prevalence
TCF19	0.00131	71%	10%
GTF2B	0.00444	86%	20%
IMAGE:4878869	0.00988	71%	20%
NAP1L3	0.00988	71%	20%
EDIL3	0.00988	71%	20%
HSH2D	0.00988	71%	20%
FGR	0.00988	71%	20%
CPEB1	0.01022	57%	10%
HCK	0.02384	57%	10%
ENG	0.02384	50%	10%
EME1	0.02384	50%	10%
CDK18	0.02384	50%	10%
ARHGAP26	0.02384	50%	10%
KCNAB2	0.02384	50%	10%
EXOC3	0.02384	50%	10%
TRIM44	0.02384	50%	10%
LCA5	0.02489	50%	10%
MAPKAPK3	0.02489	64%	20%
APIP	0.02489	64%	20%
IMAGE:3503500	0.02489	64%	20%
TUBG1	0.02489	64%	20%
CPSF3	0.02489	64%	20%
ZADH2	0.02489	64%	20%
FRK	0.02489	64%	20%

Supplementary Table S3 | Comparison of the human and murine amino acid sequences of the TAAs tested by ELISA in the study.

TAA	Identity (%) *	Similarity (%) *
p53	76	83
MST1/STK4	97	98
EDIL3	95	97
GTF2B	99	100
HCK	89	95
MAPKAPK3	94	96
SRC	98	98
NY-ESO-1	29	42
Annexin IV	91	96

* Identity, number of identical amino acid residues at each position in both sequences.

* Similarity, number of similar (same properties) amino acid residues at each position in both sequences.

Supplementary Table S4 | Oligonucleotides designed for analysis of the mRNA expression of the TAAs in AOM/DSS- and vehicle-treated control mice.

TAA	Primer*	Sequence	DNA amplified (bp)
p53	Fw	CGCTGCCCCCACCATGAGCG	227
	Rv	GGTCGGCGGTTCATGCCCCC	
STK4	Fw	GAGGAACCCACCGCGCAGGC	236
	Rv	CGTGAGGGCTGTCACACTGCTGC	
EDIL3	Fw	ACTCAGCCTCGGGGTGCCCC	330
	Rv	CCGCCATTTCTGCAAGGCTCAGCT	
GTF2B	Fw	CCCGCAAGGCAGTGGAGCTGG	181
	Rv	CTGAAGGGAAGAGATCCGGAGCCCG	
MAPKAPK3	Fw	CTCCGGCCCCAGGTGTGCGA	266
	Rv	ACAGCGCTTGCCGTGGTGCA	
PAK1	Fw	GCCAGAACAGTGGGCTCGCTTGC	208
	Rv	TGACACTGGTGGTACTGCTGGGGT	
β-actin	Fw	CATGTACGTAGCCATCCAGGC	251
	Rv	CTCTTTGATGTCACGCACGAT	

* Fw, forward. Rv, reverse.

Supplementary Methods

Clinical Information and Serum Samples. For microarray screening, serum samples from 12 Duke's stage D CRC patients collected after diagnosis and eight control sera from healthy individuals matching both the median age and the sex of the patients were used (Supplementary Table S1). For ELISA validation, an independent cohort of serum samples (Supplementary Table S1) comprising 50 CRC samples (representative of A-D Dukes stages), 54 samples from related diseases (bladder, breast, lung, pancreas, and stomach cancer; ulcerative colitis, hyperplastic polyp, and familial or personal antecedents of cancer), and 49 healthy reference control sera were obtained from the Bellvitge University Hospital (Barcelona), Hospital Puerta de Hierro (Madrid), Hospital Universitario de Salamanca, and Hospital de Cabueñes (Gijón). Sample collection was approved by the Ethical Review Boards of these institutions. Written informed consent was obtained from all patients. All sera were processed as previously described[7, 8, 9]. Samples were handled anonymously according to ethical and legal guidelines at the CSIC.

Protein Arrays. Twenty serum samples (Supplementary Table S1) were probed with the Human ProtoArray v4.1 (Invitrogen). Each microarray contains more than 8000 human recombinant proteins expressed in *Sf*9 insect cells and printed in duplicate. ProtoArray microarrays were used according to the manufacturer's instructions. Briefly, microarrays were blocked for 1 h at 4°C with 50 mM HEPES, 200 mM NaCl, 0.08% Triton X-100, 25% glycerol, 20 mM reduced glutathione, 1 mM DTT, 40 mM NaOH and 1% casein at pH 7.5. The microarrays were then incubated for 90 min at 4°C with 2 mL of the CRC and control serum samples diluted 1:500 in 1x PBS, 1% casein and 0.1% Tween at pH 7.5 (PBST-1% casein) using a different Atlas® Glass hybridization chamber (Clontech Laboratories) for every array and sample. The microarrays were washed five times for 5 min with PBST-1% casein and incubated for 90 min at 4°C with goat AlexaFluor 647-conjugated anti human IgG diluted 1:2000 (Invitrogen). The

microarrays were washed again with PBST five times for 5 min, rinsed with deionized water and dried by centrifugation at 1000 rpm for 1 min. Microarray data were acquired using a ScanArray ™ 5000 MicroArray scanner (Packard BioChip Technologies). The arrays were scanned using 635 nm and 532 nm lasers and a resolution of 10 µm / pixel to produce red (Alexa Fluor 647) or green images (Alexa Fluor 555). Finally, the Genepix Pro 7 (Molecular Devices) image analysis software was used for quantification.

Statistical Analysis. Microarrays were analyzed by three approaches to identify tumor associated antigens differentially recognized by the sera of CRC patients' sera and reference controls' sera: i) paired t tests using Pomelo II (http://pomelo2.bioinfo.cnio.es/) considering two classes (CRC and control)[10], ii) T-Rex tool for analyzing differential expression between two conditions (CRC and control) after preprocessing for normalization of data, which is included in the web-based tool GEPAS 4.0 (http://www.gepas.org/)[11], and iii) the array manufacturer's software ProtoArray Prospector Analyzer 4.0 (Invitrogen), which relies on Chebyshev's inequality principle[7], which determines if a fluorescent signal on the microarray is significant using the negative controls printed in the array. We select as colorectal cancer-specific antigens those proteins that appeared in all three bioinformatics analysis (Supplementary Fig. S1b).

Supplementary References

1. Hidai, C., Zupancic T., Penta K*., et al.* Cloning and characterization of developmental endothelial locus-1: an embryonic endothelial cell protein that binds the alphavbeta3 integrin receptor. *Genes Dev* **12,** 21-33 (1998).

2. Sun, J. C., Liang X. T., Pan K*., et al.* High expression level of EDIL3 in HCC predicts poor prognosis of HCC patients. *World J Gastroenterol* **16,** 4611-4615 (2010).

3. Ha, I., Lane W. S. & Reinberg D. Cloning of a human gene encoding the general transcription initiation factor IIB. *Nature* **352,** 689-695 (1991).

4. Poincloux, R., Al Saati T., Maridonneau-Parini I. & Le Cabec V. The oncogenic activity of the Src family kinase Hck requires the cooperative action of the plasma membrane- and lysosome-associated isoforms. *Eur J Cancer* **45,** 321-327 (2009).

5. Guiet, R., Poincloux R., Castandet J*., et al.* Hematopoietic cell kinase (Hck) isoforms and phagocyte duties - from signaling and actin reorganization to migration and phagocytosis. *Eur J Cell Biol* **87,** 527-542 (2008).

6. Montero, J. C., Seoane S., Ocana A. & Pandiella A. Inhibition of SRC family kinases and receptor tyrosine kinases by dasatinib: possible combinations in solid tumors. *Clin Cancer Res* **17,** 5546-5552 (2011).

7. Babel, I., Barderas R., Diaz-Uriarte R*., et al.* Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. *Mol Cell Proteomics* **8,** 2382-2395 (2009).

8. Babel, I., Barderas R., Diaz-Uriarte R*., et al.* Identification of MST1/STK4 and SULF1 proteins as autoantibody targets for the diagnosis of colorectal cancer by using phage microarrays. *Mol Cell Proteomics* **10,** M110 001784 (2011).

9. Barderas, R., Babel I., Diaz-Uriarte R*., et al.* An optimized predictor panel for colorectal cancer diagnosis based on the combination of tumor-associated antigens obtained from protein and phage microarrays. *J Proteomics* **75,** 4647-4655 (2012).

10. Morrissey, E. R. & Diaz-Uriarte R. Pomelo II: finding differentially expressed genes. *Nucleic Acids Res* **37,** W581-586 (2009).

11. Montaner, D., Tarraga J., Huerta-Cepas J*., et al.* Next station in microarray data analysis: GEPAS. *Nucleic Acids Res* **34,** W486-491 (2006).