Integrative System Biology Strategies for Disease Biomarker Discovery

Haiyuan Zhang[*,1,§], Hao Hu[2,§], Cao Deng[2], Yeona Chun[3], Shengtao Zhou[4], Fuqiang Huang[2] and Qin Zhou[*,1]

[1]*The Medical School, Yangtze University, Hubei 434023, P.R. China*

[2]*The State Key Laboratory of Biotherapy, West China Hospital and College of Life Science, Sichuan University, Chengdu 610041, P.R. China*

[3]*Massachusetts Institute of Technology BS 2010 Cambridge, MA 02139, USA*

[4]*Department of Gynecology and Obstetrics, West China Second Hospital, Sichuan University, Chengdu 610041, P.R. China*

Abstract: Biomarkers are currently widely used to diagnose diseases, monitor treatments, and evaluate potential drug candidates. Research of differential Omics accelerate the advancements of biomarkers' discovery. By extracting biological knowledge from the 'omics' through integration, integrative system biology creates predictive models of cells, organs, biochemical processes and complete organisms, in addition to identifying human disease biomarkers. Recent development in high-throughput methods enables analysis of genome, transcriptome, proteome, and metabolome at an unprecedented scale, thus contributing to the deluge of experimental data in numerous public databases. Several integrative system biology approaches have been developed and applied to the discovery of disease biomarkers from databases. In this review, we highlight several of these approaches and identify future steps in the context of the field of integrative system biology.

Keywords: Disease biomarkers, high-throughput, integrative system biology.

1. INTRODUCTION

The Human Genome Project (HGP) and Model Organism Genome Project lead biological research to a new area of genomics. Since the accomplishment of HGP, the current goal of biological research focuses on integrative functional genomic studies on newly identified genes and crosstalk between genes and proteins and regulatory mechanisms, which can be defined as post-genomics. Based on this, 'omic' studies have become powerful tools for the analyses of genome; transcriptome, which measures mRNA transcript levels; proteome, which quantifies protein abundance and metabolome, which determines abundance of small cellular metabolites. With various highly developed 'omic' approaches and integration of these omic-data, system biology was set up as a powerful new paradigm for research in the life science.

A biomarker is in general substance used as an indicator of a biological state, which implicates an alteration in the expression state of a protein or other substances that correlate with the risk or progression of certain diseases. For most diseases, malignant tumors in particular, early diagnosis has the potential to significantly improve patients' outcomes. The recent progress of high-throughput approaches has brought an exciting promise of biomarker discovery such as cDNA microarray; serial analysis of gene expression (SAGE), mass spectrometry (MS) and nuclear magnetic resonance (NMR) based gene and metabolome profiles as well as integration with immunological technologies, biochemical technologies and bioinformatics (Table **1**). We endeavored to highlight some of these approaches used for biomarker discovery, thus making possible deeper understanding of human physiology and pathology through determination of the variations between pathologic cells and normal phenotypes.

2. THE ORIGIN OF SYSTEM BIOLOGY

At the turn of the 21th century, the word 'system' was highly popularized. Based on 'genetics' proposed in 1980s and then 'proteomics' in 1990s, a promising subject focused on 'system biology' was built up. ***Science*** presented a special issue with emphasis on system biology in 2002 and put the concept of 'system' to the peak, leading to a global understanding of 'system biology'. However, the notion of 'system' was not completely understood for decades. In fact, the concept was taken as rather emerging in Ludwig von Bertalanffy's book which was completed in 1940. In the meantime, Norbert Weiner also raised the understanding of 'system-level', and thus 'system theory' has existed for more than 70 years [1]. Nevertheless, it was late in the 20th century that 'system theory' was applied to biology science recurrently owing to the progression of molecular biology, particularly in genome sequencing and high-throughput approaches. This provided the science community a large scale of data and will be likely to build up a more comprehensive map of whole cell system's structure and dynamics.

*Address correspondence to these authors at the Medical School, Yangtze University, Jingzhou Hubei 434000, People's Republic of China; Tel: +86-13419625818, +86-18980601823;
E-mails: zhanghaiyuan1972@hotmail.com, pkdzhou@126.com

[§]These authors contributed equally to this work.

1875-5402/12 $58.00+.00

A system study on biology is not a simple assembly of various genes or proteins. It is not just a beautiful but static painting, so merely a diagram of interactions is insufficient to display the real activities that a cell conducted. To understand this concept, Hiroaky gave a vivid example in which he drew an analog image on an organism and a plane [2]. Suppose that all the genes and proteins are like each standalone part of an airplane. It is not sufficient for us to understand the complexity underlying the engineered object just listing all of the individual components. However, when these components were constituted as an integral whole with dynamic connections with each other, a structural and functional airplane was formed. Similarly, explorers have finished the first step in the system-level studies since the accomplishment of Human Genome Project (HGP), and the current goal of biological research will focus on dynamic integrative 'omics' studies in assessing newly defined genes, crosstalk between genes or proteins, epigenetic modifications and regulatory mechanisms. An integral system-level understanding of biology mainly includes four key properties.

1) **System structure**. This is a basic knowledge for system biology as the purpose of this is to establish a network of gene interactions and biochemical pathways. After sequencing the whole genome of human and other model organisms and with the rapid development on technologies combined chemistry, physics and establishment of various bioinformatics website such as KEGG [3], GenMAPP [4], and Reactome [5], it is a far more achievable knowledge even though further improvements might be needed.

2) **System dynamics**. To understand how a system behaves over time under certain condition, metabolic analysis such as NMR might be utilized. With an isotope model, NMR can be conducted to quantify metabolite concentration as well as metabolic flux and this may elucidate essential mechanisms underlying specific behaviors.

3) **Modeling construction**. Biological and functional modules provide the conceptual framework from which models may be constructed and simulation tools then can be able to perform a mathematical simulation on a model retrieved from a file, database or graphic user interface and output the simulation results.

4) **Data integration**. Either with assembling various data resources into a single database or building an environment around distributed data sources, both of them are ideal ways for integration of diverse data which enable multiple features of a biological system to be analyzed and incorporated into a testable dynamic model (reviewed in Ref. [2, 6]).

3. HIGH-THROUGHPUT ANALYSIS FOR PROFILING OF CANCER TRANSCRIPTOME FOR BIOMARKER DISCOVERY

Throughout the twentieth century, much of cancer research has been devoted to analyses of genes that are expressed differently between cancer cells and their normal counterparts. However, it was not until the application of advanced technologies to gene expression analysis when a more comprehensive understanding of complex differences between the two types of cells was achieved. The high-throughput gene expression profiles, such as cDNA microarray and serial analysis of gene expression (SAGE) allows simultaneous analysis of thousands of genes (Table **1**). One important difference between these two methods lies in whether it is possible to identify unknown genes. Microarray analysis selects 'probes' directly from databases including GenBank, dbEST and UniGene. In contrast, SAGE undertakes a gene expression profiling without prior knowledge of the presence and sequence of genes to be analyzed [7].

Using cDNA microarray, Ray *et al.* identified 260 candidate genes and 259 ESTs which showed aberrant expression under HBx induction and played key roles in various aspects of intracellular activities [8]. The genes analyzed by array method demonstrated a close interaction between each other. For example, it has been already proved that altered expression levels of many MAPK family members were responsible for HBV-related hepatocarcinogenesis [9] and in this study *Map4k3* and *Map4k4* were up-regulated whereas *Map2k4* was down-regulated. As correlated with other studies, showing that sustained activation of Ras-Raf-MAPK pathways led to hepatocyte transformation in differentiated HBx-expressing AML12 cell line [10] and that Map2k4 functions as a metastasis suppressor gene in many types of cancers [11]. Additionally, STAT members were also found to be deregulated as well. Down-regulation of STAT1 can be in favor of cancer proliferation while in contrast, up-regulation of STAT6 can make animals become immuno-deficient and tumor sensitive. Proteins related to cell cycle regulation pathways were also found differentially expressed as a result of HBx interaction. Among these candidates, cyclin D1 and Dmtf1 showed an interesting link in the progression of HCC as HBx can cause cyclin D1 gene over-expression and then lead to the promotion of G1 to S phase transition. Furthermore, Dmtf1, which can bind and interfere with cyclin D1, was found to be down-regulated, emphasizing the critical function of cyclin D1 in HBx-mediated transformation process. Hajjou *et al.* identified a functional category of up-regulated genes correlated with lipid synthesis using cDNA microarray analysis, including ATP citrate lyase (ACL), fatty acid synthase (FASN), sterol-responsive element-binding protein 2 (SREBP2) and retinol binding protein 1 (RBP1). It was the first trial to specifically identify enzymes involved in fatty acid pathways that were altered by HBV replication [12].

It has been reported that patients with type 2 diabetes mellitus suffer from increased risks of colorectal cancer development [13]. As the Wnt signaling pathway has been identified to play a critical role in initiation and progression of colorectal cancer (CRC), it raised a hypothesis that molecular crosstalk between insulin and Wnt signaling pathway might exist [14]. A cDNA microarray was utilized to assess the effect of insulin on gene expression profiles in rat intestinal non-cancer IEC-6 cell line before and after insulin treatment [15], finding that insulin up-regulated Wnt receptor Fzd-4 and TCF-4, effectors of Wnt signaling pathway. This study renewed the understanding of insulin-Wnt signaling pathways crosstalk, including not only the stimulation of β-catenin nuclear translocation [16], but also up-regulation of certain Wnt receptors and effectors.

For gastric cancer, microarray analysis was also applied for biomarker discovery and anticancer drug discovery. Arginine has been demonstrated to modulate cell growth and proliferation in different cancer cell lines but functions specifically due to the types of cancer. For example, it was reported that arginine stimulated cell growth in lung cancer while played a contradictory role to inhibit cell growth in a breast cancer cell line. Shayanthan *et al.* evaluated the effects of arginine on the growth of a human gastric cancer cell line AGS through cDNA microarray analysis [17]. The study found that arginine caused a decrease of AGS cell growth *via* increasing the expression of caspase 8, which is one of the important proteins involved in apoptosis at both mRNA level and protein level. The relative low-level of caspase 8 may be regarded as a biomarker but not so typical, however, arginine may act as a potent antineoplastic. Similarly, using cDNA microarray-based comparative genomic hybridization (GCH), one group selected 43 statistically significant differentially expressed genes and several of them were also previously reported as being deleted or being tumor suppressor or highly homologous to known oncogenes [18]. For instance, GATA4, MAOA, CCNC mRNA synthesis were diminished dramatically in this study, correlated with other diseases such as ovarian cancer, Norrie disease, and lymphoblastic leukemia respectively, suggesting these genes may be candidate tumor suppressors. Likewise, MASL1, truncated form of HMGA2, PPARBP and GRB7 were shown to be candidate oncogenes. This study provided a list of several disease biomarkers for further exploring the development and progression of gastric cancer.

In addition to mRNA-based profiling methods for quantifying mRNA level, the non-coding RNAs such as microRNAs have also been suggested as potential biomarkers for tissue classification and identification [19, 20]. One microarray analysis applying to microRNAs was performed to distinguish hepatocellular carcinoma from metastatic tumors in the liver [21]. The study demonstrated that two microRNAs, hsa-miR-141 and hsa-miR-200c had significant higher level in non-hepatic epithelial tumors to promote epithelial phenotypes while endothelial-associated hsa-miR-126 showed higher expression levels in hepatocellular carcinomas in contrast. Thus, the expression patterns of microRNAs can be regarded as useful biomarkers for diagnosis of liver malignancies.

As mentioned at the beginning, SAGE can analyze thousands of genes simultaneously even though the function of them is totally unknown. The principle of SAGE is to quantify a 'tag' which represents the transcription product of a gene rather than measuring the intact gene of interest (Fig. **1**). A tag, for purpose of SAGE, is a nucleotide sequence of a defined length, directly 3'-adjacent to the 3'-most restriction site for a particular restriction enzyme. Using this method, Felix and coworkers identified 46 up-regulated genes from multiple myeloma compared with normal plasma cells [22]. Among them, four genes (p53csv, ddx, maokapk2 and ranbp2) were considered to be critical for myeloma tumorigenesis and could potentially be used as biomarkers for the diagnosis of this disease. SAGE also contributed to distinguish intrahepatic cholangiocarcinoma (ICC) from hepatocellular carcinoma (HCC) and metastatic adeno-carcinoma by identifying three specific genes as markers for ICC:biglycan, insulin-like growth factor-binding protein 5 (IGFBP 5) and claudin-4 [23]. It is widely known that DNA is read from 3' to 5' during transcription; mean-while, the complementary RNA is created from 5' to 3' direction and there exists heterogeneity in human mRNA start sites [24], leading to alteration of transcription start sites in different types of cells.

Cancer cells may also possess a similar phenomenon and benefit for cancer development and progression. Recent advances in molecular biology have enabled genome-wide analysis of the 5'-end region of mRNA that revealed the variation in transcriptional start sites based on 5'-end serial analysis of gene expression (5'SAGE) [24, 25] rather than traditional 3'SAGE. With the help of this strategy, YuJi Hodo *et al.* identified transcripts initiated from introns of a gene which encoded acyl-coenzyme A oxidase 2 (ACOX2) and it was confirmed that the level of ACOX2 was increased with these transcript variants compared with background non-tumor tissues. ACOX2 is a rate limiting enzyme involved in the degradation of long branched fatty acid and bile acid intermediates in peroxisomes [26]. It is down-regulated in several types of cancers [27, 28]. However, HCC cells generate the same consequence *via* a distinct approach through over-transcription of intronic-ACOX2 genes because it encoded the C-terminal of canonical ACOX2, lacking the active sites to bind FAD and substrate, thus functionally departed [29].

4. PROTEOMICS BASED APPROACHES FOR BIOMARKER VALIDATION

Global comparative proteomics has long been used for new biomarker identification that improves early diagnosis and contributes to potential therapeutic targets [30, 31] (Table **1**). The main strategies of proteomics comprise: traditional 2-DE to 2D difference gel electrophoresis (2D-DIGE) and stable isotope labeling by amino acids in cell culture (SILAC); quantitative comparative proteomics such as isotope-coded affinity tagging (ICAT), isobaric tags for relative and absolute quantitation (iTRAQ), non-isotope quantitative proteomics strategies and multiple-reaction monitoring (MRM) based proteomics (reviewed in Ref. [32]). Several examples are given below.

4.1. 2-DE-MS/2D-DIGE/Based Comparative Proteomics

2-DE is a form of gel electrophoresis commonly used to separate proteins by two properties in two dimensions on 2D gels. Coupled with MS, chemical structures of the separated proteins are determined. 2-DE-MS is the most widely used proteomic approach for bioresearch. With 2-DE-MS/MS profiling of HepG2-HBx cell line and control cell lines, Tong's group found differential expressed proteins between the two lines. Among them, three major protein clusters including calcium homeostasis, retinol metabolism and proteolysis system-related proteins regulated by HBx were identified [33]. As it has been shown that increased cytosolic calcium, mediated by HBx protein or replicating HBV virus, was a fundamental requirement for HBV replication and core viral particle assembly [34-37]. Cytosolic calcium (Ca+) may act as a key biomarker of HBV infection. Three other

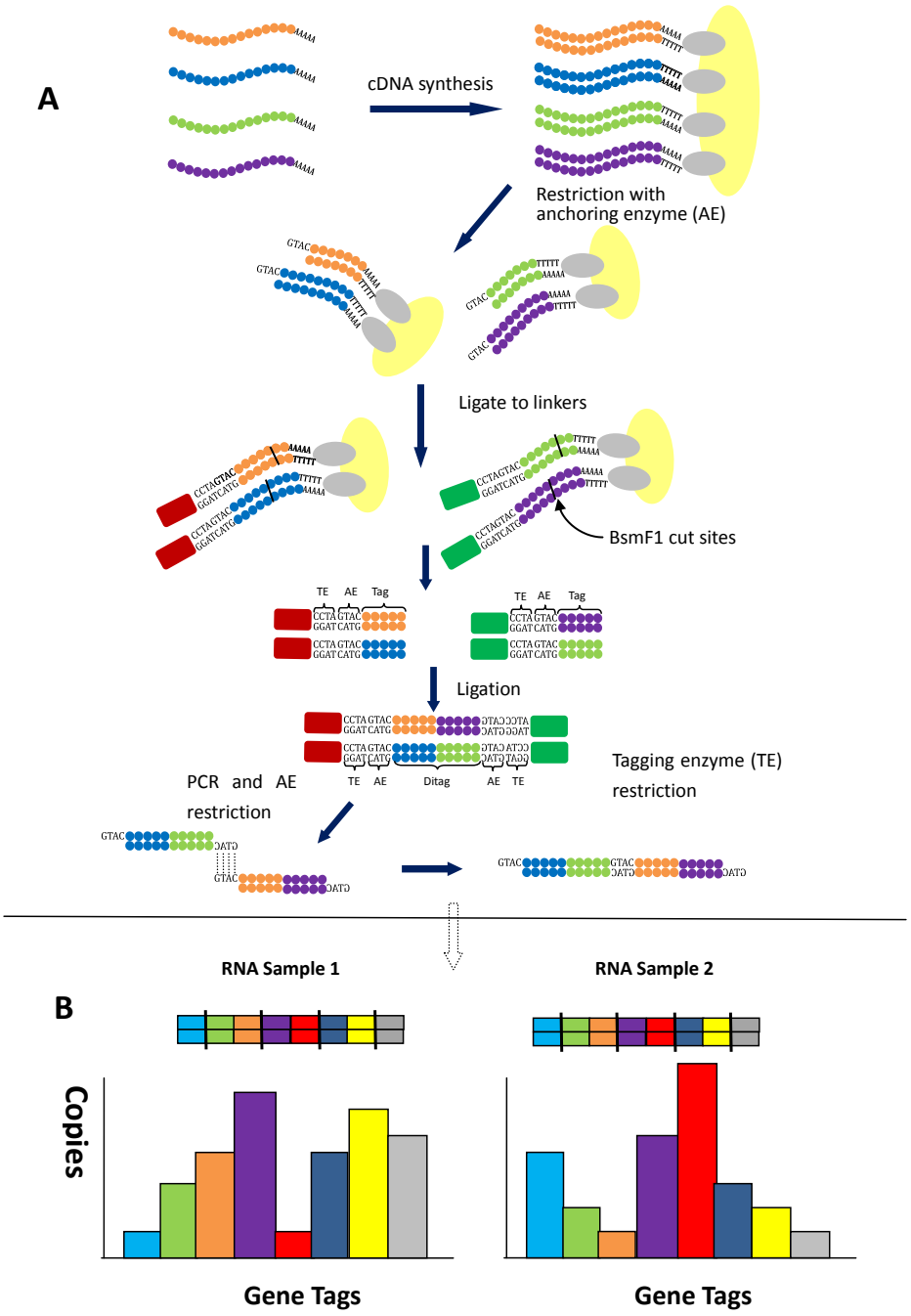

Fig. (1). A schematic diagram of SAGE. The anchoring enzyme and the tagging enzyme are NIa III and Fok I respectively. The circles with different colors stand for different tags. The tag is often a short sequence nucleotides containing about 10-14bp with sufficient information to uniquely identify a transcript for the reason that the tag is obtained from a unique position within each transcript. Then the sequence tags can be linked together to form a long serial molecules which are able to be sequenced and cloned. After analysis several times to quantify the particular tag, the expression level of the transcript of interest and its corresponding transcript can be observed (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

proteins retinal dehydrogenase 1 (RDH 1), plasma retinol-binding protein (pRBP) precursor and cellular retinol-binding protein I (CRBP I), involved in retinol (vitamin A) metabolism were strongly up-regulated *via* promoter hypomethylation due to down-regulation of DNA methytransferases (DNMTs) and the proteins responsible for proteolysis pathways showed alterations as well. It is worthwhile to notice that deregulation of DNMTs at the early stage of cancer development including hepatocarcinoma may also be an area of interest for biomarker identification [38-40]. The interaction between HBx and DNA methyltransferase (DNMT) or histone deacetylase (HDAC) facilitated epigenetic modification *via* regional hypermethylation or hypomethylation, such as silencing by DNA hypermethylation of several tumor suppressor genes, including IGFBP, RASSF1A [41], $p16^{INK14}$, E-cadherin [42, 43], GSTP [44] and DLEC1 [45]. This protein-protein interaction may be determined through many classic

methods such as yeast two-hybrid system, GST pull-down assay and immuno-precipitation [46].

As for bladder, 2-DE coupled MS/MS indicated seven proteins were down-regulated and three up-regulated in grade III transitional cell carcinoma (TCC) as compared with those of grade II [47]. Of these deregulated proteins, fatty acid binding proteins (FABP), annexinV, heat-shock protein (HSP) 27 and lactated dehydrogenase were shown to be relevant to bladder cancer. Besides, some newly-identified proteins that were not documented previously were also found in bladder, including annexin I, 15-hydroxyprostaglandin dehydroge-nase, galectin-1, lysophospholipase and mitochondrial short-chain enoyl-coenzyme A hydratase 1 precursor.

Applied to gastric cancer research, 2-DE-MS/MS identified a novel thyroid hormone-mediated gastric carcinogenic signaling pathway [48]. Concretely, a total of 107 proteins were identified as differentially expressed, and of these 57 were up-regulated whereas 50 were down-regulated and by further validation, four proteins (transthyretin (TTHY), pyruvate kinase muscle isozyme (KPYM), fumarate hydratase (FUMH), hypoxia up-regulated protein 1 (HYOU)) led to fumarate accumulation and then induced HIF1-α expression. The up-regulation of HIF1-α plays an essential role in sustaining glycolytic metabolism [49-51] which is mainly adopted by most cancer cells [52]. Thus, targeting these four proteins may exert anti-tumor effect.

2D-DIGE, a new generation of 2DE, dramatically improves the sensitivity and accuracy by adding a highly accurate quantitative dimension. By means of this technique, Langereis *et al*. discussed about proteins involved in 'inside-out' signaling which is essential for adequate migration of leukocytes to inflammatory sites [53]. Compared with two variants of a mouse acute lymphocytic leukemia cell line (L1210), a suspension (L1210-S) and an adherent (L1210-A) variant, gelsolin, L-plastin, and Rho GTPase dissociation inhibitor 2 were demonstrated as regulatory proteins functioned in integrin inside-out control and alteration of these three proteins may indicate early pathological state of lymphocytic leukemia cells.

4.2. SILAC/ICAT/iTRAQ Based Quantitative Comparative Proteomics

SILAC is a popular method for quantitative proteomics that detects differences in protein abundance among samples using non-radioactive isotopic labeling based on mass spectrometry [54]. Protein abundance is then calculated as ratios of the peak intensity of the fragment ions from the labeled versus unlabeled peptides. With high-throughput and quantitative features of SILAC, Ren *et al.* identified phosphoglycerate mutase 1 (PGAM1) as a novel therapeutic target in hepatocellular carcinoma [55]. On the other hand, SILAC was also applied in human pancreatic cancer [56], it not only validated proteins which were previously reported as deregulation in cancerous pancreatic cells, but also found several proteins that have not been correlated previously with pancreatic cancer such as perlecan (HSPG2), CD9 antigen, fibronectin receptor (integrin β1) and a novel protein designed as predicted osteoblast protein (FAM3C). These newly identified proteins were subsequently validated by other techniques such as western blot analysis, tissue microarray and further pursued as potential biomarkers.

ICAT and iTRAQ are both non-gel based techniques that can well circumvent the limitations of SILAC, furthermore, iTRAQ reagents even allow for the identification and quantitation of up to four different samples simultaneously [57]. With the help of these two high-throughput approaches, several achievements on biomarker discovery were accomplished, such as determination of the over-expressed vitamin D-binding protein in tumor-bearing breasts by using ICAT labeling [58] as well as identification of S100A6 and Annexin A2 by iTRAQ-coupled 2-D LC/MS-MS which acted as target of HBV-mediated epigenetic regulation and accelerated HCC development [59].

5. SYSTEM-LEVEL METABOLOMICS FOR THE IDENTIFICATION OF DISEASE BIOMARKER

Metabolomics is an emerging field which globally analyzes all or a large number of cellular metabolites and profiles the entire metabolome of a cell. Metabolomics is a potent approach for identifying pathways that are perturbed in a given pathology (Table **1**). Similar to other 'omics' approaches, it generates large amount of data and has been used for aging [60], immunology [61], toxicology [62, 63], nutrition [64], diabetes [65, 66], and, most importantly, cancer [67-69]. Analytical techniques that are most often used for metabolite profiling include nuclear magnetic resonance (NMR), gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), fourier-transform infrared (FT-IR) spectrometry and so forth (reviewed in [70]).

Since metabolomics is a comprehensive method and it can be expected that alterations in compound structures would disable metabolites target-driven analytical approaches. Hence, an unbiased way is needed to achieve through detecting as many metabolic features as possible and many software packages are commercially available for this purpose and biomarker detection as well [71]. As disease biomarkers is in favor of early diagnosis which has the potential to markedly improve patient survival [72] and due to the fact that the urine contains metabolic signatures of many biochemical pathways, it is ideal for metabolomic analysis serving as biomarker discovery tools. Of the human malignancies, urologic cancers seem to be most suitable for developing a urinary diagnostic assay given that this organ system is in intimate contact with this biofluid. Correspondingly, a study used hydrophilic interaction chromatography (HILIC–LC–MS), reversed-phase ultra performance liquid chromatography (RP–UPLC–MS), and gas chromatography time-of-flight mass spectrometry (GC–TOF–MS) to synthetically detect renal cell carcinoma (RCC) in urine of affected patients [73]. Although the glomerular pores might be too small to allow potential metabolites to be filtered into the urine, it is still readily appreciated that urine in RCC patients may contain abundant candidate metabolites as potential tumor markers. As a result, several significant components were generated, although still needed to be further validated, to lead to discrimination between RCC patients and the control.

Another case is closely related to RCC as well, in which LC-MS/MS was performed to discriminate the disparity among different grading systems as well as to acquire grade-specific mechanistic information [74]. As opposite to the study we showed above, the sample of this study was 50 clear cell kidney cancers equally distributed among normal tissues and Fuhrman grades. A total of 1,470,313 spectra were analyzed, resulting in the identification of 777 proteins containing two or more peptide fragments and in 777 proteins, 105 showed significant differences among the four Fuhrman grades and the normal kidney tissue. Among these proteins, several glycolytic enzymes were up-regulated, for example, phosphoglycerate kinase 1 (PGK1) and its elevation mainly occurred in grades 1 and 2, suggesting it is a relative early event in cancer progression. Additionally, grade-dependent changes in RCC were also associated with altered expression level of proteins involved in apoptosis pathway. One typical protein was apoptosis-inducing factor (AIF), a mitochondrial membrane –associated protein whose translocation from permeabilized mitochondrial to nucleus causing DNA fragmentation but its magnitude decreased in an opposite manner in terms of grades of the cancer. Intriguingly, the study also identified proteins of dehydrogenase class that had a significant high degree and these proteins played an important role in xenobiotic metabolism which is a marker for oncogenesis of kidney cancer [75]. The greatest contribution this study made for RCC was the generation of a heat map consisted of altered proteins. Since an individual marker cannot be well used for cancer prognosis and unable to be applied to regular clinical use for pathological classification either, this map of grade-specific protein SOM clusters will definitely contribute to prediction of tumor behavior.

Though researchers can monitor all the protein or metabolites alterations with these high-throughput approaches, many properties of the complex network still can not be understood extensively. Albeit extremely informative, the static component concentrations still lack of information *per se*. Instead, these components exert their functions through interacting with each others across multiple metabolic and regulatory layers and this dynamic activity will be able to reflect what is actually taking place in a living cell. Thus new approaches toward biomarker discovery have been proposed, where pathways instead of individual proteins should be monitored and targeted.

By means of LC-MS applying to secretome analysis of six cell lines in a pathways-based biomarker search study [76], 300 proteins and about 1500 unique peptides were generated. Among them, 71proteins out of the 300 were identified to be secreted or plasma membrane proteins and then integrated into a cellular signaling context. In the case of breast cancer cell lines of this research, several proteins linked to canonical signal pathways can be candidates for biomarkers and may be potential drug targets, including cathepsins and insulin growth factor binding protein (IGFBP) etc. Cathepsins are proteases located in lysosomes and shift to plasma membrane or extracellular space during neoplasia. This translocation is responsible for apoptosis, angiogenesis, cell proliferation and invasion [77, 78]. Meanwhile, overexpression of IGFBP can also result in a more aggressive breast cancer phenotype.

As to intracellular reaction, detection of metabolic fluxes is the functional end points of metabolic networks though the reaction rates can not be measured directly. Thus, there must be a need to establish an isotope model for dynamic metabolism interpretation. Take ^{13}C-based metabolic flux analysis as example, we will give a simple view below how high-throughput approaches quantify concentration of massive metabolites and try to provide a comprehensive under-standing of specific metabolic alteration caused by viral infection and cancer cells.

The study concerning virus infection developed a high-throughput measurement method based on liquid chromatography-tandem mass spectrometry to quantify changes and kinetics in metabolic fluxes of MRC-5 fibroblasts infected by human cytomegalovirus (HCMV) compared with uninfected cells [79]. We shall introduce the principle of this research first in prior of stating the result and contribution of it to biomarker discovery of HCMV-induced diseases and drug screening. As mammalian cells presented a complex metabolism and complicated metabolic input such as glucose, glutamine or essential amino acids, this complexity rendered analysis of steady-state labeling pattern inadequate for resolving fluxes in viable cells. Based on this, an isotope model was built by feeding cells with ^{13}C-labeled forms of glucose and glutamine. Through determination of labeled intermediates in HCMV-infected than in mock-infected cells, the data of metabolite concentration change can be generated. Take glycolysis and PPP as an example, lactate is the end product of glycolysis pathway, however, lactate can also be formed when glucose is metabolized by the non-oxidative PPP instead of glycolysis. Moreover, the composition of carbon atom of lactate derived from glucose through these two pathways is different, more specifically, when fed with [1, 2-^{13}C] glucose, nonoxidative PPP flux yields lactate containing just one ^{13}C atom, whereas, glycolysis flux does not. So, PPP/glycolytic flux ratio can be captured and this is meaningful since the two pathways are both important for cell metabolism and any alteration can lead to an abnormal condition of cells. HCMV infection increased the uptake of glucose and glutamine, enhanced glycolytic flux and TCA cycle fluxes. Most notably, efflux of citrate for fatty acid biosynthesis was approximately 20 fold elevated, consistent with upregulation of malonyl-CoA---a high regulated molecule in fatty acid synthesis from undetectable levels to amounts greater than ten-fold above the detecting limit. Given to this, fatty acid biosynthesis may be regarded as an antiviral target. Some pharmacological inhibitors of fatty acid biosynthetic enzymes were also proposed (i.e. TOFA, ACC inhibitor, C75, FAS inhibitor) and shown to be a safe and effective human therapeutics. Since there exists a potential link between cancer and viral infection and cells infected by virus (at least by HCMV), a parallel presentation of metabolism transformation such as elevated glycolysis is an analog of the Warburg effect and upregulation of lipid biosynthesis also shared between cancer cells and viral infection. Because of these considerations, understanding the mechanisms of virus-induced alteration in metabolic flux will not only help in viral infection, but also be beneficial to inform cancer research.

As described in the beginning of this part, NMR is the most often used metabolomic approach since it allows for

simultaneous detection, identification and quantification of a large scale of smaller-molecular-weight molecules and will facilitate monitoring of metabolite concentration. Unlike MS based metabolomics, NMR is non-selective, non-destructive, more robust, and able to be used for liquid and solid samples with minimal sample preparation (reviewed in Ref. [80]). Isotope-labeling based NMR can be used for metabolic flux detection which may directly serve as specific biomarker. Recently, a study was conducted by employing noninvasive hyperpolarized ^{13}CNMR to investigate on the metabolic flux of pyruvate to lactate [81]. Being well maintained by continuous perfusion method under controlled physiological conditions, kinetics and mechanisms of conversion of hyperpolarized pyruvate to lactate in human breast cancer cells were reliably characterized. Data generated in this perfused model indicated that the rate of hyperpolarized pyruvate did not reflect the activity of LDH, but rather the activity of monocarboxylate transporters (MCTs). It highlighted that lactate transportation would be the rate limiting step in converting glucose to lactate, resulting in altered process of glycolysis. That was to say that elevated expression of MCTs in cancers were highly associated with high-levels of glycolysis. The accumulation of lactated in extracellular microenvironment led to hypoxia conditions, thus, induced HIF1α, MCTs were over-expressed in a HIF-dependent manner and then fed back to this cycle [82]. Overall, these results suggested the kinetics of ^{13}C-pyruvate's conversion to lactate, shown to be a new biomarker for the up-regulation of MCTs in malignant breast transformations.

Additionally, Chen *et al.* introduced ^{13}C labeling NMR spectroscopy strategy for a comprehensive analysis of metabolic fluxes including glycolysis, pentose phosphate pathway, tricarboxylic acid cycle, anaplerotic reactions and biosynthetic pathways of fatty acids and amino acids in cancer cells [83] (Fig. **2**). As mentioned above, cancer cells harbor a unique metabolic phenotype compared with non-proliferated cells, which has been confirmed for decades. However, to date, the cause-effect relationship between this metabolic alteration and flux determination remains elusive, furthermore, increasing evidence indicated that various types of cancer cells associate with diverse substantial rewiring of metabolic networks. From these considerations, it is of importance to track metabolic dynamics of cancer cells to uncover specific cancer-related biomarkers. Chen's work mainly used human breast cancer as models to elucidate metabolic profiles of cancer metabolome, yet it was not clear, however, whether this global flux dysregulation was a

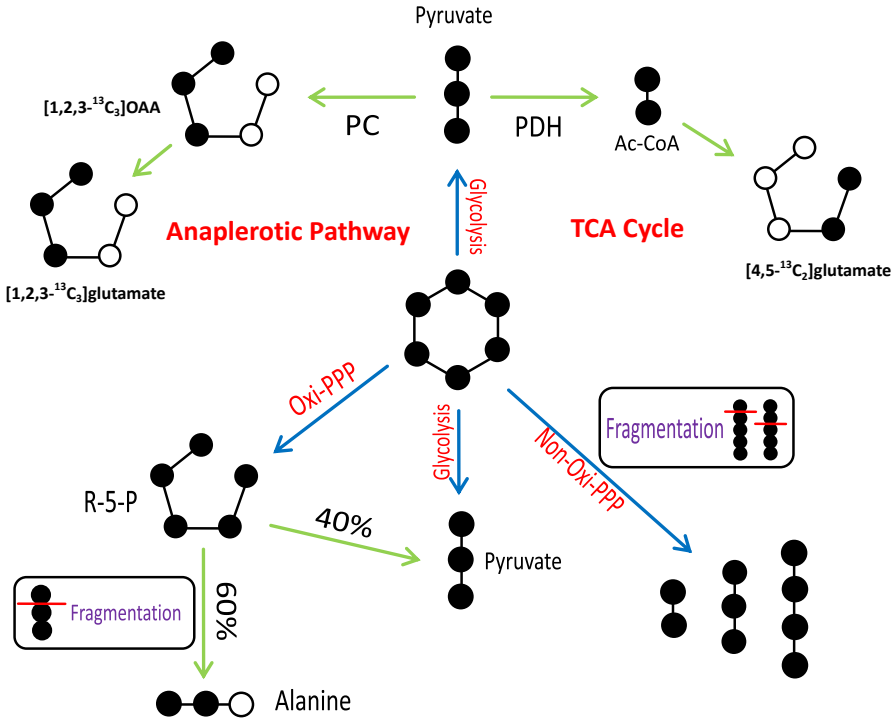

Fig. (2). Schematic illustration of ^{13}C isotopomer model for quantitative interpretation of relative metabolic fluxes in cancer cells. The principle is briefly described as follows. The analyses of the expression of [2,3-^{13}C] alanine can be used to estimate the relative activity of glycolysis vurses PPP in cancer cells. According to the carbon rearrangement that governs the PPP, three intact C1-C2-C3-C4-C5 fragments eventually yield five molecules of pyruvate. Three of them retain an intact C1-C2-C3 carbon bone while two truncated C2-C3 fragments come out of the rest molecules of pyruvate owing to a cleavage of C1-C2 carbon bond. For distinguish oxidative and non-oxidative PPP pathway, the ^{13}C labeling data of C2 of UTP or UDP that was converted from R5P was used. An intact C5 fragments were generated by R5P through oxidation of glucose-6-phosphate (Oxidative PPP pathway), however, in non-oxidative steps of PPP, the process catalyzed by the reversible transketolase and transaldolase, R5P was cleaved between C2-C3 and C1-C2 respectively. The same with above, [4,5-$^{13}C_3$] glutamine reflects the flux through pyruvate dehydrogenase and [1,2,3-$^{13}C_3$] glutamine can be utilized for the determination of pyruvate carboxylase activity which stands for anaplerotic pathway. Abbreviations: PC: pyruvate carboxylase; PDH: pyruvate dehydrogenase; Oxi-PPP: oxidative phosphopentose pathway; Non-oxi-PPP: non-oxidative phosphopentose pathway; OAA: oxaloacetate.

result of over-expression of glycolytic enzymes or ther effects. When compares MCF-10A cell line (a spontaneously immortalized non-malignant cell line) with MCF-10AT cell line (derivative of MCF-10A and mildly hyperplastic), the result showed a significant increase of glycolysis in the transformed MCF-10AT, however, the total flux from glucose from pyruvate is similar in both lines, indicating this glycolytic elevation in breast cancer is not caused by enzymes directly involved in glycolysis, instead, molecules in PPP might be candidate possibilities.

Before setting up the determination, there was a need to acquire a marker that can make a distinction between glycolysis and PPP. Alanine seemed to be a good candidate. Since pyruvate derived from these two pathways carried different carbon fragment, specifically, pyruvate converted from PPP yielded alanine with a distinct fraction Ala-C2, therefore, the result of a relative high intensity of the doublet of alanine C-2 was indicative the use of PPP for pyruvate synthesis, demonstrating that a change carbon flow from PPP to glycolysis present in breast cancer cells.

More accurately, The PPP pathway has two branch, namely oxidative and non-oxidative branch, they are generates ribose-5-phosphate from glucose-6-phosphate and glyceraldehydes-3-phosphate respectively and the difference between them mainly lies on the synthesis of NADPH. As ribose C-2 signal shows a sharp increase in oxidative PPP flux and in line with the result of a high intensity of alanine C-2, the increased level of glycolysis in MCF-10AT cells was achieved by passing through PPP, indicating the analysis of Ala-C2 could serve as a new biomarker for breast cancer (at least in this model).

As with glycolysis, ^{13}C NMR was also used for detecting the truncated TCA cycle. Upon low oxygen tension, hypoxic cells are surmised to suppress metabolism through the TCA cycle [82, 84] and anaplerotic reactions will allow intermediates to enter into TCA cycle and replenish fuel for cell proliferation [85]. There are two ways: first is to generate two ^{13}C isotopomers of ([1, 2, 3-^{13}C] and [2, 3, 4-^{13}C]) oxaloacetate from $^{13}C_3$ pyruvate and unlabeled CO_2 through pyruvate carboxylase. The second is glutamine, an important energy fuel in cancer cells (known as glutaminolysis [86]). According to rearrangement of carbon in the two pathways above, the ratio of [1, 2, 3-^{13}C]/[2,3-^{13}C] and non-labeled glutamate can determine which pathway the breast cancer cell line MCF-7 actually take. Since ^{13}C isotopomer NMR analysis performed in this study demonstrated the existence of two compartmental pools of glutamate in MCF-7 cells. One is directly derived from 2-ketoglutarate which could generate from oxaloacetate through TCA cycle, another is a non-enriched carbon source (i.e. glutamine). Glutamine can be converted into glutamate and then 2-ketoglutarate gets involved in TCA cycle. The cancer cells use a truncated TCA cycle from 2-ketoglutarate to malate, then malate is converted to pyruvate and complete oxidation. If glutamine was fed into TCA cycle, an isotopic dilution of TCA cycle intermediates would be observed. Among these intermediates, γ-glutamyl moiety of glutathione seemed to be most appropriate to determine which or whether both ways cancer cell would choose for its development because it is an irreversible one-way reaction from glutamate to glutathione. It is notable that γ-glutamyl moiety of glutathione are solely derived from acetyl-CoA. This phenomenon showed no dilution occurred at the level of TCA cycle-derived 2-ketoglutarat; hence, the malic enzyme activity is absent, together, indicating the lack of glutaminolysis and glucose as the only energy source in MCF-7 and also in MCF10 cells lines. This conclusion might run slightly against previous study which reported that glutamine acts as an important energy fuel in cancer cells and such a contravention may act as a specific biomarker for this case.

6. BIOINFORMATICS FOR INTEGRATIVE SYSTEM BIOLOGY

A comprehensive understanding of pathophysiology and disease biomarker detection requires a system level view based on the human genome, which is contrary to the reductionist approach focusing on individual molecules used to address complex biological system [87]. With rapid developments in high-throughput methodologies, massive amounts of data are generated and thus lead to establishment of numerous public databases. Based on different 'omic' study, these databases can be categorized into several layers and contribute to system biology (Fig. **3**). For gene expression data, several public repositories including Gene Expression Omnibus (GEO), Array Expression, CIBEX, Stanford Microarray Database (SMD) and Oncomine were built up and can be shared by anyone with internet. As referred to proteomic data, SWISS-2DPAGE, DCTB 2D-PAGE, GELBANK and 2D-PAGE/DIFF were developed based on the data generated by 2DE, while, ProteinProspector, Mascot, X!Tandem and Peptide-Atlas based on MS. The databases related to metabolomics embodies the Human Metabolite Database whose metabolites contained have a linkage to other public databases such as KEGG, PubChem, MetaCyc, ChEBI, PDB, SwissProt and GenBank and several other databases which were established with data generated from different technology platform, for example, the Golm Metabolome Database (GMD) is specific to data derived from GC-MS profiling experiments (details for above see Ref. [6])

7. CHALLENGES AND FUTURE PROSPECT

In order to obtain a useful understanding of diseases, it is necessary to take a systematic view of diseases with development of these high-throughput techniques. Identification of disease biomarkers can be achieved in multiple layers from genome to metabolome of abnormal cells. With the support of bioinformatics and numerals mathematical models, the mass of experimental data can be acquired and integrated.

However, new challenges have unavoidably emerged. Since alteration in transcriptional level can not fully represent a real change which can lead to phenotype transformation for a cell upon extracellular stresses or intracellular metabolism alteration, thus the techniques applied to measure mRNA levels may be not sufficient. Proteomics allows enhanced protein identification and represents a potentially powerful technique for large-scale qualitative and quantitative proteome research. Although it has been proven a promising tool for biomarker discovery,

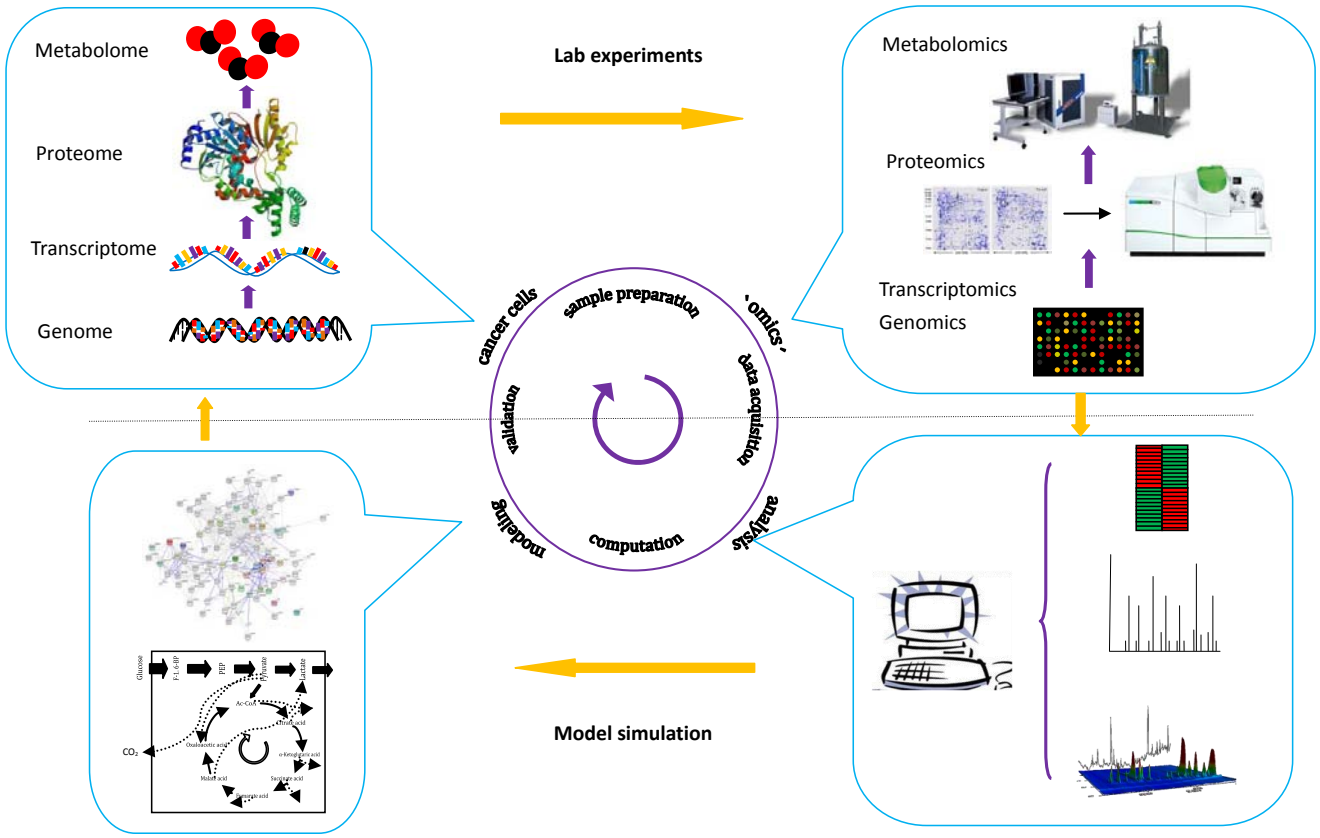

Fig. (3). A schematic model of system biology for biomarker discovery. Data generated through various 'omics' studies, intracellular network can be described with mathematical modeling techniques and bioinformatics. Biomarkers used to predict a pathological state of cells can be detected.

further work is still required to improve performance and reproducibility. On the other hand, most proteomic technologies are based on complex instrumentation, critical computing power and expensive consumables and this may be expensive. Actually, with the help of these high-performance instruments, a large quantity of data can be generated [88]. However, an exhaustive search on these multifarious data can not be really implemented. Most data were neglected by researchers and treated as waste in academic studies.

As for metabolomics, there are also many challenges. To better understand global changes in cancer metabolome, it is essential to identify and quantify metabolites during cancer development and the need for measuring metabolic flux and integration of flux data is clearly recognized. Based on this requirement, the isotope model was built up and among them, ^{13}C labeled glucose metabolism seemed to be an ideal model to elucidate cancer cells' metabolic activities. However, the available experimental flux methods themselves suffer from some limitations calling for improvement. For a tracer experiment, a metabolic steady state is a prior matter [89] and the microenvironment in cancer cells is kept changing, leading to interference of metabolic flux analysis. Furthermore, direct measurement of isotope-labeling distribution of the metabolites was impeded by their high turnover rates and low concentrations [90]. To resolve this problem, NMR or MS based measurements can detect the isotope patterns of primary metabolites in order to reduce the time which is needed for steady-state condition and this concept is well applicable to LC-MS compared to NMR-and GC-MS based isotope-labeling analysis [91]. However, this may raise another problem, which is, as many intermediates were inferred from the measurements of primary metabolites *via* known or assumed pathways, thus there is a need to complete all the possible multicellular networks in which the tracer atom is involved. Another challenge is to acquire and integrate the data of the metabolome in cancer cells. As this would be achieved through mathematical modeling, we are eagerly looking forward to a great progression on computational methods that will contribute to improve the current situation on data integration and processing.

But there are more than technical challenges for system biology. Because it is a discipline that requires a global insight into biological function from phenotype down to the molecular or even atomic level, the intense data integration [92] needs the concerted efforts made by not only experts of biology, but also talented chemists, physicists, computer scientists, mathematicians, software engineers. It could be anticipated that system biology will definitely take a lead in life science, unavoidably including disease biomarker discovery.

ACKNOWLEDGEMENTS

This work was supported by the National 973 Program of China (2010CB529900) and the National 863 High Tech Foundation (2007AA021205).

Table 1. Disease Biomarkers Discovery by Different 'Omics' Studies

Biomarkers	Types of Disease	Methods	Advantages	Disadvantages
Transcriptomics				
Map4k3, Map4k4, Map2k4, STATs, cyclin D1, Dmtf1	HCC [8]	microarray	1) high-throughput, analysis many genes simutaneously; 2) high-speed; 3) relatively cheap; 4) user-friendly, neither radioactive nor toxic; 5) adaptable and comprehensive	1) extremely large data cannot organized on simple spreadsheets; 2) difficult to distinguish from the same gene family 3) rarely expressed genes (important regulatory genes will be overlooked
ACL, FASN, SREBP 2, RBP 1	HBV-infection [12]			
Fzd-4 and TCF-4	CRC [15]			
Arginine	gastric cancer [17]			
GATA4, MAOA, CCNC, MASL1, PPARBP, GRB7	gastric cancer [18]			
hsa-miR-141, hsa-miR-200c	Epithelial cancer [21]			
p53csv, ddx, maokapk2, ranbp2	Myeloma [22]	SAGE	1) Test the expression of thousands of gene simutaneously 2) identify the gene of interest automatically	1) length of a tag is short 2) the restriction enzyme does not always yield same length fragments
biglycan, IGFBP 5, claudin-4	ICC [23]			
ACOX2	HCC [26]	5'SAGE		
Proteomics				
Ca^{+}, RDH 1, pRBP precursor, CRBP I, DNMTs	HCC [33]	2-DE-MS	1) 2-DE-MS is most widely used simultaneously profiles hundreds of proteins; 2) 2D-DIGE high sensitivity and accuracy; 3) SILAC, quantitative proteomic technologies with higher sensitivity; 4) ICAT, apply to two individual populations; 5) iTRAQ simultaneously quantify protein samples from 4 individual groups	1) 2-DE-MS, relative low sensitivity; 2) Need to enhance the performance and reproducibility; 3) Instruments and reagents can be expensive; 4) Difficult to fully use the large amounts of data
FABP, annexinV, HSP-27, annexin I, galectin-1 etc.	bladder cancer [47]	2-DE-MS/MS		
TTHY, KPYM, FUMH, HYOU	gastric cancer [48]			
L-plastin, Rho GTPase dissociation inhibitor 2	ALL [53]	2D-DIGE		
PGAM1	HCC [55]	SILAC		
HSPG2, CD9 antigen, integrin β1, FAM3C	pancreatic cancer [56]			
vitamin D-binding protein	breast cancer [58]	ICAT		
S100A6, Annexin A2	HCC [59]	iTRAQ		
Metabolomics				
PGK1, AIF, proteins of dehydrogenase class	RCC [74]	LC-MS/MS	1) high-sensitivity; 2) robust and reproducible; 3) wide analytical range	1) more costly; 2) suffer from ion suppression
cathepsins, IGFBP	breast cancer [76]	LC-MS		
fatty acid biosynthesis (such as ACC)	HCMV-infection [79]	^{13}C-LC/MS/MS		
MCTs	breast cancer [81]	^{13}C NMR	1) nondestructive; 2) sample preparation is straight-forward and automated	1) low sensitivity and small analytical range; 2) expensive
Ala-C2 etc.	breast cancer [83]			

ABBREVIATIONS

SAGE = Serial analysis of gene expression

MS = Mass spectrum

NMR = Nuclear magnetic resonance

HCC = Hepatocellular carcinoma

ICC = Intrahepatic cholangiocarcinoma

ALL = Acute lymphocytic leukemia

RCC = Renal cell carcinoma

RDH = Retinal dehydrogenase 1

pRBP = Plasma retinol-binding protein

TTHY = Transthyretin

FUMH = Fumarate hydratase

HYOU = Hypoxia up-regulated protein1

KPYM = Pyruvate kinase muscle isozyme

MCT = Monocarboxilate transporters

ACL = ATP citrate lyase

SREBP = Sterol regulatory element binding protein

REFERENCES

[1] Weiner, N. *Cybernetics or Control and Communication in the Animal and the Machine*. MIT Press: Cambridge, **1948.**

[2] Kitano, H. Systems biology: a brief overview. *Science*, **2002,** *295*(5560), 1662-1664.

[3] Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **2004**, *32*(Database issue), D277-280.

[4] Dahlquist, K.; Salomonis, N.; Vranizan, K.; Lawlor, S.; Conklin, B. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, **2002**, *31*(1), 19-20.

[5] Joshi-Tope, G.; Gillespie, M.; Vastrik, I. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **2005**, *33*, D428-432.

[6] Ng A, Bursteinas B, Gao Q, Mollison E, Zvelebil M. Resources for integrative systems biology: from data through databases to networks and dynamic system models. *Brief Bioinform.*, **2006**, *7*(4), 318-330.

[7] Velculescu, V.; Zhang, L.; Vogelstein, B.; Kinzler, K. Serial analysis of gene expression. *Science*, **1995**, *270*(5235), 484-487.

[8] Ng, R.; Lau, C.; Lee, S.; Tsui, S.; Fung, K.; Waye, M. cDNA microarray analysis of early gene expression profiles associated with hepatitis B virus X protein-mediated hepatocarcinogenesis. *Biochem. Biophys. Res. Commun.*, **2004**, *322*(3), 827-835.

[9] Xu, X.; Huang, J.; Xu, Z.; Qian, B.; Zhu, Z.; Yan, Q.; Cai, T.; Zhang, X.; Xiao, H.S.; Qu, J.; Liu, F.; Huang, Q.H.; Cheng, Z.H.; Li, N.G.; Du, J.J.; Hu, W.; Shen, K.T.; Lu, G..; Fu, G.; Zhong, M.; Xu, S.H.; Gu, W.Y.; Huang, W.; Zhao, X.T.; Hu, G.X.; Gu, J.R.; Chen, Z.; Han, Z.G. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*(26), 15089-15094.

[10] Lee, S.; Tarn, C.; Wang, W.; Chen, S.; Hullinger, R.; Andrisani, O. Hepatitis B virus X protein differentially regulates cell cycle progression in X-transforming versus nontransforming hepatocyte (AML12) cell lines. *J. Biol. Chem.*, **2002**, *277*(10), 8730-8740.

[11] Arbuthnot, P.; Capovilla, A.; Kew, M. Putative role of hepatitis B virus X protein in hepatocarcinogenesis: effects on apoptosis, DNA repair, mitogen-activated protein kinase and JAK/STAT pathways. *J. Gastroenterol. Hepatol.*, **2000**, *15*(4), 357-368.

[12] Hajjou, M.; Norel, R.; Carver, R.; Marion, P.; Cullen, J.; Rogler, L.E.; Rogler, C.E. cDNA microarray analysis of HBV transgenic mouse liver identifies genes in lipid biosynthetic and growth control pathways affected by HBV. *J. Med. Virol.*, **2005**, *77*(1), 57-65.

[13] Tran, T.; Naigamwalla, D.; Oprescu, A. Hyperinsulinemia, but not other factors associated with insulin resistance, acutely enhances colorectal epithelial proliferation *in vivo*. *Endocrinology*, **2006**, *147*(4), 1830-1837.

[14] Gregorieff, A.; Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes & Dev.*, **2005** *19*, 877-890.

[15] Sun, J.; Wang, D.; Jin, T. Insulin alters the expression of components of the Wnt signaling pathway including TCF-4 in the intestinal cells. *Biochim. Biophys. Acta.*, **2010**, *1800*(3), 344-351.

[16] Sun, J.; Khalid, S.; Rozakis-Adcock, M.; Fantus, I.; Jin, T. P-21-activated protein kinase-1 functions as a linker between insulin and Wnt signaling pathways in the intestine. *Oncogene*, **2009**, *28*(35), 3132-3144.

[17] Nanthakumaran, S.; Brown, I.; Heys, S.; Schofield, A. Inhibition of gastric cancer cell growth by arginine: molecular mechanisms of action. *Clin. Nutr.*, **2009**, *28*(1), 65-70.

[18] Yang, S.; Jeung, H.C.; Jeong, H.J.; Choi, Y.H.; Kim, J.E.; Jung. J.J.; Rha, S.Y.; Yang, W.I.; Chung, H.C. Identification of genes with correlated patterns of variations in DNA copy number and gene expression level in gastric cancer. *Genomics*, **2007**, *89*(4), 451-459.

[19] Volinia, S.; Calin, G.A.; Liu, C.G.; Ambs, S.; Cimmino, A.; Petrocca, F.; Visone, R.; Iorio, M.; Roldo, C.; Ferracin, M.; Prueitt, R.L.; Yanaihara, N.; Lanza, G.; Scarpa, A.; Vecchione, A.; Negrini, M.; Harris, C.C.; Croce, C.M. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*(7), 2257-2261.

[20] Rosenfeld, N.; Aharonov, R.; Meiri, E.; Rosenwald, S.; Spector, Y.; Zepeniuk, M.; Benjamin, H.; Shabes, N.; Tabak, S.; Levy, A.; Lebanony, D.; Goren, Y.; Silberschein, E.; Targan, N.; Ben-Ari, A.; Gilad, S.; Sion-Vardy, N.; Tobar, A.; Feinmesser, M.; Kharenko, O.; Nativ, O.; Nass, D.; Perelman, M.; Yosepovich, A.; Shalmon, B.; Polak-Charcon, S.; Fridman, E.; Avniel, A.; Bentwich, I.; Bentwich, Z.; Cohen, D.; Chajut, A.; Barshack, I.. MicroRNAs accurately identify cancer tissue origin. *Nat. Biotechnol.*, **2008**, *26*(4), 462-469.

[21] Barshack, I.; Meiri, E.; Rosenwald, S.; Lebanony, D.; Bronfeld, M.; Aviel-Ronen, S.; Rosenblatt, K.; Polak-Charcon, S.; Leizerman, I.; Ezagouri, M.; Zepeniuk, M.; Shabes, N.; Cohen, L.; Tabak, S.; Cohen, D.; Bentwich, Z.; Rosenfeld, N. Differential diagnosis of hepatocellular carcinoma from metastatic tumors in the liver using microRNA expression. *Int. J. Biochem. Cell Biol.*, **2010** *42*(8), 1355-1362.

[22] Felix, R.S.; Colleoni, G.W.; Caballero, O.L.; Yamamoto, M.; Almeida, M.S.; Andrade, V.C.; Chauffaille Mde, L.; Silva, W.A.; Jr., Begnami, M.D.; Soares, F.A.; Simpson, A.J.; Zago, M.A.; Vettore, A.L. SAGE analysis highlights the importance of p53csv, ddx5, mapkapk2 and ranbp2 to multiple myeloma tumorigenesis. *Cancer Lett.*, **2009**, *278*(1), 41-48.

[23] Nishino, R.; Honda, M.; Yamashita, T.; Takatori, H.; Minato, H.; Zen, Y.; Sasaki, M.; Takamura, H.; Horimoto, K.; Ohta, T.; Nakanuma, Y.; Kaneko, S. Identification of novel candidate tumour marker genes for intrahepatic cholangiocarcinoma. *J. Hepatol.*, **2008**, *49*(2), 207-216.

[24] Kimura, K.; Wakamatsu, A.; Suzuki, Y.; Ota, T.; Nishikawa, T.; Yamashita, R.; Yamamoto, J.; Sekine, M.; Tsuritani, K.; Wakaguri, H.; Ishii, S.; Sugiyama, T.; Saito, K.; Isono, Y.; Irie, R.; Kushida, N.;Yoneyama, T.; Otsuka, R.; Kanda, K.; Yokoi, T.; Kondo, H.; Wagatsuma, M.; Murakawa, K.; Ishida, S.; Ishibashi, T.; Takahashi-Fujii, A.; Tanase, T.; Nagai, K.; Kikuchi, H.; Nakai, K.; Isogai, T.; Sugano, S. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome. Res.*, **2006**, *16*(1), 55-65.

[25] Kasai, Y.; Hashimoto, S.; Yamada, T.; Sese, J.; Sugano, S.; Matsushima, K.; Morishita, S. 5'SAGE: 5'-end Serial Analysis of Gene Expression database. *Nucleic Acids Res,*, **2005**, *33*(Database issue), D550-552.

[26] Hodo, Y.; Hashimoto, S.; Honda, M.; Yamashita, T.; Suzuki, Y.; Sugano, S.; Kaneko, S.; Matsushima, K. Comprehensive gene expression analysis of 5'-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma. *Genomics*, **2010**, *95*(4), 217-223.

[27] Zha, S.; Ferdinandusse, S.; Hicks, J.L.; Denis, S.; Dunn, T.A.; Wanders, R.J.; Luo, J.; De Marzo, A.M.; Isaacs, W.B. Peroxisomal branched chain fatty acid beta-oxidation pathway is upregulated in prostate cancer. *Prostate*, **2005**, *63*(4), 316-323.

[28] Stier, H.; Fahimi, H.; Van Veldhoven, P.; Mannaerts, G.; Völkl, A.; Baumgart, E. Maturation of peroxisomes in differentiating human hepatoblastoma cells (HepG2): possible involvement of the peroxisome proliferator-activated receptor alpha (PPAR alpha). *Differentiation*, **1998**, *64*(1), 55-66.

[29] Tokuoka, K.; Nakajima, Y.; Hirotsu,K.; Miyahara, I.; Nishina, Y.; Shiga, K.; Tamaoki, H.; Setoyama, C.; Tojo, H.; Miura, R. Three-dimensional structure of rat-liver acyl-CoA oxidase in complex with a fatty acid: Insights into substrate-recognition and reactivity toward molecular oxygen. *J. Biochem.*, **2006**, *139*(4), 789-795.

[30] Umar, A.; Jaremko, M.; Burgers, P.; Luider, T.; Foekens, J.; Pasa-Tolic, L. High-throughput proteomics of breast carcinoma cells: a focus on FTICR-MS. *Expert Rev. Proteomic.*, **2008**, *5*(3), 445-455.

[31] Zhao, Y.C.; Lee, W.N.P.; Xiao, G.G. Quantitative proteomics and biomarker discovery in human cancer. *Expert Rev. Proteomic.*, **2009**, *6*(2), 115-118.

[32] Liu, R.; Wang, K.; Yuan, K.F.; Wei, Y.Q.; Huang, C.H. Integrative oncoproteomics strategies for anticancer drug discovery. *Expert Rev. Proteomic.*, **2010**, *7*(3), 411-429.

[33] Tong, A.; Gou, L.; Lau, Q.C.; Chen, B.; Zhao, X.; Li, J.; Tang, H.; Chen, L.; Tang, M.; Huang, C.; Wei, Y.Q. Proteomic profiling identifies aberrant epigenetic modifications induced by hepatitis B virus X protein. *J. Proteome Res.*, ***2009***, *8*(2), 1037-1046.

[34] McClain, S.L.; Clippinger, A.J.; Lizzano, R.; Bouchard, M.J. Hepatitis B virus replication is associated with an HBx-Dependent mitochondrion-regulated increase in cytosolic calcium levels. *J. Virol.*, **2007**, *81*(21), 12061-12065.

[35] Xie, H.Y.; Xia, W.L.; Zhang, C.C.; Wu, L.M.; Ji, H.F.; Cheng, Y.; Zheng, S.S. Evaluation of hepatitis B virus replication and proteomic analysis of HepG2.2.15 cell line after cyclosporine A treatment. *Acta Pharm. Sinic.*, **2007**, *28*(7), 975-984.

[36] Bouchard, M.J.; Wang, L.H.; Schneider, R.J. Calcium signaling by HBx protein in hepatitis B virus DNA replication. *Science*, **2001**, *294*(5550), 2376-2378.

[37] Choi, Y.W.; Park, S.G; Yoo, J.; Jung, G.H. Calcium ions affect the hepatitis B virus core assembly. *Virology*, **2005**, *332*(1), 454-463.

[38] Hamilton, J.P. Epigenetic mechanisms involved in the pathogenesis of hepatobiliary malignancies. *Epigenomics*, **2010**, *2*(2), 233-243.

[39] Tsai, W.; Chung, R. Viral hepatocarcinogenesis. *Oncogene*, **2010**, *29*, 2309-2324.

[40] Niu, D.D.; Feng, H.X.; Chen, W.N. Proteomic analysis of HBV-associated HCC: Insights on mechanisms of disease onset and biomarker discovery. *J. Proteomics*, **2010**, *73*(7), 1283-1290.

[41] Park, I.; Sohn, B.; Yu, E. Aberrant epigenetic modifications in hepatocarcinogenesis induced by hepatitis B virus X protein. *Gastroenterology*, **2007**, *132*(4), 1476-1494.

[42] Liu, J.; Lian, Z.; Han S. Downregulation of E-cadherin by hepatitis B virus X antigen in hepatocellullar carcinoma. *Oncogene*, **2006**, *25*(7), 1008-1017.

[43] Oh, B.; Kim, H.; Park, H. DNA methyltransferase expression and DNA methylation in human hepatocellular carcinoma and their clinicopathological correlation. *Int. J. Mol. Med.*, **2007**, *20*(1), 65-73.

[44] Sun, W.; Zhong, F.; Zhi, L.; Zhou, G.; He, F. Systematic -omics analysis of HBV-associated liver diseases. *Cancer Lett.*, **2009**, *286*(1), 89-95.

[45] Qiu, G.; Salto-Tellez, M.; Ross, J.; Yeo, W.; Cui, Y.; Wheelhouse, N.; Chen, G.; Harrison, D.; Lai, P.; Tao, Q. The tumor suppressor gene DLEC1 is frequently silenced by DNA methylation in hepatocellular carcinoma and induces G1 arrest in cell cycle. *J. Hepatol*., **2008**, *48*(3), 433-441.

[46] Zheng, D.L.; Zhang, L.; Cheng, N.; Xu, X.; Deng, Q.; Teng, X.; Wang, K.; Zhang, X.; Huang, J.; Hai, Z. Epigenetic modification induced by hepatitis B virus X protein via interaction with de novo DNA methyltransferase DNMT3A. *J. Hepatol*., **2009**, *50*(2), 377-387.

[47] Sheng KH, Yao YC, Chuang SS, Wu H, Wu TF. Search for the tumor-related proteins of transition cell carcinoma in Taiwan by proteomic analysis. *Proteomics*, **2006**, *6*(3), 1058-1065.

[48] Obach, M.; Navarro-Sabate, A.; Caro, J.; Kong, K.; Duran, J.; Marta, G.; Perales, C. 6-Phosphofructo-2-kinase (pfkfb3) gene promoter contains hypoxia-inducible factor-1 binding sites necessary for transactivation in response to hypoxia. *J. Biol. Chem.*, **2004**, *279*(51), 53562-53570.

[49] Ryan, H.E.; Lo, J.; Johnson, RS. HIF-1 alpha is required for solid tumor formation and embryonic vascularization. *EMBO J*., **1998**, *17*(11), 3005-3015.

[50] Minchenko, A.; Leshchinsky,I.; Opentanova, I.; Sang, N.; Vickram, S.; Armstead, V.; Caro, J. Hypoxia-inducible factor-1-mediated expression of the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase-3 (PFKFB3) gene. Its possible role in the Warburg effect. *J. Biol. Chem.*, **2002**, *277*(8), 6183-6187.

[51] Liu, R.; Li, Z.; Bai, S.; Zhang, H.; Tang, M.; Lei, Y.; Chen, L.; Liang, S.; Zhao, Y.; Wei, Y.; Huang, C. Mechanism of cancer cell adaptation to metabolic stress: proteomics identification of a novel thyroid hormone-mediated gastric carcinogenic signaling pathway. *Mol. Cell Proteomics*, **2009**, *8*(1), 70-85.

[52] Warburg, O. On the origin of cancer cells. *Science*, **1956**, *123*, 309-314.

[53] Langereis, J.; Prinsen, B.; de Sain-van der Velden, M.; Coppens, C.; Koenderman, L.; Ulfman, L. A 2D-DIGE approach to identify proteins involved in inside-out control of integrins. *J. Proteome Res*., **2009**, *8*(8), 3824-3833.

[54] Ibarrola, N.; Molina, H.; Iwahori, A.; Pandey, A. A novel proteomic approach for specific identification of tyrosine kinase substrates using [13C] tyrosine. *J. Biol. Chem.*, **2004**, *279*(16), 15805-15813.

[55] Ren, F.; Wu, H.; Lei, Y.; Zhang, H.; Liu, R.; Zhao, Y.; Chen, X.; Zeng, D.; Tong, A.; Chen, L.; Wei, Y.; Huang, C. Quantitative proteomics identification of phosphoglycerate mutase 1 as a novel therapeutic target in hepatocellular carcinoma. *Mol. Cancer*, **2010**, *9*, 81.

[56] Grønborg, M.; Kristiansen, T.; Iwahori, A. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol. Cell Proteomics*, **2006**, *5*(1), 157-171.

[57] Zieske, L. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J. Exp. Bot.*, **2006**, *57*(7), 1501-1508.

[58] Timothy, M.P.; David, H.H.; Yanna, L. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. *BMC Cancer*, **2006**, *6*, 68.

[59] Niu, D.; Sui, J.; Zhang, J.; Feng, H.; Chen, W. iTRAQ-coupled 2-D LC-MS/MS analysis of protein profile associated with HBV-modulated DNA methylation. *Proteomics*, **2009**, *9*(15), 3856-3868.

[60] Kristal, B.; Shurubor, Y. Metabolomics: opening another window into aging. *Sci. Aging Knowledge Environ*., **2005**, *2005*(26), pe19.

[61] Wikoff, W.; Kalisak, E.; Trauger, S.; Manchester, M.; Siuzdak, G. Response and recovery in the plasma metabolome tracks the acute LCMV-induced immune response. *J. Proteome Res.*, **2009**, *8*(7), 3578-3587.

[62] Qi, G.; Jasmin, K.S.; Timothy, M.D.; Faisal, R.; David, J.S.; Claus, S.; Stephen, R.S.; Peter K, Morgan, A.J.; Jacob, G.B. Validation of metabolomics for toxic mechanism of action screening with the earthworm *Lumbricus rubellus*. *Metabolomics*, **2009**, *5*(1), 72-83.

[63] Chen, C.; Krausz, K.; Idle, J.; Gonzalez, F. Identification of novel toxicity-associated metabolites by metabolomics and mass isotopomer analysis of acetaminophen metabolism in wild-type and Cyp2e1-null mice. *J. Biol. Chem.*, **2008**, *283*(8), 4543-4559.

[64] Gibney, M.; Walsh, M.; Brennan, L.; Roche, H.; German, B.; van Ommen, B. Metabolomics in human nutrition: opportunities and challenges. *Am. J. Clin. Nutr.*, **2005**, *82*(3), 497-503.

[65] Bain, J.; Stevens, R.; Wenner, B.; Ilkayeva, O.; Muoio, D.; Newgard, C. Metabolomics applied to diabetes research: moving from information to knowledge. *Diabetes*, **2009**, *58*(11), 2429-2443.

[66] Griffin, J.; Vidal-Puig, A. Current challenges in metabolomics for diabetes research: a vital functional genomic tool or just a ploy for gaining funding? *Physiol. Genomics*, **2008**, *34*, 1-5.

[67] Serkova, N.; Glunde, K. Metabolomics of cancer. *Methods Mol. Biol*., **2009**, *520*, 273-295.

[68] Kim, K.; Aronov, P.; Zakharkin, S.O. Urine metabolomics analysis for kidney cancer detection and biomarker discovery. *Mol. Cell Proteomics*, **2009**, 8(3), 558-570.

[69] Griffiths, J.; Stubbs, M. Opportunities for studying cancer by metabolomics: preliminary observations on tumors deficient in hypoxia-inducible factor 1. *Adv. Enzyme Regul.*, **2003**, 4(3), 67-76.

[70] Griffin, J.; Shockcor, J. Metabolic profiles of cancer cells. *Nat. Rev. Cancer*, **2004**, *4*(7), 551-561.

[71] Katajamaa, M.; Miettinen, J.; Oresic, M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics*, **2006**, *22*(5), 634-636.

[72] Weiss, R.; Lin, P. Kidney cancer: identification of novel targets for therapy. *Kidney Int.*, **2006**, *69*(2), 224-232.

[73] Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. A comprehensive urinary metabolomic approach for identifying kidney cancerr. *Anal. Biochem.*, **2007**, *363*(2), 185-195.

[74] Perroud, B.; Ishimaru, T.; Borowsky, A.; Weiss, R. Grade-dependent proteomics characterization of kidney cancer. *Mol. Cell Proteomics*, **2009**, *8*(5), 971-985.

[75] Yang, C. Dietary factors may modify cancer risk by altering xenobiotic metabolism and many other mechanisms. *J Nutr*. **2006**, *136*(10), 2685S-2686S.

[76] Lawlor, K.; Nazarian, A.; Lacomis, L.; Tempst, P.; Villanueva, J. Pathway-based biomarker search by high-throughput proteomics profiling of secretomes. *J. Proteome Res.*, **2009**, *8*(3), 1489-1503.

[77] Le Gall, C.; Bellahcène, A.; Bonnelye, E.; Gasser, J.; Castronovo, V.; Green, J.; Zimmermann, J.; Clézardin, P. A cathepsin K inhibitor reduces breast cancer induced osteolysis and skeletal tumor burden. *Cancer Res*., **2007**, *67*(20), 9894-9902.

[78] Gocheva, V.; Zeng, W.; Ke, D. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. *Genes & Dev.*, **2006**, *20*(5), 543-556.

[79] Munger, J.; Bennett, B.; Parikh, A.; Feng, X.; McArdle, J.; Rabitz, H.; Shenk, T.; Rabinowitz, J. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nat. Biotechnol.*, **2008**, *26*(10), 1179-1186.

[80] Serkova, N.; Niemann, C. Pattern recognition and biomarker validation using quantitative 1H-NMR-based metabolomics. *Expert Rev. Mol. Diagn*., **2006**, *6*(5), 717-731.

[81] Harris, T.; Eliyahu, G.; Frydman, L.; Degani, H. Kinetics of hyperpolarized 13C1-pyruvate transport and metabolism in living human breast cancer cells. *Proc. Natl. Acad. Sci. USA*, **2009**, *106*(43), 18131-18136.

[82] Denko, N. Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat. Rev. Cancer*, **2008**, *8*(9), 705-713.

[83] Chen, Y.; Adam, D.R.; Andrei, O.; Jeffrey, W.S. Profiling of central metabolism in human cancer cells by two-dimensional NMR, GC-MS analysis, and isotopomer modeling. *Metabolomics*, **2008**, *4*(1), 13-29.

[84] Gogvadze, V.; Orrenius, S.; Zhivotovsky, B. Mitochondria in cancer cells: what is so special about them? *Trends Cell Biol.*, **2008**, *18*(4), 165-173.

[85] Quash, G.; Fournet, G.; Reichert, U. Anaplerotic reactions in tumour proliferation and apoptosis. *Biochem. Pharmacol.*, **2003**, *66*(3), 365-370.

[86] Yang, C.; Sudderth, J.; Dang, T.; Bachoo, R.; McDonald, J.; DeBerardinis, R. Glioblastoma cells require glutamate dehydrogenase to survive impairments of glucose metabolism or Akt signaling. *Cancer Res.*, **2009**, *69*(20), 7986-7993.

[87] Ahn, A.; Tewari, M.; Poon, C.; Phillips, R. The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med.*, **2006**, *3*(6), e208.

[88] Cho W. Contribution of oncoproteomics to cancer biomarker discovery. *Mol. Cancer*, **2007**, 6, 25.

[89] Wiechert, W.K.N. From stationary to instationary metabolic flux analysis. *Adv. Biochem. Eng. Biotechnol.*, **2005**, *92*, 145-172.

[90] Dauner, M.; Sauer, U. GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. *Biotechnol. Prog.*, **2000**, *6*(4), 42-49.

[91] Van Winden, W.; van Dam, J.; Ras, C.; Kleijn, R.; Vinke, J.; Van Gulik, W.; Heijnen, J. Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of (13)C-labeled primary metabolites. *FEMS Yeast Res.*, **2005**, *5*, 559-568.

[92] Thiel, K.; Portland, Oregon. Systems biology, Iincorporated? *Nat. Biotechnol.*, **2006**, *24*(9), 1055-1057.

Received: February 25, 2011 **Revised: March 30, 2011** **Accepted: April 6, 2011**