Identification of tumor-associated antigens with diagnostic ability of colorectal cancer by in-depth immunomic and seroproteomic analysis

María Garranzo-Asensio, Pablo San Segundo-Acosta, Carmen Povés , María Jesús Fernández-Aceñero, Javier Martínez-Useros, Ana Montero-Calle, Guillermo Solís-Fernández, Maricruz Sanchez-Martinez, Nuria Rodríguez, María Ángeles Cerón, Servando Fernandez-Diez, Gemma Domínguez, Vivian de Los Ríos, Alberto Peláez-García, Ana Guzmán-Aránguez, Rodrigo Barderas.

J Proteomics. 2020 Mar 1;214:103635.

Elsevier Editorial System(tm) for Journal of Proteomics

Manuscript Draft

Manuscript Number: JPROT-D-19-00483R2

Title: Identification of tumor-associated antigens with diagnostic ability of colorectal cancer by in-depth immunomic and seroproteomic analysis

Article Type: Full Length Article

Section/Category: Original Article

Keywords: Colorectal cancer; immune response; Seroproteomics; Diagnostic autoantibody biomarkers; Mass spectrometry

Corresponding Author: Dr. Rodrigo Barderas, Chemistry

Corresponding Author's Institution: Instituto de Salud Carlos III

First Author: María Garranzo-Asensio

Order of Authors: María Garranzo-Asensio; Pablo San Segundo-Acosta; Carmen Poves; Maria Jesus Fernández-Aceñero; Javier Martínez-Useros, PhD; Ana Montero-Calle; Guillermo Solís-Fernández; Maricruz Sanchez-Martinez; Nuria Rodriguez; María Ángeles Cerón; Servando Fernandez-Diez; Gemma Domínguez; Vivian de los Ríos; Alberto Pelaez-Garcia; Ana Guzmán-Aránguez; Rodrigo Barderas, Chemistry

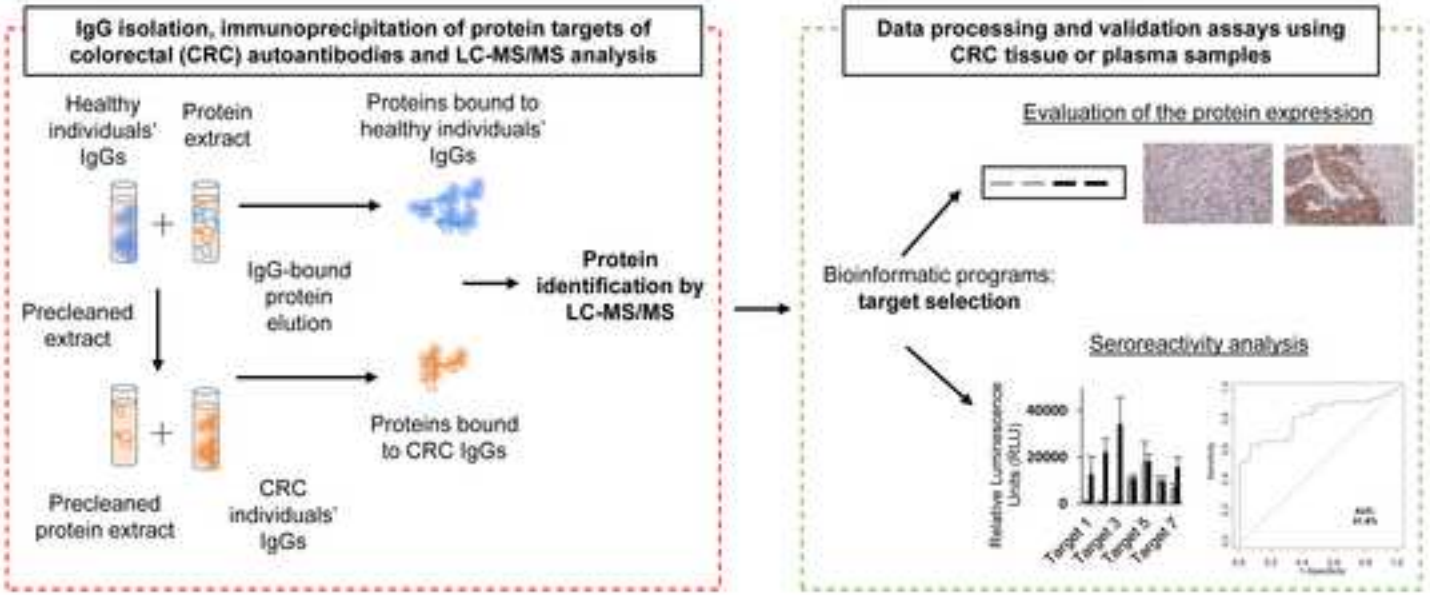
Abstract: Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer related death worldwide. Its diagnosis at early stages would significantly improve the survival of CRC patients. The humoral immune response has been demonstrated useful for cancer diagnosis, predating clinical symptoms up to 3 years. Here, we employed an in-depth seroproteomic approach to identify proteins that elicit a humoral immune response in CRC patients. The seroproteomic approach relied on the immunoprecipitation with patient-derived autoantibodies of proteins from CRC cell lines with different metastatic properties followed by LC-MS/MS. After bioinformatics, we focused on 31 targets of CRC autoantibodies. After WB and IHC validation, ERP44 and TALDO1 showed potential to discriminate disease-free and metastatic CRC patients, and time to recurrence of CRC patients in stage II. Using plasma samples of 30 healthy individuals, 28 premalignant individuals, and 32 CRC patients, nine out of 13 selected targets for seroreactive analysis showed significant diagnostic ability to discriminate either CRC patients or premalignant subjects from controls. Our results suggest that the here defined panel of CRC autoantibodies and their target proteins should be included in CRC blood-based biomarker panels to get a clinically useful blood-based diagnostic signature for CRC detection.

***Significance**

Significance: Colorectal cancer is one of the deadliest cancer types mainly due to its late diagnosis. Its early diagnosis, therefore, is of great importance since it would significantly improve the survival of CRC patients. In our work, the in-depth seroproteomic analysis of colorectal cancer using isolated IgGs from colorectal cancer patients and controls and protein extract of colorectal cancer cells provide the identification of valuable biomarkers with diagnostic and prognostic ability of the disease

*Graphical Abstract

Click here to download high resolution image



***Highlights (for review)**

- Colorectal cancer (CRC) associated death is mainly related to its late diagnosis
- The humoral immune response predates clinical symptoms up to 3 years
- Immunomic and seroproteomic analyses identified valuable autoantibody CRC biomarkers
- Identified targets of autoantibodies discriminated CRC and premalignant individuals from controls

Identification of tumor-associated antigens with diagnostic ability of colorectal cancer by in-depth immunomic and seroproteomic analysis

María Garranzo-Asensio[a], Pablo San Segundo-Acosta[a,b], Carmen Povés[c], María Jesús Fernández-Aceñero[d], Javier Martínez-Useros[e], Ana Montero-Calle[a], Guillermo Solís-Fernández[a], Maricruz Sanchez-Martinez[a], Nuria Rodríguez[f], María Ángeles Cerón[d], Servando Fernandez-Diez[c], Gemma Domínguez[g], Vivian de los Ríos[h], Alberto Peláez-García[i], Ana Guzmán-Aránguez[b], Rodrigo Barderas[a,*]

[a]Chronic Disease Programme, UFIEC, Instituto de Salud Carlos III, Majadahonda E-28220, Madrid, Spain

[b]Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain

[c]Gastroenterology Unit, Hospital Universitario Clínico San Carlos, E-28040 Madrid, Spain

[d]Surgical Pathology Department, Hospital Universitario Clínico San Carlos, E-28040 Madrid, Spain

[e]Translational Oncology Division, OncoHealth Institute, Fundacion Jimenez Diaz University Hospital, E-28040 Madrid, Spain

[f]Medical Oncology Department, Hospital Universitario La Paz, E-28046 Madrid, Spain

[g]Departamento de Medicina, Facultad de Medicina, Instituto de Investigaciones Biomédicas "Alberto Sols", CSIC-UAM, E-28029 Madrid, Spain

[h]Centro de Investigaciones Biológicas, CSIC, E-28040 Madrid, Spain

[i]Surgical Pathology Department, Hospital Universitario La Paz, E-28046 Madrid, Spain

[*]To whom correspondence should be addressed:

Rodrigo Barderas.

Chronic Disease Programme, UFIEC, Functional Proteomics Unit, Instituto de Salud Carlos III, E-28222 Majadahonda, Madrid, Spain; Tel.: 34-91-8223231; E-mail: r.barderasm@isciii.es

Running title

CRC biomarker identification by seroproteomic analysis

Disclosure of Potential Conflicts of Interest

The authors declare no potential conflicts of interest

Abstract word count: 200
Significance word count: 73
Text word count: 5901
Tables and Figures: 7
Supplementary Tables and Figures: 8

Abstract

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer related death worldwide. Its diagnosis at early stages would significantly improve the survival of CRC patients. The humoral immune response has been demonstrated useful for cancer diagnosis, predating clinical symptoms up to 3 years. Here, we employed an in-depth seroproteomic approach to identify proteins that elicit a humoral immune response in CRC patients. The seroproteomic approach relied on the immunoprecipitation with patient-derived autoantibodies of proteins from CRC cell lines with different metastatic properties followed by LC-MS/MS. After bioinformatics, we focused on 31 targets of CRC autoantibodies. After WB and IHC validation, ERP44 and TALDO1 showed potential to discriminate disease-free and metastatic CRC patients, and time to recurrence of CRC patients in stage II. Using plasma samples of 30 healthy individuals, 28 premalignant individuals, and 32 CRC patients, nine out of 13 selected targets for seroreactive analysis showed significant diagnostic ability to discriminate either CRC patients or premalignant subjects from controls. Our results suggest that the here defined panel of CRC autoantibodies and their target proteins should be included in CRC blood-based biomarker panels to get a clinically useful blood-based diagnostic signature for CRC detection.

1. Introduction

Colorectal cancer is the second cause of cancer death in Europe, mainly due to its late diagnosis. If CRC patients were early diagnosed, the percentage of treatable cured patients would greatly increase to 80-95% of patients. However, only around 37% of CRC patients are diagnosed at an early stage, leading to a drop in patient survival from 80-95% to 50 or even 6-10% for CRC diagnosis at stage III and IV respectively [1]. It is vital thus to improve the current scenario for early CRC detection consequently increasing patient survival. In this sense, evidence from several studies has shown that CRC screening is effective and economical in average-risk population [2]. Therefore, the development of accurate blood-based biomarker panels is mandatory for the early diagnosis of the disease.

The humoral immune response has been proven to play an important role in CRC. Indeed, numerous tumor-associated antigens (TAAs) target of autoantibodies produced by the immune system of CRC patients have been identified by protein microarrays-based proteomic techniques [3]. Since autoantibodies can be detected at early cancer stages, they can be exploited to increase the percentage of CRC patients early diagnosed [4]. However, new biomarkers and diagnostic platforms are needed to reach enough sensitivity and specificity to get a clinically useful blood-based diagnostic signature of CRC based on autoantibodies.

Up to date, CRC TAAs have been mainly identified through protein microarrays-based proteomic techniques such as natural and recombinant protein microarrays and phage microarrays [5, 6]. Such techniques are mainly hindered by the limited identification of dysregulated proteins to those printed onto the array. Unfortunately, the set of proteins that can be found in commercial microarrays encompasses only between 50% to 66% of all human proteins without their PTMs and proteoforms. Then, whether the missing proteins are potential CRC biomarkers cannot be evaluated.

In our aim to extend the number of identified CRC TAAs, we set up an immunomic and seroproteomic approach to find protein targets of CRC patients' autoantibodies (IgGs) using CRC cell lines protein extracts. Immunoprecipitated proteins bound to CRC IgGs were identified by LC-MS/MS using a Q-exactive and screened by bioinformatics to find actual

targets of CRC autoantibodies. A total of 13 out of 79 proteins identified as potential targets of CRC IgGs were selected for validation by: i) extensive meta-analysis to determine their potential dysregulation in CRC, ii) WB and IHC to determine if their potential dysregulation could be associated to prognosis of CRC patients, and iii) seroreactive analysis to determine their diagnostic effectiveness using a collection of 80 individual plasma samples of healthy individuals, premalignant subjects with low- and high-grade colorectal adenomas, and CRC patients. In this sense, it was demonstrated that 9 out of the 13 selected proteins for validation were actual targets of CRC IgGs, with significant diagnostic ability to discriminate either CRC patients at late stages or premalignant subjects from control individuals.

2. Material and Methods

2.1. Colorectal cancer cell culture and lysate production

SW480 and SW620 cell lines from the American Type Culture Collection cell repository, and KM12C and KM12SM cell lines from I. Fidler's laboratory (MD Anderson Cancer Center), were grown according to established protocols. For protein extract preparation, cells were lysed with RIPA buffer (Sigma) containing protease inhibitors.

2.2. Immunoprecipitation of protein targets of IgGs from CRC patients

Sera pools (100 µl) containing three and five sera from CRC stage III and IV, respectively (CRC pool), or eight sera from healthy individuals with negative colonoscopy (healthy pool) were used (Table 1; Supplementary Table S1). For IgG isolation, sera pools were separately incubated in 600 µl 20 mM sodium phosphate pH 7 with 200 µl of protein G agarose beads for 90 min at room temperature and gentle rotation. After incubation, protein G agarose beads were washed six times with 600 µl 20 mM sodium phosphate pH 7. Bound IgGs were then eluted with 300 µl 0.1 M glycine pH 2.7, and neutralized with 1 M Tris pH 8.8. Isolated IgGs were visualized by Coomasie Blue staining after SDS-PAGE, and its concentration measured.

Next, 200µg of purified IgGs from healthy and CRC pools were separately incubated during 2 h with 100 µl of protein G agarose beads in 600 µl of TBS buffer pH 7.4. Then, beads were washed with 200 mM methylamine pH 8.9 prior to the covalent binding of the IgGs to the agarose beads with 50 mM DMP for 1 h at room temperature at 25 rpm. Then, beads were washed with 600 µl of triethanolamine and blocked with 600 µl of ethanolamine, and non-covalently bound IgGs eluted with 400 µl of 0.1 M glycine in 2 M urea pH 2.7, and equilibrated afterwards with 100 µl of TBS buffer pH 7.4. Then, for immunoprecipitation of seroreactive proteins, 5 mg of CRC cells protein extracts were pre-cleaned using 20 µl of Protein G agarose beads previously equilibrated in RIPA buffer, and then incubated overnight with the healthy

IgGs covalently bound to beads. Next, the pre-cleaned protein extract was collected and the beads with IgGs from healthy individuals were washed ten times with 600 µl of washing buffer (TBS-0.2% triton). Beads were then centrifuged, and bound proteins eluted with 100 µl 0.1 M glycine pH 2.7 and subsequently neutralized with 20 µl of 2 M ammonium bicarbonate pH 8.8.

Finally, for the identification of proteins seroreactive to CRC patients IgGs, the pre-cleaned extract was incubated with the agarose beads covalently bound with the CRC patients IgGs, and proteins were eluted as above described. For protein identification by LC-MS/MS, 50% of the eluted content was concentrated onto a SDS-PAGE and in-gel digested with trypsin according to established protocols [7-11].

2.3. LC-MS/MS analysis

All peptide separations were carried out on an Easy-nLC 1000 nano system (Thermo Scientific). For each analysis, samples were loaded into a precolumn Acclaim PepMap 100 (Thermo Scientific) and eluted in a RSLC PepMap C18, 15 cm long, 50 µm inner diameter and 2 µm particle size (Thermo Scientific). The mobile phase flow rate was 300 nL/min using 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The gradient profile was set as follows: 0%-35% solvent B for 90 min, 35%-100% solvent B for 4 min, 100% solvent B for 8 min. Four microliters of each sample were injected.

MS analysis was performed using a Q Exactive mass spectrometer (Thermo Scientific). For ionization, 2000 V of liquid junction voltage and 270 °C capillary temperature were used. The full scan method employed a m/z 400-1500 mass selection, an Orbitrap resolution of 70,000 (at m/z 200), a target automatic gain control (AGC) value of 3e6, and maximum injection times of 100 ms. After the survey scan, the 15 most intense precursor ions were selected for MS/MS fragmentation. Fragmentation was performed with a normalized collision energy of 27 and MS/MS scans were acquired with a starting mass of m/z 100, AGC target was 2e5, resolution of 17,500 (at m/z 200), intensity threshold of 8e3, isolation window of 2 m/z units and maximum IT was 100 ms. Charge state screening was enabled to reject unassigned, singly charged, and

151 greater than or equal to seven protonated ions. A dynamic exclusion time of 20s was used to
152 discriminate against previously selected ions.
153
154 *2.4. MS data analysis*
155
156 MS data were analyzed with Proteome Discoverer (version 1.4.1.14) (Thermo Scientific)
157 using standardized workflows. Mass spectra *.raw files were searched against
158 SwissProt_2016_10.fasta, Homo sapiens (human) database (20121 sequences protein entries)
159 using Mascot (version 2.6.0, Matrix Science) search engine. Precursor and fragment mass
160 tolerance were set to 10 ppm and 0.02 Da, respectively, allowing 2 missed cleavages,
161 carbamidomethylation of cysteines as fixed modification, and methionine oxidation, Acetyl (N-
162 term) and Phosphorylation (S, T, Y) as variable modifications. Identified peptides were filtered
163 using Percolator algorithm with a q-value threshold of 0.01 [12] .
164
165 *2.5. Bioinformatics analysis*
166
167 Identified proteins appearing in more than 10% of agarose-based immunoprecipitation
168 experiments reported in CRAPome (Contaminant Repository for Affinity Purification Mass
169 Spectrometry Data) data base were considered false positives [13]. Bioinformatics analysis of
170 proteins passing this criterion was performed using STRING and IPA Analysis (Ingenuity
171 Systems) to identify altered networks and pathways [14]. STRING Version 9.1 and MCL
172 clustering enrichment 2 with the default 0.4 confidence score were used to identify the
173 interacting partners in the dataset.
174 To examine whether protein or genetic alterations associated to CRC had been previously
175 associated with identified proteins out of CRAPome, we performed extensive meta-analysis
176 using the Oncomine [15], UALCAN [16], and cBioPortal [17] databases, which give
177 information on differential protein expression, heatmaps of selected dysregulated genes, and
178 identified genetic alterations in tumor tissue samples, respectively. To evaluate whether the

identified proteins had been previously associated with disease prognosis, information was retrieved from the Human Protein Atlas [18, 19].

2.6. Gateway plasmid construction, gene cloning, DNA preparation and protein expression

Sequence-verified, full-length cDNA plasmid containing selected targets for validation (Supplementary Table S2) in flexible pDONR221 or pENTR223 vectors was obtained from the publicly available DNASU Plasmid Repository [20]. The ORFs were transferred by LR clonase reactions (Invitrogen, Carlsbad, CA) to the pANT7_cHalo expression vector for *in vitro* protein expression to get the autoantigens expressed as fusion proteins to HaloTag in the C-terminal [21-23]. As controls, we also obtained the expression vectors for producing HaloTag protein. All donor and expression plasmids were sequence verified prior to a subsequent use.

Proteins were expressed using the 1-Step Human Coupled IVT Kit HeLa cell lysate (Thermo Fisher Scientific, Waltham, MA) as per manufacturer's recommendations.

2.7. RNA Extraction and semi-quantitative PCR

RNA was extracted from $5x10^7$ cells with the RNeasy Mini Kit (Qiagen) and quantified with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific). cDNA was synthesized using the NZY First-Strand cDNA Synthesis Kit, separate oligos (NZYTech). For semi-quantitative reverse transcriptase-PCR (RT-PCR), reactions were performed using specific primers (Supplementary Table S3) and the Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific), according to the manufacturer recommendations. PCR products were separated on 1.5% agarose gel containing GelRed (Biotium). GAPDH was used as internal control.

204 *2.8. SDS-PAGE and WB analysis*

205

206 SDS-PAGE and WB analyses were performed as previously reported [24]. Alternatively,
207 0.67 µl of the *in vitro* protein extracts or 10µg of protein extracts were run in 10% SDS-PAGE
208 and transferred to nitrocellulose membranes (Hybond-C extra). After blocking, membranes were
209 incubated with optimized dilutions of indicated specific monoclonal or polyclonal antibodies
210 (Supplementary Table S4). Immunodetection on membranes was achieved by using HRP-
211 conjugated secondary antibodies (Supplementary Table S4). Chemiluminescence signal was
212 developed with ECL Western Blotting Substrate (Thermo Scientific) and detected on an
213 Amersham Imager 680 (GE Healthcare).

214

215 *2.9. Patient plasma*

216

217 The Institutional Ethical Review Boards of the Instituto de Salud Carlos III, Hospital Clínico
218 San Carlos (Madrid), and La Paz Hospital (Madrid) approved this study on biomarker discovery
219 and validation (CEI PI 45). Plasma samples were obtained from the biobanks of the Hospital
220 Clínico San Carlos (IdISSC) and La Paz Hospital (IdIPAZ), which belongs to the National
221 Biobank Net (ISCIII) cofounded with FEDER funds, after approval of the Ethical Review
222 Boards of these institutions. Written informed consent was obtained from all patients. Plasma
223 samples were collected using a standardized sample collection protocol and stored at -80°C until
224 use [4, 25-27].
225 For seroreactive analysis, 32 plasma samples from CRC patients, 18 from premalignant
226 subjects (low- or high-grade adenomas in the colon), and 30 from control individuals
227 (individuals asymptomatic and negative colonoscopy individuals, or individuals with fecal
228 occult blood test -FOBT- positive and negative colonoscopy) were used (Table 1;
229 Supplementary Table S1).

2.10. Seroreactive analysis

HaloTag fusion proteins coupled to Magne HaloTag beads (MBs, Promega) and ELISA-like immunoassay was previously optimized to measure autoantibodies [11, 22]. Briefly, for each reaction, a relation of 0.67 µL of the *in vitro* protein expression was incubated with 0.5 µL of MBs suspension overnight at 4ºC for protein capture to MBs according to manufacturer instructions. For covalent binding, a mixture of the required amount of protein and MBs was made taking into account the number of replicates and measurements to be performed. After three washing steps with PBS, Triton X-100 0.05%, Tween 20 0.1% rocking for 5 min at 250 rpm, unbound proteins were removed adding 100 µL of glycine 0.1 M pH 2.7 for 5 min at RT and 250 rpm shaking. Then, HaloTag fusion proteins immobilized in MBs were blocked with SuperBlock for 1h and then transferred onto separate wells of Bio-Plex 96-well plates to perform subsequent incubations overnight with pooled or individual plasma samples at indicated dilutions in PBS supplemented with Tween 20 0.1% (v/v) and BSA 3% (w/v). After extensive washing as above, HRP-conjugated anti-human IgG antibody (Jackson) diluted 1:10000 in PBS supplemented with Tween 20 0.1% (v/v) and BSA 3% (w/v) was incubated to detect the presence of autoantibodies. Alternatively, to verify the TAA-HaloTag immobilization, HaloTag was detected with the monoclonal anti-HaloTag antibody (Promega) diluted 1:1000, followed by 1h of incubation with HRP- conjugated anti-mouse IgG diluted 1:2500. To develop the chemiluminescence signal, MBs were incubated on black Maxisorp 96-well plates (Nunc) with 50 µL of SuperSignal ELISA Pico Maximum Sensitivity Substrate (Pierce) for 1 min. Then, signal was recorded on The Spark multimode microplate reader (Tecan Trading AG).

2.11. Tissue Microarrays and immunohistochemistry

Core tumor tissue samples from 50 metastatic and non-metastatic CRC patients and core tumor tissue samples from 94 recurrent or non-recurrent CRC stage II patients composed the

two TMAs used in the study. Immunohistochemical staining using optimized antibody dilutions (Supplementary Table S4), visualization and immunoreactivity was conducted according to established protocols [4, 14, 25, 27].

2.12. Statistical analysis

Statistical analyses were performed with Microsoft Office Excel. For the seroreactive and immunohistochemistry analysis, data distribution using the Shapiro-Wilk test and variance homogeneity using the Bartlett test was first evaluated. Since data normality was discarded in all cases, we then assessed whether the means of control individuals, and premalignant individuals and/or CRC patients were statistically different from each other using the non-parametric U-Mann Whitney test assuming unequal variances. p-values <0.05 were considered statistically significant. The diagnostic capacity of each individual protein as well as their combination was evaluated by a receiver operating characteristic (ROC) curve. ROC curves and the corresponding area under the curve (AUC) and the maximized sensitivity and specificity values were obtained with the R program (version 3.2.3) using the R package Epi [28].

3. Results

In the present study, we aimed at developing a methodological approach to identify tumor-associated antigens through immunoprecipitation of potential autoantibody targets using IgGs from CRC patients at stages III and IV and healthy individuals. The isolated IgGs were separately incubated with protein extracts from CRC isogenic cell lines with different metastatic properties and, then, immunoprecipitated and subsequently identified by mass spectrometry. SW480 and KM12C CRC cells were used as model of primary tumor and SW620 and KM12SM cells as model of metastatic CRC to lymph nodes and liver, respectively. A schematic representation of the workflow of the study is depicted (Fig. 1).

3.1. Identification of seroreactive proteins to IgGs from CRC patients by immunoprecipitation followed by LC-MS/MS

For IgG isolation, plasma samples from healthy individuals with negative colonoscopy and unknown pathology (control group), and plasma samples from CRC patients at stages III and IV obtained before surgery (CRC group) were separately pooled (Table 1).

The analyzed protein extract (10 mg) was obtained by mixing equal amounts of cell lysates from KM12C, KM12SM, SW480, and SW620 cells. The IgGs obtained from healthy individuals were used to pre-clean the protein extract, which was then incubated with IgGs from CRC patients to identify specific CRC seroreactive proteins.

Both protein fractions seroreactive to IgGs from healthy and/or CRC patients were firstly analyzed by SDS-PAGE followed by silver staining, and then, by mass spectrometry for the identification of CRC TAAs.

First, by SDS-PAGE and silver-staining, a differential staining at high- and low-molecular weight was observed when comparing protein bands from the eluted seroreactive proteins to IgGs from healthy and CRC patients (Fig. 1). Then, the eluted proteins from both IgGs were concentrated in parallel by SDS-PAGE and in-gel digested with trypsin for subsequent mass

spectrometry analysis (Fig. 1). In total, 442 proteins were identified in both experiments. From the set of identified proteins, 203 were found to be common to both control and CRC conditions, while 160 and 79 were identified only in the control or in the CRC immunoprecipitation, respectively (Supplementary Table S5).

3.2. Bioinformatics of the seroreactive proteins to CRC IgGs

Next, the set of CRC specific proteins was subjected to an extensive meta-analysis with different bioinformatic tools to identify those that were more likely to be actual CRC autoantibody targets, and thus potential disease biomarkers.

First, CRAPome was used to remove false positive seroreactive proteins target of autoantibodies. Using a cut-off of <10% of appearance in recorded mass spectrometry analysis of immunoprecipitation experiments using agarose beads, 31 proteins were found more prone to be actual CRC autoantibody targets (Table 2). Remarkably, the well-documented cancer autoantibody p53 appeared among the 31 proteins [29], suggesting that our approach would be useful for the identification of CRC-specific TAAs.

It has been reported that proteins target of autoantibodies are usually overexpressed and/or mutated. Therefore, we used the cBioPortal, Oncomine and UALCAN databases to investigate whether alterations related to these proteins at a genetic and/or protein level had been previously associated to CRC. We searched for recorded genetic alterations in the TGCA CRC database in cBioPortal. Interestingly, 69% and 80% of the identified targets had genetic alterations (mutation, amplification, deep deletion, and other alteration) in colon and rectal adenocarcinomas, respectively (Fig. 2A). We then used the Oncomine and UALCAN databases to find any dysregulation in the expression of the proteins at tissue level in CRC. We found that 22 out of the 31 identified proteins showed a statistically significant altered expression ($p<0.05$) in CRC in comparison to adjacent paired healthy tissue (Fig. 2B), with 17 of them showing upregulation in colorectal tumoral tissue (Fig. 2B). Moreover, one of the identified proteins,

CTTNBP2NL, was found to be a CRC prognosis marker in the Pathology Atlas from the Human Protein Atlas (Fig. 2C).

Finally, String database and IPA were used to identify possible altered cell functions and molecular pathways. With String, proteins were clustered by direct and indirect interactions, obtaining 5 interacting groups composed of proteins related to i) GTPases and actin-interacting proteins; ii) transcription, immunity and cell cycle functions; iii) pentose pathway and endoplasmic reticulum; iv) respiratory chain and caspase activity; and v) respiratory chain and cell messengers (Supplementary Fig. 1A). These functions, related to immune response, actin interaction and cell death, have been closely associated to cancer. Among the networks found to be altered with IPA, some were related to highly altered processes in cancer such as cell cycle, DNA replication, recombination, protein synthesis and carbohydrate metabolism (Supplementary Fig. 1B).

Collectively, these data show that the identified proteins that are potential targets of CRC autoantibodies are dysregulated at a genetic and/or protein level in CRC. Furthermore, their dysregulation plays an important role in the alteration of molecular pathways and cellular functions related to the pathology.

3.3. Genetic and protein expression verification in CRC cell lines

From the 31 remaining proteins after CRAPome screening, 13 were selected for validation at mRNA and protein level (Table 2). The criteria for this selection were antibody and cDNA availability together with existing information (i.e. whether they had already been described as TAAs in other cancer types). Plasma or secreted proteins were discarded for further validation.

We verified the presence of the targets either at mRNA or protein level in the cell lines used in the study. The presence of all 13 selected targets for validation was confirmed at mRNA level (Supplementary Fig. 2A). ERP44, SCGD1B2, TALDO1, and TRIM29 were also found at the protein level (Supplementary Fig. 2B).

3.4. Evaluation of protein expression dysregulation in CRC tissue samples as biomarker prognostic assessment

Next, since TAAs are usually highly dysregulated at protein level in CRC tissue, we evaluated if those proteins validated at protein level in CRC cells might show any dysregulation in tissue and whether they show any diagnostic or prognostic ability. To this end, ERP44, SCGD1B2, TALDO1, and TRIM29 protein abundance was evaluated by IHC using two different TMAs related to metastasis or recurrence of CRC. All proteins showed staining in CRC tissue in both TMAs, indicating a clear presence of the proteins in the pathological tissue samples. TALDO1 showed nuclear staining, while ERP44, SCGD1B2 and TRIM29 showed cytoplasmic staining (Fig. 3A). Moreover, both epithelial and stromal staining was observed for TALDO1, ERP44, and SCGDIB2, while TRIM29 showed only epithelial staining.

Interestingly, ERP44 could statistically discriminate between CRC liver metastatic patients according to their survival status ($p = 0.049$) (Fig. 3B). In addition, both TALDO1 and ERP44 could significantly discriminate between stage II CRC patients according to their recurrence status ($p=0.024$) or according to the time to recurrence ($p=0.039$), respectively (Fig. 3B). Next, to evaluate whether these targets could be useful markers of the disease at protein level, ROC curves were calculated. For the metastatic TMA, when grouping patients according to survival status ERP44 showed an AUC of 70% (40% sensitivity, 100% specificity). For the stage II CRC TMA, when grouping patients according to recurrence status TALDO1 showed an AUC of 63.4% (56.9% sensitivity, 66.7% specificity). Finally, when grouping patients according to time to recurrence, the AUC of ERP44 was 61.4% (59.1% sensitivity, and 65.9% specificity) (Fig. 3C).

Collectively, these results highlight that the dysregulation of proteins seroreactive to IgGs from CRC patients show potential to discriminate metastatic CRC patients, recurrence and time to recurrence of CRC patients in stage II.

383 *3.5. Cloning, in vitro protein expression and seroreactive analysis of the potential CRC TAAs*

384

385 Our research hypothesis relied on the identification of proteins target of autoantibodies in
386 CRC with diagnostic ability. To address this question, we evaluated the presence of cancer
387 autoantibodies in plasma samples from CRC patients, premalignant subjects, and control
388 individuals, and determine their diagnostic ability.

389 To this end, the 13 selected targets were cloned in the pANT7_cHalo expression vector for
390 their *in vitro* expression as HaloTag fusion proteins [22]. Then, we confirmed the *in vitro*
391 expression of the TAAs as fusion proteins to HaloTag, and the correct covalent immobilization
392 of the fusion proteins in MBs (Fig. 4A). All proteins were visualized by immunostaining after
393 SDS-PAGE at the theoretical molecular mass of the fusion protein. In addition, all of them were
394 able to covalently bind to chloroalkane functionalized MBs (data not shown), and thus, they
395 should be able to be recognized by CRC autoantibodies in an ELISA-like test [11, 22, 30].

396 Next, we analyzed the seroreactivity to these 13 potential TAAs using 80 individual plasma
397 samples (30 from control individuals, 28 from premalignant subjects, and 32 from stage III and
398 IV CRC patients). We found that CRC patients possess significantly higher levels of
399 autoantibodies than control individuals, whose levels were always lower or undetectable (Fig.
400 4B). However, only autoantibodies against CHCHD3, CTTNBP2NL, FKBP4, MGST3,
401 THSD7A, TRIM29 -as well as p53- showed a statistically significant difference (p-values
402 <0.01) comparing the asymptomatic healthy control group and the CRC group (Fig. 4C). In
403 addition CTTNBP2NL, FKBP4, and TALDO1 autoantibody levels were able to significantly
404 discriminate between asymptomatic healthy controls and premalignant subjects (p-value < 0.01)
405 (Fig. 4D). Finally, ERP44, TALDO1, CHCHD3, CTTNBP2NL, FKBP4, MGST3, THSD7A,
406 and TRIM29 -as well as p53- were able to discriminate between the asymptomatic healthy
407 individuals and the colorectal pathological group (p-value<0.04 (Fig. 4E).

3.6. Evaluation of the diagnostic value of CRC autoantibodies

Finally, the diagnostic value of the CRC TAAs was evaluated by means of ROC curves. AUC, sensitivity and specificity values were calculated individually or in combination using those autoantibodies showing significant levels in above comparisons.

CHCHD3, CTTNBP2NL, FKBP4, MGST3, THSD7A, and TRIM29 showed an individual AUC higher than 70% to discriminate between CRC patients and control individuals (Supplementary Table S6). In combination, the AUC increased to 99.7%, with a sensitivity of 96.9% and specificity of 100% (Fig. 5A). In addition, autoantibodies against CTTNBP2NL, FKBP4, and TALDO1 could discriminate between control individuals and premalignant subjects with an individual AUC higher than 71.9% (Supplementary Table S6), and in combination the AUC increased to 88.9% with a sensitivity of 83.3% and specificity of 86.7% (Fig. 5B). Finally, CHCHD3, CTTNBP2NL, ERP44, FKBP4, MGST3, TALDO1, THSD7A, and TRIM29 showed an individual AUC higher than 59.2% to discriminate between control individuals and the pathological group, and their combination with p53 led to a combined AUC of 95.3%, with a sensitivity of 92.0% and a specificity of 90.0% (Fig. 5C).

Collectively, these results highlight the usefulness of the followed approach to identify TAAs with significant diagnostic ability. In addition, our results suggest that the here defined CRC TAAs should be included in a CRC blood-based biomarker panel to get a clinically useful blood-based diagnostic signature for CRC detection.

4. Discussion

Protein microarrays are a highly renowned used technique for identifying protein-protein interactions, drug and small-molecule screening, epitope mapping, host-pathogen interactions … [31-34]. However, protein microarrays have been mainly used for the identification of tumor-associated antigens in cancer malignancies [31-35]. Although theoretically the whole proteome might be printed in protein microarrays, the highest density achieved so far is about 75% of the human proteome -considering one ORF per protein but without multiple proteoforms of the same protein- [36-38]. Alternatively, high-density PrESTs arrays covering approximately the whole proteome has been produced [38]. However, in this case medium to large peptides of about 90% of all human ORF proteins are printed in the arrays, with the limitation that in these arrays 3D non-linear relevant epitopes, and thus relevant TAAs, might be lost during the screening.

Therefore, although protein microarrays have been demonstrated highly useful in identifying CRC autoantibodies and their target proteins, new methodologies need to be applied to achieve the completion in the identification of TAAs. Here, we investigated whether a modified SASI methodology without SILAC labeling of cells based on mass spectrometry [39], which have been useful for identifying targets of IgGs in vaccinated pancreatic cancer patients with a granulocyte-macrophage colony stimulating factor (GM-CSF) secreting pancreatic cancer vaccine (GVAX) in phase II clinical trials [39], might help in elucidating new CRC TAAs using IgGs isolated from the plasma of CRC patients and protein extracts from CRC cell lines.

Previous studies applying mass spectrometry approaches have permitted the identification of autoantibodies against TAAs in different types of cancer, including CRC [40-42]. However, these studies have been mainly focused on combining immunoblotting of protein extracts separated by 2DE-gels using patients' sera for identification of reactive spots and their subsequent protein identification by mass spectrometry, limiting the identification to medium or highly abundant TAAs. Here, the SASI modified approach combining immunoprecipitation directly coupled to LC-MS/MS allowed for the identification of autoantibodies and their target

proteins in colorectal cancer. The use of immunoprecipitation to capture the reactive TAAs in solution having the proteins their *a priori* 3D conformation instead of blotting them after their separation by 2DE-gel confers some advantages. Mainly, there is no limitation on protein isolation by immunoprecipitation for further identification *via* LC-MS/MS. All reactive proteins to the immobilized antibodies are bound independently on their concentration, while 2DE-MS is not useful to identify low abundant proteins. Moreover, reactive spots proceeding from 2DE have conformational limitations since they are bound onto membranes and discontinuous reactive epitopes are lost. Since previous studies have revealed that the immune response against proteins, apart from being patient specific, is highly polyclonal and mainly produced against 3D discontinuous epitopes [23, 35]. This last limitation could produce the misidentification of relevant TAAs, which could be overcome by immunoprecipitation. Here, among the identified proteins as potential autoantibody targets in CRC (Table 2), we observed the presence of at least one protein containing continuous and discontinuous epitopes -p53-, as it has been previously demonstrated by ELISA, WB and protein microarrays [23, 29, 43], and thus supporting our approach to identify TAAs avoiding conformational limitations.

Here, we have validated our protein dataset by meta-analysis of databases related to CRC. Twenty-two out of 31 identified proteins were observed to present a genetic alteration or altered expression at mRNA or protein level in the disease. In addition, by IHC TALDO1, and ERP44 protein abundance in CRC tissue could discriminate between metastatic and non-metastatic patients and between recurrent and non-recurrent patients, or according to the time to recurrence. These alterations have been defined as the ones producing a humoral immune response in cancer patients to overproduce autoantibodies against them [35].

Of our identified validated dataset, several markers have been found to play a potential role in cancer. CHCHD3, a component of the mitochondrial inner membrane complex MICOS, plays an important role in the maintenance of its stability and in mitochondrial morphological stability [44]. To our knowledge, there is only one study that has linked its dysregulation with early cancer diagnosis [45], but in pancreatic cancer. Regarding CTTNBP2NL, it has been shown dysregulated in a genetic study of oral squamous cell carcinoma [46]. ERP44, a protein

484 that reportedly regulates calcium release into the endoplasmic reticulum, has shown that it
485 modulates lung cancer cell migration [47], and that its knock-down inhibited mammosphere
486 growth of breast cancer cells [48]. On the other hand, MGST3, TALDO1, TRIM29, and
487 THSD7A had been previously reported as potential cancer markers in different studies by the
488 evaluation of their expression levels. However, no studies focusing on them as target of
489 autoantibodies and/or their potential use as blood-based biomarkers for cancer diagnosis have
490 been described. Only autoantibodies against THSD7A have been previously described to
491 classify membranous nephropathy and to be associated to outcome and treatment response [49].
492 MGST3, a protein with antioxidant functions by reducing GSH, has been reported as biomarker
493 for oral squamous cancer [50]. It is highly up-regulated in basal-like breast cancer cell lines, and
494 thus, makes it a possible triple-negative breast cancer biomarker [51]. TALDO1, an important
495 protein for the balance of metabolites in the pentose pathway, has been previously identified by
496 proteomic techniques to be a potential biomarker in non-small cell lung cancer [52]. In addition,
497 it has also been identified to be upregulated in pancreatic cancer [53]. However, there is no
498 information on its possible role in CRC. On the other hand, TRIM29, which plays an important
499 role on the regulation of macrophage activation against bacterial and viral infections within the
500 respiratory tract, has been associated to multiple cancer types [54, 55]. Its upregulation has been
501 described to promote cell proliferation, poor prognosis, and aggressiveness in CRC [56].
502 THSD7A is involved in vascular invasion, metastasis and angiogenesis and plays a role in actin
503 cytoskeleton rearrangement [57]. Lastly, the immunophilin protein FKBP4 has been the only
504 protein identified in this study previously reported in other studies as CRC serum antigen [58,
505 59].

506 One of the goals of the study consisted of the discovery of TAAs useful for CRC detection,
507 which might be used as an effective and efficient strategy for the blood-based CRC screening
508 for population at risk (>50 years old). We used pooled samples from CRC patients to avoid
509 potential aberrations and biological variations in the humoral immune response of CRC patients
510 appearing in individual samples, which might increase the capacity to identify the most

significant and consistent proteins target of autoantibodies between pooled samples from well-characterized CRC patients in comparison to control individuals. Interestingly, although most of the identified CRC TAAs have been described as dysregulated in other cancers, there are no studies focusing on them as target of autoantibodies (except for FKBP4) and/or their potential as blood-based biomarkers for cancer diagnosis, including colorectal cancer. Therefore, we could achieve the specific CRC detection at early cancer stages by a simple blood test, avoiding at the same time the detection of other cancer malignancies by integrating them in a multiplexed panel including (if needed) other CRC-specific TAAs targets of autoantibodies to increase as much as possible the specificity and sensitivity of the test, especially for the detection of premalignant individuals.

Regarding the comparison of the here described panel for CRC detection with that used in clinics, it is worth to remark that there is no useful blood-based colorectal cancer marker for diagnosis and that the actual clinical marker is only recommended for recurrence (CEA). Moreover, the recommended population screening strategies for CRC are based on stool test (FOBT) followed, if positive, by structural examination by colonoscopy. While structural exams detect both cancer and premalignant lesions, stool tests mainly identify cancer because of their limited sensitivity to detect low- or high-grade adenomas. Moreover, the relatively low specificity of the initial examination results by means of FOBT produces a high false-positive rate and leads to a significant number of unnecessary colonoscopies. Here, with the identified TAA panel for CRC detection, AUC values of 99.7% and 88.9% were achieved for the detection of CRC patients and premalignant subjects in comparison to healthy control individuals, respectively. Regarding the clinical detection by FOBT followed by colonoscopy, usual clinical AUC values are about 70% for CRC detection and 60% for premalignant subjects. In addition, in a recent report performing FOBT and measuring at the same time two miRNAs -not yet clinically established-, AUC values for detecting CRC and premalignant subjects increase to 93% and 64% [60], respectively. Remarkably, in both cases, our approach was able to improve these results. Therefore, the here described approach might serve for a more accurate

detection of those patients that should undergo colonoscopy or surgery for the resection of colorectal cancer or premalignant lesions.

On the other hand, although it has been provided the unbiased discovery of autoantibody biomarkers validated using individual plasma samples from CRC patients and colorectal premalignant subject in comparison to control individuals, their diagnostic ability should be further validated using large number of samples and independent patient cohorts, including blinded samples and samples from other cancer types to assure their CRC specificity and establish their utility for the diagnosis of the disease. In this sense, further research is guaranteed to try to improve the results for the early diagnosis of the disease by multiplexing the here described TAAs and by implementing these biomarkers into a panel containing other previously validated CRC autoantigens in multiplexed bio(sensing) platforms (i.e., biosensors [22], Luminex or protein microarrays) to reach enough sensitivity, specificity, and AUC for the blood-based clinical diagnosis of the disease.

5. Conclusions

In summary, in this work we propose a CRC blood-based diagnostic signature based on the identification of non-previously defined CRC TAAs by immunoprecipitation of proteins from CRC cell lines and IgGs isolated from plasma of CRC patients followed by LC-MS/MS analysis. These candidate TAAs should be part of a multiplex biomarker panel for the blood-based diagnosis of the disease to be further analyzed with samples obtained from longitudinal studies in subjects at the preclinical or early clinical stages of the disease to assure its usefulness for early CRC detection.

Supplementary data to this article can be found online.

Acknowledgments

This work was supported by the Ramon y Cajal programme of the MINECO and the financial support of the PI17CIII/00045 grant from the AES-ISCIII program to R.B.. M.G-A. was supported by a contract of the Programa Operativo de Empleo Juvenil y la Iniciativa de Empleo Juvenil (YEI) with the participation of the Consejería de Educación, Juventud y Deporte de la Comunidad de Madrid y del Fondo Social Europeo. The FPU predoctoral contracts to P.S.S.-A. and A.M-C. are supported by the Spanish Ministerio de Educación, Cultura y Deporte. G.S-F. is recipient of a predoctoral contract (grant number 1193818N) supported by The Flanders Research Foundation (FWO).

Authors' Contributions

Conception and design: R.B. Development of methodology: M.G-A., P. SS-A., A.G-A., G.S-F., R.B. Perform Research: C.P., MJ.F-A., J. M-U., A. M-C., G. S-F., N.R., MA.C., G. S-F., V.R., A.P-G. Analysis and interpretation of data: M.G-A., A.G-A., C.P., MJ.F-A., J. M-U., V.R., A.P-G., R.B. Writing, review, and/or revision of the manuscript: M.G-A., A.G-A., R.B. Technical, obtaining and processing of samples, or material support: C.P., MJ.F-A., A. M-C., MA.C., N.R., A.M-C., A.G-A., R.B.

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

[1] Holleczek B, Rossi S, Domenic A, Innos K, Minicozzi P, Francisci S, et al. On-going improvement and persistent differences in the survival for patients with colon and rectum cancer across Europe 1999-2007 - Results from the EUROCARE-5 study. Eur J Cancer. 2015;51:2158-68.

[2] Steele RJ, Rey JF, Lambert R, International Agency for Research on C. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Professional requirements and training. Endoscopy. 2012;44 Suppl 3:SE106-15.

[3] Casal JI, Barderas R. Identification of cancer autoantigens in serum: toward diagnostic/prognostic testing? Mol Diagn Ther. 2010;14:149-54.

[4] Barderas R, Villar-Vazquez R, Fernandez-Acenero MJ, Babel I, Pelaez-Garcia A, Torres S, et al. Sporadic colon cancer murine models demonstrate the value of autoantibody detection for preclinical cancer diagnosis. Sci Rep. 2013;3:2938.

[5] Barderas R, Babel I, Casal JI. Colorectal cancer proteomics, molecular characterization and biomarker discovery. Proteomics Clin Appl. 2010;4:159-78.

[6] Barderas R, Villar-Vazquez R, Casal JI. Colorectal Cancer Circulating Biomarkers. Biomarkers in Cancer Springer Series. 2014;1:1-21.

[7] Barderas R, Mendes M, Torres S, Bartolome RA, Lopez-Lucendo M, Villar-Vazquez R, et al. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. Mol Cell Proteomics. 2013;12:1602-20.

[8] Mendes M, Pelaez-Garcia A, Lopez-Lucendo M, Bartolome RA, Calvino E, Barderas R, et al. Mapping the Spatial Proteome of Metastatic Cells in Colorectal Cancer. Proteomics. 2017;17.

[9] Pelaez-Garcia A, Barderas R, Batlle R, Vinas-Castells R, Bartolome RA, Torres S, et al. A proteomic analysis reveals that Snail regulates the expression of the nuclear orphan receptor Nuclear Receptor Subfamily 2 Group F Member 6 (Nr2f6) and interleukin 17 (IL-17) to inhibit adipocyte differentiation. Mol Cell Proteomics. 2015;14:303-15.

[10] Pelaez-Garcia A, Barderas R, Mendes M, Lopez-Lucendo M, Sanchez JC, Garcia de Herreros A, et al. Data from proteomic characterization of the role of Snail1 in murine mesenchymal stem cells and 3T3-L1 fibroblasts differentiation. Data Brief. 2015;4:606-13.

[11] San Segundo-Acosta P, Montero-Calle A, Fuentes M, Rabano A, Villalba M, Barderas R. Identification of Alzheimer's Disease Autoantibodies and Their Target Biomarkers by Phage Microarrays. J Proteome Res. 2019.

[12] Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 2007;4:923-5.

[13] Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods. 2013;10:730-6.

[14] Garranzo-Asensio M, San Segundo-Acosta P, Martinez-Useros J, Montero-Calle A, Fernandez-Acenero MJ, Haggmark-Manberg A, et al. Identification of prefrontal cortex protein alterations in Alzheimer's disease. Oncotarget. 2018;9:10847-67.

[15] Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004;6:1-6.

[16] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017;19:649-58.

[17] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401-4.

[18] Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010;28:1248-50.

[19] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347:1260419.

[20] Seiler CY, Park JG, Sharma A, Hunter P, Surapaneni P, Sedillo C, et al. DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. Nucleic Acids Res. 2014;42:D1253-60.
[21] Ramachandran N, Raphael JV, Hainsworth E, Demirkan G, Fuentes MG, Rolfs A, et al. Next-generation high-density self-assembling functional protein arrays. Nat Methods. 2008;5:535-8.
[22] Garranzo-Asensio M, Guzman-Aranguez A, Poves C, Fernandez-Acenero MJ, Torrente-Rodriguez RM, Ruiz-Valdepenas Montiel V, et al. Toward Liquid Biopsy: Determination of the Humoral Immune Response in Cancer Patients Using HaloTag Fusion Protein-Modified Electrochemical Bioplatforms. Anal Chem. 2016;88:12339-45.
[23] Katchman BA, Barderas R, Alam R, Chowell D, Field MS, Esserman LJ, et al. Proteomic mapping of p53 immunogenicity in pancreatic, ovarian, and breast cancers. Proteomics Clin Appl. 2016;10:720-31.
[24] Barderas R, Shochat S, Timmerman P, Hollestelle MJ, Martinez-Torrecuadrada JL, Hoppener JW, et al. Designing antibodies for the inhibition of gastrin activity in tumoral cell lines. Int J Cancer. 2008;122:2351-9.
[25] Babel I, Barderas R, Diaz-Uriarte R, Martinez-Torrecuadrada JL, Sanchez-Carbayo M, Casal JI. Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. Mol Cell Proteomics. 2009;8:2382-95.
[26] Babel I, Barderas R, Diaz-Uriarte R, Moreno V, Suarez A, Fernandez-Acenero MJ, 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. 2011;10:M110 001784.
[27] Barderas R, Babel I, Diaz-Uriarte R, Moreno V, Suarez A, Bonilla F, 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. 2012;75:4647-55.
[28] Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.
[29] Lubin R, Schlichtholz B, Bengoufa D, Zalcman G, Tredaniel J, Hirsch A, et al. Analysis of p53 antibodies in patients with various cancers define B-cell epitopes of human p53: distribution on primary structure and exposure on protein surface. Cancer Res. 1993;53:5872-6.
[30] Montero-Calle A, San Segundo-Acosta P, Garranzo-Asensio M, Rabano A, Barderas R. The Molecular Misreading of APP and UBB Induces a Humoral Immune Response in Alzheimer's Disease Patients with Diagnostic Ability. Mol Neurobiol. 2019.
[31] Yu X, Petritis B, Duan H, Xu D, LaBaer J. Advances in cell-free protein array methods. Expert Rev Proteomics. 2018;15:1-11.
[32] Cretich M, Damin F, Chiari M. Protein microarray technology: how far off is routine diagnostics? Analyst. 2014;139:528-42.
[33] Atak A, Mukherjee S, Jain R, Gupta S, Singh VA, Gahoi N, et al. Protein microarray applications: Autoantibody detection and posttranslational modification. Proteomics. 2016;16:2557-69.
[34] Duarte JG, Blackburn JM. Advances in the development of human protein microarrays. Expert Rev Proteomics. 2017;14:627-41.
[35] Anderson KS, LaBaer J. The sentinel within: exploiting the immune system for cancer biomarkers. J Proteome Res. 2005;4:1123-33.
[36] Gupta S, Mukherjee S, Syed P, Pandala NG, Choudhary S, Singh VA, et al. Evaluation of autoantibody signatures in meningioma patients using human proteome arrays. Oncotarget. 2017;8:58443-56.
[37] Syed P, Gupta S, Choudhary S, Pandala NG, Atak A, Richharia A, et al. Autoantibody Profiling of Glioma Serum Samples to Identify Biomarkers Using Human Proteome Arrays. Sci Rep. 2015;5:13895.
[38] Sjoberg R, Mattsson C, Andersson E, Hellstrom C, Uhlen M, Schwenk JM, et al. Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. N Biotechnol. 2016;33:582-92.

[39] Jhaveri DT, Kim MS, Thompson ED, Huang L, Sharma R, Klein AP, et al. Using Quantitative Seroproteomics to Identify Antibody Biomarkers in Pancreatic Cancer. Cancer Immunol Res. 2016;4:225-33.
[40] Lin L, Zheng J, Yu Q, Chen W, Xing J, Chen C, et al. High throughput and accurate serum proteome profiling by integrated sample preparation technology and single-run data independent mass spectrometry analysis. J Proteomics. 2018;174:9-16.
[41] Liu XX, Ye H, Wang P, Zhang Y, Zhang JY. Identification of 1433zeta as a potential biomarker in gastric cancer by proteomicsbased analysis. Mol Med Rep. 2017;16:7759-65.
[42] Rezaei M, Nikeghbalian S, Mojtahedi Z, Ghaderi A. Identification of antibody reactive proteins in pancreatic cancer using 2D immunoblotting and mass spectrometry. Oncol Rep. 2018;39:2413-21.
[43] Garranzo-Asensio M, Guzman-Aranguez A, Poves C, Fernandez-Acenero MJ, Montero-Calle A, Ceron MA, et al. The specific seroreactivity to Np73 isoforms shows higher diagnostic ability in colorectal cancer patients than the canonical p73 protein. Sci Rep. 2019;9:13547.
[44] Ott C, Dorsch E, Fraunholz M, Straub S, Kozjak-Pavlovic V. Detailed analysis of the human mitochondrial contact site complex indicate a hierarchy of subunits. PLoS One. 2015;10:e0120213.
[45] Ning L, Pan B, Zhao YP, Liao Q, Zhang TP, Chen G, et al. [Immuno-proteomic screening of human pancreatic cancer associated membrane antigens for early diagnosis]. Zhonghua Wai Ke Za Zhi. 2007;45:34-8.
[46] Cha JD, Kim HJ, Cha IH. Genetic alterations in oral squamous cell carcinoma progression detected by combining array-based comparative genomic hybridization and multiplex ligation-dependent probe amplification. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2011;111:594-607.
[47] Huang X, Jin M, Chen YX, Wang J, Zhai K, Chang Y, et al. ERP44 inhibits human lung cancer cell migration mainly via IP3R2. Aging (Albany NY). 2016;8:1276-86.
[48] Wise R, Duhachek-Muggy S, Qi Y, Zolkiewski M, Zolkiewska A. Protein disulfide isomerases in the endoplasmic reticulum promote anchorage-independent growth of breast cancer cells. Breast Cancer Res Treat. 2016;157:241-52.
[49] Zaghrini C, Seitz-Polski B, Justino J, Dolla G, Payre C, Jourde-Chiche N, et al. Novel ELISA for thrombospondin type 1 domain-containing 7A autoantibodies in membranous nephropathy. Kidney Int. 2019;95:666-79.
[50] Pedro NF, Biselli JM, Maniglia JV, Santi-Neto D, Pavarino EC, Goloni-Bertollo EM, et al. Candidate Biomarkers for Oral Squamous Cell Carcinoma: Differential Expression of Oxidative Stress-Related Genes. Asian Pac J Cancer Prev. 2018;19:1343-9.
[51] Privat M, Rudewicz J, Sonnier N, Tamisier C, Ponelle-Chachuat F, Bignon YJ. Antioxydation And Cell Migration Genes Are Identified as Potential Therapeutic Targets in Basal-Like and BRCA1 Mutated Breast Cancer Cell Lines. Int J Med Sci. 2018;15:46-58.
[52] Kim YJ, Sertamo K, Pierrard MA, Mesmin C, Kim SY, Schlesser M, et al. Verification of the biomarker candidates for non-small-cell lung cancer using a targeted proteomics approach. J Proteome Res. 2015;14:1412-9.
[53] Davalieva K, Kostovska IM, Kiprijanovska S, Markoska K, Kubelka-Sabit K, Filipovski V, et al. Proteomics analysis of malignant and benign prostate tissue by 2D DIGE/MS reveals new insights into proteins involved in prostate cancer. Prostate. 2015;75:1586-600.
[54] Dukel M, Streitfeld WS, Tang TC, Backman LR, Ai L, May WS, et al. The Breast Cancer Tumor Suppressor TRIM29 Is Expressed via ATM-dependent Signaling in Response to Hypoxia. J Biol Chem. 2016;291:21541-52.
[55] Yuan Z, Villagra A, Peng L, Coppola D, Glozak M, Sotomayor EM, et al. The ATDC (TRIM29) protein binds p53 and antagonizes p53-mediated functions. Mol Cell Biol. 2010;30:3004-15.
[56] Glebov OK, Rodriguez LM, Soballe P, DeNobile J, Cliatt J, Nakahara K, et al. Gene expression patterns distinguish colonoscopically isolated human aberrant crypt foci from normal colonic mucosa. Cancer Epidemiol Biomarkers Prev. 2006;15:2253-62.
[57] Stahl PR, Hoxha E, Wiech T, Schroder C, Simon R, Stahl RA. THSD7A expression in human cancer. Genes Chromosomes Cancer. 2017;56:314-27.

748 [58] Kiyamova R, Garifulin O, Gryshkova V, Kostianets O, Shyian M, Gout I, et al. Preliminary
749 study of thyroid and colon cancers-associated antigens and their cognate autoantibodies as
750 potential cancer biomarkers. Biomarkers. 2012;17:362-71.
751 [59] Garifulin OM, Kykot VO, Gridina NY, Kiyamova RG, Gout IT, Filonenko VV.
752 Application of serex-analysis for identification of human colon cancer antigens. Exp Oncol.
753 2015;37:173-80.
754 [60] Duran-Sanchon S, Moreno L, Auge JM, Serra-Burriel M, Cuatrecasas M, Moreira L, et al.
755 Identification and validation of microRNA profiles in fecal samples for detection of colorectal
756 cancer. Gastroenterology. 2019.

Table 1

Summary of the plasma samples used in the study.

Autoantigen Discovery

	Sample number	Age (Median ± SD in years)	Male / Female
CRC (Stage III)	3	79 ± 5	3/0
CRC (Stage IV)	5	68 ± 5	1/4
Control	8	41 ± 4	3/5

Autoantigen Validation*

	Sample number	Age (Median ± SD in years)	Male / Female
CRC	32	69 ± 11	15/17
Premalignant	18	59 ± 7	12/6
Control	30	51± 10	12/18

* For a full description of these samples *see* Supplementary Table S1.

Table 2

Identified proteins as potential autoantibody targets in CRC that appear in <10% of agarose-based immunoprecipitation registered experiments in CRAPome.

Accession	Protein Name	Number of experiments * (found/total)	Ave SC	Max SC	Ratio
P07305	H1F0	23/281	5.1	71	0.08
P52434	POLR2H	20/281	2.3	11	0.07
O95969	SCGB1D2	1/281	1	1	0.00
Q15005	SPCS2	0	1.1	2	0.00
O14880	MGST3	0	1.7	4	0.00
Q16718	NDUFA5	4/281	1.2	2	0.01
P14174	MIF	25/281	3.9	59	0.09
P49720	PSMB3	13/281	3.1	16	0.05
P61019	RAB2A	15/281	1.9	5	0.05
P52565	ARHGDIA	17/281	2.2	10	0.06
Q14019	COTL1	0	1	1	0.00
P61224	RAP1B	17/281	2.7	11	0.06
P37837	TALDO1	10/281	2.7	13	0.04
O75489	NDUFS3	1/281	2.3	9	0.00
P00403	MT-CO2	2/281	2.7	14	0.01
P30740	SERPINB1	0/281	1	1	0.00
P04637	TP53	25/281	5.4	31	0.09
Q9NX63	CHCHD3	3/281	4.6	31	0.01
Q9BS26	ERP44	0/281	1.3	2	0.00
P61158	ACTR3	10/281	3.1	17	0.04
O43464	HTRA2	1/281	1.5	2	0.00
Q9UBU8	MORF4L1	1/281	1.5	2	0.00
Q02790	FKBP4	28/281	6.7	52	0.10
Q9UPZ6	THSD7A	0	1	1	0.00
P35228	NOS2	1/281	1	1	0.00
Q8N3J5	PPM1K	4/281	2	2	0.01
Q14134	TRIM29	1/281	1.2	2	0.00
Q9P2B4	CTTNBP2NL	1/281	1	1	0.00
P33992	MCM5	22/281	5	34	0.08
Q92794	KAT6A	0	2	2	0.00
Q9HD67	MYO10	0	1	1	0.00

* Agarose beads immunoprecipitations. Proteins highlighted in grey were used for validation.

766 **Legend to the Figures**

767

768 **Fig. 1.** Schematic work-flow of the modified SASI approach used for the identification and
769 validation of CRC autoantigens. Immunoprecipitation coupled to mass spectrometry of a CRC
770 cell protein extract using isolated IgGs from plasma samples was used to identify tumor-
771 associated antigens target of CRC autoantibodies. The identified potential targets were analyzed
772 using bioinformatics tools and validated through different techniques. Arrows depicted in the
773 SDS-PAGE highlight those protein bands that seem to differ between the immunoprecipitation
774 using control IgGs and CRC IgGs.

775

776 **Fig. 2.** Bioinformatics analysis of the autoantibody targets. (A) cBioPortal data on CRC showed
777 that 69% and 80% of the identified targets had genetic alterations in tissue samples from colon
778 and rectal adenocarcinomas, respectively. (B) The Oncomine and UALCAN databases showed
779 that 22 out of the 31 identified proteins had a statistically significant altered expression ($p<0.05$)
780 in CRC in comparison to adjacent paired healthy tissue. (C) The Human Protein Atlas [18, 19]
781 showed that CTTNBP2NL has been described as a CRC prognosis marker.

782

783 **Fig. 3.** The identified targets of CRC autoantibodies are dysregulated at protein level in tissue
784 samples. (A) IHC analysis showed that indicated proteins were present in colorectal
785 pathological tissue. (B) TALDO1 and ERP44 could significantly discriminate between stage II
786 CRC patients (p-value<0.05). (C) ROC curves analysis of the IHC data showed that ERP44 had
787 an AUC of 70% (40% sensitivity, 100% specificity) to discriminate CRC patients according to
788 survival status. TALDO1 showed an AUC of 63.4% (56.9% sensitivity, 66.7% specificity)
789 when grouping patients according to recurrence status. When grouping patients according to
790 disease free survival, the AUC of ERP44 was 61.4% (59.1% sensitivity, and 65.9% specificity).

791

792 **Fig. 4.** Cloning, *in vitro* protein expression and evaluation of seroreactivity levels of 13
793 potential autoantibody targets of CRC autoantibodies. (A) Confirmation of the *in vitro* protein

794 expression of the 13 selected targets as HaloTag fusion proteins. (B) Seroreactivity levels of the
795 13 proteins using the indicated 80 individual plasma samples (Table 1; Supplementary Table
796 S1). (C) Autoantibody levels against CHCHD3 (p-value = 4.57e-05), CTTNBP2NL (p-value =
797 0.00074), FKBP4 (p-value = 0.00079), MGST3 (p-value = 4.36e-05), THSD7A (p-value =
798 9.23e-05), and TRIM29 (p-value = 0.0027) as well as p53 (p-value = 0.011) could discriminate
799 with statistical significance between the asymptomatic healthy control group and the CRC
800 group. (D) Autoantibody levels of CTTNBP2NL (p-value = 0.00028), FKBP4 (p-value =
801 0.0014), and TALDO1 (p-value = 0.010) were able to significantly discriminate between
802 asymptomatic healthy control individuals and premalignant subjects. (E) Autoantibody levels
803 against ERP44 (p-value = 0.045), TALDO1 (p-value = 0.018), CHCHD3 (p-value = 0.0024),
804 CTTNBP2NL (p-value = 6.68e-05), FKBP4 (p-value = 0.00015), MGST3 (p-value = 0.002),
805 THSD7A (p-value = 0.014), and TRIM29 (p-value = 0.0045)-as well as p53 (p-value = 0.041)
806 were able to discriminate between the asymptomatic healthy control individuals and the
807 colorectal pathological group.

808

809 **Fig. 5.** Evaluation of the diagnostic value of CRC autoantibodies. (A) The combination of
810 CHCHD3, CTTNBP2NL, FKBP4, MGST3, THSD7A, and TRIM29 showed an AUC of 99.7%,
811 with a sensitivity of 96.9% and specificity of 100%. (B) Autoantibodies against CTTNBP2NL,
812 FKBP4, and TALDO1 showed an AUC of 88.9% with a sensitivity of 83.3% and specificity of
813 86.7% when discriminating asymptomatic healthy control individuals and premalignant
814 subjects. (C) Combined detection of autoantibodies against CHCHD3, CTTNBP2NL, ERP44,
815 FKBP4, MGST3, TALDO1, THSD7A, TRIM29, and p53 showed an AUC of 95.3%, with a
816 sensitivity of 92.0% and a specificity of 90.0% when discriminating the asymptomatic healthy
817 control individuals and the pathological group.

Legend to the Supplementary Figures

Supplementary Fig. 1. Bioinformatic analysis of the identified targets by STRING and IPA. (A) Clustering of the direct and indirect interactions of the identified autoantibody targets performed with STRING. Direct and indirect clustering of proteins showed five interacting groups composed of proteins related to GTPases and actin-interacting proteins, transcription, immunity, and cell cycle functions; pentose pathways and endoplasmic reticulum, respiratory chain and caspase activity, and respiratory chain and cell messengers. (B) IPA analysis of the identified targets. Among the identified altered networks appeared the cell cycle, DNA replication, recombination, protein synthesis; and cancer, carbohydrate metabolism, and cell cycle networks.

Supplementary Fig. 2. Verification of the presence of the autoantibody targets at mRNA and protein level in the CRC cells used in the study. (A) mRNA level. (B) Protein level. GAPDH and tubulin were used as loading controls at mRNA and protein level, respectively.

Figure 1

Click here to download high resolution image

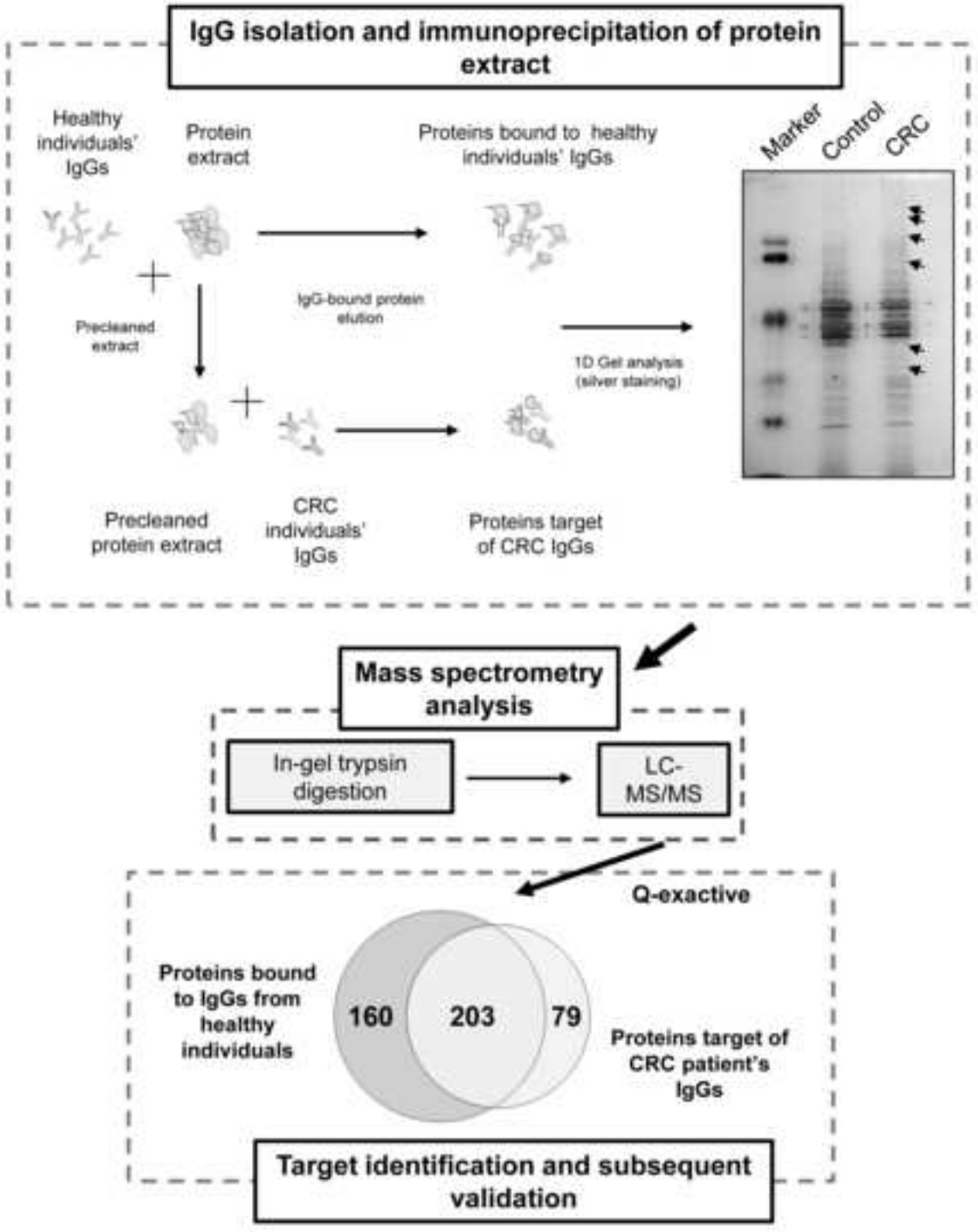


Figure 1

Figure 2

Click here to download high resolution image

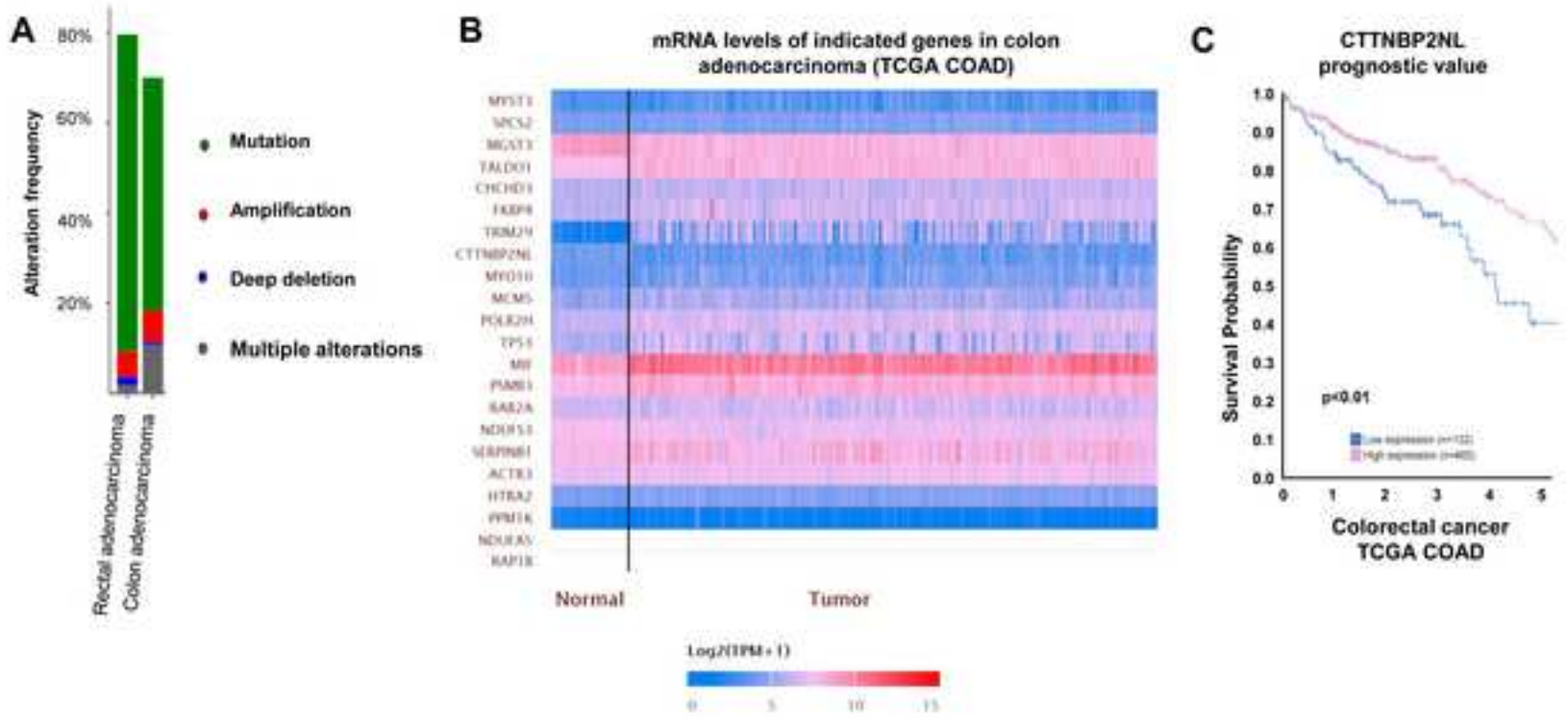


Figure 2

Figure 3

Click here to download high resolution image

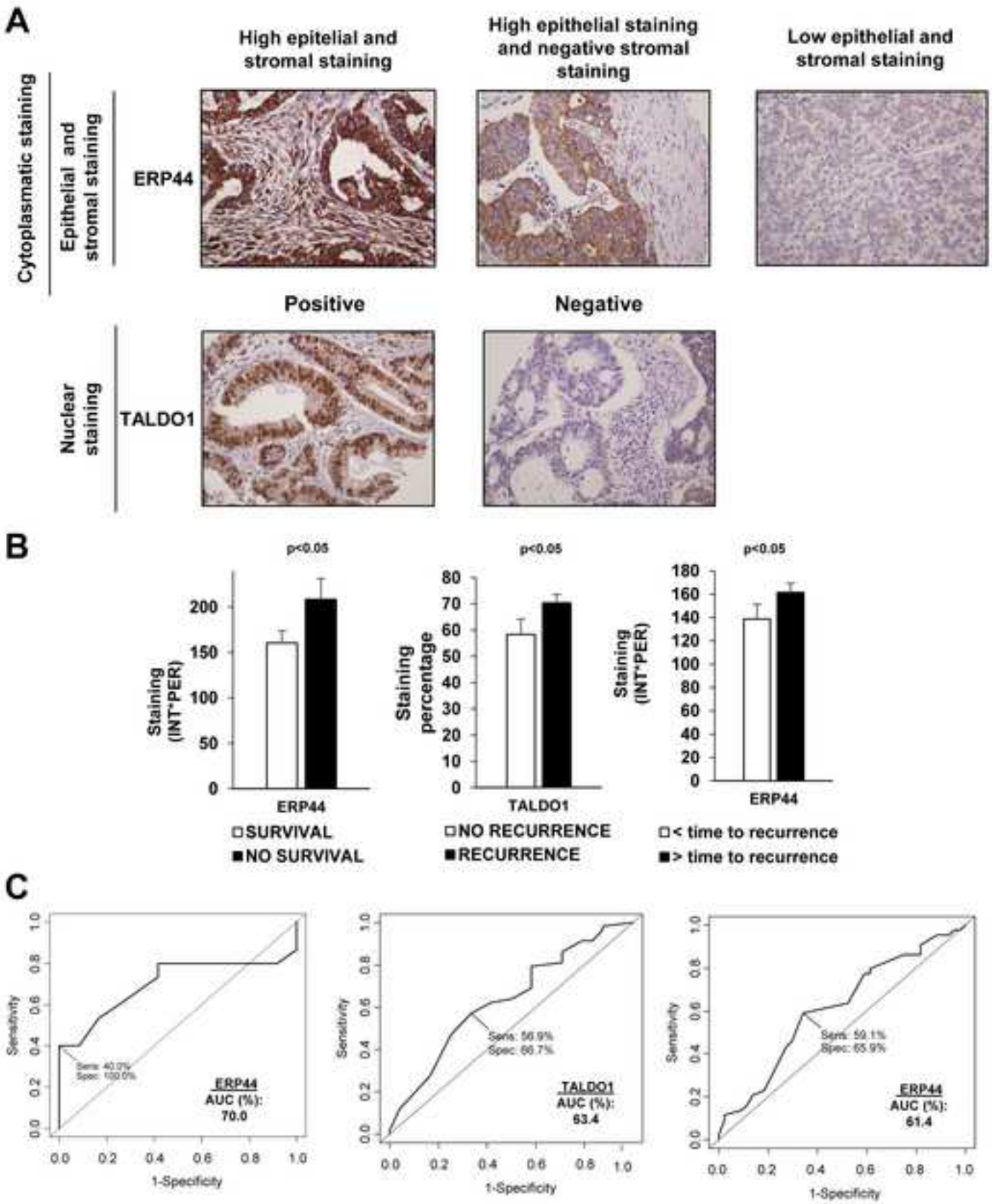


Figure 3

Figure 4

Click here to download high resolution image

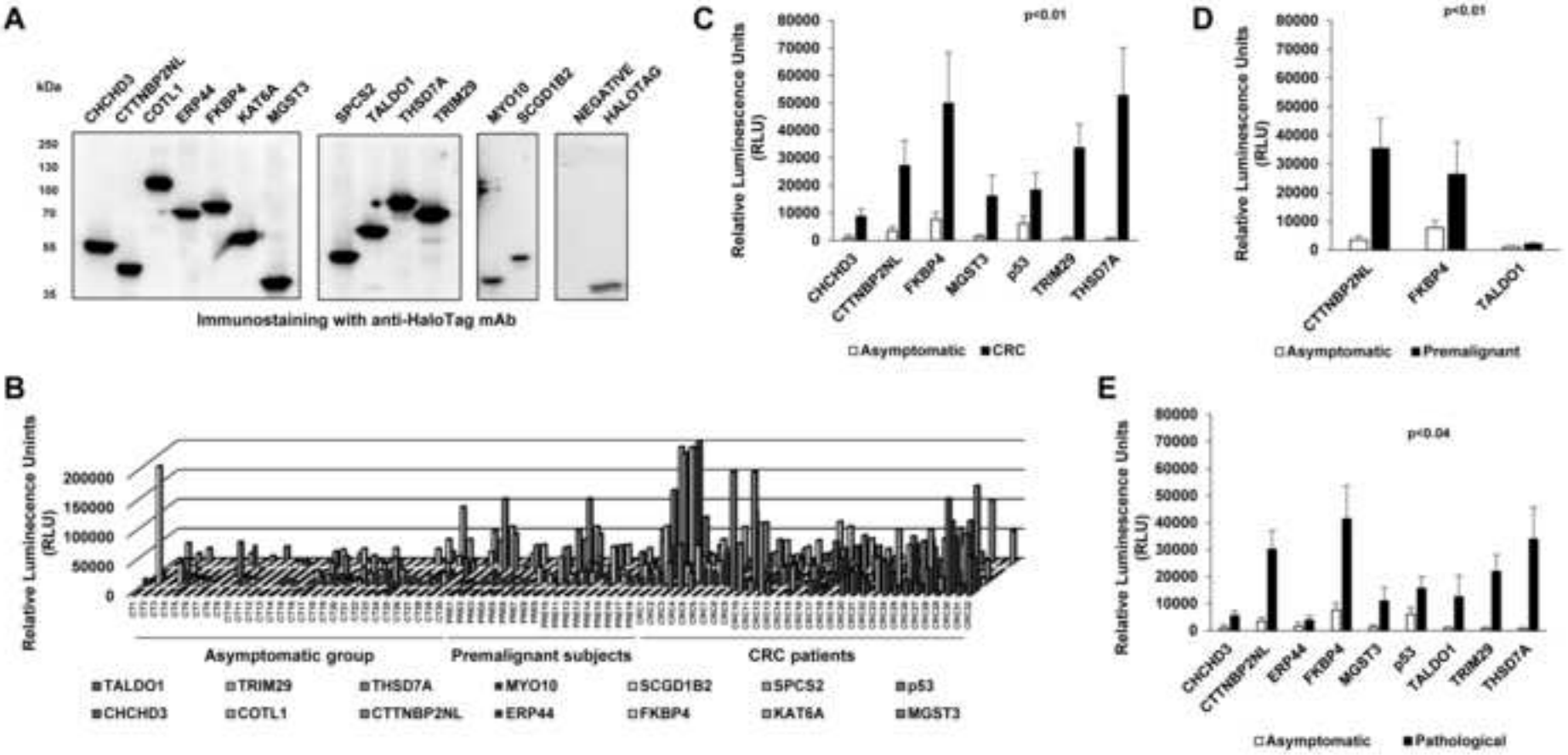


Figure 4

Figure 5

Click here to download high resolution image

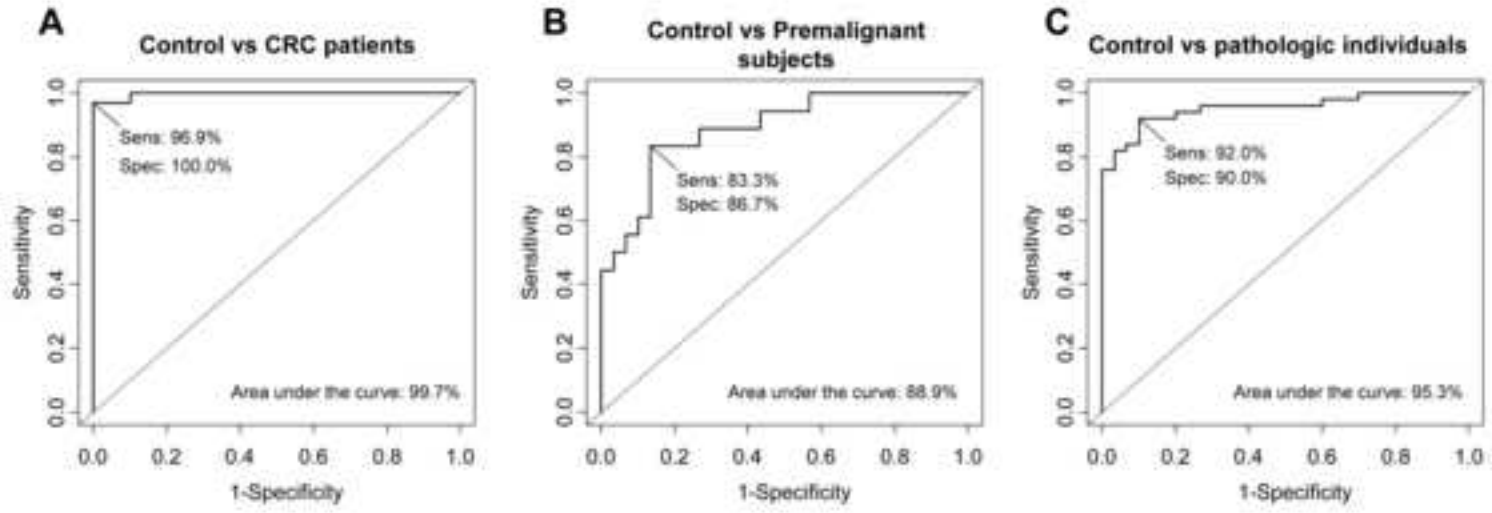


Figure 5

Supplementary Fig S1

Click here to download Supplementary material: Supplementary Fig. S1.tif

Supplementary Fig S2

Click here to download Supplementary material: Supplementary Fig. S2.tiff

Supplementary Table S1

Click here to download Supplementary material: Supplementary Table S1.xlsx

Supplementary Table S2

Click here to download Supplementary material: Supplementary Table S2.xlsx

Supplementary Table S3

Click here to download Supplementary material: Supplementary Table S3.xlsx

Supplementary Table S4

Click here to download Supplementary material: Supplementary Table S4.xlsx

Supplementary Table S5

Click here to download Supplementary material: Supplementary Table S5.xlsx

Supplementary Table S6

Click here to download Supplementary material: Supplementary Table S6.xlsx

***Conflict of Interest**

Click here to download Conflict of Interest: coi_disclosure All Authors.pdf