



## Review

# An “Omics” based survey of human colon cancer

Prashant R. Nambiar<sup>a,\*</sup>, Rishi R. Gupta<sup>b</sup>, Vikas Misra<sup>c</sup>

<sup>a</sup> Toxicologic Pathology, Drug Safety Research and Development, Pfizer Inc., Eastern point Road, Groton, CT 06340, USA

<sup>b</sup> Groton Structural Biology, Antibacterial Chemistry/Discovery Technologies, Pfizer PharmaTx Research and Development, Eastern Point Road, Groton, CT 06340, USA

<sup>c</sup> Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA

## ARTICLE INFO

### Article history:

Received 6 November 2009

Received in revised form 9 July 2010

Accepted 19 July 2010

Available online 5 August 2010

### Keywords:

Colon cancer  
 Transcriptomics  
 Metabolomics  
 Proteomics  
 Systems biology  
 Colorectal cancer

## ABSTRACT

Despite an increased understanding of the molecular pathogenesis of colorectal cancer (CRC) during the past two decades, reliable and robust biomarkers to enable screening, surveillance, and primary prevention of this disease are lacking. CRC diagnosis and therapy remain dependent upon descriptive classification and staging systems, based primarily on morphology and histology. The traditional approach of understanding complex biological systems by studying smaller, discrete units of the whole system has been less fruitful for understanding complex diseases. The implicit assumption of traditional methods, that a single or even only a few factors, play a dominant role in a complex disease might be inadequate when studying multifactorial diseases such as cancer. The burgeoning field of systems biology adopts a holistic approach, wherein the integration of individual parts of the system is sought. The cornerstone of a systems biology approach has been the development of a variety of high-throughput “omics” sciences, including genomics, transcriptomics, proteomics, and metabolomics. This review will focus on the “omics” literature in the field of sporadic human CRC and present examples of how a systems approach has been extremely useful in understanding concepts that would have been difficult to develop using traditional methods.

© 2010 Elsevier B.V. All rights reserved.

## Contents

1. Introduction .....	4
2. Transcriptomics .....	4
2.1. What is transcriptomics? .....	4
2.2. Transcriptomics in colorectal cancer .....	5
3. Proteomics .....	11
3.1. What is proteomics? .....	11
3.2. Technology involved in proteomics .....	11
3.3. Proteomics of CRC .....	11
3.4. Advantages and disadvantages .....	13
4. Metabolomics .....	13
4.1. What is metabolomics? .....	13
4.2. Technologies involved in metabolomics .....	14
4.3. Metabolomics in colorectal cancer .....	14
5. Conclusion .....	15
Conflict of interest .....	16
Acknowledgments .....	16
Appendix A. Supplementary data .....	16
References .....	16

\* Corresponding author.

E-mail address: [prashant.nambiar@pfizer.com](mailto:prashant.nambiar@pfizer.com) (P.R. Nambiar).

## 1. Introduction

Colorectal cancer (CRC) is projected to be the third most common type of cancer in the US with 148,810 new cases and 49,960 deaths expected in 2008 [1,2]. Precursor lesions of CRC, adenomas will develop in the gastrointestinal tract of approximately 50% of men and women in the industrialized western world by the seventh decade. The yearly conversion rate of these adenomas to carcinomas has been estimated to range between 0.1 and 0.25% [3].

Despite an increased understanding of molecular pathogenesis of CRC in the past two decades, reliable and robust biomarkers that would enable screening, surveillance, and primary prevention of CRC are lacking. In fact, the only available tumor marker for early detection of CRC is the serial fecal occult blood test (FOBT). The FOBT that has shown effectiveness in CRC screenings is fraught with issues of false-negative and false-positive rates [4–6]. DNA based analyses of fecal samples for a markers such as a panel of mutated genes, methylated vimentin, etc. have yielded high sensitivity, are expensive, and have not yet replaced the old warhorse FOBT [7–10]. Furthermore, the few prognostic markers such as 18q LOH (loss of heterozygosity, MSI (microsatellite instability), thymidine phosphorylase, dihydropyrimidine dehydrogenase, thymidine synthase, and K-ras are not recommended by the American Society of Clinical Oncology (ASCO) and European Group of Tumor Markers (EGTM). The only acceptable tumor marker for recurrence is the carcinoembryonic antigen (CEA) that was first discovered in 1965 [11,12]. In other words, CRC diagnosis and therapy are still dependent upon descriptive classification and staging systems based on morphology/histology [13].

One reason for the lack of reliable and robust biomarkers in CRC is that cancer is a complex multifactorial disease in which interaction of host genetics with the macro- and microenvironment plays an important role in the pathogenesis. Although, the traditional approach of studying complex biological systems (by studying smaller, simpler units of the whole system) has been successful tremendously in the 19th–20th century, the efforts have been less fruitful for complex diseases/systems. The implicit assumption of traditional methods that a single or few factors play a dominant role in a complex disease might be inadequate while studying multifactorial diseases such as cancer [14]. Therefore, the extensive amounts of molecular/genetic data in different stages of CRC represent pieces of the puzzle and not the entire puzzle.

The alternate burgeoning field of systems biology attempts to address the limitation of the conventional reductionist approach by integrating the individual parts of the system and studying them in its entirety (holistic approach) while maintaining the temporal and spatial context of the data. The power of this holistic, non-reductionist, systems approach is becoming apparent only now and numerous examples exist in the literature. A case in point is an elegant study in which the entire genome from both colon and breast cancer were sequenced using a robust high-throughput sequencing strategy. A two-dimensional view of the mutation map in these two types of tumors revealed that only a few genes were mutated in high frequency in a large proportion of tumors (“Gene mountains”) while most of the genes were mutated infrequently (“Gene hills”) [15]. Historically, the focus has been on the “Gene Mountains” which was partly due to the limited technology available in the past. However, the use of whole genome sequencing technology had shown that the vast majority of cancer mutations, including those that drive tumor evolution occur infrequently and do not necessarily fall in the “gene mountain” category [15].

The lack of consistency or overlap in the panel of “critical” genes mutated in tumors of the same tissue type can possibly explain (1) the heterogeneous molecular profile and malignant potential of human cancers of similar histologic type [13] and (2) the difficulty in correlating changes in a small set of genes to behavior,

prognosis, or response to therapy. The poor predictability of the altered gene(s) is primarily because such alterations represent only a small component of each tumor’s mutational composition [16,17]. However, a multi-dimensional analysis of unrelated and distinct sets of genetic mutations identified in a cancer genome sequencing study revealed interesting results [15,18]. This multi-dimensional analysis of the cancer genome databases was done at four different levels: (1) sequence similarity, (2) functional annotation, (3) protein–protein interaction and (4) molecular pathways [18]. The analysis revealed that although distinct sets of genes were mutated in tumors of the same type, the affected genes belonged to a limited number of pathways and processes [15,18]. Thus, a large number of these infrequent mutations reflect alterations in a much smaller number of cell signaling pathways thereby indicating that pathways rather than individual genes govern tumorigenesis [15,18].

A systems biology approach has great promise as it can, as shown above, provide valuable insights into processes that were previously impossible to visualize or understand. The integration of different high-throughput “omics” sciences, including genomics, transcriptomics, proteomics, and metabolomics have been the cornerstone for systems biology. Systems biology approaches allow simultaneous examination of changes in the genome (genomics), transcriptome (transcriptomics), proteome (proteomics) or metabolome (metabolomics) in biological samples, with the ultimate goal of understanding physiologic or pathologic mechanisms [14]. Such a collective approach allows integration and definition of the relationships of several and theoretically all the elements in a system, rather than study each element in isolation. Each of the “omics” sciences is different with different levels of feasibility and inherent advantages and disadvantages. Genomics is the study of all the nucleotide sequences, including structural genes, regulatory sequences, and non-coding DNA segments in the chromosomes of an organism. Transcriptomics is the study of global mRNA expression of a particular tissue [14]. Proteomics involves identification of the proteins in the body and their role in physiologic and pathophysiologic functions. Metabolomics, a new emerging field, is the study of global metabolite profiles in biologic samples, isolated cells, and importantly urine, saliva, and blood [19].

This review will focus on the transcriptomics, proteomics and metabolomics literature in the field of sporadic human CRC.

## 2. Transcriptomics

### 2.1. What is transcriptomics?

Transcriptomics, also known as global gene expression profiling, is a tool for evaluating gene expression levels of thousands of genes in parallel. It is among the early “omics” sciences, with the earliest forms being spotted nylon arrays with a few representative genes. The general idea of transcriptomics is that genes involved in common processes are often co-expressed and therefore genes following similar expression patterns may be functionally linked and/or under similar genetic control mechanisms. Although mRNA transcript levels might not correlate directly with the protein level, the general assumption is that patterns of gene expression levels reflect the physiologic status of the cell.

The concept of a microarray experiment is based on the principle of highly specific binding of complementary single-stranded nucleic acid sequences, a concept that has been exploited in the past by way of northern and southern blots. This review will not focus on the technical aspects of the microarray and the reader is directed to numerous excellent review articles that cover various aspects of microarray technology [20].

## 2.2. Transcriptomics in colorectal cancer

The multistage progression model of adenoma–carcinoma–metastasis sequence in CRC with specific alterations at each stage is similar to the Darwinian concept of advantageous phenotype conferring advantages in growth and survival. Some of these stage-specific mutations include APC mutations (normal epithelium → adenoma), K-Ras mutations (described in adenomas), and p53 and SMAD changes (adenoma → carcinoma transition).

Despite the extensive molecular and genetic characterization of the different stages of CRC, robust classification and prognostication in CRC is still an unmet medical need. The availability of high-throughput transcriptomics tools (genome-level gene expression) is now allowing the evaluation of potential gene signatures that could open a window in our understanding of tumor biology.

One can obtain patterns of gene expression at specific stages of tumorigenesis and thus understand molecular perturbations driving de novo tumor development and progression. Therefore, application of this high-throughput technology has been particularly relevant in human CRC as it follows the multistage progression paradigm of preneoplasia → adenoma → dysplastic adenoma → carcinoma. Since the first genome-wide oligonucleotide array analysis on 40 tumors and 22 normal samples, numerous microarray studies have been published on adenomas, adenocarcinomas, and metastatic carcinomas [21]. The primary objectives of most of microarray studies are (1) to accurately classify histologically distinct stages of CRC via their specific expression profiles and (2) develop predictive biomarkers that can be utilized for prognostication and assessment of biologic behavior with an ultimate aim of developing personalized medicine.

Indeed, application of the transcriptome platform to distinct histologic stages of CRC (early polyp to advanced cancer) has enabled successful stage-specific differentiation of tumors and from the adjacent/matched normal mucosa via discriminating gene expression profiles/signatures. For example, Lin et al., using a cDNA microarray representing 23,040 genes, were successful in identifying a panel of 50 genes that was differentially expressed in adenomas and adenocarcinomas [22]. A molecular diagnosis score, based on this 50-gene panel, was used to correctly classify an independent set of adenomas and adenocarcinomas. In another study, 20 genes were noted to be upregulated in adenomas and adenocarcinomas, and many of the expression changes noted in the cancers were presaged by those in the adenomas. Furthermore, a clustering algorithm could successfully distinguish normal tissue, adenoma, and adenocarcinomas [23]. Unique gene expression signatures (panel of 65 genes) were also noted in microdissected (via laser-capture) cells from early adenomatous crypts relative to adjacent normal crypts via a low complexity cDNA nylon array with 588 gene probes [24]. In another study, clustering analysis allowed the distinct separation of flat-adenomas from villous adenomas, thereby indicating that these morphologically distinct adenomas have distinct gene expression profiles [25]. Distinct molecular differences between normal and tumor tissues, tumors with/without lymph node invasion or genetic instability, and tumors from the right or left colon were identified with supervised clustering analyses of the microarray-generated data [26]. This approach was also successful in identification of a gene set that divided patients with significantly different 5-year survival (100% in one group and 40% in the other group;  $P < 0.005$ ) [26].

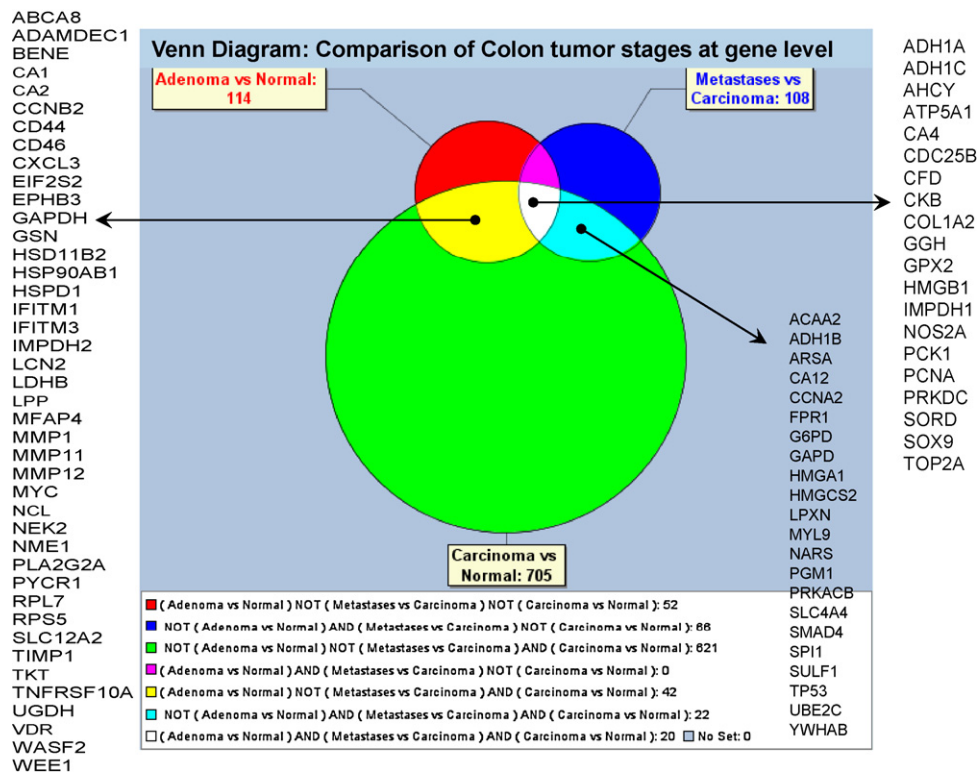
An interesting aspect of many of these gene expression studies is that most of the altered gene expression values for adenomas were intermediate between normal samples and carcinomas [13]. This is consistent with the existing histopathological placement of adenomas as an intermediate stage between normal epithelium and carcinoma.

Gene expression analyses can be used to generate tumor markers that help determine treatment options and prognosis. The two commonly used CRC staging systems are the more recent Tumor-Node-Metastasis (TNM) and older Duke's and Astler-Coller staging systems. In these systems, tumors are staged depending on the extent of invasion of the different segments of the bowel wall (primary site) and presence of metastasis in local lymph nodes or distant organs. The treatment options for these tumors are dependent upon the morphologic stage of the detected tumor. Patients with stage IIIA–C (Dukes C) tumors are subjected to post-surgical chemotherapy. However, with only 25–30% tumor recurrence rate in stage IIA–B (Dukes B), the benefits of post-surgical chemotherapy in patients with Dukes B are unclear.

The inability of the classic staging systems to accurately predict tumor outcome has led to the application of transcriptomics for prognostication and guidance for effective therapy. For example, gene expression profiling done with a 32,000 cDNA microarray on 78 human CRC specimens yielded a molecular diagnosis tool of 43 core genes (including osteopontin and neuregulin). This diagnostic gene panel was 90% accurate (93% sensitivity, 84% specificity) in predicting 36-month overall survival in 78 patients and was significantly better than Dukes staging [27]. One of the altered genes, osteopontin, was observed to be consistently overexpressed in CRC and metastases compared to adenomas in a separate study using Astler-Coller staging system [28]. In a separate study, discrimination of survival difference was possible in an independent set of patients, especially those with Dukes B and C tumors, thus highlighting the value of molecular staging in accurate patient prognosis [27]. In a different study, cluster analysis of gene expression profiles allowed grouping of early Dukes A tumors with normal tissue, Dukes B with Dukes C, and the more advanced Dukes D to each other [29].

Microarray analysis has served not only as a tool in correct identification of the tumor stage, but also as an excellent tool that allows informed use of adjuvant chemotherapy in specific stages of CRC. For example, a 23-gene signature panel was found to predict recurrence in patients with Dukes B class of CRC (with a survival rate of 75%) with 78% accuracy and odds ratio of 13 [30]. Such patients with 13-fold risk of tumor recurrence, similar to Dukes C patients would be ideal candidates to receive adjuvant therapy. In patients with Dukes C stage tumors (with 50% rate of recurrence post-surgery), expression profiling outperformed previously reported genetic markers such as TP53, K-RAS, and 18q allelic loss in prognostication [31]. Subsequent validation of this data using one of the most differentially expressed genes *RHOA* (RAS homologue) via immunohistochemistry in an independent cohort of Dukes C tumors revealed a direct correlation between *RHOA* expression and survival [31]. Thus, transcriptomics platform has been applied to CRC with various tumor classifying criteria including Dukes staging system, Astler–Coller stages, presence or absence of metastasis, positive for metastases in lymph nodes or not, difference in survival rates, and TNM [32–35].

Application of microarray technology to pre-surgical biopsies of tumors has provided an opportunity for accurate tumor classification and prognostication. Conventional pathology based evaluation of tumor biopsies may not provide information on the existence of synchronous lymph node metastasis or provide insights for responses to chemo- or radiotherapy. In fact, microarray analysis of tumor biopsies were successful in differentiating polyps from adenomas, low grade dysplastic adenomas from high grade dysplastic adenomas, adenomas from CRC, and early from advanced stage CRC with specific discriminatory 9, 65, 61, and 34 probe sets, respectively [36,37]. However, a major concern with a single representative biopsy of tumor is that intra-tumor heterogeneity can confound interpretation of data. This issue was successfully addressed by Komori et al. [38]. The authors were successful in dif-



**Fig. 1.** Comparison of differentially expressed genes in different human colon tumor classes. Data was compiled from microarray studies that were published in the public domain.

differentiating cancers from adenomas with one biopsy because the intra-tumor heterogeneity of the evaluated biopsies was smaller than the inter-tumor heterogeneity.

Examples of single studies that evaluate human CRC in multiple dimensions, i.e. genome, transcriptome, proteome, and metabolome, are few. An elegant study by Habermann et al. identified sequential alteration of the genome, transcriptome, and proteome during CRC progression [39]. Gene expression analysis on 9K cDNA arrays identified 58 genes differentially expressed between normal mucosa and adenoma, 116 genes between adenoma and carcinoma, and 158 genes between primary carcinoma and liver metastasis. The chromosome-specific average gene expression levels in this study correlated directly with chromosome copy number changes (evaluated via comparative genome hybridization). Analysis of the two-dimensional gel electrophoresis and subsequent mass spectrometry-based protein expression revealed a lack of direct correlation between differentially expressed genes and proteins. However, pathway analysis revealed that the apparent lack of correlation was resolved since the majority of the altered genes belonged to identical pathways or networks. Thus, the application of “omics” tools on distinct stages of colon tumors in the same study allowed the conclusion that chromosomal aneuploidies directly affect average resident gene expression levels, thereby contributing to a massive deregulation of the cellular transcriptome and proteome [39].

As seen above, numerous microarray studies on different stages of human CRC have been published using highly variable gene array platforms, methods of data collection and analysis, and most importantly variable patient population sets. Most of these studies suggest that microarrays may provide means to improve the classification of CRC, and provide novel diagnostic and/or prognostic markers and therapeutic targets. However, a consistent problem has been the lack of overlap of these gene expression signatures across multiple studies. This has prevented the development of

“discriminatory gene expression panel(s)” for clinical use. This lack of correlation in differentially expressed genes from multiple studies can be (1) a function of the limited number of genes represented in the microarrays (as noted in earlier studies with cDNA arrays which contained a few thousand probe sets), (2) inter-study variation because of the non-standardized implementation of microarray experiments (a common problem), and (3) true biological phenomenon. Although, the first two points are valid, the role of biology in this discrepancy cannot be ruled out. For example, the integrative systems biology approach of Sjöblom has shown that despite wide inter-study variation in panels of mutated genes, the broad functional categories or pathways these genes represent are more common [16].

To apply this integrated approach to the currently available transcriptomics data, we compiled and analyzed transcriptional data in human colon cancer that was available in the public domain. Our goal here was not to provide a comprehensive analysis of all the studies on human CRC in the public domain, but to present methods of analysis that one can employ to effectively implement principles of systems biology in transcriptomics. We either retrieved complete datasets when available or manually extracted the gene IDs or accession numbers or molecule links from individual research articles. The data sets were comprised of genes that were significantly altered (both upregulated and downregulated) in three groups—adenoma vs. normal, carcinoma vs. normal, and carcinoma vs. metastatic carcinoma. We identified 134, 792, and 130 unique gene IDs from these three groups, respectively. Differentially expressed (included both upregulated and downregulated) genes comparing adenoma and normal (aka adenoma), primary carcinoma and normal (aka carcinoma), and metastatic carcinoma and primary carcinoma (aka metastatic carcinoma) were compiled from 10, 20, and 14 studies, respectively. It is important to note that the number of studies listed here include review articles that contain numerous individual microarray studies.

**Table 1**  
Identification and description of genes that were significantly altered in different human colon tumor classes.

Adenoma vs. carcinoma vs. metastasis		Adenoma vs. carcinoma		Carcinoma vs. metastasis	
Gene ID	Gene description	Gene ID	Gene description	Gene ID	Gene description
ADH1A	Class I alcohol dehydrogenase (ADH1) alpha subunit	ABCA8	ATP-binding cassette, subfamily A, member 8	ACAA2	Acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)
ADH1C	Alcohol dehydrogenase class I gamma subunit (ADH3)	ADAMDEC1	A disintegrin and metalloproteinase domain-like protein Decysin 1	ADH1B	Alcohol dehydrogenase IB (class I), beta polypeptide
AHCY	S-adenosylhomocysteine hydrolase (AHCY)	BENE	mal, T-cell differentiation protein-like	ARSA	Arylsulfatase A
ATP5A1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	CA1	G-protein coupled receptor 103 (GPR1093)	CA12	Carbonic anhydrase XII
CA4	Carbonic anhydrase IV	CA2	Aspartate carbamoyltransferase (CAD)	CCNA2	Cyclin A2
CDC25B	M-phase inducer phosphatase 2 (cdc25b)	CCNB2	cyclin B2	FPR1	Formyl peptide receptor 1
CFD	Complement factor D (C3 convertase activator) (Adipsin)	CD44	CD44 molecule (Indian blood group)	G6PD	Glucose-6-phosphate dehydrogenase
CKB	Creatine kinase B chain	CD46	CD46 molecule, complement regulatory protein	GAPD	Glyceraldehyde-3-phosphate dehydrogenase
COL1A2	Uridine nucleotide receptor (UNR)	CXCL3	Leukocyte tyrosine kinase receptor (LTK)	HMGA1	High-mobility group AT-hook 1
GGH	Gamma-glutamyl hydrolase	EIF2S2	Eukaryotic translation initiation factor 2, subunit 2 beta, 38 kDa	HMGCS2	3-Hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)
GPX2	Glutathione peroxidase 2 (gastrointestinal)	EPHB3	Ephrin type-B receptor 3 (HEK2)	LPXN	Leupaxin
HMGB1	High-mobility group box 1	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	MYL9	Myosin, light chain 9, regulatory
IMPDH1	IMP (inosine monophosphate) dehydrogenase type 1	GSN	Gelsolin (amyloidosis, Finnish type)	NARS	Asparaginyl-tRNA synthetase
NOS2A	Inducible nitric oxide synthase (NOS2A)	HSD11B2	Hydroxysteroid (11-beta) dehydrogenase 2	PGM1	Phosphoglucomutase 1
PCK1	Phosphoenolpyruvate carboxykinase-Cytosolic (PEPCK)	HSP90AB1	Heat shock protein 90 kDa alpha (cytosolic), class B member 1	PRKACB	Protein kinase, cAMP-dependent, catalytic, beta
PCNA	Proliferating cell nuclear antigen	HSPD1	Heat shock 60 kDa protein 1 (chaperonin)	SLC4A4	Solute carrier family 4, sodium bicarbonate cotransporter, member 4
PRKDC	DNA-dependent protein kinase catalytic subunit (DNA-PKcs)	IFITM1	Interferon induced transmembrane protein 1 (9–27)	SMAD4	SMAD family member 4
SORD	Sorbital dehydrogenase	IFITM3	Interferon induced transmembrane protein 3 (1–8 U)	SPI1	Spleen focus forming virus (SFFV) proviral integration oncogene spi1
SOX9	SRY (sex determining region Y)-box 9	IMPDH2	IMP (inosine monophosphate) dehydrogenase 2	SULF1	sulfatase 1

Table 1 (Continued)

Adenoma vs. carcinoma vs. metastasis		Adenoma vs. carcinoma		Carcinoma vs. metastasis	
Gene ID	Gene description	Gene ID	Gene description	Gene ID	Gene description
TOP2A	Topoisomerase II	LCN2	Lipocalin 2	TP53	Tumor protein p53 (Li-Fraumeni syndrome)
		LDHB	Lactate dehydrogenase B	UBE2C	Ubiquitin-conjugating enzyme E2C
		LPP	LIM domain containing preferred translocation partner in lipoma Microfibrilla associated protein 4	YWHAB	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
		MFAP4			
		MMP1	MMP-1 (collagenase 1)		
		MMP11	MMP-11 (stromelysin 3)		
		MMP12	Secretin receptor (SCT-R)		
		MYC			
		NCL	Nucleolin		
		NEK2	Serine/threonine protein kinase NEK2 (NIMA-related protein kinase 2)		
		NME1	Nucleoside diphosphate kinase A (NDK A) (NM23-H1)		
		PLA2G2A	Phospholipase A2, membrane associated		
		PYCR1	Pyrroline-5-carboxylate reductase 1		
		RPL7	Ribosomal protein L7		
		RPS5	Ribosomal protein S5		
		SLC12A2	Sodium/potassium/chloride transporter		
		TIMP1	Tissue inhibitor of metalloproteinase 1		
		TKT	Transketolase		
		TNFRSF10A	Tumor necrosis factor receptor superfamily, member 10a		
		UGDH	UDP-glucose dehydrogenase		
		VDR	Vitamin D3 receptor (VDR/1,25-dihydroxyvitamin D3 receptor)		
		WASF2	WAS protein family, member 2		
		WEE1	WEE1-like protein kinase		

The paucity of “metastatic carcinoma vs. normal” studies resulted in the “metastatic carcinoma vs. primary carcinoma” as the only viable comparison. The significantly greater number of differentially regulated genes in the “carcinoma vs. normal” group was also a function of the larger number of studies available for this group. The number of genes that were unique or common between different stages was visualized with a Venn diagram (Figs. 1 and 2).

The next step was to evaluate or identify gene networks that were common or unique to each of the three comparisons. Significance of biological pathways ( $P < 0.05$ ) was estimated through a variation of Fisher's exact test and adjusted for multiple test-

ing using Benjamini-Hochberg FDR analysis (built-in function of GeneGo Metacore® software). For this, gene lists for each of the three comparisons were curated using MetaCore® (GeneGo, Inc.). The powerful “Enrichment analysis workflow tool” of MetaCore® analyzes experimental data via a gene enrichment process using MetaCore's biological pathways, molecular function, and GO terms which are overrepresented in the initial gene list. In the current work we made use of the Pathway Maps engine which is a collection of manually created pathway maps. Maps are grouped hierarchically into folders according to main cell processes, protein functions, and diseases. We also made use of MetaCore's Map fold-

**Table 2**

Pathways lists obtained using Metacore® enrichment analysis populated from differentially expressed genes from multiple microarray studies from various stages of human colon tumors.

Adenoma vs. normal and carcinoma vs. normal and metastasis vs. carcinoma	Adenoma vs. normal and carcinoma vs. normal	Carcinoma vs. metastasis
<b>Cell cycle.Role of 14-3-3 proteins in cell cycle regulation</b>	Pyruvate metabolism/Rodent version	<b>Cell cycle.Nucleocytoplasmic transport of CDK/Cyclins</b>
<b>Cytoskeleton remodeling.TGF, WNT and cytoskeletal remodeling</b>	Signal transduction.AKT signaling	<b>Cytoskeleton remodeling.Role of Activin A in cytoskeleton remodeling</b>
<b>Immune response.Oncostatin M signaling via MAPK in mouse cells</b>	Development.PIP3 signaling in cardiac myocytes	<b>Cell cycle.Regulation of G1/S transition (part 1)</b>
<b>Development.Hedgehog and PTH signaling pathways in bone and cartilage development</b>		<b>Cell cycle.Regulation of G1/S transition (part 2)</b>
<b>DNA damage.ATM/ATR regulation of G2/M checkpoint</b>		<b>Apoptosis and survival.Apoptotic Activin A signaling</b>
<b>Immune response.Oncostatin M signaling via MAPK in human cells</b>		Development.Slit-Robo signaling
<b>Cell cycle.Transition and termination of DNA replication</b>		Immune response.Alternative complement pathway
<b>Immune response.Histamine H1 receptor signaling in immune response</b>		Immune response.MIF – the neuroendocrine-macrophage connector
<b>Cell cycle.Cell cycle (generic schema)</b>		Bile Acid Biosynthesis
<b>Pyruvate metabolism</b>		Immune response.CCR3 signaling in eosinophils
<b>Cell adhesion.Chemokines and adhesion</b>		Cytoskeleton remodeling.Regulation of actin cytoskeleton by Rho GTPases
<b>Cytoskeleton remodeling.Cytoskeleton remodeling</b>		Transcription.Transcription regulation of aminoacid metabolism
<b>Cell cycle.Initiation of mitosis</b>		Proteolysis.Putative SUMO-1 pathway
<b>Transcription.Role of VDR in regulation of genes involved in osteoporosis</b>		Cytoskeleton remodeling.RalA regulation pathway
<b>Immune response.IL-17 signaling pathways</b>		Immune response.MIF in innate immunity response
<b>Cell adhesion.Cadherin-mediated cell adhesion</b>		Cytoskeleton remodeling.ACM3 and ACM4 in keratinocyte migration
<b>Immune response.MIF-mediated glucocorticoid regulation</b>		G-protein signaling.G-Protein alpha-12 signaling pathway
<b>DNA damage.ATM/ATR regulation of G1/S checkpoint</b>		Cytoskeleton remodeling.Role of PKA in cytoskeleton reorganisation
Normal and pathological TGF-beta-mediated regulation of cell proliferation		Signal transduction.Calcium signaling
DNA damage.Brca1 as a transcription regulator		Apoptosis and survival.BAD phosphorylation
Cell cycle.Role of APC in cell cycle regulation		Transcription.CREB pathway
		Immune response.IL-1 signaling pathway
		Muscle contraction.Regulation of eNOS activity in endothelial cells

Bold pathways indicate  $P < 0.05$ .

ers which is again manually created pathway maps that are grouped hierarchically into folders according to main biological processes. Furthermore, the analysis method we used was “Experiment Intersections” method which considers mappings of the intersections of the selected experiments (gene lists), rather than of the experiments themselves. In this way, the set of genes that are common to all active experiments (gene lists), or similar in some of the three experiments, or unique to a given experiment, were utilized in identifying relevant pathways.

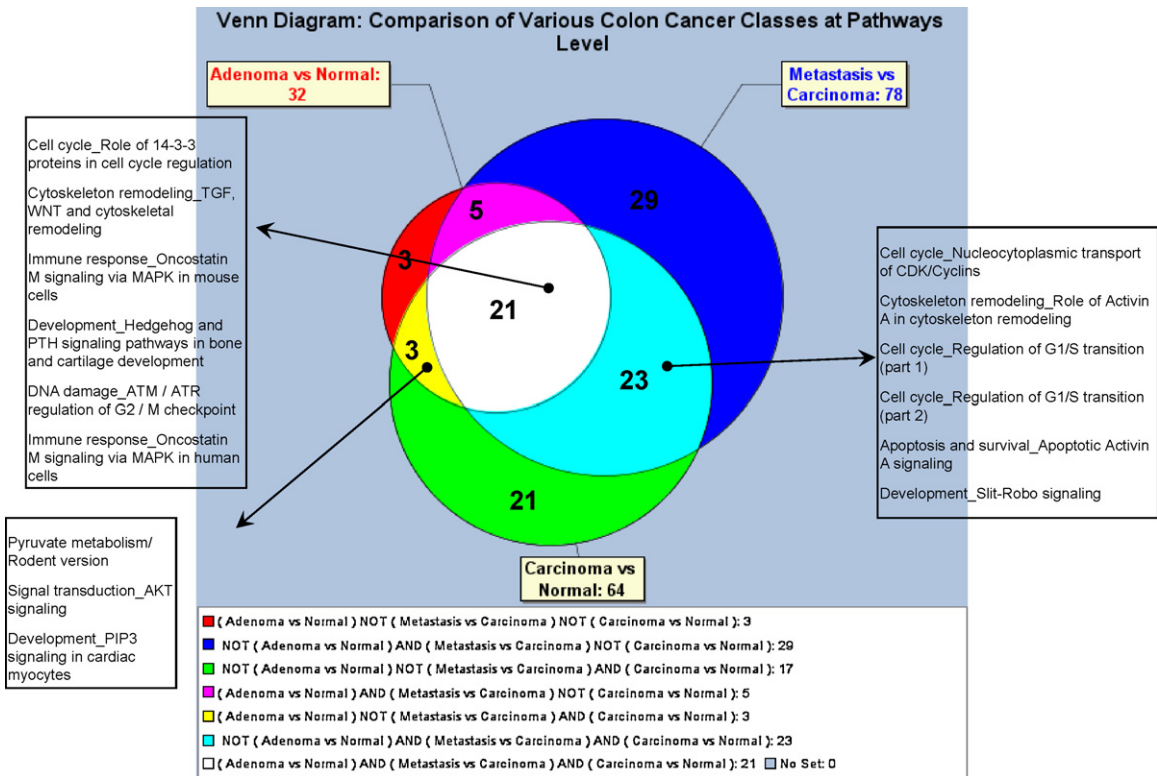
We considered the three gene lists for the three comparison sets and obtained biological processes for each list (Fig. 1). The number of differentially expressed genes that were common to only adenoma and carcinomas, common to carcinoma and metastatic carcinoma, and common to all three groups was 42, 22, and 20, respectively (Fig. 1, Table 1). These corresponded to 3, 23, and 21 pathways for adenoma and carcinomas alone, carcinoma and metastatic carcinoma, and common to all three groups, respectively (Fig. 2, Table 2).

Although detailed analysis of each overrepresented pathway is beyond the scope of this review article, brief discussion of the most overrepresented pathways in all three comparisons (i.e. adenoma, carcinoma, and metastasis) is warranted. One such example is the 14-3-3 protein pathway (Fig. 3). 14-3-3 proteins regulate the cell cycle via cell cycle checkpoint kinase 1 (Chk1), and by phosphorylation of cell division cycle 25 phosphatases (CDC25s), tyrosine

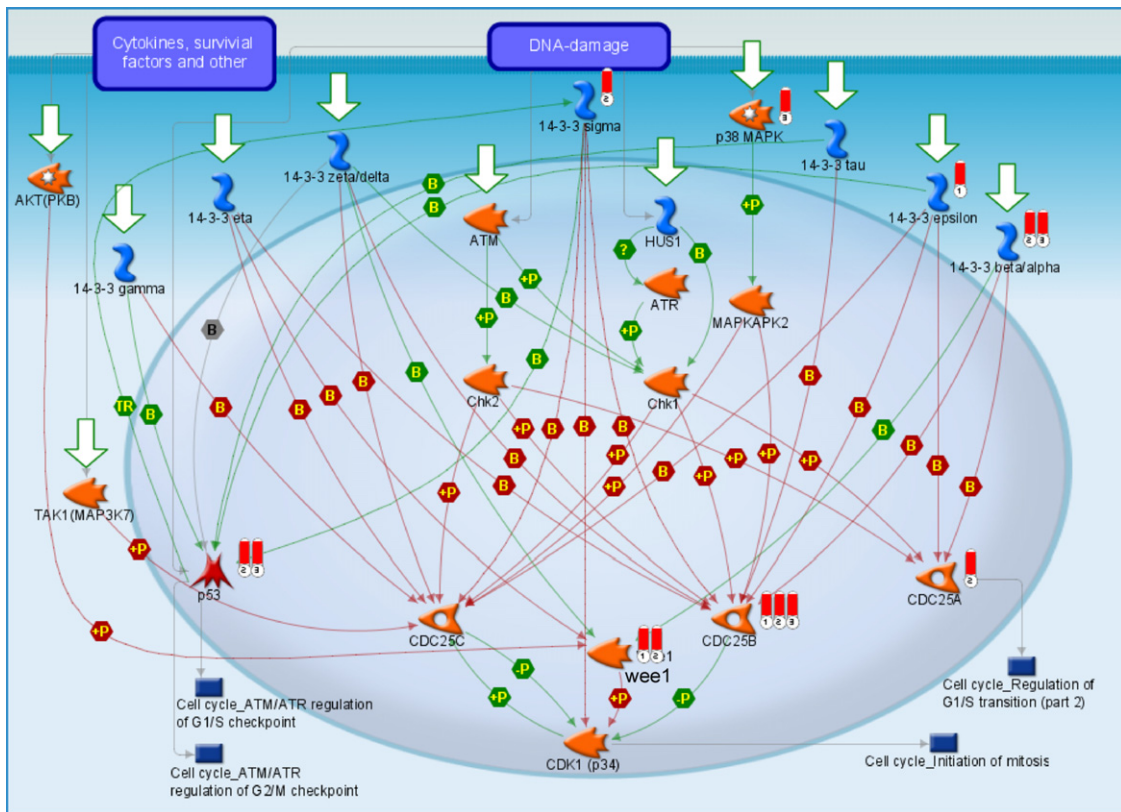
kinase Wee1, tumor suppressor p53 and cyclin-dependent kinase 1 (CDK1) [40–44]. Indeed, CDC25B is important for nucleocytoplasmic transport of cyclin–cyclin-dependent kinase complexes, a step important for cell cycle-initiation of mitosis. Wee1 is common to adenoma and adenocarcinoma and it is also responsible for the inhibition of CDK1. Other pathways that were significantly overrepresented in specific comparisons are presented in Fig. 2 and Table 2.

It is important to note that the studies comparing carcinomas to normal were much higher in number than the other groups and therefore, differentially expressed genes in the carcinoma group were very likely to be overrepresented. Moreover, unlike the adenoma and carcinoma comparisons, the metastatic carcinomas are compared to the primary carcinomas and not to normal tissue. Therefore, genes or pathways that were not flagged in this comparison (i.e. carcinoma vs. metastasis) could be due to either because the gene/pathway is comparably altered in this group or not represented on the array or not altered at all.

An important biologic phenomenon at the level of mRNA that might be overlooked with a microarray designed for regular transcripts is the phenomenon of alternative splicing. Alternative splicing is a means for expanding the limited repertoire of genes into very complex proteomes. Traditional approaches to the study of alternative splicing have focused on individual genes. The advent of whole genome based microarrays has allowed comprehensive



**Fig. 2.** Comparison of pathways that were overrepresented in different colon tumor classes. The gene lists were compiled from selected microarray studies that were published in the public domain. Selected pathway lists that were annotated by Metacore® enrichment analysis are listed.



**Fig. 3.** Alterations in the 14-3-3 protein pathway in human colon adenoma, carcinoma and metastasis. Differentially expressed genes for each stage (1=adenoma, 2=carcinoma, 3=metastatic carcinoma) is depicted with the red bar. Data was compiled from microarray studies that were published in the public domain. Selected pathway lists that were annotated by Metacore® enrichment analysis are listed. The legend for GeneGo Pathway maps is given in Supplemental Fig. 1.

scanning of the entire genome for alternatively spliced gene transcripts. Although, it is beyond the scope of this review to discuss the findings in this field in detail; briefly, specific alterations in splicing patterns have been found in many cancers often with critical role in transformation, motility, and metastasis [45,46]. The cancer-specific splice variants have been proposed to have potential as excellent diagnostic, prognostic, and predictive biomarkers as well as potential drug targets. Some of the pathways targeted by alternatively spliced gene products in human cancer appear to be similar to the ones shown in Table 2 (integrin signaling, extracellular matrix proteins, and cytoskeletal organization).

To conclude, one of the major disadvantages to the microarray platform is the specific assumptions of the mRNA expression data. Specifically, some of the common assumptions include direct correlation between mRNA and protein levels, co-regulated genes are functionally associated and hybridization signals reflect intracellular biological processes. To fully understand the pathogenesis of tumors, the gene expression data from microarray experiments needs to be complemented with information such as the status of proteins, activities and reaction rates of enzymes, transport processes, etc. This gap in our knowledge base can be filled in with parallel array platforms for proteins, enzymes, metabolites, and antibodies [47].

### 3. Proteomics

#### 3.1. What is proteomics?

Proteomics can be defined as the science and technologies associated with mapping, visualizing, and/or quantitating the expression of all or a majority of the proteins in living systems. Proteomics has been around for a few decades in the form of two-dimensional polyacrylamide gels [48,49]. However, unlike those available for nucleic acids, the tools in the field of proteomics had reproducibility issues and also lacked sensitive analytical techniques. The use of mass spectrometry (MS)-based proteomics really revolutionized the field since the mid-1990s—especially after the ability to ionize peptides was discovered [50]. The two primary reasons for the rapid growth in the field of proteomics are (1) the availability of the entire genome sequences from the simple *E. coli* genome to the most complex, human genome [51,52] and (2) the availability of rapid-scanning mass spectrometers which can scan on the same timescales as liquid chromatography (LC) thereby allowing their integration into a single powerful workflow.

#### 3.2. Technology involved in proteomics

Proteomics generally involves gel-dependent (one-dimensional and 2D) [53] and gel-independent LC-based separations [54], both of which depend on a downstream step of MS-based identification. For gel-based approaches, intact proteins are resolved on a gel matrix followed by an in-gel tryptic digestion, which is performed to generate short fragments called peptides. Such gel-dependent LC–tandem MS (MS/MS) procedures are carried out on peptides instead of whole proteins. This type of strategy is known as bottom-up or shotgun proteomics because of the rapid nature of analysis of the whole protein (top level) from the peptide level (bottom level) [4–7]. In contrast, gel-independent LC–MS/MS approaches query the intact native protein (topdown). These gel-independent approaches can be useful for detecting endogenous perturbations such as post-translational modifications (PTM) but are severely deficient in terms of providing sequence coverage of the proteins being queried [9].

The standard shotgun proteomics experiment has three stages: sample preparation, LC–MS(/MS) analyses, and proteome infor-

matics. The very first step of sample preparation is usually cellular/tissue lysis via sonication, homogenizer, or osmotic shock. The extracted proteins are denatured, reduced (to enable downstream proteolysis), and digested into peptides followed by a 2D LC-based separation by combining strong cation exchange (based on charge) and reverse phase (based on hydrophobicity) columns [55,56]. The charged peptides are then transferred into the gas phase for MS measurements, either by electrospray ionization (ESI-MS/MS) for solution phase or matrix-assisted laser desorption and ionization (MALDI) [57,58]. Accurate peptide and protein identification is done by software such as SEQUEST and MASCOT, which compare both the intact and fragmented peptide patterns to those derived *in silico* from human genome databases.

Some of the problems associated with this type of protein identification include non-uniform ionization of peptides, increased loss of some peptides types (for example, adhesion of hydrophobic peptides to surfaces), and challenges in direct quantitation of peptides. One solution is to use stable isotope labeling, which enables quantitation of the abundance of labeled protein in one tissue type (for example, normals) relative to the abundance of unlabeled protein from another condition (for example, carcinoma).

Surface-enhanced laser desorption/ionization (SELDI) originally developed by T. William Hutchens in 1993 is another ionization method in mass spectrometry which has become very popular for biomarker detection. SELDI is conventionally used with time-of-flight (TOF) mass spectrometers and is useful for detection of biomarkers in tissue samples such as blood, urine, or other clinical samples. SELDI-MS is a variation of MALDI wherein the surface has an “inert, unbiased” nature in the case of the latter whereas the surface in the former has some sort of biochemical affinity for the proteins in the sample, thus turning the spotting step itself into a fractionation. Therefore, depending on their particular biochemical nature, proteins in the sample may either adsorb or get washed away from the surface. After the washing step, the matrix is applied to the surface followed by crystallization with the sample peptides. Several studies have used SELDI to profile samples from CRC patients in order to discover biomarkers associated with this disease [59–61].

In summary, there are several different approaches for the identification and quantitation of peptides/proteins under the umbrella term of “proteomics”.

#### 3.3. Proteomics of CRC

In contrast to the extensive literature on genomic and transcriptional analyses of human CRC, data on proteome-based analysis in human CRC is limited. The handful of proteome-based analyses of human CRC reported in the literature mostly used serum as the test sample. The objective of such studies has been the identification of biomarker(s) that can non-invasively discriminate patients with CRC (in different stages) from normal individuals. This is especially important because of the lack of good biomarker(s) for CRC discovered using other platforms. For example, the commonly used serum biomarker CEA has limited sensitivity of 30–40% for early CRC [62].

Proteome-based analysis of samples acquired with non-invasive methods such as blood collection with good sensitivity and specificity and low cost would be highly desirable and acceptable to both healthcare professionals and patients. While the above remains the goal, the following findings present the progress already made in the field of biomarker discovery in CRC proteomics. Surprisingly, despite the use of different types of protein arrays on distinct sets of patient populations, several common biomarkers have been identified across studies. This is very promising for the future of proteome-based applications in diagnosis and prognosis of CRC. A summary of the different studies is provided in Table 1. An analysis of sera from 62 CRC (Dukes stages A–D) and 31 normal individuals

**Table 3**  
Summary of proteomics studies for classification of human colorectal cancer.

Authors	Technique(s)	Comparison(s)	Classification	Biomarker(s)
Ward	SELDI arrays (IMAC30), LC-MS/MS + neural networks	Sera: 62 CRC (Dukes A/B, C/D) vs. 31 normal	Successful separation of CRC vs. normal	Complement system: C3a-des-arg, $\alpha$ 1-antitrypsin, transferrin
Habermann	SELDI arrays (IMAC3 and WCX2), QTOF-MS/MS	Sera: CRA vs. CRC vs. normals  58 Patients + 32 normals and an independent set of 59 serum samples	95% Sensitivity and 91% specificity Validation set: C3a-des-arg predicted CRC with 96.8% sensitivity and 96.2% specificity C3a-des-arg also predicted CRA in 86.1% of 36 CRA patients while only 5.6% were false-negative HNP 1–3 were markedly up in both sera and tissue from CRC relative to normals	C3a-des-arg (similar to Ward)
Albrethsen	IMAC30 and NP20 SELDI arrays for sera and tissue screens, respectively, MALDI-TOF	Sera: 125 CRC vs. 100 normals  Tissue: paired adjacent normals and tumors from same patient ( $n = 40$ )	No sensitivity and specificity data reported	Defensin family: human neutrophil peptides aka alpha-defensins 1–3
Yu	SELDI hydrophobic (H4 arrays), neural networks, support vector machine	Sera: 55 CRC (Dukes A–D), 35 CRA, and 92 normals	CRA vs. CRC: sensitivity was 89%, Specificity was 83% (moderate) CRC vs. normal: sensitivity was 89%, specificity was 92% (high)	No robust biomarker
Melle	SAX2 and WCX2 SELDI arrays.	LCM of CRC (39) vs. adjacent normals (40)	Elevated HNP 1–3  ELISA-based evaluation of HNP1–3 in sera yielded sensitivity of 100% and specificity of 69%	HNP1–3 (similar to Albrethsen)
Wu	MALDI-TOF, ELISA	Secretomes of 21 cancer cell lines yielded CRMP-2  This was used in 169 CRC (stages I–IV) vs. adjacent normals + plasma from 201 CRC and 201 normals	CRC vs. normals: a combination of CEA and CRMP-2 could yield sensitivity of 76.8% and specificity of 95.1%	Collapsin response mediator protein-2, CRMP-2; Carcinoembryonic Antigen, CEA
Li	MALDI-TOF	Microdissected adenomas (24), CRC (71), 55 normals	Greater heterogeneity in CRC when compared to adenomas highlighting commonality of mechanisms in adenomas vs. carcinomas.	No biomarker reported

successfully differentiated between each other based on profiles of three proteins from the complement system: C3a-des-arg which is the stable form of C3a anaphylatoxin,  $\alpha$ 1-antitrypsin and transferrin [61]. The SELDI chips used were Cu<sup>2+</sup>-loaded IMAC30 arrays and

artificial neural networks were used to classify samples as cancer or non-cancer. This study achieved a high sensitivity and specificity of 95% and 91%, respectively, using artificial neural networks however no stage-specific classification was reported.

In another study by Habermann et al., the same biomarker C3a-des-arg was found to be elevated in the sera from patients ( $n = 58$ ) with colorectal adenomas and carcinomas when compared to healthy normal individuals ( $n = 32$ ) [60]. Evaluation of C3a-des-arg in an independent set of 59 serum samples from CRC patients could accurately predict CRC with 96.8% sensitivity and 96.2% specificity. Interestingly, elevated levels of this protein were also noted in 86.1% of independently collected sera from 36 adenoma patients with a false-negative rate of 5.6%.

The proteomics approach using sera from 125 CRC patients and 100 healthy individuals in addition to paired samples of colon tumors and adjacent normal colon tissues from the same patient ( $n = 40$ ) revealed specific biomarkers. These were from the defensin family including Human Neutrophil Peptides (HNP 1–3), also known as alpha-defensins 1–3. Although, the sensitivity and specificity are not reported in this study, HNP 1–3 were found to be significantly upregulated in both sera and tissue samples in patients with CRC [59]. Similarly, HNP 1–3 levels were elevated in microdissected (via Laser Capture Microdissection) tumor tissue ( $n = 39$ ) relative to adjacent normal tissue ( $n = 40$ ) [55].

In a study set similar to the Habermann et al. study, Yu et al. [56] obtained sera from 55 CRC patients ( $n = 8, 22, 13,$  and  $12$  for Dukes' A, B, C, and D, respectively), 35 colorectal adenoma patients, and 92 healthy normals and used these for CRC diagnosis. SELDI hydrophobic surface (H4) chips were used and the bioinformatic tools of artificial neural networks were used to distinguish CRC from healthy normals whereas support vector machine algorithm was used to distinguish between CRC and CRA. The specificity and sensitivity for CRA vs. CRC were moderate with these values being 83% and 89%, respectively, in the test set. However, these same parameters were high for CRC vs. healthy normals at 92% and 89%, respectively. No potential biomarker proteins were identified in this study nor were there any stage-specific differences reported.

In the Wu et al. study [58], secretomes of 21 cancer cell lines of varying tissue origin were first used for biomarker screening. Using secretome analysis, collapsin response mediator protein-2 (CRMP-2) was identified in only two cell lines out of a total of 21 cancer cell lines, both of which were CRC in origin. In addition, immunohistochemical staining with anti-CRMP-2 antibody was significantly higher for the early stages (I and II; 68.7%) compared to the later stages (III and IV; 48.8%). Thus, CRMP-2 was chosen for biomarker validation in 169 paired CRC tumors and adjacent normals. These 169 CRC samples were classified as 150 adenocarcinomas and 19 mucinous adenocarcinomas. This study set was also classified into 16, 67, 57, and 29 CRC tumors from stages I to IV, respectively. Serum was collected from 201 healthy normals and CRC patients, the latter group having a spread of 27, 72, 60, and 40 patients in stages I–IV, respectively. Protein identification was done using a TOF/TOF mass spectrometer. The authors observed a combination of CEA and CRMP-2 could yield a reasonable sensitivity of 76.8% and specificity of 95.1% for detection of CRC vs. healthy normals.

The Li et al. study [57], consisted of 24 adenomas, 71 CRC and 55 normals from microdissected tissues. Protein identification was done using MALDI-TOF. One of the interesting findings in this study was the greater heterogeneity observed in primary carcinoma when compared to adenomas because of the possible commonality of mechanism in adenoma formation such as the frequent inactivating mutation of the *Apc* gene. This helps in providing guidance to future CRC proteomic studies when power calculations are made for study size and design.

Thus, studies using the proteomics approach in CRC have yielded some interesting leads with respect to biomarkers of diagnosis and prediction (summarized in Table 3). However, the data is still preliminary and far from being employed as a robust set of biomarker(s) capable of accurate diagnosis, prognosis, stage-specific identification, etc. However, with an increase in the number

of proteome-based studies, we can hope for a robust panel of biomarkers in the near future.

### 3.4. Advantages and disadvantages

The advantages of profiling the whole proteome are obvious: (i) protein levels may not significantly corroborate with mRNA levels detected from transcriptomic profiling, due to a whole host of post-translational regulatory mechanisms which are opaque to transcriptome detection methods, (ii) post-translational modifications of proteins such as phosphorylation, glycosylation or ubiquitylation which are extremely important to protein function cannot be detected by genomics and (iii) since proteins actually carry out biological function, it is absolutely critical that we sample and assay at this level.

The disadvantages of proteomic profiling are the apparent lack of a very high degree of reproducibility in protein identification from identical samples run in the same laboratory or in different laboratories. For example, a recent systematic study involving shipping 20 different highly pure proteins to different groups resulted in only 7 groups identifying all 20 proteins accurately [63]. However, after some feedback and reanalysis, all 20 groups identified all the supplied proteins. This seemed to point at the skill level of the laboratories involved – which required some level of “guidance” before accurate identification – as a potential source of the lack of this reproducibility. For this reason, we have decided to only discuss studies that had at least 20 CRC samples so that this robust size would provide enough statistical power and counter the noise or lack of reproducibility intrinsic to proteomics platforms.

Other disadvantages that have been documented in the field of biomarkers in cancer detection, stratification, and prognosis are: (i) presence of highly abundant proteins in the tissue sample – for example albumin in serum or amylase in saliva – that mask the signals arising from low-abundance proteins, (ii) the right type of tissue to use, i.e. blood or urine (for ease of extraction) or tumor biopsy (which maybe hard to access), and (iii) difficulty or inability to profile for membrane proteins (or hydrophobic proteins for some platforms) because of difficulties in solubilizing such proteins.

## 4. Metabolomics

### 4.1. What is metabolomics?

Metabolomics is the study of global metabolite profiles in a biological system (isolated cells, tissue, urine, saliva, blood plasma, etc.) under a given set of conditions. These metabolites are a summation of the total metabolic processes in a cell including anabolism and catabolism along with all the other processes such as absorption, distribution, detoxification of natural and xenobiotic materials and energy utilization [19,64]. A metabolite has been defined as “a native small molecule (nonpolymeric compound) that participates in general metabolic reactions and is required for the maintenance, growth, and normal function of cells.” [14]. Metabolites can belong to diverse classes such as organic acids, amino acids, fatty acids, sugars, sugar alcohols, steroids, nucleic acid bases, etc. [65]. Although the exact number of endogenous metabolites in humans is unknown, it has been estimated to range from 3000–20,000. In contrast, the estimated number of genes and protein in humans are 23,000 and 60,000–100,000, respectively [19].

Although a systems biology approach in understanding disease involves a comprehensive evaluation of all the “omics” platforms, metabolomics have distinct advantages over other “omics” plat-

forms. One of the most obvious ones is that metabolites, despite their constant fluxes in living systems, reflect the functional status of cells as they are downstream to DNA, RNA, and protein, and thus are truly reflective of “the pulse” of cells when compared to the above macromolecules. The metabolome is at the intersection of the genome and the environment and therefore, metabolomics often allows the detection of subtle changes in metabolic pathways and deviation from homeostasis before the detection of phenotypic changes. Because the metabolite profiles in a living system are highly dynamic with constant fluxes, it is quite imperative to be cognizant of the dimension of time, unlike the other “omics” platforms. In other words, global metabolite profiles evaluated in a temporal manner can be said to be a summary manifestation of all the other upstream “omics” profiles.

Metabolite profiling has been performed since 1970s, but these were mostly restricted to quantitative analysis of a small group of compounds in a specific pathway [66]. With the advent of new technologies, it is theoretically possible to profile most of the metabolites in a biological sample (metabolite profiling) which allows a snapshot of the metabolic/functional state under a given condition.

#### 4.2. Technologies involved in metabolomics

A sequential array of different technologies is needed for a successful metabolomics experiment—starting from sample preparation, storage, data capture, and analysis. To ensure comprehensive sampling of cellular metabolites, use of acid and organic solvent extraction procedures facilitates extraction of most cellular and membranous pools of metabolites. Some of the commonly used extraction agents include perchloric acid, water/acetonitrile mixture, chloroform/methanol, etc. Unlike transcriptomics or proteomics, it is often important in metabolomics to consider the intent of the scientific experiment. This is because global profiling of metabolites using open platforms is less sensitive than closed platforms designed to study specific classes of metabolites. Two of the commonly used tools in the detection of metabolites are nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS). NMR technology can be subdivided into solution state NMR, *in vivo* NMR and high-resolution magic angle spinning (HRMAS) NMR. MS can be subdivided into gas chromatography– or liquid chromatography–mass spectrometry (GC– or LC–MS), Fourier transform infrared (FT-IR) and Raman spectroscopy.

Detailed reviews of these techniques and data analysis are described elsewhere [19,66,67]. Briefly, solution state NMR utilizes aqueous samples and is more sensitive than *in vivo* NMR while HRMAS-NMR is a bridge between solution state and *in vivo* NMR. HRMAS has become an excellent tool for evaluation of metabolites in solid tissues because it can generate high resolution spectra from solid tissues. It is comparable to solution state and enables simultaneous detection of both aqueous and lipid metabolites [67]. The advantage provided by NMR is the unbiased and universal detection of metabolites of different classes. However, it is not as sensitive as MS and cannot be coupled to upstream chromatography-based separation techniques.

MS is more sensitive than NMR and is typically preceded by prior separation of metabolites via chromatography (GC or LC). The advantage with GC–MS is the huge knowledge base that is available which allows identification of a wide spectrum of metabolites. However, one disadvantage of GC-based approaches is that most human metabolites are not volatile and hence these non-volatile metabolites have to be extensively derivatized for making them volatile and amenable to the MS-based detection system. In contrast, LC-based approaches can directly analyze solutions without the need for derivatization.

#### 4.3. Metabolomics in colorectal cancer

While genetic alterations have been extensively characterized in CRC, the metabolic events downstream to genomics, transcriptomics, and proteomics have not been analyzed in depth, so far. A major reason for this is the lack of tools that can successfully evaluate low molecular weight metabolites in a high-throughput manner (metabolomics). Recent advances in technology have led to the application of metabolomics to research into different types of cancer including brain, breast, ovary, and colon using cell lines, *ex vivo* models, animal models, and clinical samples. Recent tissue-based metabolomics approaches have shown considerable promise in identifying biomarkers for human ovarian carcinomas, prostate cancer, cervical cancer, brain cancer, etc. [68].

Successful applications of metabolomics to the study of CRC require demonstration of metabolites in human tissues collected during routine surgery (i.e. biopsies and surgical resection), identification of as many metabolites as possible to enable robust pattern analysis that can yield functional information, and development of databases that can help build relationships between the observed metabolite fluxes and pathways in the test samples [69].

The literature on metabolite profiles in CRC, although limited, has recently been on the increase in the past few years. Several *in vitro*, *ex vivo*, and clinical studies with CRC cell lines, human tumor explants, and primary tumors have been attempted. These aim to define metabolic profiles in normal and precancerous/cancerous tissues and also evaluate prognostic/predictive biomarkers in blood that can link metabolism to clinical outcome and treatment response [69]. The ultimate objective of these studies is to classify tumors based on metabolic biochemistry so as to complement the genetic and proteomic classification. Examples of few of these emerging studies applying metabolomics technology to colon tumors are presented below.

Denkert et al., using a GC–MS approach, identified a panel of 82 metabolites that could differentiate CRC from normal mucosa in both the training and independent sets with a high degree of sensitivity. Of these 82 metabolites, 25 (metabolites in amino acid and purine synthesis) were upregulated, while 57 (mostly TCA-cycle intermediates) were down regulated in CRC [69].

Because metabolomics is an emerging field, different technologies have been used with different results. Chan et al. combined NMR (HRMAS-NMR) and GC–MS to analyze metabolites in CRC biopsies and their matched normal mucosa. The profiles obtained from both techniques were used to develop a metabolic profile that could differentiate malignant from normal samples. The identified metabolites belonged to cellular phenomena that are the usual suspects associated with tumorigenesis, i.e. increased tissue hypoxia, glycolysis, nucleotide biosynthesis, lipid/steroid metabolism, and inflammation. This discriminatory metabolite profile could distinguish tumors from different regions of the large intestine, i.e., colon and rectum, but was unsuccessful in separating tumors belonging to different Dukes stages [65].

In a small but interesting study by Mal et al., comparison of CRC to normal colon tissue yielded a discriminatory panel of 12 metabolites included organic acids (lactate, glycine), monosaccharides (mannose, galactose, glucose), fatty acids, etc. [68]. This pattern of metabolite levels was consistent with higher glucose uptake and conversion to lactate and increased glycolysis within the CRC relative to normal tissue. This phenomenon of cancer cells producing energy via glycolysis instead of TCA-cycle dependent oxidative phosphorylation even in presence of an oxygen supply was proposed by Warburg in the middle of the 20th century (Warburg effect) [70].

A reiteration of the theme of reduced tissue glucose and increased glycolysis was noted in a similar study using capillary electrophoresis time-of-flight MS. In this study, evaluation of 16

CRC and adjacent normal-appearing tissues yielded 94 metabolites that were differentially expressed [71]. These 94 metabolites were involved in glycolysis, pentose phosphate pathway, TCA and urea cycles, and amino acid and nucleotide metabolism. The extremely low levels of glucose, elevated lactate and other glycolytic pathway intermediates in tumor tissue further confirmed Warburg's hypothesis of an enhanced glycolytic mechanism and reduced oxidative phosphorylation in cancer tissues relative to normal tissues. Tumor tissues were also noted to have significant accumulation of all amino acids except glutamine which indicated increased protein catabolism (autophagic breakdown) and utilization of glutamine for energy synthesis via glutaminolysis. Intriguingly, there were no significant differences between CRC and normal tissues for most metabolites involved in nucleotide synthesis which implied that the selective advantage of tumor cells over normal cells was in their ability to exploit ancillary energy generating processes such as anaerobic glycolysis, glutaminolysis, and local catabolism of proteins to generate amino acids, with an ultimate objective to survive despite limited resources to fuel growth.

Metabolite profiling has been applied to a few *in vitro* and *ex vivo* colon cancer models [67]. In an animal study, differences between "normal mucosa" from APC<sup>Min/+</sup> mice and mucosa from wild type were identified at the metabolite level using HRMAS and solution-<sup>1</sup>H NMR [72]. The discriminatory profile included metabolites such as glycerophosphocholine, myoinositol, dimethyl amine, lactate, etc. Some of these metabolites are derived during lipid metabolism, gut microbial activity, or glycolysis. The differences in metabolite profiles in histologically similar mucosa from two distinct genotypes of mice highlight the concept of field cancerization, wherein early genetic and metabolic perturbations noted within apparently normal tissue lead to preneoplastic and overt neoplastic phenotype [73].

One distinct advantage of metabolomics is the ready application to body fluids viz. blood, plasma, urine, fecal water, etc. Monleón et al. successfully profiled metabolites within fecal water extracts from patients with CRC and identified lower concentration of short chain fatty acids (acetate and butyrate) and some alteration in the concentration of amino acids proline and cysteine (major components of colonic epithelial glycoproteins) [74].

Despite the limitation of detecting relevant metabolite signals from a background of constant metabolite fluxes responsive to diet, microenvironment, etc., specific studies have shown overlapping findings. For example, specific differences were noted in fecal water extracts and sera between patients with CRC and normal individuals. Moreover, sera from patients with CRC and adenomas were successfully discriminated from normal sera in 70% of CRC with a 5% false-positive rate in control samples [75]. The chemical uniqueness of a metabolite and its detection in biofluids obtained by non-invasive methods allow metabolic biomarkers to translate more readily from laboratory to clinic, and between animal models and man.

In summary, some of the key advantages of metabolomics are: it is relatively inexpensive after the initial investment in the technology, it is high-throughput and automated, and allows for interspecies comparison of data as metabolites are conserved across species (in contrast to gene and protein sequences). This eliminates the need for recalibration between studies spanning different species [66].

However, there are still numerous technical challenges in our ability to profile the entire spectrum of metabolites generated *in vivo*. Some of the key difficulties with metabolomics-based approaches are (1) technical, wherein a single separation method does not work for all metabolites as they come in widely variable forms. This limitation is evident from the study by Wishart et al. wherein they found only 8% of the total metabolites to be identified by NMR, GC-MS, and LC-FTMS [76]. Furthermore, current technol-

ogy only allows visualization of 10–15% of the metabolome and the possibility of a single method scanning the entire metabolome appears remote at this point [77], (2) inter-individual differences are markedly high because of the highly transient and sensitive nature of metabolic fluxes and the close interaction with micro- and macroenvironment including diet and microflora, (3) lack of uniformity in the technology and data analyses systems across studies, and (4) limited amount of accumulated data via databases that limits comprehensive and thorough identification of metabolites.

## 5. Conclusion

Although the reductionist approach of studying diseases have been effective in diseases caused by single or a few factors, this approach is highly limiting in complex multifactorial diseases. In complex, multifactorial diseases such as cancer, the most highly expressed genes or the most reproducible differences in genes often turn out to be downstream effects of deregulation in processes and pathways elsewhere. Moreover, even with identification of specific causal or contributory gene mutations, effective gene specific treatments have not been successful in general. This is most likely due to the complex downstream consequences propagated at the level of transcriptional, translational and post-translational circuits [78]. An integrative approach as advocated by systems biology considers entire networks of interconnected genes and is likely to detect subtle perturbations that might be more causal in nature.

Although, we are far from the "desired state" in this integrative approach, we have attained tremendous progress in this field, especially technical, in the last decade or so. Transcriptomics, proteomics, metabolomics, interactomics (study of protein-protein interactions on a global scale) have helped further our understanding of data that would have been impossible to interpret with common intuition.

What have we achieved so far by the application of employing transcriptomics, proteomics, and metabolomics to cancer? Although, high-throughput "omics" based applications have not yet been successful in identifying bonafide and consistent biomarkers that can accurately diagnose or prognosticate CRC, they have yielded interesting insights into the biology of CRC. For example, evaluation of genome, transcriptome, and proteome in distinct stages of CRC allowed the detection of chromosomal aneuploidies that could be correlated to average gene expression levels on the affected chromosomes. Such correlation between chromosomal aberrations/changes in DNA copy number and corresponding gene expression have not been restricted to colon cancer alone, but have been described in other solid tumors (breast, prostate, head and neck) and even shown in yeasts [79–86]. Moreover, such integrative approaches have led to the discovery of several novel candidate genes (*PLCG1* on 20q, *DBC1* on 8q21, and *NDGR1* on 8p24) in colon cancer [87].

Integrative "wholistic" approaches can now explain the lack of reproducibility of mutated and/or differentially expressed genes in CRC with pathway analysis tools. These tools show that the seemingly unrelated genes often fall in the same or similar pathway(s) suggesting that the numbers of affected pathways are less than previously thought. This concept of a set of pathways governing cell phenotype is important because pathway(s) can be therapeutically targeted through any of numerous comprised components irrespective of actual genetic defects in the intended target(s) [78]. This concept has been successful tested in renal cell carcinomas using a class of mammalian target rapamycin (mTOR) inhibitors that are effective for renal cell carcinomas. A common genetic defect associated with renal carcinomas is the loss of Von-Hippel Lindau tumor suppressor genes, a gene involved in degradation of HIF1/2/3, rather than any mutations related to mTOR gene [88].

Visualization of the complex pathways that represent a cancer cell for an individual patient might revolutionize personalized medicine. Herein, specific therapies can be designed to correct dysfunctional pathway in individual patients. Often times, apparently heterogeneous tumors of same type can be classified into smaller subsets based on the set of affected pathways. With such a classification system based on the system level would hopefully yield relevant therapies and diagnostic tools. Examples of such possibilities are minimal in the literature currently; but this is bound to change with time.

Therapeutic responses in breast and ovarian cancer to HER2-targeting receptor tyrosine kinase (RTK) inhibitors, such as trastuzumab and pertuzumab, are poorly correlated with HER2 protein expression [89]. Successful prediction of the therapeutic response can eliminate ineffective and expensive treatments with toxic side effect. In fact, systems biology approaches have allowed the stratification of patients based on PTEN expression thereby eliminating ineffective and potentially toxic treatment in approximately 25% of HER2<sup>+</sup> cancer patients [89].

By studying pathway interactions in three kinds of Hodgkins lymphoma, it was concluded that approximately 80% of 65,000 interactions in these tumors were similar—indicating a network backbone [90]. It was also noted that dysregulated pathway interactions could arise without component gene mutations in the pathway [90]. Targeting the “tumor network bone” could technically be a valuable therapeutic strategy to combat all three types of Hodgkins lymphoma. Such conclusions or approaches would have been impossible with the standard reductionist approaches.

Despite these advances, there exist numerous rate limiting steps in this field such as data collection, integration and assimilation into pathways and networks, and development of accurate model. These hurdles are being rapidly overcome by current technologies and strategies. In fact, in a recent joint EU-USA workshop on advancement of cancer research with systems biology had identified multiple areas that need work for the continued success of the systems biology approach. Some of the identified areas include “cell-context specific molecular interaction maps in cancer (cancer interactomes, widely available experimental platforms for rapid biochemical validation, cellular network based contexts for the integration of orthogonal data modalities including gene expression, SNPs, gene copy number, epigenetic data, etc., information on pathway synergy for therapeutic intervention, assembly and validation of cell-context specific molecular interaction maps for cancer cells (genes, proteins, miRNA, lipids, metabolites, etc.)” among others [91].

#### Conflict of interest

None.

#### Acknowledgments

The authors acknowledge the analytic input provided by Dennis Pelletier in Drug Safety Research Development at Pfizer Inc. VM is supported by the Claudia Adams Barr Program in Innovative Basic Cancer Research (PI: Dana H Gabuzda) at Dana-Farber Cancer Institute. We would also like to acknowledge scientists at GeneGo Inc. specifically Dr. Lellean JeBailey for helping us with MetaCore® analysis.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mrfmmm.2010.07.008.

#### References

- [1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M.J. Thun, Cancer statistics, 2008, *CA Cancer J. Clin.* 58 (2) (2008) 71–96.
- [2] C.J. Kahi, D.K. Rex, T.F. Imperiale, Screening, surveillance, and primary prevention for colorectal cancer: a review of the recent literature, *Gastroenterology* 135 (2) (2008) 380–399.
- [3] J.R. Jass, Limitations of the adenoma–carcinoma sequence in colorectum, *Clin. Cancer Res.* 10 (17) (2004) 5969–5970, author reply 5970.
- [4] J.D. Hardcastle, W.M. Thomas, J. Chamberlain, G. Pye, J. Sheffield, P.D. James, T.W. Balfour, S.S. Amar, N.C. Armitage, S.M. Moss, Randomised, controlled trial of faecal occult blood screening for colorectal cancer. Results for first 107,349 subjects, *Lancet* 1 (8648) (1989) 1160–1164.
- [5] O. Kronborg, C. Fenger, J. Olsen, O.D. Jorgensen, O. Sondergaard, Randomised study of screening for colorectal cancer with faecal-occult-blood test, *Lancet* 348 (2) (1996) 1467–1471.
- [6] J.S. Mandel, J.H. Bond, T.R. Church, D.C. Snover, G.M. Bradley, L.M. Schuman, F. Ederer, Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study, *N. Engl. J. Med.* 328 (19) (1993) 1365–1371.
- [7] D.A. Ahlquist, J.E. Skoletsky, K.A. Boynton, J.J. Harrington, D.W. Mahoney, W.E. Pierceall, S.N. Thibodeau, A.P. Shuber, Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel, *Gastroenterology* 119 (5) (2000) 1219–1227.
- [8] W.D. Chen, Z.J. Han, J. Skoletsky, J. Olson, J. Sah, L. Myeroff, P. Platzer, S. Lu, D. Dawson, J. Willis, et al., Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene, *J. Natl. Cancer Inst.* 97 (15) (2005) 1124–1132.
- [9] S.M. Dong, G. Traverso, C. Johnson, L. Geng, R. Favis, K. Boynton, K. Hibi, S.N. Goodman, M. D’Alessio, P. Paty, et al., Detecting colorectal cancer in stool with the use of multiple genetic targets, *J. Natl. Cancer Inst.* 93 (11) (2001) 858–865.
- [10] M. Li, W.D. Chen, N. Papadopoulos, S.N. Goodman, N.C. Bjerregaard, S. Laurberg, B. Levin, H. Juhl, N. Arber, H. Moinova, et al., Sensitive digital quantification of DNA methylation in clinical samples, *Nat. Biotechnol.* 27 (9) (2009) 858–863.
- [11] P. Gold, S.O. Freedman, Specific carcinoembryonic antigens of the human digestive system, *J. Exp. Med.* 122 (3) (1965) 467–481.
- [12] S. Huerta, Recent advances in the molecular diagnosis and prognosis of colorectal cancer, *Expert Rev. Mol. Diagn.* 8 (3) (2008) 277–288.
- [13] J. Cardoso, J. Boer, H. Morreau, R. Fodde, Expression and genomic profiling of colorectal cancer, *Biochim. Biophys. Acta* 1775 (1) (2007) 103–137.
- [14] R. Romero, J. Espinoza, F. Gotsch, J.P. Kusanovic, L.A. Friel, O. Erez, S. Mazaki-Tovi, N.G. Than, S. Hassan, G. Tromp, The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome, *BJOG* 113 (Suppl. 3) (2006) 118–135.
- [15] L.D. Wood, D.W. Parsons, S. Jones, J. Lin, T. Sjoblom, R.J. Leary, D. Shen, S.M. Boca, T. Barber, J. Ptak, et al., The genomic landscapes of human breast and colorectal cancers, *Science* 318 (5853) (2007) 1108–1113.
- [16] T. Sjoblom, Systematic analyses of the cancer genome: lessons learned from sequencing most of the annotated human protein-coding genes, *Curr. Opin. Oncol.* 20 (1) (2008) 66–71.
- [17] T. Sjoblom, S. Jones, L.D. Wood, D.W. Parsons, J. Lin, T.D. Barber, D. Mandelker, R.J. Leary, J. Ptak, N. Silliman, et al., The consensus coding sequences of human breast and colorectal cancers, *Science* 314 (5797) (2006) 268–274.
- [18] J. Lin, C.M. Gan, X. Zhang, S. Jones, T. Sjoblom, L.D. Wood, D.W. Parsons, N. Papadopoulos, K.W. Kinzler, B. Vogelstein, et al., A multidimensional analysis of genes mutated in breast and colorectal cancers, *Genome Res.* 17 (9) (2007) 1304–1318.
- [19] W.M. Claudino, A. Quattrone, L. Biganzoli, M. Pestrin, I. Bertini, A. Di Leo, Metabolomics: available results, current research projects in breast cancer, and future applications, *J. Clin. Oncol.* 25 (19) (2007) 2840–2846.
- [20] P.R. Nambiar, S.R. Boutin, R. Raja, D.W. Rosenberg, Global gene expression profiling: a complement to conventional histopathologic analysis of neoplasia, *Vet. Pathol.* 42 (6) (2005) 735–752.
- [21] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, A.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, *Proc. Natl. Acad. Sci. U.S.A.* 96 (12) (1999) 6745–6750.
- [22] Y.M. Lin, Y. Furukawa, T. Tsunoda, C.T. Yue, K.C. Yang, Y. Nakamura, Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas, *Oncogene* 21 (26) (2002) 4120–4128.
- [23] D.A. Notterman, U. Alon, A.J. Sierk, A.J. Levine, Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays, *Cancer Res.* 61 (7) (2001) 3124–3130.
- [24] S. Lechner, U. Muller-Ladner, B. Renke, J. Scholmerich, J. Ruschoff, F. Kullmann, Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia, *Gut* 52 (8) (2003) 1148–1153.
- [25] K. Noshu, H. Yamamoto, Y. Adachi, T. Endo, Y. Hinoda, K. Imai, Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis, *Br. J. Cancer* 92 (7) (2005) 1193–1200.
- [26] F. Bertucci, S. Salas, S. Eysteries, V. Nasser, P. Finetti, C. Ginestier, E. Charafe-Jauffret, B. Lloriod, L. Bachelart, J. Montfort, et al., Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters, *Oncogene* 23 (7) (2004) 1377–1391.
- [27] S. Eschrich, I. Yang, G. Bloom, K.Y. Kwong, D. Boulware, A. Cantor, D. Coppola, M. Kruhoffer, L. Aaltonen, T.F. Orntoft, et al., Molecular staging for survival prediction of colorectal cancer patients, *J. Clin. Oncol.* 23 (15) (2005) 3526–3535.

- [28] D. Agrawal, T. Chen, R. Irby, J. Quackenbush, A.F. Chambers, M. Szabo, A. Cantor, D. Coppola, T.J. Yeatman, Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling, *J. Natl. Cancer Inst.* 94 (7) (2002) 513–521.
- [29] K. Birkenkamp-Demtroder, L.L. Christensen, S.H. Olesen, C.M. Frederiksen, P. Laiho, L.A. Aaltonen, S. Laurberg, F.B. Sorensen, R. Hagemann, T.F. Ørntoft, Gene expression in colorectal cancer, *Cancer Res.* 62 (15) (2002) 4352–4363.
- [30] Y. Wang, T. Jatko, Y. Zhang, M.G. Mutch, D. Talantov, J. Jiang, H.L. McLeod, D. Atkins, Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer, *J. Clin. Oncol.* 22 (9) (2004) 1564–1571.
- [31] D. Arango, P. Laiho, A. Kokko, P. Alhopuro, H. Sammalkorpi, R. Salovaara, D. Nicorici, S. Hautaniemi, H. Alazzouzi, J.P. Mecklin, et al., Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer, *Gastroenterology* 129 (3) (2005) 874–884.
- [32] A. Barrier, P.Y. Boelle, A. Lemoine, C. Tse, D. Brault, F. Chiappini, F. Lacaine, S. Houry, M. Huguier, A. Flahault, et al., Gene expression profiling of nonneoplastic mucosa may predict clinical outcome of colon cancer patients, *Dis. Colon Rectum* 48 (12) (2005) 2238–2248.
- [33] A. Barrier, P.Y. Boelle, F. Roser, J. Gregg, C. Tse, D. Brault, F. Lacaine, S. Houry, M. Huguier, B. Franc, et al., Stage II colon cancer prognosis prediction by tumor gene expression profiling, *J. Clin. Oncol.* 24 (29) (2006) 4685–4691.
- [34] A. Barrier, A. Lemoine, P.Y. Boelle, C. Tse, D. Brault, F. Chiappini, J. Breittschneider, F. Lacaine, S. Houry, M. Huguier, et al., Colon cancer prognosis prediction by gene expression profiling, *Oncogene* 24 (40) (2005) 6155–6164.
- [35] A. Barrier, F. Roser, P.Y. Boelle, B. Franc, C. Tse, D. Brault, F. Lacaine, S. Houry, P. Callard, C. Penna, et al., Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling, *Oncogene* 26 (18) (2007) 2642–2648.
- [36] O. Galamb, B. Gyorffy, F. Sipos, S. Spisak, A.M. Nemeth, P. Miheller, Z. Tulassay, E. Dinya, B. Molnar, Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature, *Dis. Markers* 25 (1) (2008) 1–16.
- [37] O. Galamb, F. Sipos, N. Solymosi, S. Spisak, T. Krenacs, K. Toth, Z. Tulassay, B. Molnar, Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results, *Cancer Epidemiol. Biomarkers Prev.* 17 (10) (2008) 2835–2845.
- [38] T. Komori, I. Takemasa, M. Yamasaki, M. Motoori, T. Kato, N. Kikkawa, N. Kawaguchi, M. Ikeda, H. Yamamoto, M. Sekimoto, et al., Gene expression of colorectal cancer: preoperative genetic diagnosis using endoscopic biopsies, *Int. J. Oncol.* 32 (2) (2008) 367–375.
- [39] J.K. Habermann, U. Paulsen, U.J. Roblick, M.B. Upender, L.M. McShane, E.L. Korn, D. Wangsa, S. Kruger, M. Duchrow, H.P. Bruch, et al., Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis, *Genes Chromosomes Cancer* 46 (1) (2007) 10–26.
- [40] S.N. Dalal, M.B. Yaffe, J.A. DeCaprio, 14-3-3 Family members act coordinately to regulate mitotic progression, *Cell Cycle* 3 (5) (2004) 672–677.
- [41] A. Forrest, B. Gabrielli, Cdc25B activity is regulated by 14-3-3, *Oncogene* 20 (32) (2001) 4393–4401.
- [42] K. Jiang, E. Pereira, M. Maxfield, B. Russell, D.M. Goudelock, Y. Sanchez, Regulation of Chk1 includes chromatin association and 14-3-3 binding following phosphorylation on Ser-345, *J. Biol. Chem.* 278 (27) (2003) 25207–25217.
- [43] K. Katayama, N. Fujita, T. Tsuruo, Akt/protein kinase B-dependent phosphorylation and inactivation of WEE1Hu promote cell cycle progression at G2/M transition, *Mol. Cell. Biol.* 25 (13) (2005) 5725–5737.
- [44] T. Samuel, H.O. Weber, P. Rauch, B. Verdoodt, J.T. Eppel, A. McShea, H. Hermeking, J.O. Funk, The G2/M regulator 14-3-3sigma prevents apoptosis through sequestration of Bax, *J. Biol. Chem.* 276 (48) (2001) 45201–45206.
- [45] P.J. Gardina, T.A. Clark, B. Shimada, M.K. Staples, Q. Yang, J. Veitch, A. Schweitzer, T. Awad, C. Sugnet, S. Dee, et al., Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array, *BMC Genomics* 7 (2006) 325.
- [46] K. Thorsen, K.D. Sorensen, A.S. Brems-Eskildsen, C. Modin, M. Gaustadnes, A.M. Hein, M. Kruhoffer, S. Laurberg, M. Borre, K. Wang, et al., Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis, *Mol. Cell Proteomics* 7 (7) (2008) 1214–1224.
- [47] G.A. Evans, Designer science and the "omic" revolution, *Nat. Biotechnol.* 18 (2) (2000) 127.
- [48] N.G. Anderson, N.L. Anderson, Twenty years of two-dimensional electrophoresis: past, present and future, *Electrophoresis* 17 (3) (1996) 443–453.
- [49] P.H. O'Farrell, High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.* 250 (10) (1975) 4007–4021.
- [50] A. Cho, D. Normile, Nobel prize in chemistry. Mastering macromolecules, *Science* 298 (5593) (2002) 527–528.
- [51] F.R. Blattner, G. Plunkett 3rd, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, et al., The complete genome sequence of *Escherichia coli* K-12, *Science* 277 (5331) (1997) 1453–1474.
- [52] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, et al., The sequence of the human genome, *Science* 291 (5507) (2001) 1304–1351.
- [53] A. Shevchenko, O.N. Jensen, A.V. Podtelejnikov, F. Saggiocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, M. Mann, Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels, *Proc. Natl. Acad. Sci. U.S.A.* 93 (25) (1996) 14440–14445.
- [54] H.J. Issaq, The role of separation science in proteomics research, *Electrophoresis* 22 (17) (2001) 3629–3638.
- [55] C. Melle, G. Ernst, B. Schimmel, A. Bleul, H. Thieme, R. Kaufmann, H. Mothes, U. Settmacher, U. Claussen, K.J. Halbhuber, et al., Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer, *Gastroenterology* 129 (1) (2005) 66–73.
- [56] J.K. Yu, Y.D. Chen, S. Zheng, An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics, *World J. Gastroenterol.* 10 (21) (2004) 3127–3131.
- [57] J.Q. Li, B.J. Xu, B. Shakhtour, N. Deane, N. Merchant, M.J. Heslin, K. Washington, R.J. Coffey, R.D. Beauchamp, Y. Shyr, et al., Variability of in situ proteomic profiling and implications for study design in colorectal tumors, *Int. J. Oncol.* 31 (1) (2007) 103–111.
- [58] C.C. Wu, H.C. Chen, S.J. Chen, H.P. Liu, Y.Y. Hsieh, C.J. Yu, R. Tang, L.L. Hsieh, J.S. Yu, Y.S. Chang, Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes, *Proteomics* 8 (2) (2008) 316–332.
- [59] J. Albrethsen, R. Bogebo, S. Gammeltoft, J. Olsen, B. Winther, H. Raskov, Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1–3) in colon cancer serum and tumours: a biomarker study, *BMC Cancer* 5 (2005) 8.
- [60] J.K. Habermann, U.J. Roblick, B.T. Luke, D.A. Prieto, W.J. Finlay, V.N. Podust, J.M. Roman, E. Oevermann, T. Schiedeck, N. Homann, et al., Increased serum levels of complement C3a anaphylatoxin indicate the presence of colorectal tumors, *Gastroenterology* 131 (4) (2006) 1020–1029, quiz 1284.
- [61] D.G. Ward, N. Suggett, Y. Cheng, W. Wei, H. Johnson, L.J. Billingham, T. Ismail, M.J. Wakelam, P.J. Johnson, A. Martin, Identification of serum biomarkers for colon cancer by proteomic analysis, *Br. J. Cancer* 94 (12) (2006) 1898–1905.
- [62] R.H. Fletcher, Carcinoembryonic antigen, *Ann. Intern. Med.* 104 (1) (1986) 66–73.
- [63] A.W. Bell, E.W. Deutsch, C.E. Au, R.E. Kearney, R. Beavis, S. Sechi, T. Nilsson, J.J. Bergeron, T.A. Beardslee, T. Chappell, et al., A HUPO test sample study reveals common problems in mass spectrometry-based proteomics, *Nat. Methods* (2009).
- [64] S. Rochfort, Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research, *J. Nat. Prod.* 68 (12) (2005) 1813–1820.
- [65] E.C. Chan, P.K. Koh, M. Mal, P.Y. Cheah, K.W. Eu, A. Backshall, R. Cavill, J.K. Nicholson, H.C. Keun, Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS), *J. Proteome Res.* 8 (1) (2009) 352–361.
- [66] J.L. Griffin, J.P. Shockcor, Metabolic profiles of cancer cells, *Nat. Rev. Cancer* 4 (7) (2004) 551–561.
- [67] J.L. Griffin, R.A. Kauppinen, Tumour metabolomics in animal models of human cancer, *J. Proteome Res.* 6 (2) (2007) 498–505.
- [68] M. Mal, P.K. Koh, P.Y. Cheah, E.C. Chan, Development and validation of a gas chromatography/mass spectrometry method for the metabolic profiling of human colon tissue, *Rapid Commun. Mass Spectrom.* 23 (4) (2009) 487–494.
- [69] C. Denkert, J. Budczies, W. Weichert, G. Wohlgemuth, M. Scholz, T. Kind, S. Niesporek, A. Noske, A. Buckendahl, M. Dietel, et al., Metabolite profiling of human colon carcinoma—deregulation of TCA cycle and amino acid turnover, *Mol. Cancer* 7 (2008) 72.
- [70] O. Warburg, On the origin of cancer cells, *Science* 123 (3191) (1956) 309–314.
- [71] A. Hirayama, K. Kami, M. Sugimoto, M. Sugawara, N. Toki, H. Onozuka, T. Kinoshita, N. Saito, A. Ochiai, M. Tomita, et al., Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry, *Cancer Res.* 69 (11) (2009) 4918–4925.
- [72] A. Backshall, D. Alferez, F. Teichert, I.D. Wilson, R.W. Wilkinson, R.A. Goodlad, H.C. Keun, Detection of metabolic alterations in non-tumor gastrointestinal tissue of the Apc(Min/+) mouse by (1)H MAS NMR spectroscopy, *J. Proteome Res.* 8 (3) (2009) 1423–1430.
- [73] B.J. Braakhuis, M.P. Tabor, J.A. Kummer, C.R. Leemans, R.H. Brakenhoff, A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications, *Cancer Res.* 63 (8) (2003) 1727–1730.
- [74] D. Monleon, J.M. Morales, A. Barrasa, J.A. Lopez, C. Vazquez, B. Celda, Metabolite profiling of fecal water extracts from human colorectal cancer, *NMR Biomed.* 22 (3) (2009) 342–348.
- [75] C. Ludwig, D.G. Ward, A. Martin, M.R. Viant, T. Ismail, P.J. Johnson, M.J. Wakelam, U.L. Gunther, Fast targeted multidimensional NMR metabolomics of colorectal cancer, *Magn. Reson. Chem.* (2009).
- [76] D.S. Wishart, M.J. Lewis, J.A. Morrissey, M.D. Flegel, K. Jeroncic, Y. Xiong, D. Cheng, R. Eisner, B. Gautam, D. Tzur, et al., The human cerebrospinal fluid metabolome, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 871 (2) (2008) 164–173.
- [77] N. Blow, Metabolomics: biochemistry's new look, *Nature* 455 (7213) (2008) 697–700.
- [78] P.K. Kreeger, D.A. Lauffenburger, Cancer systems biology: a network modeling perspective, *Carcinogenesis* 31 (1) (2010) 2–8.
- [79] E.J. Douglas, H. Fiegler, A. Rowan, S. Halford, D.C. Bicknell, W. Bodmer, I.P. Tomlinson, N.P. Carter, Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas, *Cancer Res.* 64 (14) (2004) 4817–4825.
- [80] M. Grade, P. Hormann, S. Becker, A.B. Hummon, D. Wangsa, S. Varma, R. Simon, T. Liersch, H. Becker, M.J. Difilippantonio, et al., Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas, *Cancer Res.* 67 (1) (2007) 41–56.

- [81] T.R. Hughes, C.J. Roberts, H. Dai, A.R. Jones, M.R. Meyer, D. Slade, J. Burchard, S. Dow, T.R. Ward, M.J. Kidd, et al., Widespread aneuploidy revealed by DNA microarray expression profiling, *Nat. Genet.* 25 (3) (2000) 333–337.
- [82] E. Hyman, P. Kauraniemi, S. Hautaniemi, M. Wolf, S. Mousset, E. Rozenblum, M. Ringner, G. Sauter, O. Monni, A. Elkhoulou, et al., Impact of DNA amplification on gene expression patterns in breast cancer, *Cancer Res.* 62 (21) (2002) 6240–6245.
- [83] B.G. Masayeva, P. Ha, E. Garrett-Mayer, T. Pilkington, R. Mao, J. Pevsner, T. Speed, N. Benoit, C.S. Moon, D. Sidransky, et al., Gene expression alterations over large chromosomal regions in cancers include multiple genes unrelated to malignant progression, *Proc. Natl. Acad. Sci. U.S.A.* 101 (23) (2004) 8715–8720.
- [84] J.L. Phillips, S.W. Hayward, Y. Wang, J. Vasselli, C. Pavlovich, H. Padilla-Nash, J.R. Pezullo, B.M. Ghadimi, G.D. Grossfeld, A. Rivera, et al., The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis, *Cancer Res.* 61 (22) (2001) 8143–8149.
- [85] J.R. Pollack, T. Sorlie, C.M. Perou, C.A. Rees, S.S. Jeffrey, P.E. Lonning, R. Tibshirani, D. Botstein, A.L. Borresen-Dale, P.O. Brown, Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors, *Proc. Natl. Acad. Sci. U.S.A.* 99 (20) (2002) 12963–12968.
- [86] D. Tsafrir, M. Bacolod, Z. Selvanayagam, I. Tsafrir, J. Shia, Z. Zeng, H. Liu, C. Krier, R.F. Stengel, F. Barany, et al., Relationship of gene expression and chromosomal abnormalities in colorectal cancer, *Cancer Res.* 66 (4) (2006) 2129–2137.
- [87] J.F. Reid, M. Gariboldi, V. Sokolova, P. Capobianco, A. Lampis, F. Perrone, S. Signoroni, A. Costa, E. Leo, S. Pilotti, et al., Integrative approach for prioritizing cancer genes in sporadic colon cancer, *Genes Chromosomes Cancer* 48 (11) (2009) 953–962.
- [88] A. Kapoor, R.A. Figlin, Targeted inhibition of mammalian target of rapamycin for the treatment of advanced renal cell carcinoma, *Cancer* 115 (16) (2009) 3618–3630.
- [89] D. Faratian, A. Goltsov, G. Lebedeva, A. Sorokin, S. Moodie, P. Mullen, C. Kay, I.H. Um, S. Langdon, I. Goryanin, et al., Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab, *Cancer Res.* 69 (16) (2009) 6713–6720.
- [90] K.M. Mani, C. Lefebvre, K. Wang, W.K. Lim, K. Basso, R. Dalla-Favera, A. Califano, A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas, *Mol. Syst. Biol.* 4 (2008) 169.
- [91] R. Aebbersold, C. Auffray, E. Baney, E. Barillot, A. Brazma, C. Brett, S. Brunak, A. Butte, A. Califano, J. Celis, et al., Report on EU-USA workshop: how systems biology can advance cancer research (27 October 2008), *Mol. Oncol.* 3 (1) (2009) 9–17.