Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2011, **7**, 2997–3005

www.rsc.org/molecularbiosystems

PAPER

Proteomic comparison of colorectal tumours and non-neoplastic mucosa from paired patient samples using iTRAQ mass spectrometry†

Lucy Jankova,[a] Charles Chan,[b] Caroline L. S. Fung,[b] Xiaomin Song,[c] Sun Y. Kwun,[b] Mark J. Cowley,[d] Warren Kaplan,[d] Owen F. Dent,[e] Elie L. Bokey,[e] Pierre H. Chapuis,[e] Mark S. Baker,[c] Graham R. Robertson,[a] Stephen J. Clarke[f] and Mark P. Molloy*[c]

Received 15th June 2011, Accepted 13th July 2011
DOI: 10.1039/c1mb05236e

Quantitative mass spectrometry using iTRAQ was used to identify differentially expressed proteins from 16 colorectal cancer (CRC) tumours compared to patient-paired adjacent normal mucosa. Over 1400 proteins were identified and quantitated, with 118 determined as differentially expressed by >1.3-fold, with false discovery rate < 0.05. Gene Ontology analysis indicated that proteins with increased expression levels in CRC tumours include those associated with glycolysis, calcium binding, and protease inhibition. Proteins with reduced levels in CRC tumours were associated with loss of ATP production through: (i) reduced β-oxidation of fatty acids, (ii) reduced NADH production by the tricarboxylic acid cycle and (iii) decreased oxidative phosphorylation activity. Additionally, biosyntheses of glycosaminoglycans and proteoglycans were significantly reduced in tumour samples. Validation experiments using immunoblotting and immunohistochemistry (IHC) showed strong concordance with iTRAQ data suggesting that this workflow is suitable for identifying biomarker candidates. We discuss the uses and challenges of this approach to generate biomarker leads for patient prognostication.

Introduction

Colorectal cancer (CRC) is the second most common malignancy by incidence, and is second only to lung cancer as the most frequent cause of cancer death in the Western world.[1] The central paradigm of colorectal cancer pathogenesis in most instances is described by the *adenoma–carcinoma sequence*, where a stepwise series of cellular events driven by genetic mutations leads to the development of malignancy.[2] The development of modern genomic and proteomic techniques has enabled greater understanding of these events through holistic analysis of fluxes in gene and protein expression that are central to CRC tumourigenesis. The advent of proteomic technology provides the potential to perform global protein expression profiling of tumours, and the ability to conduct comprehensive systematic searches for new prognostic biomarkers. These techniques also provide information about the translated products of the genetic and epigenetic molecular changes, which are the effector molecules in tumourigenesis, invasion and metastasis and which may be targeted by different therapies. The traditional mainstay of proteomic studies has been two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), which has been successfully applied over many years to identify many up- and down-regulated proteins in CRC.[3–13] Examples of proteins up-regulated in tumour tissue include annexins (A1, A3, A4, A5), S100 isoforms (A6, A8, A9, A11), cytokeratins (CK-8, CK-10, CK-19), tropomyosin, vimentin, and lactate dehydrogenase. Additionally, these experiments have shown that proteins such as liver fatty acid binding protein, selenium binding protein, creatine kinase, cyclooxygenase 2, ATP synthase, carbonic anhydrase, and NADH-ubiquinone oxidoreductase are down-regulated in CRC tumour tissues. Shotgun proteomics employing liquid chromatography and multi-dimensional mass spectrometry as utilised here have some advantages over gel-based techniques

[a] *Cancer Pharmacology Unit, ANZAC Research Institute and Discipline of Medicine, The University of Sydney, Sydney, Australia*
[b] *Department of Anatomical Pathology, Concord Repatriation General Hospital, Sydney, Australia*
[c] *Australian Proteome Analysis Facility, Dept. Chemistry & Biomolecular Sciences, Macquarie University, Sydney, Australia. E-mail: mmolloy@proteome.org.au; Fax: + 61 2 9850 6200; Tel: + 61 2 9850 6218*
[d] *Peter Wills Bioinformatics Centre, The Garvan Institute of Medical Research, Sydney, Australia*
[e] *Department of Colorectal Surgery, Concord Repatriation General Hospital and Discipline of Surgery, The University of Sydney, Sydney, Australia*
[f] *Department of Medicine, Concord Repatriation General Hospital and Discipline of Medicine, The University of Sydney, Sydney, Australia*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c1mb05236e

in terms of speed, sensitivity, scope of analysis and dynamic range. A relatively new approach couples peptide mass spectrometry with isobaric tags for relative and absolute quantitation (iTRAQ-MS) reagents, which enables multiplexed, simultaneous differential measurement and identification of proteins, thereby avoiding many of the limitations of 2D-PAGE. iTRAQ-MS has been utilised in an attempt to identify biomarkers in samples from patients with endometrial,[14] head and neck,[15] low-grade breast,[16] prostate[17] and hepatocellular cancers,[18] among others, but its application to CRC has been limited. The potential value of iTRAQ-MS would initially be to provide the capacity to utilise a single technique to simultaneously characterise prognostic and predictive biomarkers for further evaluation and validation, whilst also providing insights into the molecular pathophysiology of CRC through comparison of the tumour proteome with that of adjacent mucosa.

We have applied iTRAQ-MS to conduct a comprehensive proteome profile of protein expression between CRC tumour tissue and paired normal mucosa from 16 patients (Australian Clinicopathological Staging System: Stages A–D). We identified that 118 proteins were differentially expressed by at least 1.3-fold independent of tumour stage. The expression changes of some of these proteins were confirmed using immunoblots and immunohistochemistry. We discuss the potential clinical uses and limitations of this workflow for cancer biomarker discovery.

Results

Colorectal tumour and normal mucosa profiling of all tumour stages with iTRAQ

iTRAQ-MS was used to identify proteins that were differentially expressed between fresh-frozen CRC and paired, non-neoplastic large bowel mucosa from 16 patients (Table S1, ESI†). The median age of the patients sampled was 75 years (range 60–92 years) and 44% were female. All stages of primary CRC were represented so as to include tumours from patients with varying prognoses. Specifically, we profiled tumours and corresponding normal tissue from 3 patients with ACPS A (early disease), 5 with ACPS B, 5 with ACPS C and 3 with ACPS D (advanced disease). There was a predominance of patients with right-sided tumours. An example of iTRAQ-MS protein identification and quantitation is shown in Fig. S1 (ESI†). In total, 1453 non-redundant proteins were quantitated in 2 or more of the 16 samples. However, the average number of proteins quantitated from each paired sample was approximately 850, representing comprehensive depth of proteome coverage across the tumours studied. Approximately 21% of the identified proteins showed $\geq$ 1.5-fold change in expression levels between paired CRC and adjacent mucosa. Unsupervised hierarchical clustering of the protein ratios provided an overall assessment of the heterogeneity of CRC–mucosa paired samples between these individuals (Fig. 1). Stage A samples and 4 of the 5 Stage C samples clustered together, however differences were observed amongst samples of other clinical stages. It was noted that protein expression ratios could not be obtained for all proteins in all tumours as numerous information-dependent data acquisition MS runs were required to profile all 32 samples using the

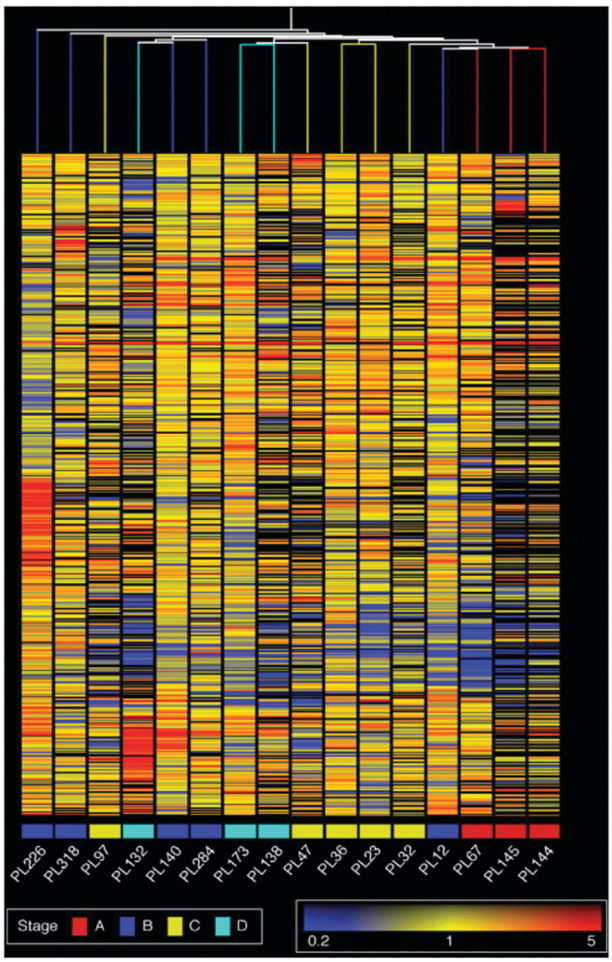

Fig. 1 Unsupervised hierarchical clustering of 1453 iTRAQ ratios observed for tumour and adjacent mucosa from 16 CRC patients. Colored scale is the representation of expression ratios between tumour and normal mucosa. Blue shades indicate proteins with reduced expression in tumours, red shades indicate proteins with increased expression in tumours, yellow shades indicate only minor changes in expression between paired samples, while black bars indicate no quantitative value obtained for the given protein. The tumour stage (A–D) is indicated in colour above the sample identifier.

iTRAQ 4-plex reagents. The "gaps" in the data sets between samples are typical of information-dependent data acquisition mass spectrometry experiments.

We conducted statistical analyses to identify proteins that were differentially expressed between tumour and paired normal mucosa independent of CRC stage. To ensure the suitability of parametric methods of analysis we first applied the Shapiro–Wilk test for normality. In addition, we inspected plots of the 160 most differentially expressed proteins, which confirmed that the proteins' ratios appeared to be derived from a single distribution. To detect differentially expressed proteins we applied an empirical Bayes moderated *t*-test,[19] with adjustment for multiple testing by the positive false discovery rate.[20] By a combination of volcano plots and sensitivity analysis (Fig. S2, ESI†), we determined that FDR < 0.05 and average ratio > 1.3 represented a good cut-off for determining differentially expressed proteins. Using these thresholds, we identified

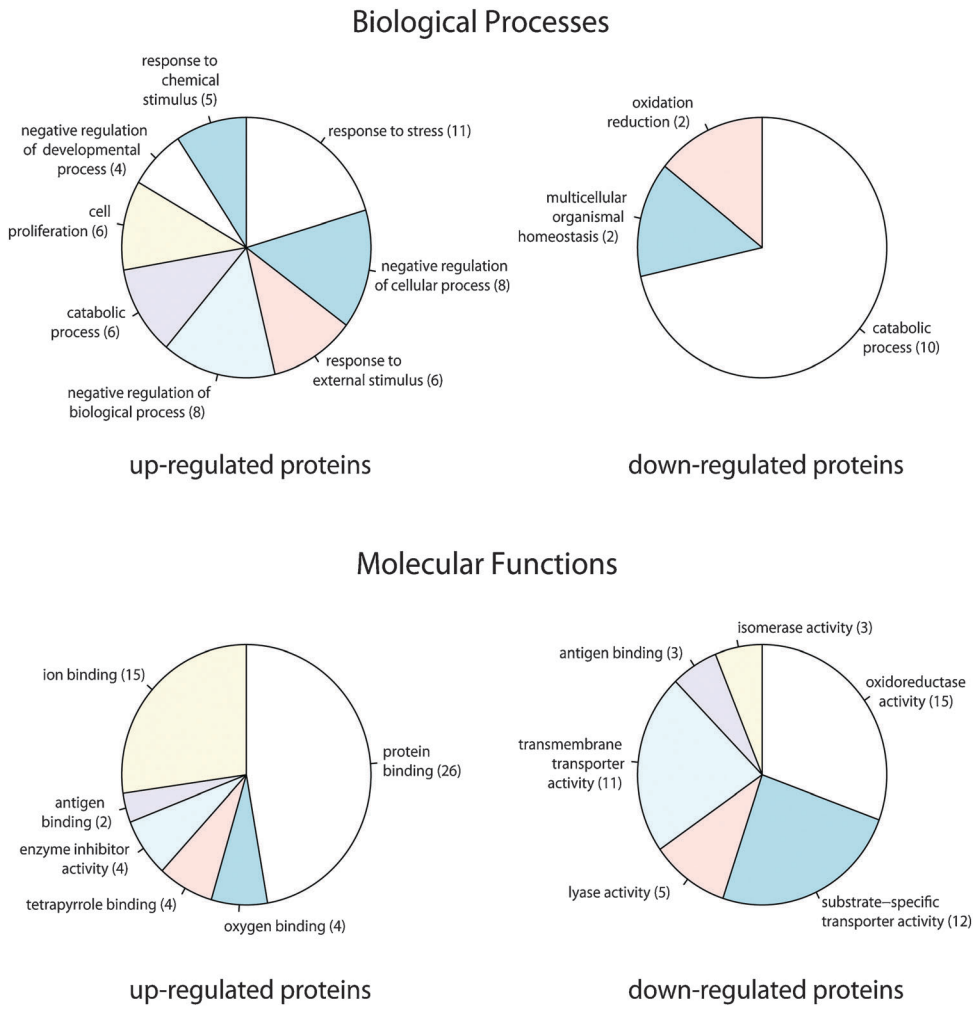

Fig. 2 Gene ontology enrichment analysis using GOminer for biological process and molecular function of differentially expressed proteins (Table S1, ESI†) using FDR < 0.15 and level 2 GO terms.

118 proteins (Table 2, ESI†), and of these, 45 proteins were elevated in tumour tissue, while 73 were repressed.

To gain a broad assessment of the types of biological processes and molecular functions that were likely to be affected by differential protein abundance, Gene Ontology (GO)[21] enrichment analysis was performed (Fig. 2). We observed striking differences in both biological processes and molecular function GO analyses for the 118 proteins differentially expressed between tumour and adjacent mucosa. A common theme amongst tumour-induced proteins was biological processes involving response to stress and stimulus, host defence response, proliferation and motility. Interestingly, annotations of the molecular functions linked to these processes were dominated by various binding activities (*e.g.* calcium-binding for S100 family proteins). Additionally, 5 proteins were involved in enzyme inhibition (*e.g.* the serpins). In contrast, proteins down-regulated in tumour compared to normal mucosa tissues were associated with anabolic and catabolic metabolism as well as energy production. Enzymes responsible for these biological processes were broadly categorised as oxidoreductases, hydrolases, and isomerases. Proteins annotated with transporter activities were also repressed.

Orthogonal verification of iTRAQ data

To evaluate the veracity of iTRAQ analysis, tissue lysates used for these experiments were evaluated by immunoblot for four representative differentially expressed proteins (namely; maspin, anterior gradient protein 2 (AGR2), transglutaminase 2 (TGM2), fatty acid binding protein liver (FABPL)) (Fig. 3). In the majority of these cases there was consistency between the observed iTRAQ ratios and the abundance of the proteins determined by immunoblotting.

Further verification was conducted by immunohistochemistry of 11 proteins detected by iTRAQ-MS that were of interest to us. Of these, iTRAQ showed that five proteins were differentially expressed between tumour tissue and mucosa (maspin, carcinoembryonic antigen (CEA), cytokeratin 20 (CK-20), galectin-3, S100A8/A9), while 6 others showed variable abundance across patients as measured by iTRAQ (β-catenin, AGR2, stat-1, caldesmon, TGM2 (Fig. S3, ESI†), HLA class II histocompatibility antigen (HLA-DR)).

IHC enabled the location of protein expression within the tumours to be confirmed. The expression of several proteins was demonstrated in the neoplastic epithelial cells (β-catenin, maspin, AGR2, TGM2, CEA, CK-20), while other proteins

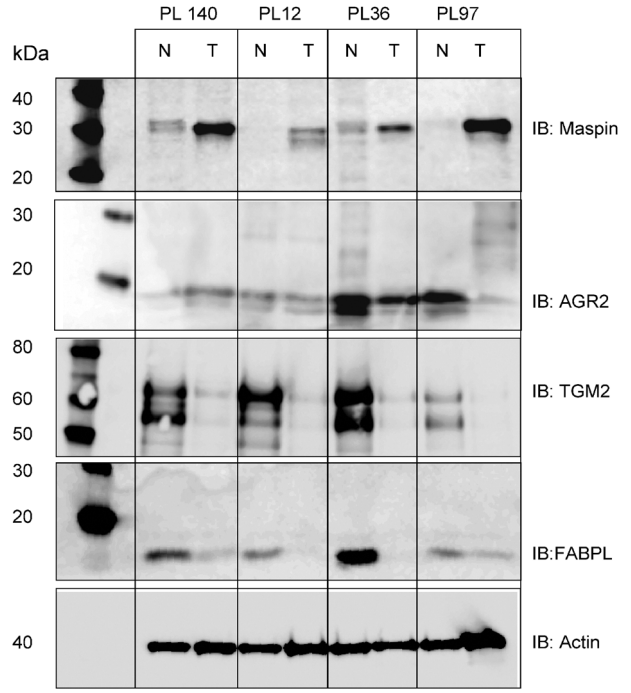

Fig. 3 Immunoblots of tumour and normal mucosa samples used for iTRAQ analysis. (N) refers to normal mucosa, (T) refers to tumour. β-Actin was used as loading control.

were expressed exclusively in stromal mesenchymal cells (caldesmon), in inflammatory cells (S100A8/9), or occasionally in both epithelial and non-epithelial cells (HLA-DR, stat-1, galectin-3). Some proteins were more highly expressed in cells at the advancing tumour front (β-catenin, maspin, HLA-DR). Fig. 4 illustrates the localisation of expression of selected proteins in various cellular compartments.

The 11 proteins examined by IHC were observed at different frequencies across the 16 sample cohort using iTRAQ-MS; from AGR2 and S100A8/A9 (quantitated in all 16 samples) to TGM2 and HLA-DR (quantitated in 4 samples). Fig. 5 displays a bar chart of the concordance between iTRAQ quantitative data and IHC quantitation (see Table S3, ESI†). The median concordance between iTRAQ and IHC was approximately 70%. Only for two proteins where the number of iTRAQ measurements was low was the concordance $\leq$ 25%. Given that iTRAQ quantitation was derived from lysates of the epithelial-cell-rich superficial part of the tumours and IHC quantitation was obtained from whole tissue sections including the deep advancing tumour front, there was remarkably solid concordance between the ability of these techniques to identify differential protein expression in the respective samples.

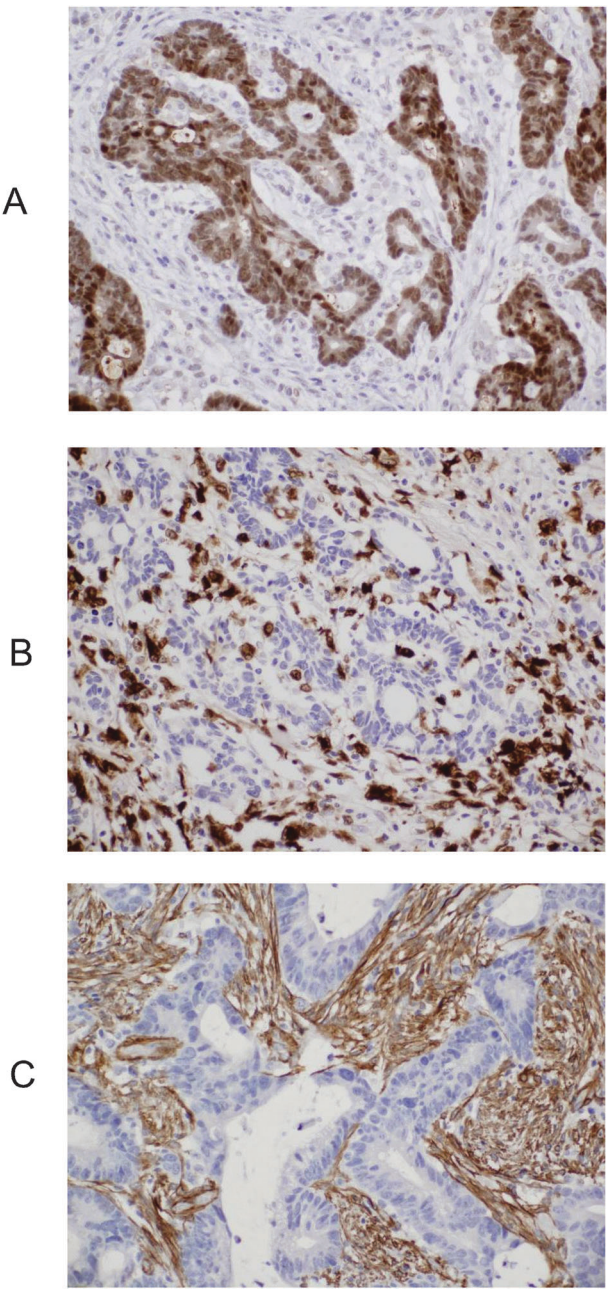

Fig. 4 Immunohistochemical localisation of protein expression in CRC. (A) Maspin is expressed in the nuclei and cytoplasm of the neoplastic epithelial cells (original magnification 200×). (B) S100A8/9 is expressed in tumour-associated inflammatory cells (histiocytes and neutrophils), but not in epithelial cells (original magnification 200×). (C) Caldesmon is expressed in tumour-associated mesenchymal stromal cells, but not in epithelial cells (original magnification 200×).

Discussion

Proteomic profile of CRC

iTRAQ-MS profiling of CRC and matched normal mucosa identified over 1400 proteins (95% confidence, ProtScore $\geq$ 1.3) across 16 CRC patient samples spanning all disease stages. Differentially expressed proteins were determined *via* a moderated *t*-test; a widely used analytical approach in the field of gene-expression microarray analysis, which has recently been shown to be superior than *t*-tests for the analysis of proteomic data.[22] 118 proteins with an average tumour *vs.* mucosa ratio of >1.3, and FDR < 0.05 were declared as differentially expressed, and chosen for further study (Table S2 and Fig. S2, ESI†). The 1.3-fold change in expression was chosen instead of a higher value to compensate for the compression of iTRAQ ratios that leads to under-estimation of fold changes. This phenomenon is known to occur as a result of background ions that are co-isolated during selection of the targeted

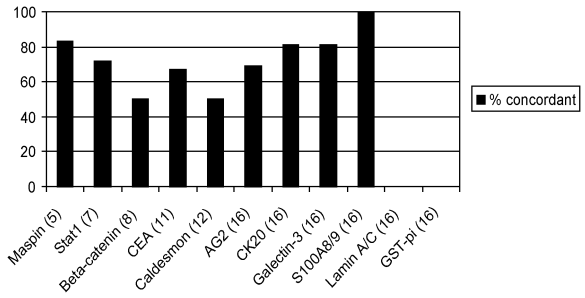

Fig. 5 Concordance of iTRAQ and IHC data for 11 proteins normalised to the number of iTRAQ observations. Numbers in parentheses is the number of patient samples where the protein was observed using iTRAQ. The maximum occurrence is 16.

iTRAQ labelled precursor.[23] Fig. 1 illustrates that the expression ratios between tumour and normal mucosa of over 1400 proteins were insufficient to accurately cluster all patient samples based on clinicopathological staging. While some reports using gene expression array profiling show good correlations with CRC stage, other reports mixed clustering of mRNA levels and pathological stage.[24–26] The absence of complete clustering here may be attributed to the limited number of samples studied per stage, the limited overlap of ratios obtained from sample to sample in the information-dependent acquisition LC/MS runs, or true inter-patient heterogeneity in protein abundance. Improved sample multiplexing offered by an 8-plex iTRAQ reagent would assist in addressing this problem.[27]

A global observation from this study is that iTRAQ analysis revealed many of the proteins commonly reported from comparative investigations using 2DE of CRC and matched normal tissue. For instance, proteins observed here as elevated in CRC and previously reported include lactate dehydrogenase, annexin-3, maspin, S100 calcium-binding protein family isoforms A6, A8, A9, A10, A11. An additional isoform, S100P has recently been shown to be elevated at the mRNA level in flat adenoma of the colon[28] and at the protein level in colorectal adenoma,[29] and we now confirm that this protein is also elevated in CRC. While epithelial cells do express S100 family proteins they are also commonly associated with leukocytes. The elevated expression of S100 proteins that we have observed is partly attributed to invasion of CRC by immune cells. This notion is supported by IHC of S100A8/A9 which showed strong staining of histiocytes and some neutrophils in the CRC tissue (Fig. 4). Furthermore, elevated levels of neutrophil elastase, neutrophil α-defensin, lymphocyte-specific protein-1 and macrophage migration inhibitory factor in human CRC compared to normal bowel tissue provide further evidence of a significant immune response in CRC.[30,31]

Examples of proteins previously reported as repressed in CRC and identified here include ATP synthase components, cytochrome c oxidase subunits, carbonic anhydrase 2, liver fatty acid binding protein, selenium binding protein. We observed reduced levels of proteins that are crucial for: (i) ATP generation through β-oxidation of fatty acids (*i.e.* FABPL, fatty acid binding protein epithelial, acetoacetyl-CoA thiolase, 3-ketoacyl CoA thiolase, hydroxyacyl-CoA dehydrogenase and butyryl-CoA dehydrogenase), (ii) supply of NADH by the tricarboxylic acid cycle (fumarase, aconitase, succinyl-CoA synthetase) and (iii) oxidative phosphorylation. Combined with the observations of increases in tumour levels of glycolytic enzymes such as lactate dehydrogenase and pyruvate kinase, this supports the notion of a shift in energy production in CRC towards anaerobic glycolytic metabolism.[10] We observed that the Na^+–K^+-ATPase α-subunit and β-subunit were repressed in CRC tumour tissue compared to normal mucosa. Mazzanti *et al.* had previously observed that the Na^+–K^+-ATPase β-subunit was decreased in CRC tumour tissue, while others have shown that this subunit is repressed in renal clear cell carcinoma[32] and early stages of bladder cancer.[33] Our data confirm that the α1-subunit is similarly repressed in CRC. Further evidence of the shift from aerobic metabolism is supported by the repression in CRC tissue samples of numerous protein subunits associated with the mitochondrial electron transport chain including NADH-ubiquinone oxidoreductase B15 subunit, sulfide:quinone oxidoreductase, cytochrome c oxidase subunits II, Va, Vb, VIc, VIIa-L, and ubiquinol-cytochrome-c reductase complex core proteins I and II.

iTRAQ analysis revealed the repression of numerous other proteins that are of central importance to glycosaminoglycan and proteoglycan biosynthesis. The oxidoreductase UDP-glucose 6-dehydrogenase that catalyses the production of UDP-glucuronate is reduced in CRC. Similarly, UDP-glucuronosyltransferase 2B17, an isoform responsible for glucuronidation was significantly repressed by more than 2.5-fold in six of the nine patients where this protein was detected. It was less than 1.5-fold repressed in the remaining 3 patients. It is interesting to speculate on the prognostic significance of UDP-glucuronosyltransferase levels given that glucuronidation is one of the chief pathways for xenobiotic metabolism, and patients with locally advanced or metastatic tumours usually receive chemotherapy. Glycosaminoglycans are added to specific protein cores to form proteoglycans. Interestingly, the proteoglycans lumican and mimecan were both observed at decreased levels in CRC tumours. Lumican is expressed in the fibroblasts and neural cells of normal mucosa, but not in epithelial cells. In advanced CRC it has been reported as diffusely located in the cytoplasm in ~60% of CRC, with increased levels observed at the invasive front and an association with survival.[34] The decrease in lumican expression observed here may be attributed to reduced concentration of lumican-positive cells in the tumour compared to normal mucosa, or impairment of the pathway needed for glycosaminoglycan biosynthesis as discussed above. Mimecan was significantly reduced in CRC tissue compared to normal mucosa in our samples, which is consistent with the recent report by Wang *et al.*,[35] that showed loss of mimecan in colorectal adenomas and CRC tumours when measured using densitometry of immunoblots. Reduced expression in CRC reported here of sialic acid synthase, the enzyme responsible for production of the most common sialic acid, *N*-acetylneuraminic acid, would reduce the available pool of sialic acid used for capping cell surface expressed glycoconjugates. While not absolute, alterations in the extent of cell surface sialylation has been correlated with metastasis in various cancers.[36–38]

Correlation of iTRAQ and immunodetection

Immunoblotting was used with a small number of samples to confirm some iTRAQ measured changes in protein levels (Fig. 3).

We then evaluated each of the 16 samples by IHC for 12 proteins using conventional tissue sections of normal mucosa and tumour, including superficial and deep advancing tumour front (Fig. 4, Fig. S3 (ESI†)). The strong level of concordance obtained between the results of iTRAQ and IHC quantitation is encouraging (Fig. 5). The limited discrepancies could be accounted for by several technical and biological factors. iTRAQ quantitation depends on the detection of relative amounts of peptides cleaved from denatured, soluble portions of proteins, while IHC depends on the affinity of antibodies to antigenic domains of whole proteins in formalin-fixed paraffin-embedded tissue. The antigenicity of these domains may be altered by cross-linking and masking during fixation and processing, and is further modified by antigen retrieval procedures performed during the staining process.

Identification of protein biomarkers using mass spectrometry relies on the extraction of proteins from fresh tissues, and the efficiency of this depends partly on the physical properties of the proteins and the cellular compartment in which they reside. The efficiency of extraction of low molecular weight cytoplasmic proteins such as the S100 proteins is likely to be greater than large relatively insoluble membrane and cytoskeleton-associated proteins. These include caldesmon which is associated with the cytoskeleton of stromal mesenchymal cells, and CK-20 which is associated with the membrane and cytoskeleton of colonic mucosal epithelial cells.

The ability to detect a difference in protein expression is dependent both on absolute levels of expression and the relative (-fold) difference in expression of the protein in question. Proteins present at high absolute levels and or showing large-fold differences are most easily detected by either of the two methods used in this study. S100A8/A9 consistently showed concordant elevated expression in tumour tissue by both techniques. Obtaining reliable iTRAQ quantitation on proteins that are expressed in lower amounts (*e.g.* TGM2) is problematic, whereas signal amplification using IHC overcomes this problem. We have previously shown that peptides with low iTRAQ signal intensity are associated with high variance in replicate experiments.[39]

It should be recognised that tumour is composed not only of transformed malignant epithelial cells, but also surrounding stromal mesenchymal cells (fibroblasts, myofibroblasts) and inflammatory cells (histiocytes, neutrophils, lymphocytes), and that these components may also be altered in the diseased state. For those proteins that are expressed in more than one cell type, the proportion of the different cell types present in tumour and normal tissue influences the overall expression level. It is important to note that iTRAQ quantitation was derived from lysates of the epithelial cell-rich superficial part of the tumours and IHC quantitation was obtained from whole tissue sections including the deep advancing tumour front. This circumstance alone is a factor that may greatly influence data concordance between iTRAQ and IHC.

Within individual cells themselves, the proteins may be localised to different cellular compartments (nucleus, cytoplasm, cell membrane, and extracellular). The translocation of certain proteins from one cellular compartment to another (*i.e.* cytoplasm to nucleus) may have important functional consequences; however this may not result in a change in absolute level of expression. An example of this is the nuclear translocation of β-catenin observed at the deep advancing tumour front in several tumours. This can only be practically identified using IHC.

Prognostic biomarker discovery

One important application of comprehensive proteome profiling of tumours is to identify candidate biomarkers of prognostication. This requires linking analysed samples with patient survival data. For rigorous analysis such an approach requires that the specimens used for proteomic profiling were collected some 5 years prior, to allow sufficient time for survival data to be accumulated. While it is a common practice to archive tumour sections for pathological investigations, the collection of specimens suitable for proteomics requires more forethought as analysis of fresh tissue is optimal. Given that the availability of fresh frozen specimens with appropriate survival data is often limiting it will be of interest to prospectively determine if the proteins that are differentially expressed between CRC and normal mucosa have value as prognostic biomarkers. Perhaps a quicker path to prognostic biomarker discovery will come from future analyses with iTRAQ-MS profiling of biobanked fresh frozen specimens that have been dichotomised based on overall survival.

In the current study the differentially expressed proteins can only be considered candidate biomarkers until orthogonally validated. Defining an appropriate validation strategy is important because it is only pragmatic and cost effective to conduct profiling studies such as iTRAQ-MS with small cohort sizes. It should be noted that our study using specimens from 16 patients (32 samples in total) is quite large for a mass spectrometry driven study when compared with peer research reported in the proteomic literature. Orthogonal methods that can be conducted in large-scale such as IHC using tumour microarrays present one solution for biomarker validation. Using this approach with many hundreds of independent samples it is possible to gain high confidence as to the utility of the candidate biomarker by increasing statistical power. While iTRAQ-MS is not well suited to the analysis of hundreds of samples due to cost and time pressures, we have nonetheless shown that interesting insights into cancer cell biology and promising lead biomarker candidates can be readily discovered with this approach.

Experimental

CRC patients

Information on consecutive patients having a resection for colorectal cancer performed by members of the Concord Hospital Department of Colorectal Surgery have been entered into a prospective computer database since 1971[40,41] which has the approval of the South Western Sydney Health Area Ethics Committee. Since 1981 all operations have followed a standardized procedure[42] and data recording has been supervised by a single surgeon (P.C.).

Pathological examination of resected specimens followed a standard protocol[40,41,43] and over 90% of specimens before 2001 were reported on by a single histopathologist (R. C. Newland)

and subsequently by C.C. Only adenocarcinomas (including mucinous and signet ring carcinomas) were included. Where multiple tumours were present, only the lesion with the most advanced stage was included. All pathological characteristics analysed were examined for in every specimen and their presence or absence recorded explicitly. There were no missing data on any variable. Tumours were staged according to the Australian Clinicopathological Staging System (ACPS) for colorectal cancer which accommodates sub-stages compatible with other clinicopathological staging systems such as TNM.[44]

Tissue samples

Tissue samples used for proteomics analysis were collected during surgery at the Concord Hospital Department of Colorectal Surgery. Fresh tumour tissue representative of all clinicopathological stages and corresponding paired normal mucosal samples (Table S1, ESI†) was harvested using a dissecting microscope. Mucosa was taken from the resection margin furthest away from the tumour (approx. 10 cm). The specimens were washed briefly in phosphate-buffered saline, and then stored in nitrogen vapour at −170 °C. Expression and subcellular localization of proteins was determined by immunohistochemistry in paraffin-embedded normal colorectal mucosa and CRC tissues from the same patients. The samples were appropriately used under the approval of the South Western Sydney Health Area Ethics Committee.

iTRAQ labelling

Thawed tissues were disrupted with a Teflon coated glass homogenizer in 1 ml ice cold lysis buffer (10 mM NaF, 15 mM NaCl, 1 mM sodium orthovanadate, 20 mM HEPES pH 7.5, 1 mM EDTA, 1 mM EGTA, 0.15% (w/v) SDS, 0.1 mM PMSF and protease inhibitor cocktail (Roche, Indianapolis, IN)). Tissue homogenates were centrifuged at 13 200*g* for 15 min at 4 °C and the relative protein concentration in the supernatant determined by UV absorbance. Equal amounts of protein (100 μg) in whole lysates from each sample were reduced, alkylated and processed with trypsin. Samples were derivatized using a 4-plex iTRAQ reagent kit (Applied Biosystems, Foster City, CA) as per the manufacturer's protocol using previously reported modifications.[39] Each iTRAQ experiment consisted of two patient's tumours and two corresponding paired normal mucosa samples.

Mass spectrometry

Strong cation exchange chromatography and nanoLC/MS/MS using a QStar XL mass spectrometer were as previously reported.[39]

Protein identification and quantitation

MS/MS data were analysed against the SWISS-PROT *Homo sapiens* database (v51.0, 14 987 entries) using ProteinPilot v1.0 (Applied Biosystems) which utilises the Paragon search algorithm for peak-picking and database matching.[45] Data were searched considering cysteine alkylation, use of trypsin, allowing for biological modifications and using the thorough identification search option. All reported data were based on 95% confidence for protein identification as determined by ProtScore $\geq$ 1.3. The default ion intensity for conducting quantitation was 40 counts. Data were normalized for loading anomalies by bias corrections calculated using ProteinPilot, and ratios log transformed for further statistical analysis.

Statistical analysis of iTRAQ quantitation

The Shapiro–Wilk test for normality was applied to the ratios from proteins observed in at least two patient samples. Proteins were tested for consistent differential expression using an empirical Bayes moderated *t*-test as implemented in the limma library, version 2.14[19] from Bioconductor version 2.2[46] in R version 2.7.1. Briefly, to estimate the magnitude of effect of colon cancer compared to normal, regardless of the tumour stage, a simple linear model was fitted to the protein ratios for each protein with at least one observation. The moderated *t*-statistic was calculated for each protein and *p*-values were corrected for multiple testing by the positive false discovery rate (FDR).[20] Following these criteria the data reported in Table S1 (ESI†) required detection in at least 2 patients regardless of tumour stage, and differential expression of $\geq$ 1.3 fold between CRC and normal mucosa.

Data visualisation

Hierarchical clustering using average linkage was performed using GeneSpring GX 7.3.1 (Agilent, Santa Clara, CA). Gene Ontology (GO) enrichment analyses were performed using the web-based, high throughput GOminer,[47] by uploading the list of differentially expressed proteins and comparing it to an automatically generated list of all proteins, sampled from the known human proteome. One hundred (100) permutations were performed to estimate the FDR, and results reported were limited to FDR $<$ 0.15. GOminer results were imported into R, and the statistically significant level 2 GO terms were selected for plotting in π-charts.

Immunoblotting

Tumour and mucosa lysates were separated by SDS-PAGE, transferred to nitrocellulose and immunoblotted with the relevant primary antibody (see below) at 1 : 1000 dilution. β-Actin was used as a loading control at 1 : 10 000.

Immunohistochemistry

Immunohistochemistry (IHC) of 12 proteins was conducted on 4 μm sections of formalin-fixed, paraffin-embedded CRC and normal mucosal tissue from the same 16 patients that were analysed by iTRAQ. Antibodies were used at the following dilutions. Transglutaminase 2 (1 : 600; Abcam, ab2972), anterior gradient 2 (1 : 600; Abcam, ab43043), β-catenin (1 : 200; Abcam, ab2365), caldesmon (1 : 500; Abcam, ab45691), maspin (1 : 400; BD Pharmingen, 554292, Clone G167-70), galectin-3 (1 : 100; Novocastra, NCL-GAL3, Clone 9C4), CEA (1 : 2000; Sigma, C-2331, Clone C6G9), cytokeratin 20 (1 : 100; Dako, M7019, Clone Ks20.8), HLA DR (1 : 1000; Abcam, ab6339), STAT1 (1 : 500; Abcam, ab31369), S100A8 + S100A9 (1 : 50; Abcam, 17050, Clone27E10).

All stains were performed by a direct immunoperoxidase method using a horseradish peroxidase-labelled polymer and visualized using a diaminobenzidine chromogen.

Briefly, deparaffinized sections underwent heat-induced epitope retrieval (HIER) in a pressure cooker (Decloaker, Biocare Medical, Concord CA) for 5 min. After cooling, the slides were incubated in normal goat serum for 10 min before the primary antibody was applied. The pretreatment for STAT1 antibody was performed in the Decloaker using Tris EDTA buffer. For S100A8 + S100A9 antibody, an enzymatic pre-treatment was performed (Dako Proteinase K Ready-to-use). After overnight incubation using Sequenza racks (Thermo Scientific), the slides were incubated with the polymeric horseradish peroxidase detection system (LabVision TL-060-HL) for 30 min. For HLA DR, after washing unbound antibody, sections were treated with polyclonal rabbit anti-rat immunoglobulin conjugated with HRP.

Examination of IHC slides

IHC staining was assessed by an experienced pathologist (C.C.) for both intensity and distribution (to account for non-homogeneous staining) using a semi-quantitative scoring system. Staining intensity was graded as follows: 0 = absent, 1 = low, 2 = moderate, and 3 = high intensity staining. Where there was a significant difference in staining intensity across a tumour or normal tissue, both a primary (A) and a secondary (B) staining pattern were recorded. A composite intensity score was derived by the following formula: intensity $A \times A\%$ + intensity $B \times B\%$. Relative protein expression was subsequently derived as the ratio of the composite intensity scores of paired tumour and normal tissue: upregulated = intensity tumour/intensity normal > 1.25, downregulated = intensity tumour/intensity normal < 0.75, and unchanged = intensity tumour/intensity normal 0.75–1.25.

Role of personnel

Study design: CC, ELB, PHC, MSB, GRR, SJC, MPM.
Provision of clinical specimens: ELB, PHC.
Performed key research tasks: LJ, CC, CLSF, XS, SYK.
Data analysis and interpretation: LJ, CC, MJC, GRR, MPM.
Statistical analysis: MJC, WK, OFD.
Wrote the paper: CC, MJC, PHC, MSB, GRR, SJC, MPM.

Abbreviations

AGR2	anterior gradient protein 2
ACPS	Australian Clinicopathological Staging System
CEA	carcinoembryonic antigen
CK	cytokeratin
CRC	colorectal cancer
FDR	false discovery rate
FABPL	fatty acid binding protein liver
TGM2	transglutaminase 2
IHC	immunohistochemistry

Acknowledgements

Before 2001 RC Newland examined and reported on over 90% of the surgical specimens and reviewed the remainder. This research was supported by the Cancer Institute NSW through its Translational Program Grant awarded to Sydney Collaborative Research Into Proteomic Testing for Colorectal Cancer (SCRIPT). M.P.M. is the recipient of a NHMRC Career Development Award (434227). Aspects of this research were facilitated by access to the Australian Proteome Analysis Facility funded by the Australian Government's NCRIS program.

References

1 A. Jemal, R. C. Tiwari, T. Murray, A. Ghafoor, A. Samuels, E. Ward, E. J. Feuer and M. J. Thun, *Ca-Cancer J. Clin.*, 2004, **54**, 8–29.
2 B. Vogelstein, E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, Y. Nakamura, R. White, A. M. Smits and J. L. Bos, *N. Engl. J. Med.*, 1988, **319**, 525–532.
3 J. Stulik, K. Koupilova, J. Osterreicher, J. Knizek, A. Macela, J. Bures, P. Jandik, F. Langr, K. Dedic and P. R. Jungblut, *Electrophoresis*, 1999, **20**, 3638–3646.
4 J. Stulik, L. Hernychova, S. Porkertova, J. Knizek, A. Macela, J. Bures, P. Jandik, J. I. Langridge and P. R. Jungblut, *Electrophoresis*, 2001, **22**, 3019–3025.
5 D. B. Friedman, S. Hill, J. W. Keller, N. B. Merchant, S. E. Levy, R. J. Coffey and R. M. Caprioli, *Proteomics*, 2004, **4**, 793–811.
6 U. J. Roblick, D. Hirschberg, J. K. Habermann, C. Palmberg, S. Becker, S. Kruger, M. Gustafsson, H. P. Bruch, B. Franzen, T. Ried, T. Bergmann, G. Auer and H. Jornvall, *Cell. Mol. Life Sci.*, 2004, **61**, 1246–1255.
7 T. Tomonaga, K. Matsushita, S. Yamaguchi, M. Oh-Ishi, Y. Kodera, T. Maeda, H. Shimada, T. Ochiai and F. Nomura, *Clin. Cancer Res.*, 2004, **10**, 2007–2014.
8 P. Alfonso, A. Nunez, J. Madoz-Gurpide, L. Lombardia, L. Sanchez and J. I. Casal, *Proteomics*, 2005, **5**, 2602–2611.
9 X. Bi, Q. Lin, T. W. Foo, S. Joshi, T. You, H. M. Shen, C. N. Ong, P. Y. Cheah, K. W. Eu and C. L. Hew, *Mol. Cell. Proteomics*, 2006, **5**, 1119–1130.
10 R. Mazzanti, M. Solazzo, O. Fantappie, S. Elfering, P. Pantaleo, P. Bechi, F. Cianchi, A. Ettl and C. Giulivi, *Am. J. Physiol.: Gastrointest. Liver Physiol.*, 2006, **290**, G1329–G1338.
11 Y. Ma, J. Peng, W. Liu, P. Zhang, L. Huang, B. Gao, T. Shen, Y. Zhou, H. Chen, Z. Chu, M. Zhang and H. Qin, *Mol. Cell. Proteomics*, 2009, **8**, 1878–1890.
12 D. Chen, F. Chen, X. Lu, X. Yang, Z. Xu, J. Pan, Y. Huang, H. Lin and P. Chi, *Int. J. Oncol.*, 2010, **37**, 355–365.
13 L. Zhao, H. Wang, X. Sun and Y. Ding, *FEBS J.*, 2010, **277**, 4195–4204.
14 L. V. DeSouza, J. Grigull, S. Ghanny, V. Dube, A. D. Romaschin, T. J. Colgan and K. W. Siu, *Mol. Cell.Proteomics*, 2007, **6**, 1170–1182.
15 A. Matta, L. V. DeSouza, N. K. Shukla, S. D. Gupta, R. Ralhan and K. W. Siu, *J. Proteome Res.*, 2008, **7**, 2078–2087.
16 P. Bouchal, T. Roumeliotis, R. Hrstka, R. Nenutil, B. Vojtesek and S. D. Garbis, *J. Proteome Res.*, 2009, **8**, 362–373.
17 S. D. Garbis, S. I. Tyritzis, T. Roumeliotis, P. Zerefos, E. G. Giannopoulou, A. Vlahou, S. Kossida, J. Diaz, S. Vourekas, C. Tamvakopoulos, K. Pavlakis, D. Sanoudou and C. A. Constantinides, *J. Proteome Res.*, 2008, **7**, 3146–3158.
18 R. Chaerkady, H. C. Harsha, A. Nalli, M. Gucek, P. Vivekanandan, J. Akhtar, R. N. Cole, J. Simmers, R. D. Schulick, S. Singh, M. Torbenson, A. Pandey and P. J. Thuluvath, *J. Proteome Res.*, 2008, **7**, 4289–4298.
19 G. K. Smyth, *Stat. Appl. Genet. Mol. Biol.*, 2004, **3**, article3.
20 J. D. Storey and R. Tibshirani, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 9440–9445.
21 M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, *Nat. Genet.*, 2000, **25**, 25–29.
22 L. Ting, M. J. Cowley, S. L. Hoon, M. Guilhaus, M. J. Raftery and R. Cavicchioli, *Mol. Cell. Proteomics*, 2009, **8**, 2227–2242.

23 S. Y. Ow, M. Salim, J. Noirel, C. Evans, I. Rehman and P. C. Wright, *J. Proteome Res.*, 2009, **8**, 5347–5355.
24 C. M. Frederiksen, S. Knudsen, S. Laurberg and T. F. Orntoft, *J. Cancer Res. Clin. Oncol.*, 2003, **129**, 263–271.
25 A. Koehler, F. Bataille, C. Schmid, P. Ruemmele, A. Waldeck, H. Blaszyk, A. Hartmann, F. Hofstaedter and W. Dietmaier, *J. Pathol.*, 2004, **204**, 65–74.
26 J. Friederichs, R. Rosenberg, J. Mages, K. P. Janssen, C. Maeckl, H. Nekarda, B. Holzmann and J. R. Siewert, *Int. J. Colorectal Dis.*, 2005, **20**, 391–402.
27 A. Pierce, R. D. Unwin, C. A. Evans, S. Griffiths, L. Carney, L. Zhang, E. Jaworska, C. F. Lee, D. Blinco, M. J. Okoniewski, C. J. Miller, D. A. Bitton, E. Spooncer and A. D. Whetton, *Mol. Cell. Proteomics*, 2008, **7**, 853–863.
28 H. Kita, Y. Hikichi, K. Hikami, K. Tsuneyama, Z. G. Cui, H. Osawa, H. Ohnishi, H. Mutoh, H. Hoshino, C. L. Bowlus, H. Yamamoto and K. Sugano, *J. Gastroenterol.*, 2006, **41**, 1053–1063.
29 F. Lam, L. Jankova, O. F. Dent, M. P. Molloy, S. Y. Kwun, C. Clarke, P. Chapuis, G. Robertson, P. Beale, S. Clarke, E. L. Bokey and C. Chan, *Proteomics: Clin. Appl.*, 2010, **4**, 60–70.
30 Y. Naito, K. Saito, K. Shiiba, A. Ohuchi, K. Saigenji, H. Nagura and H. Ohtani, *Cancer Res.*, 1998, **58**, 3491–3494.
31 J. Galon, A. Costes, F. Sanchez-Cabo, A. Kirilovsky, B. Mlecnik, C. Lagorce-Pages, M. Tosolini, M. Camus, A. Berger, P. Wind, F. Zinzindohoue, P. Bruneval, P. H. Cugnenc, Z. Trajanoski, W. H. Fridman and F. Pages, *Science*, 2006, **313**, 1960–1964.
32 S. A. Rajasekaran, W. J. Ball, Jr., N. H. Bander, H. Liu, J. D. Pardee and A. K. Rajasekaran, *J. Urol.*, 1999, **162**, 574–580.
33 C. Espineda, D. B. Seligson, W. J. Ball, Jr., J. Rao, A. Palotie, S. Horvath, Y. Huang, T. Shi and A. K. Rajasekaran, *Cancer (N. Y.)*, 2003, **97**, 1859–1868.
34 T. Seya, N. Tanaka, S. Shinji, K. Yokoi, M. Koizumi, N. Teranishi, K. Yamashita, T. Tajiri, T. Ishiwata and Z. Naito, *Oncol. Rep.*, 2006, **16**, 1225–1230.
35 Y. Wang, Y. Ma, B. Lu, E. Xu, Q. Huang and M. Lai, *Exp. Biol. Med. (Maywood, NJ, U. S.)*, 2007, **232**, 1152–1159.
36 J. G. Steele, C. Rowlatt, J. K. Sandall and L. M. Franks, *Int. J. Cancer*, 1983, **32**, 769–779.
37 J. W. Dennis, S. Laferte, M. Fukuda, A. Dell and J. P. Carver, *Eur. J. Biochem.*, 1986, **161**, 359–373.
38 E. Gorelik, U. Galili and A. Raz, *Cancer Metastasis Rev.*, 2001, **20**, 245–277.
39 X. Song, J. Bandow, J. Sherman, J. D. Baker, P. W. Brown, M. T. McDowell and M. P. Molloy, *J. Proteome Res.*, 2008, **7**, 2952–2958.
40 R. C. Newland, P. H. Chapuis, M. T. Pheils and J. G. MacPherson, *Cancer (N. Y.)*, 1981, **47**, 1424–1429.
41 R. C. Newland, P. H. Chapuis and E. J. Smyth, *Cancer (N. Y.)*, 1987, **60**, 852–857.
42 E. L. Bokey, P. H. Chapuis, O. F. Dent, B. J. Mander, I. P. Bissett and R. C. Newland, *Dis. Colon Rectum*, 2003, **46**, 860–866.
43 N. C. Davis and R. C. Newland, *Aust. N. Z. J. Surg.*, 1983, **53**, 211–221.
44 L. P. Fielding, P. A. Arsenault, P. H. Chapuis, O. Dent, B. Gathright, J. D. Hardcastle, P. Hermanek, J. R. Jass and R. C. Newland, *J. Gastroenterol. Hepatol.*, 1991, **6**, 325–344.
45 I. V. Shilov, S. L. Seymour, A. A. Patel, A. Loboda, W. H. Tang, S. P. Keating, C. L. Hunter, L. M. Nuwaysir and D. A. Schaeffer, *Mol. Cell. Proteomics*, 2007, **6**, 1638–1655.
46 R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang and J. Zhang, *Genome Biol.*, 2004, **5**, R80.
47 B. R. Zeeberg, H. Qin, S. Narasimhan, M. Sunshine, H. Cao, D. W. Kane, M. Reimers, R. M. Stephens, D. Bryant, S. K. Burt, E. Elnekave, D. M. Hari, T. A. Wynn, C. Cunningham-Rundles, D. M. Stewart, D. Nelson and J. N. Weinstein, *BMC Bioinf.*, 2005, **6**, 168.