



Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com



ORIGINAL RESEARCH

The Immunome of Colon Cancer: Functional *In Silico* Analysis of Antigenic Proteins Deduced from IgG Microarray Profiling



Johana A. Luna Coronell^{1,a}, Khulan Sergelen^{1,b}, Philipp Hofer^{2,c},
István Gyurján^{1,d}, Stefanie Brezina^{2,e}, Peter Hettegger^{1,f}, Gernot Leeb^{3,g},
Karl Mach^{3,h}, Andrea Gsur^{2,i}, Andreas Weinhäusel^{1,*j}

¹ Molecular Diagnostics, AIT – Austrian Institute of Technology, A-1190 Vienna, Austria

² Department of Medicine I, Institute of Cancer Research, Comprehensive Cancer Center, Medical University Vienna, A-1090 Vienna, Austria

³ Hospital Oberpullendorf, A-7350, Oberpullendorf, Austria

Received 2 June 2017; revised 1 October 2017; accepted 19 October 2017

Available online 2 March 2018

Handled by Quan-Zhen Li

KEYWORDS

Autoantibody tumor biomarker;
Cancer immunology;
Colorectal cancer;
Immunomics;
Protein microarray

Abstract Characterization of the colon cancer immunome and its autoantibody signature from differentially-reactive antigens (DIRAGs) could provide insights into aberrant cellular mechanisms or enriched networks associated with diseases. The purpose of this study was to characterize the antibody profile of plasma samples from 32 **colorectal cancer** (CRC) patients and 32 controls using proteins isolated from 15,417 human cDNA expression clones on microarrays. 671 unique DIRAGs were identified and 632 were more highly reactive in CRC samples. Bioinformatics analyses reveal that compared to control samples, the immunoproteomic IgG profiling of CRC samples is mainly associated with cell death, survival, and proliferation pathways, especially proteins involved in EIF2

* Corresponding author.

E-mail: andreas.weinhaeusel@ait.ac.at (Weinhäusel A).

^a ORCID: 0000-0003-4625-5378.

^b ORCID: 0000-0002-0546-3587.

^c ORCID: 0000-0003-2550-6019.

^d ORCID: 0000-0002-9880-8475.

^e ORCID: 0000-0001-5238-6900.

^f ORCID: 0000-0001-8557-588X.

^g ORCID: 0000-0003-3889-5594.

^h ORCID: 0000-0002-5555-9798.

ⁱ ORCID: 0000-0002-9795-1528.

^j ORCID: 0000-0002-9318-1204.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2017.10.002>

1672-0229 © 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

and mTOR signaling. Ribosomal proteins (*e.g.*, RPL7, RPL22, and RPL27A) and CRC-related genes such as *APC*, *AXIN1*, *E2F4*, *MSH2*, *PMS2*, and *TP53* were highly enriched. In addition, differential pathways were observed between the CRC and control samples. Furthermore, 103 DIRAGs were reported in the SEREX antigen database, demonstrating our ability to identify known and new reactive antigens. We also found an overlap of 7 antigens with 48 “CRC genes.” These data indicate that immunomics profiling on protein microarrays is able to reveal the complexity of immune responses in cancerous diseases and faithfully reflects the underlying pathology.

Introduction

Colorectal cancer (CRC) is reported worldwide as the second most common cancer in women and third in men, which makes it a leading cause of cancer-associated mortality in developed countries [1,2]. Various screening methods for CRC are available, such as fecal occult blood tests (FOBT), colonoscopy, and flexible sigmoidoscopy [3]. Implementation of nationwide screening programs, and minimal invasive and early diagnostic methods could help to reduce the high mortality rate of CRC. Early diagnostic methods would enable prompt detection of cancer at early stages, which is essential for therapeutic success and a higher patient survival rate. Therefore, discovery, and identification of sensitive as well as specific markers that could be exploited at the earliest possible stage is needed. Ideally, the identification of biomarkers shall be established with easy sample access [4] from body fluids like serum, plasma or saliva in a minimally invasive manner, which are generally preferred than undergoing colonoscopy.

In cancer, altered protein expression during neoplastic transformation and tumor progression can elicit immune responses and induce the formation of tumor autoantibodies [5]. Besides the involvement in inhibiting tumor growth, immune responses could also promote tumor growth through a process called immunoediting consisting of elimination, equilibrium, and escape phases [6,7]. Immunoediting may affect the composition and quantity of circulating antibodies. The reactivity of these antibodies toward recognized or unrecognized tumor-associated antigens (TAAs) can be affected by multiple factors related to cancer growth, such as aberrant expression of differentiation genes, accumulation of mutations, inaccurate post-translational modifications, alternative splicing, as well as deregulated necrotic or apoptotic processes [8,9]. These TAAs usually have key functions in tumorigenesis, for instance, regulation of cell proliferation and cycle, differentiation, and apoptosis [10,11]. Antibodies are very stable and can be detected months or even years before a clinical cancer diagnosis [12], which makes it possible to determine the differentially-reactive antigens (DIRAGs) among patients as well as relative to control samples by analyzing the immunome (antibody profile) [13,14]. Therefore, autoantibodies could be used as a serological tool for early diagnosis of cancer.

Autoantibody signatures for several cancer types have been reported, including colon cancer, prostate cancer, breast cancer, liver cancer, ovarian cancer, renal cancer, head and neck cancer, esophageal cancer, lymphoma, and leukemia [15–22]. Autoantibodies in cancer can be identified using various methods, such as phage display [12,23], serological analysis of recombinant cDNA expression libraries (SEREX) [15,24], and serological proteomics analysis (SERPA, also known as Proteomex) [25–27]. However, these techniques require complex steps [28,29]. There exists significance and need for

identifying new protein biomarkers in CRC, as reviewed by us [3] and lately by Coughlin and Murray [30].

Protein arrays, which comprise recombinant proteins, protein fractions, or purified proteins, offer a potent tool for both definition and identification of immune profiles [31]. Proteins included in the arrays are known, which are printed with a comparable concentration in a highly-multiplex manner. Therefore, there exists no bias in identification of biomarkers with great sensitivity [32]. Additionally, high-density protein arrays increase the chance of discovering novel autoantibodies against low abundance proteins while also allowing testing of thousands of proteins simultaneously [33]. Thus, detection of diagnostic autoantibody signatures by testing patient samples from, *e.g.*, cancer patients versus control samples, can be conducted in a cost-effective manner [3,34]. A review on protein-based approaches for biomarker discovery was recently done by Huang and Zhu [35].

In this study, we have produced and tested protein microarrays from 15,417 human cDNA expression clones presenting 6369 unique human proteins for the identification of DIRAGs [36]. Our previous work has demonstrated that using purified IgG does avoid artifacts caused by the matrix of serum or plasma samples, and is thus an ideal way to analyze DIRAG profiles [37]. Thus IgG derived from heparin-plasma of 32 CRC patients and 32 controls was used in the current study. As a result, we show that biological profiles can be illustrated via antibody profiling.

Results

IgG profiling on protein-microarrays

To identify DIRAGs from IgG profiling, we performed the immunoprofiling of CRC and control samples using our in-house protein microarray as previously described [36]. All plasma samples used were collected in the ongoing molecular epidemiology “Colorectal Cancer Study of Austria” (CORSA), targeted to inhabitants of the Austrian province Burgenland aged 40–80 years, as described in the Material and Methods section. After the data were normalized using distance weighted discrimination (DWD), we used *t*-test ($P = 0.01$ as cut-off) to evaluate the differences in antibody profiles between the CRC and control samples. Consequently, 671 unique antigenic proteins were identified as DIRAGs based on the median fold-change between classes. Among them, 632 antigenic proteins were found to be higher reactive in CRC samples, whereas 31 antigenic proteins were more reactive in the control samples. However, we also found that 8 antigenic proteins exhibited unclear immunoreactivity, *i.e.*, two different antigen clones expressing the same proteins were found to be significantly different in immunoreactivity but in opposite

directions (one clone with increased immunoreactivity and the other with reduced immunoreactivity) between CRC and control samples. The list of significant antigenic proteins can be found in Table S1. The technical performance and reliability of the protein array analysis is provided as described in the methods and shown in Figure S1.

These 671 DIRAGs were subjected to bioinformatics analyses as outlined in **Figure 1**.

Functional analysis for associated pathways and networks

To identify the associated canonical pathways and network functions, we then performed functional analysis using Ingenuity Pathway Analysis (IPA, Ingenuity® Systems, www.ingenuity.com) [38]. As shown in Table S2, DIRAGs were involved in 50 canonical pathways ($P < 0.01$; $-\log P$ value > 2). The top 5 pathways include the eukaryotic initiation factor 2 (EIF2) signaling pathway, mTOR signaling, growth hormone signaling, virus entry via endocytic pathways, and 14-3-3-mediated signaling (**Table 1**).

IPA analysis revealed that the EIF2 signaling pathway was the most overrepresented canonical pathway between CRC and control samples ($P = 4 \times 10^{-6}$). A total of 24 proteins were represented in the EIF2 signaling pathway, including three proteins from the phosphoinositide 3-kinase (PI3K)

family, namely phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit δ (PIK3CD), PIK3C type 2 β (PIK3C2B), and PI3K regulatory subunit 2 β (PIK3R2). PI3Ks are involved in signaling pathways such as cell motility, cell migration, vesicle transport, and apoptosis [39].

To identify interactions at the molecular level between the DIRAGs found (Table S1) and how they might work together, we then analyzed mechanistic bionetworks using IPA. As shown in **Tables 2, 3** out of the 5 bionetworks found are related to cell death and survival, with one related to cancer as well. In addition, 3 bionetworks are involved in cellular growth and proliferation (Table 2). The detailed list of related DIRAGs can be found in Table S3.

Protein interaction enrichment analysis with WebGestalt

Comparison of the DIRAGs with the protein list from the annotated genes presented in the UniPEx library in pre-defined functional categories was performed for a hierarchical protein interaction module enrichment analysis. The hierarchical relationship of the enriched phenotype terms can be observed in the directed acyclic graph (DAG) found in Figure S2.

Among the 19 enriched network modules, three modules are found to contain 14-36 proteins. These include Module 1 (36 proteins), Module 2 (26 proteins), and Module 3 (14 proteins).

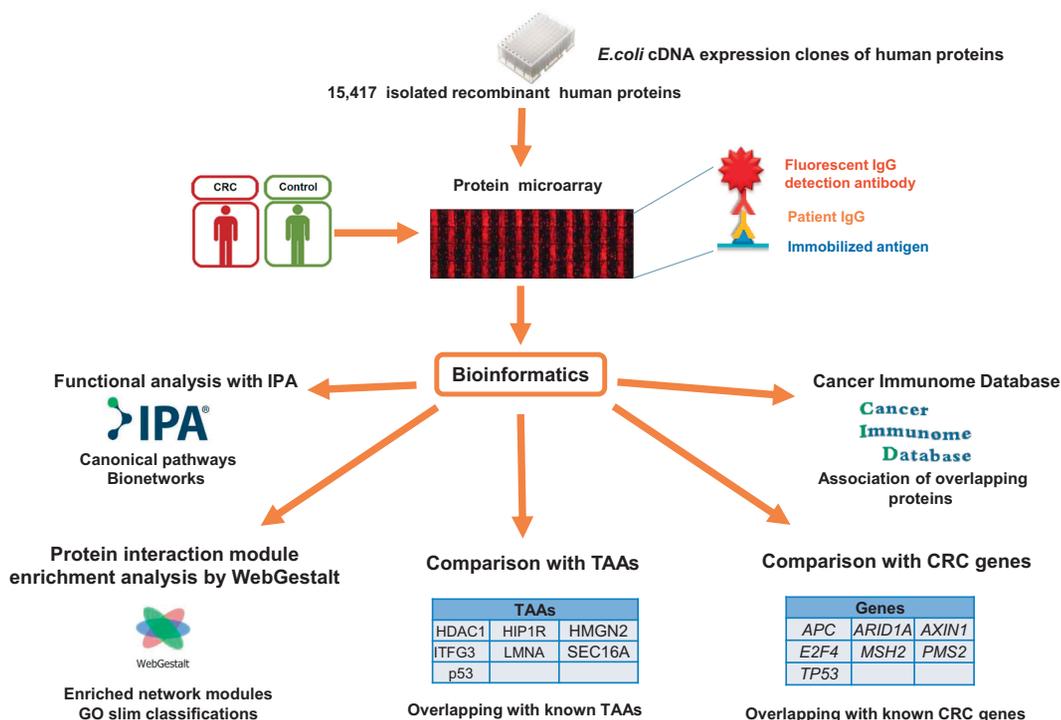


Figure 1 Procedure overview

The described procedure exemplifies the methodology used in this study. An expression library consisting of 15,417 cDNA clones was used to produce recombinant human proteins. The recombinant proteins were isolated and used for printing protein microarrays. IgG was isolated from a total of 64 samples (32 CRC samples and 32 healthy control samples) and tested on the protein microarrays. Bioinformatics analyses (t -tests) were performed to identify the DIRAGs between the groups of arrays. Subsequently, the list of DIRAGs were subjected to functional analysis with IPA, hierarchical protein interaction module enrichment analysis with WebGestalt, association of overlapping proteins with the Cancer Immunome Database analysis, and analysis of overlap with known CRC and TAAs. CRC, colorectal cancer; TAA, tumor-associated antigen; GO, Gene Ontology.

Table 1 Top 5 pathways enriched with DIRAGs

Pathway	-Log (P value)	Ratio	Proteins
EIF2 signaling	5.39	0.162	PABPC1, PIK3C2B, RPL22, RPL27A, RPL37A, RPS19, PDPK1, PPP1R15A, RPS17/RPS17L, EIF4G1, RPL7, RPS7, EIF3G, EIF3F, RPS27, EIF4G2, RPL28, RPL36AL, RPL19, RPS25, PIK3CD, PIK3R2, RPS10, RPL18
mTOR signaling	4.63	0.150	PIK3C2B, ULK1, DDIT4, RPS19, PDPK1, RPS17/RPS17L, EIF4G1, PRKCZ, EIF3G, RPS7, DGKZ, EIF3F, RPS27, EIF4G2, PRKCD, TSC2, RPS6KB2, RPTOR, RPS25, PRKCH, PIK3CD, PIK3R2, RPS10
Growth hormone signaling	4.41	0.226	PIK3C2B, PRKCD, RPS6KB2, PLCG1, PDPK1, PRKCH, PIK3CD, STAT3, PIK3R2, STAT1, ELK1, PRKCZ
Virus entry via endocytic pathway	3.71	0.183	PIK3C2B, FLNB, AP1G2, HLA-C, HLA-A, PRKCD, CLTA, HLA-B, PLCG1, PIK3CD, PRKCH, PIK3R2, PRKCZ
14-3-3-mediated signaling	3.45	0.158	PIK3C2B, TUBB3, YWHAE, PDIA3, YWHAZ, PLCG1, VIM, PRKCZ, PRKCD, TSC2, PIK3CD, PRKCH, PIK3R2, ELK1, PDCD6IP

Note: The ratio is the number of proteins in a given pathway that meet the cutoff criteria ($P < 0.01$), divided by the total number of proteins that make up that pathway. The complete list of 50 pathways can be found in Table S2.

Table 2 Top 5 associated network functions obtained with IPA

Associated network functions	Score	No. of DIRAGs found
Cell death and survival, cell cycle, cellular growth and proliferation	40	35
Cellular movement, cellular growth and proliferation, cell cycle	11	16
Cell cycle, cellular development, cellular growth and proliferation	11	18
Cell death and survival, cell cycle, cellular development	10	17
Cell death and survival, cancer, reproductive system disease	8	15

Note: The score indicates the likelihood of the focus genes in a network being found together due to random chance and is used to rank networks according to their degree of relevance to the network eligible molecules in a dataset, based on the connectivity of the molecules in a given network. The score is calculated with the right-tailed Fisher's Exact test. The maximum network size is set at 35 by default.

As shown in Figure S3, some higher antigenic reactive proteins are overexpressed (up-regulated, in red) in Module 1. These include proteins involved in translation factors, *e.g.*, ISG15 ubiquitin-like modifier (ISG15), as well as transport and cytoskeleton, *e.g.*, dynein cytoplasmic 1 heavy chain 1 (DYNC1H1) and filamin B (FLNB). Proteins in Module 2 are mostly transcription factors, or proteins associated with double-strand break repair and DNA binding (Figure S4).

Figure 2 shows the node-link diagram for Module_3, which contains 12 ribosomal proteins including 5 L ribosomal proteins (RPLs), *i.e.*, RPL7, RPL18, RPL19, RPL22, RPL27A, RPL28, and RPL37A, and 5 ribosomal protein S, *i.e.*, RPS7, RPS10, RPS17, RPS19, and RPS25. In addition, signal recognition particle receptor (SRPR) and signal sequence receptor subunit 2 (SSR2) in the endoplasmic reticulum were found in Module 3 as well.

Of note, we found that ribosomal proteins were also enriched in the EIF2 signaling pathway obtained with IPA (Table 1). Therefore, we compared the proteins from Module_3 and the proteins from the EIF2 signaling pathway. We thus found an overlap of 12 ribosomal proteins, including RPL7, RPL18, RPL19, RPL22, RPL27A, RPL36AL, RPL37A, RPS7, RPS10, RPS17, RPS19, and RPS25. This result indicates that complex cellular structures (especially ribosomes) are a frequent target of autoantibodies.

To gain further understanding of the biological meaning of the DIRAGs, we performed Gene Ontology (GO) slim classifications [40]. Molecular function analysis indicated that DIRAGs are predominantly involved in binding functions (394 of 671), including protein, ion, nucleic acid and nucleotide acid binding (**Figure 3A**). The biological process analysis showed that 66% of DIRAGs were found in metabolic processes (441 proteins), while 58% were involved in biological regulation (387 proteins) (Figure 3B). Furthermore, cellular component analysis revealed that the classified proteins were mainly found in nuclear components (327 DIRAGs), macromolecular complexes (237 DIRAGs), membrane function (226 DIRAGs), membrane enclose lumen (219 DIRAGs), and cytosol (189 DIRAGs) (Figure 3C).

Comparison with the cancer immunome database

To better understand the 671 unique DIRAGs identified in our microarray study (Table S1) in alignment with known tumor antigens, we compared our data with the Cancer Immunome Database. Among the 1545 known antigens from the SEREX database (<http://ludwig-sun5.unil.ch/CancerImmunomeDB/>), 568 antigens were included in the UniPex library. Aligning the 671 unique DIRAGs obtained in this study with these 568 SEREX antigens revealed that 103 antigens were found

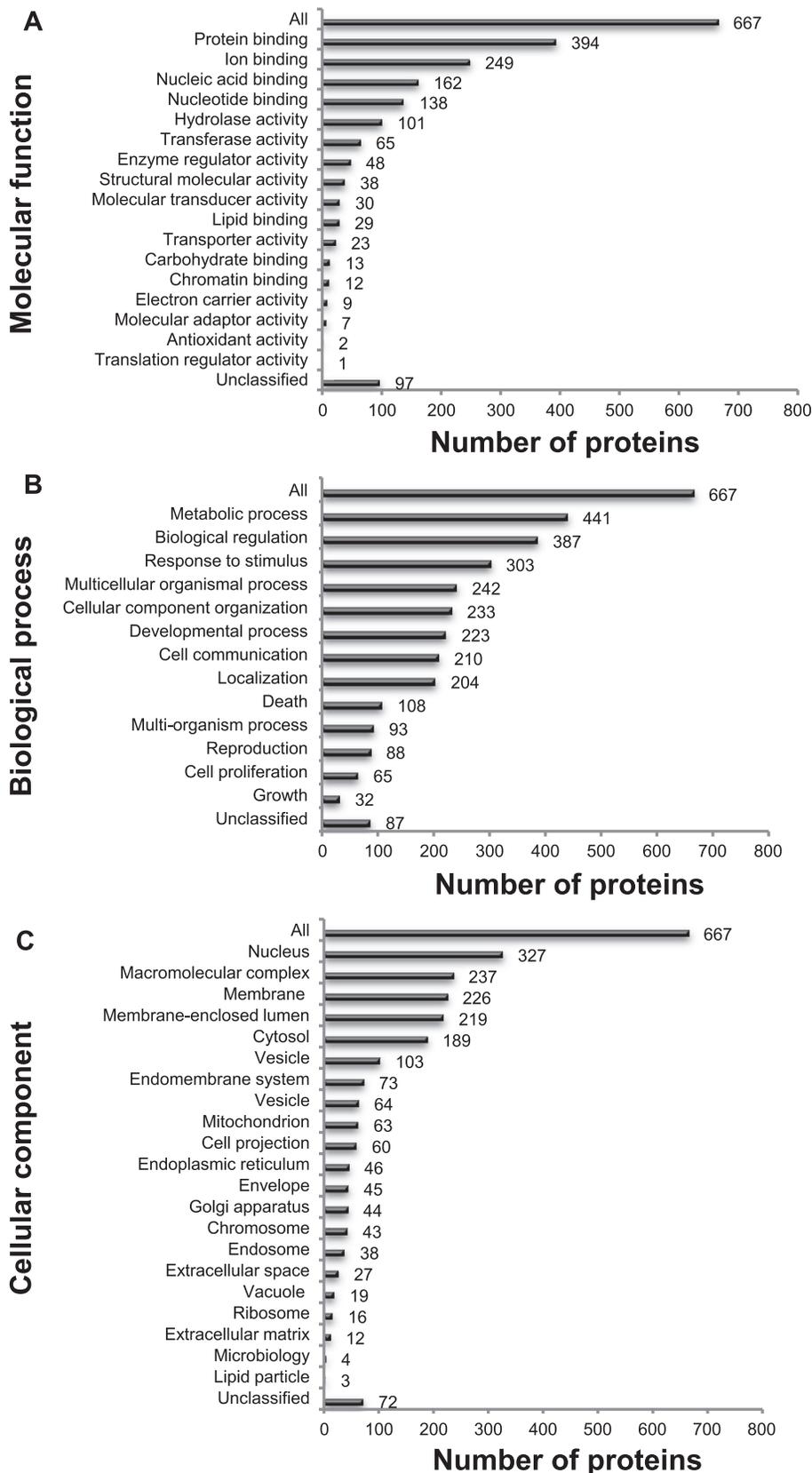


Figure 3 GO Slim classification analysis of the 671 DIRAGs identified

Histogram of functional annotations of DIRAGs in CRC samples in comparison with control samples ($P = 0.01$) was generated based on the WebGestalt derived GO slim charts in the three GO functional categories. **A.** Molecular function. **B.** Biological process. **C.** Cellular component. More than half of the proteins are nuclear proteins. DIRAG, differentially-reactive antigens; CRC, colorectal cancer; GO, Gene Ontology.

Table 3 List of CRC DIRAGs overlapping with published TAAs

DIRAG	Fold change	Upregulation/downregulation	Refs.
HDAC1	1.36	↑	[32]
HIP1R	1.66	↑	[43]
HMGN2	1.71	↑	[44]
ITFG3	1.32	↑	[31]
LMNA	1.66	↑	[43]
SEC16A	0.54	↓	[31]
p53	1.75	↑	[31,41,43–46]

Note: CRC DIRAGs are found to overlap with the published TAAs. The upward and downward arrows indicate that expression of the DIRAG was found up-regulated and down-regulated, respectively, in this study. TAA, tumor-associated antigen.

Discussion

Information obtained from the colon cancer immunome is of great significance, as the immune system plays a crucial role in cancer advancement [49]. Obtaining information on the molecular mechanisms in which the TAAs are involved is of great aid in understanding the biology and the mechanisms underlying the development of cancer. Furthermore, the changes in immunoreactivity or antibody-profiles provide disease-specific molecular signatures, which could be used for diagnostics and probably have additional significance to the clinical parameters currently in use for disease management. For autoantibody profiling, high-density protein arrays are a good tool for discovery, enabling a high-throughput test of many samples especially when using customized microarrays presenting selected proteins. Moreover, we have previously demonstrated (and recently Negm and colleagues used a very similar approach [50]) that purified IgG optimally conserves DIRAG profiles, thus circumventing matrix artifacts found in serum or plasma samples [36,37].

Further bioinformatics analysis reveals that the EIF2 signaling pathway was the most overrepresented canonical pathway, which could be explained by the fact that this pathway is required to initiate protein synthesis. In addition, the EIF2 signaling pathway can also induce PI3K; in agreement, PI3K was found to be overrepresented in our study as well (Table S2).

Amplification of PI3K plays a role in the transduction of signals from extracellular stimuli, such as hormones, mitogens, growth factors and cytokines, to cellular pathways controlling cell growth, proliferation, and survival [51]. PI3K is well known to promote tumorigenesis in a variety of experimental models of cancer [49] including CRC [52]. One of the pathways activated by the amplification of PI3K is the mTOR pathway, which was also found in the top 5 canonical pathways, with 23 molecules represented in the pathway (Table 1). It is known that the mTOR pathway is activated during various cellular processes such as tumor formation and is deregulated in cancer [53]. Our results are in line with whole-exome sequencing and integrative data from TCGA network. Through analyzing the mRNA expression changes from 195 tumor samples, it was demonstrated that the PI3K, p53, and WNT pathways are deregulated in CRC [54].

One key finding of the colon cancer immunome is that factors involved in protein synthesis are enriched and overexpressed, which is confirmed by analyses performed using both IPA and WebGestalt (Hierarchical Protein Interaction

Module Enrichment Analysis). Functional analyses also showed overexpression of ribosomal proteins and translation initiation factor proteins involved in the EIF2 signaling pathway (Figure 2). Furthermore, CRC DIRAGs identified were found to be enriched in proteins involved in binding functions, such as protein, ion, nucleic acid and nucleotide acid binding (Figure 3). These results are in accordance with results from Yu and colleagues [55]. In their study using CRC and adjacent normal tissues, they employed gene expression microarray analysis and also found that metabolic processes are the most common biological processes from the differential proteins analyzed (Figure 3B) [55]. As mentioned before, our results suggest that complex cellular structures are a frequent target of autoantibodies. This is further supported by the finding of enriched proteins that are known to be implicated in protein binding [9,56], folding [57], and cell proliferation [58].

Differentially-reactive antibodies are reporters of the immune system targeting cellular as well as secreted proteins from tumors. Our results are corroborated by results obtained from Emmink and collaborators [56], who found that both extensive survival and anti-oxidant networks are represented in the secretome of colon cancer stem cells. Consistent with our findings (Figure 2), they found several ribosomal proteins and translation initiation factors and, most significantly, enriched proteins governing cell death. As a consequence, the immune response as seen in changed antibody profiles might also be driven by secreted proteins from tumors [5], as our DIRAGs are in concordance with the proteins identified in Emmink's study [56].

In line with existing knowledge, we also found a highly significant overlap between DIRAGs (671) and the SEREX-derived antigens listed in the Cancer Immunome Database, of which 568 antigens were also present in our protein array (Fisher's exact test: $P = 8.5 \times 10^{-8}$). This result demonstrates the reliability of antigenic proteins defined by our protein array. The identified antigenic proteins are mainly associated with the cell cycle, connective tissue development, transcription factors, and cell-to-cell signaling interaction networks (Table 2). It is well known that tumors reside in a microenvironment that is associated with aberrantly-altered cancer-associated cells, inflammation, hypoxia, and loss of normal tissue architecture [49,59], further supporting our findings. The results further advocate that our approach toward identifying and characterizing antibody profiles has the potential to identify biomarkers displaying the complexity of such antigenic responses.

Screening plasma samples using our protein microarrays leads to the identification of both known (Table 3) and new

TAAAs, which may serve as new biomarkers. For instance, HDAC1, which plays a role in cell proliferation, survival, and inhibition of differentiation, shows higher antigenic reactivity in our study, which has been corroborated in CRC tumor studies [60,61]. Besides, 7 known CRC genes were found by comparing the DIRAGs (Table S1) with the literature-defined 48 CRC genes (Table S6). The tumor suppressor *APC*, listed on top, interacts with *AXIN1*, which, in addition, interacts with other Wnt/ β -catenin signaling pathway components [62] and is essential for degradation of β -catenin in the Wnt/ β -catenin signaling cascade, an important signal transduction pathway in CRC [63]. Moreover, E2F4 is an important transcription factor in cell cycle control [64], while *MSH2*, an MMR gene like *ARID1* and *PMS2* [46,65], is highly associated with hereditary non-polyposis CRC. The MMR system recognizes and repairs mismatches between base pairs during DNA replication. *PMS2* has been found to interact with p53 [66], a transcription factor that activates apoptotic, autophagial, cell cycle arresting and cellular metabolism genes, which confers its tumor suppressor activity [46].

Taken together, our data provide a comprehensive view on the colon cancer immunome as an additional pathological layer worth considering in more detail when both bioinformatics analyses such as IPA and WebGestalt have provided overlapping information as complementary evidence. Moreover, analysis of tumor-associated antigenic proteins found in the Cancer Immunome Database provides insights into associations with cancer antigens, as well as the differentially reactive activity of antigens that are known in CRC. Further experiments to address to what extent the mechanisms involved in the antigenicity of autoantigens operate within malignancies need to be performed to deepen our understanding of interactions and networks in cancer formation. Although the antigenicity of autoantibodies to TAAAs has been acknowledged in various elements of cancer growth [8,9], additional understanding can be gathered with the aid of network and functional analyses as exemplified herein.

Materials and methods

Clinical information and samples

All plasma samples were collected in the ongoing molecular epidemiology “Colorectal Cancer Study of Austria” (CORSA). Since May 2002, 11,657 individuals have participated in CORSA (01/2014). The screening program “Burgenland Prevention Trial of Colorectal Disease with Immunological Testing” (B-PREDICT), which is a province-wide program, invites the public to participate in fecal occult blood testing (FOBT) annually. This invitation is open to all inhabitants of the Austrian province Burgenland, as long as they are between 40 and 80 years old. FOBT-positive individuals are offered a complete colonoscopy and, at the time of colonoscopy, are asked to take part in CORSA. A blood sample from the participants is collected as well as information in a short questionnaire. The questionnaire includes information regarding anthropometric and demographic factors, smoking status, alcohol consumption, and basic dietary habits. After sample acquisition, the heparinized plasma was centrifuged at 2000g for 10 min, and the resulting supernatant was stored as plasma samples at -80 °C until further use.

Table 4 Demographics of the study population

Variable	CRC (n = 32)	Control (n = 32)
Age	65.9 (48–82)	63.7 (40–78)
Sex		
Male	18	18
Female	14	14
Meat consumption		
Very frequent	6	5
Frequent	11	17
Seldom	13	8
None	2	2
Smoking		
Current	3	5
Former	10	8
Never	17	17
No information	2	2
Clinical tumor stage		
0	1	NA
I	8	NA
II	8	NA
III	5	NA
IV	3	NA
Missing	7	NA
Lymph node metastasis	6	NA

Note: Age (years) refers to the age of patients at the time of CRC diagnosis or the age of controls at the time of being recruited to the study, indicated as mean (range).

Clinical data of CORSA participants were processed in a central database following regulated documentation guidelines. All subjects provided written informed consent. The institutional local ethics review board “Ethikkommission Burgenland” authorized the study. Further information of the study cohort is described previously [67,68]. According to histopathology, individuals were classified as CRC cases ($n = 32$) and controls ($n = 32$). All individuals with serrated adenomas have been excluded. Controls underwent a complete colonoscopy and were found to be free of CRC and free of polyps. Persons with severe medical conditions including any other malignant condition at the initial study point were disqualified from the study (Table 4).

IgG purification of blood samples

IgG purification was performed as previously described [36]. Briefly, the Melon™ Gel IgG Purification Spin Plate Kit (Thermo Scientific, Waltham, MA) was used to purify all samples according to the manufacturer’s instructions using 30 μ l of plasma, followed by determination of IgG concentration as previously described [37]. Sample integrity was determined by running each purified sample on a NuPAGE® Novex 4%–12% Bis-Tris Precast Gel (Life Technologies, Carlsbad, CA).

Protein microarray production and processing

Protein expression, purification, and microarray production were performed as previously described [36]. Briefly, the UniPEX – human in-frame cDNA protein expression library consisting of 15,417 *E. coli* cDNA expression clones and

presenting 6369 unique, distinct human proteins was purchased from Imagenes (Berlin, Germany). *E. coli* culture as well as induction of protein expression were conducted in 96-well plates with minor modifications as previously described [69]. Protein expression was induced by cultivation in autoinduction medium or by adding isopropyl β -D-1-thiogalactopyranoside (IPTG). After cultivation, purification and elution of the His-tagged recombinant proteins was performed. Protein estimation was performed by running a gel [70]. Clarified *E. coli* lysates and plain buffer were used as positive and negative controls, respectively. ARChip Epoxy glass slides [71] were used to spot the protein arrays in duplicate using an Omnigridd arrayer. An illustration of the protein microarray design can be found in [70].

Protein microarray processing was done as previously described [36]. Due to array processing handling capacity, arrays were processed in 4 runs on different days with a balanced design (Day 1: 17 CRC vs. 17 controls; Day 2: 15 CRC vs. 15 controls; Day 3: 9 CRC vs. 9 controls; and Day 4: 9 CRC vs. 8 controls). A total of 50 arrays (including replicate analyses) were tested for the 32 CRC samples, and a total of 49 arrays (including replicate analyses) were tested for the 32 control samples.

Technical performance of protein microarray analysis

We have conducted a technical study to confirm the reliability of the findings on protein microarrays by cross-wise serial mixing of 2 individuals' IgG and then testing for significant DIRAGs of 4-fold replicates from the "pure samples at 100%" (Figure S1A). The samples from the 2 individuals show 4638 significant DIRAGs ($P < 0.05$) as illustrated by the volcano plot (Figure S1B). Correlation analysis of these 4638 significant DIRAGs with the IgG-relative amount of both individuals using the mixing-series showed that 97% DIRAGs have Pearson's correlations > 0.5 (corr = 0.5 to 1.0 or -0.5 to -1.0) and that 72% are found with a corr = 0.812 to 1.00 or -0.812 to -1.00 (histogram in Figure S1C).

Data acquisition and statistical analysis

Array imaging, scanning, feature aligning, and gridding of spotted proteins were performed as described earlier [36] using the GenePix Pro 6.0 (Molecular Devices, Sunnyvale, CA). Briefly, correction for the systematic bias that may have been introduced using different batches of arrays was performed prior to arrangement of the protein microarray data and statistical analysis using "Distance Weighted Discrimination/DWD" as described [72]. Statistical analysis of the microarray data was carried out using R 3.0.1 and BRB-ArrayTools 4.3.1 [73]. Sample size calculation was conducted using the BRB-ArrayTools plug-in with a significance level of 0.001 (α), a power of 0.75 ($1 - \beta$), and a fold change of 2. An expected sample size in each class of 32 was determined by applying the 50th percentile of the variance distribution. DIRAGs between the patients and controls were defined using the Class Comparison tool in BRB-ArrayTools with $P < 0.01$. A file was prepared for the resulting analysis data, which included both a list of DIRAGs ID annotations and its ratio of the geometric means between sample groups (Table S1).

Pathway analysis

IPA was used for the generation of "Core Analyses" to interpret the data in relation to biological networks, biological processes, and pathways using the Ingenuity Knowledge Base reference set. The analyzed canonical pathways were ordered by the ratio (features in a given pathway meeting the selection criteria, divided by the total number of features included in that pathway) and the Fisher's Exact test P value.

Web-based gene set analysis toolkit

The Web-based GENE SeT AnaLysis Toolkit (WebGestalt; <http://bioinfo.vanderbilt.edu/webgestalt/>) [74,75] was used for hierarchical enrichment analyses of protein interaction networks and Gene Ontology (GO) slim classification for creating bar charts with respect to biological processes, molecular functions, and cellular components. The hypergeometric test was used for enrichment analysis, and adjustment for multiple testing was achieved using the Benjamini & Hochberg procedure. The significance level was adjusted to the top 10 pathways ($P < 0.05$), and a minimal base amount of two genes for a category was set.

Cancer immunome database comparison

Comparisons with the entire 1545 SEREX antigens enlisted in the Cancer Immunome Database were performed (<http://ludwig-sun5.unil.ch/CancerImmunomeDB/>). The complete unique list of SEREX antigens was matched to the DIRAGs table (Table S1), searching for already known antigens. Fisher's exact test (two-tailed) was used to test for significant enrichment of antigens found in our study and the SEREX antigens present in the UniPEX library. Statistical tests were done using RStudio software (version 0.97.551).

TAA literature review and comparison with CRC genes

To identify DIRAGs from this study that are possibly acknowledged as TAAs, a table of the known CRC TAAs was compiled from several publications [5,31,32,41–45]. This list (Table S5) was intersected with the DIRAGs (Table S1). Additionally, a compilation of the most recognized CRC genes listed in the literature was generated from reviews [46,48,76]. To further complement this list, we added information of the cancer census gene mutation data from the Catalogue Of Somatic Mutations In Cancer (COSMIC) website, <http://www.sanger.ac.uk/cosmic> [77]. Only somatic gene mutations detected in CRC were considered. With the found CRC genes, a table was generated (Table S6) and the overlapping antigens with those identified in our experiments were examined.

Authors' contributions

JL performed the study, analyzed and interpreted the data, and wrote the manuscript. JL, KS and PeH processed the protein microarrays and performed data analysis. IG interpreted data and helped with the manuscript writing. AG, PhH, and SB conducted biobanking, sample, and data management, and helped with study design. GL and KM coordinated patient recruitment

and clinical examination of patients. AW designed the study, coordinated, interpreted, supervised and corrected the manuscript. All authors read and approved the final manuscript.

Competing interest

The authors have declared no competing interests.

Acknowledgments

We want to thank Ronald Kulovics and Michael Stierschneider (AIT) for protein purification and array printing, and Eva Kristandl (Landeskrankenhaus Wiener Neustadt) for biobanking of plasma samples. This work was supported by the Life Science Krems Fund (Project No. 30), Jubiläumsfonds of the Austrian National Bank (Project No. 15192), and the Vienna Science and Technology Fund (Project No LS11-026) of Austria.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.gpb.2017.10.002>.

References

- [1] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–917.
- [2] Edwards BK, Ward E, Kohler BA, Ehemann C, Zaubler AG, Anderson RN, et al. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010;116:544–73.
- [3] Luna Coronell JA, Syed P, Sergelen K, Gyurján I, Weinhäusel A. The current status of cancer biomarker research using tumour-associated antigens for minimal invasive and early cancer diagnostics. *J Proteomics* 2012;76:102–15.
- [4] Yamashita K, Watanabe M. Clinical significance of tumor markers and an emerging perspective on colorectal cancer. *Cancer Sci* 2009;100:195–9.
- [5] Barderas R, Babel I, Diaz-Urriarte R, Moreno V, Suárez 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.
- [6] Jiang X, Shapiro DJ. The immune system and inflammation in breast cancer. *Mol Cell Endocrinol* 2014;382:673–82.
- [7] Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural innate and adaptive immunity to cancer. *Annu Rev Immunol* 2011;29:235–71.
- [8] Syed P, Gyurján I, Kriegner A, Vierlinger K, Singer CF, Rappaport-Fürhauser C, et al. *In silico* design and performance of peptide microarrays for breast cancer tumour-auto-antibody testing. *J Mol Biochem* 2012;1:129–43.
- [9] Backes C, Ludwig N, Leidinger P, Harz C, Hoffmann J, Keller A, et al. Immunogenicity of autoantigens. *BMC Genomics* 2011;12:340.
- [10] Ullah MF, Aatif M. The footprints of cancer development: cancer biomarkers. *Cancer Treat Rev* 2009;35:193–200.
- [11] Chang S, Kohrt H, Maecker HT. Monitoring the immune competence of cancer patients to predict outcome. *Cancer Immunol Immunother* 2014;63:713–9.
- [12] Ran Y, Hu H, Zhou Z, Yu L, Sun L, Pan J, et al. Profiling tumor-associated autoantibodies for the detection of colon cancer. *Clin Cancer Res* 2008;14:2696–700.
- [13] Schwenk JM, Igel U, Kato BS, Nicholson G, Karpe F, Uhlén M, et al. Comparative protein profiling of serum and plasma using an antibody suspension bead array approach. *Proteomics* 2010;10:532–40.
- [14] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845–67.
- [15] Syed P, Vierlinger K, Kriegner A, Sergelen K, Luna-Coronell J, Rappaport-Fürhauser C, et al. Evaluation of auto-antibody serum biomarkers for breast cancer screening and *in silico* analysis of sero-reactive proteins. *J Mol Biochem* 2012;1:116–28.
- [16] Diesinger I, Bauer C, Brass N, Schaefer HJ, Comtesse N, Sybrecht G, et al. Toward a more complete recognition of immunoreactive antigens in squamous cell lung carcinoma. *Int J Cancer* 2002;102:372–8.
- [17] Krackhardt AM, Witzens M, Harig S, Hodi FS, Zauls AJ, Chessia M, et al. Identification of tumor-associated antigens in chronic lymphocytic leukemia by SEREX. *Blood* 2002;100:2123–31.
- [18] Wang Y, Han KJ, Pang XW, Vaughan HA, Qu W, Dong XY, et al. Large scale identification of human hepatocellular carcinoma-associated antigens by autoantibodies. *J Immunol* 2002;169:1102–9.
- [19] Qian F, Odunsi K, Blatt LM, Scanlan MJ, Mannan M, Shah N, et al. Tumor associated antigen recognition by autologous serum in patients with breast cancer. *Int J Mol Med* 2005;15:137–44.
- [20] Stone B, Schummer M, Paley PJ, Thompson L, Stewart J, Ford M, et al. Serologic analysis of ovarian tumor antigens reveals a bias toward antigens encoded on 17q. *Int J Cancer* 2003;104:73–84.
- [21] Devitt G, Meyer C, Wiedemann N, Eichmüller S, Kopp-Schneider A, Haferkamp A, et al. Serological analysis of human renal cell carcinoma. *Int J Cancer* 2006;118:2210–9.
- [22] Vaughan HA, St Clair F, Scanlan MJ, Chen YT, Maraskovsky E, Sizeland A, et al. The humoral immune response to head and neck cancer antigens as defined by the serological analysis of tumor antigens by recombinant cDNA expression cloning. *Cancer Immunol* 2004;4:5.
- [23] Zhong L, Ge K, Zu J, Zhao L, Shen W, Wang J, et al. Autoantibodies as potential biomarkers for breast cancer. *Breast Cancer Res* 2008;10:R40.
- [24] Scanlan MJ, Chen YT, Williamson B, Gure AO, Stockert E, Gordan JD, et al. Characterization of human colon cancer antigens recognized by autologous antibodies. *Int J Cancer* 1998;76:652–8.
- [25] Lichtenfels R, Kellner R, Atkins D, Bukur J, Ackermann A, Beck J, et al. Identification of metabolic enzymes in renal cell carcinoma utilizing PROTEOMEX analyses. *Biochim Biophys Acta* 2003;1646:21–31.
- [26] He Y, Wu Y, Mou Z, Li W, Zou L, Fu T, et al. Proteomics-based identification of HSP60 as a tumor-associated antigen in colorectal cancer. *Proteomics Clin Appl* 2007;1:336–42.
- [27] Hamrita B, Chahed K, Kabbage M, Guillier CL, Trimeche M, Chaïeb A, et al. Identification of tumor antigens that elicit a humoral immune response in breast cancer patients' sera by serological proteome analysis (SERPA). *Clin Chim Acta* 2008;393:95–102.
- [28] Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics* 2004;3:367–78.
- [29] Pan S, Shi M, Jin J, Albin RL, Lieberman A, Gearing M, et al. Proteomics identification of proteins in human cortex using multidimensional separations and MALDI tandem mass spectrometer. *Mol Cell Proteomics* 2007;6:1818–23.

- [30] Coghlin C, Murray GI. Biomarkers of colorectal cancer: recent advances and future challenges. *Proteomics Clin Appl* 2015;9:64–71.
- [31] Kijanka G, Hector S, Kay EW, Murray F, Cummins R, Murphy D, et al. Human IgG antibody profiles differentiate between symptomatic patients with and without colorectal cancer. *Gut* 2010;59:69–78.
- [32] Babel I, Barderas R, Diaz-Uriarte R, Martínez-Torrecuadrada JL, Sánchez-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.
- [33] Ramachandran N, Raphael JV, Hainsworth E, Demirhan G, Fuentes MG, Rolfs A, et al. Next-generation high-density self-assembling functional protein arrays. *Nat Methods* 2008;5:535–8.
- [34] Tomizaki K, Usui K, Mihara H. Protein-detecting microarrays: current accomplishments and requirements. *ChemBiochem* 2005;6:782–99.
- [35] Huang Y, Zhu H. Protein array-based approaches for biomarker discovery in cancer. *Genomics Proteomics Bioinformatics* 2017;15:73–81.
- [36] Luna-Coronell JA, Vierlinger K, Gamperl M, Hofbauer J, Berger I, Weinhäusel A. The prostate cancer immunome: in silico functional analysis of antigenic proteins from microarray profiling with IgG. *Proteomics* 2016;16:1204–14.
- [37] Roskopf S, Gyurján I, Luna-Coronell JA, Vierlinger K, Singer CF, Rappaport C, et al. The pre-analytical processing of blood samples for detecting biomarkers on protein microarrays. *J Immunol Methods* 2015;418:39–51.
- [38] Krämer A, Green J, Pollard Jr J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2014;30:523–30.
- [39] Liu Z, Sun C, Zhang Y, Ji Z, Yang G. Phosphatidylinositol 3-Kinase-C2 β inhibits cisplatin-mediated apoptosis via the Akt pathway in oesophageal squamous cell carcinoma. *J Int Med Res* 2011;39:1319–32.
- [40] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000;25:25–9.
- [41] Liu W, Wang P, Li Z, Xu W, Dai L, Wang K, et al. Evaluation of tumour-associated antigen (TAA) miniarray in immunodiagnosis of colon cancer. *Scand J Immunol* 2009;69:57–63.
- [42] Carpelan-Holmström M, Haglund C, Lundin J, Järvinen H, Roberts P. Pre-operative serum levels of CA 242 and CEA predict outcome in colorectal cancer. *Eur J Cancer* 1996;32A:1156–61.
- [43] Scanlan MJ, Welt S, Gordon CM, Chen YT, Gure AO, Stocker E, et al. Cancer-related serological recognition of human colon cancer: identification of potential diagnostic and immunotherapeutic targets. *Cancer Res* 2002;62:4041–7.
- [44] Barderas R, Babel I, Casal JI. Colorectal cancer proteomics, molecular characterization and biomarker discovery. *Proteomics Clin Appl* 2010;4:159–78.
- [45] Chan CC, Fan CW, Kuo YB, Chen YH, Chang PY, Chen KT, et al. Multiple serological biomarkers for colorectal cancer detection. *Int J Cancer* 2010;126:1683–90.
- [46] Al-Sohaily S, Biankin A, Leong R, Kohonen-Corish M, Warusavitarne J. Molecular pathways in colorectal cancer: pathways of colorectal carcinogenesis. *J Gastroenterol Hepatol* 2012;27:1423–31.
- [47] Jones S, Li M, Parsons DW, Zhang X, Wesseling J, Kristel P, et al. Somatic mutations in the chromatin remodeling gene *ARID1A* occur in several tumor types. *Hum Mutat* 2012;33:100–3.
- [48] Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. *Nat Rev Cancer* 2004;4:177–83.
- [49] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74.
- [50] Negm OH, Hamed MR, Schoen RE, Whelan RL, Steele RJ, Scholfield J, et al. Human blood autoantibodies in the detection of colorectal cancer. *PLoS One* 2016;11:e0156971.
- [51] de Freitas Junior JCM, Morgado-Díaz JA. The role of N-glycans in colorectal cancer progression: potential biomarkers and therapeutic applications. *Oncotarget* 2016;7:19395–413.
- [52] Malinowsky K, Nitsche U, Janssen KP, Bader FG, Späth C, Drecoll E, et al. Activation of the PI3K/AKT pathway correlates with prognosis in stage II colon cancer. *Br J Cancer* 2014;110:2081–9.
- [53] Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell* 2012;149:274–93.
- [54] Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330–7.
- [55] Yu J, Li X, Zhong C, Li D, Zhai X, Hu W, et al. High-throughput proteomics integrated with gene microarray for discovery of colorectal cancer potential biomarkers. *Oncotarget* 2016;7:75279–92.
- [56] Emmink BL, Verheem A, Van Houdt WJ, Steller EJA, Govaert KM, Pham TV, et al. The secretome of colon cancer stem cells contains drug-metabolizing enzymes. *J Proteomics* 2013;91:84–96.
- [57] Shen H, Huang J, Pei H, Zeng S, Tao Y, Shen L, et al. Comparative proteomic study for profiling differentially expressed proteins between Chinese left- and right-sided colon cancers. *Cancer Sci* 2013;104:135–41.
- [58] Zhang J, Wang K, Zhang J, Liu SS, Dai L, Zhang JY. Using proteomic approach to identify tumor-associated proteins as biomarkers in human esophageal squamous cell carcinoma. *J Proteome Res* 2011;10:2863–72.
- [59] Paltridge JL, Belle L, Khew-Goodall Y. The secretome in cancer progression. *Biochim Biophys Acta* 2013;1834:2233–41.
- [60] Mariadason JM. HDACs and HDAC inhibitors in colon cancer. *Epigenetics* 2008;3:28–37.
- [61] Wilson AJ, Byun DS, Popova N, Murray LB, L'Italien K, Sowa Y, et al. Histone deacetylase 3 (HDAC3) and other class I HDACs regulate colon cell maturation and p21 expression and are deregulated in human colon cancer. *J Biol Chem* 2006;281:13548–58.
- [62] Nakamura T, Hamada F, Ishidate T, Anai K, Kawahara K, Toyoshima K, et al. Axin, an inhibitor of the Wnt signalling pathway, interacts with beta-catenin, GSK-3 β and APC and reduces the beta-catenin level. *Genes Cells* 1998;3:395–403.
- [63] Sue Ng S, Mahmoudi T, Li VSW, Hatzis P, Boersema PJ, Mohammed S, et al. MAP3K1 functionally interacts with Axin1 in the canonical Wnt signalling pathway. *Biol Chem* 2010;391:171–80.
- [64] Matsubara N, Yoshitaka T, Matsuno T, Ikeda M, Isozaki H, Tanaka N, et al. Multiple tumors and a novel E2F-4 mutation. A case report. *Digestion* 2000;62:213–6.
- [65] Bardhan K, Liu K. Epigenetics and colorectal cancer pathogenesis. *Cancers (Basel)* 2013;5:676–713.
- [66] Fedier A, Ruefenacht UB, Schwarz VA, Haller U, Fink D. Increased sensitivity of p53-deficient cells to anticancer agents due to loss of Pms2. *Br J Cancer* 2002;87:1027–33.
- [67] Hofer P, Baierl A, Feik E, Führlinger G, Leeb G, Mach K, et al. MNS16A tandem repeats minisatellite of human telomerase gene: a risk factor for colorectal cancer. *Carcinogenesis* 2011;32:866–71.
- [68] Hofer P, Baierl A, Bernhart K, Leeb G, Mach K, Micksche M, et al. Association of genetic variants of human telomerase with colorectal polyps and colorectal cancer risk. *Mol Carcinog* 2012;51:E176–82.
- [69] Stempfer R, Syed P, Vierlinger K, Pichler R, Meese E, Leidinger P, et al. Tumour auto-antibody screening: performance of protein microarrays using SEREX derived antigens. *BMC Cancer* 2010;10:627.
- [70] Brezina S, Soldo R, Kreuzhuber R, Hofer P, Gsur A, Weinhaeusel A. Immune-signatures for lung cancer diagnostics: evaluation of

- protein microarray data normalization strategies. *Microarrays (Basel)* 2015;4:162–87.
- [71] Preininger C, Bodrossy L, Sauer U, Pichler R, Weilharter A. ARChip epoxy and ARChip UV for covalent on-chip immobilization of *pmoA* gene-specific oligonucleotides. *Anal Biochem* 2004;330:29–36.
- [72] Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, et al. Adjustment of systematic microarray data biases. *Bioinformatics* 2004;20:105–14.
- [73] Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. Analysis of gene expression data using BRB-ArrayTools. *Cancer Inform* 2007;3:11–7.
- [74] Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 2013;41:W77–83.
- [75] Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 2005;33:W741–8.
- [76] Kudryavtseva AV, Lipatova AV, Zaretsky AR, Moskalev AA, Fedorova MS, Rasskazova AS, et al. Important molecular genetic markers of colorectal cancer. *Oncotarget* 2016;7:53959–83.
- [77] Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 2004;91:355–8.