

Systems biology approaches to disease marker discovery

Donald Sharon^{a,1}, Rui Chen^{b,1} and Michael Snyder^{b,*}

^a*Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, USA*

^b*Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA*

Abstract. Our understanding of human disease and potential therapeutics is improving rapidly. In order to take advantage of these developments it is important to be able to identify disease markers. Many new high-throughput genomics and proteomics technologies are being implemented to identify candidate disease markers. These technologies include protein microarrays, next-generation DNA sequencing and mass spectrometry platforms. Such methods are particularly important for elucidating the repertoire of molecular markers in the genome, transcriptome, proteome and metabolome of patients with diseases such as cancer, autoimmune diseases, and viral infections, resulting from the disruption of many biological pathways. These new technologies have identified many potential disease markers. These markers are expected to be valuable to achieve the promise of truly personalized medicine.

Keywords: Systems biology, disease markers, cancer, autoimmune disease, protein microarray, mass spectrometry, Next-generation sequencing

1. Introduction

Disease markers are of vital importance to clinicians and their patients as early detection, accurate prognosis/diagnosis and monitoring of therapy can lead to increased overall survival and cure rates. As our knowledge of diseases quickly expands, the field of disease marker discovery will play increasingly important roles in the delivery of improved diagnosis and treatment. These markers, such as protein (including autoantibodies, which are antibodies specific to self-antigens [43]), hormonal markers (such as lack of insulin in Type I diabetic patients [89]), and genetic/genomic markers (such as *BRCA1* mutation in breast cancer patients [52]), enable clinicians to diagnose the disease while it is still at early stages, to ensure appropriate surgical intervention, efficient drug treat-

ment and monitoring, and to predict an individual's risk of developing specific diseases before they experience symptoms. Traditionally, discovery and detection of these disease markers relied on low throughput technologies such as Enzyme-Linked Immunosorbent Assay (ELISA) or 2D-gel plus Edman degradation for protein markers, Reverse Transcription-Polymerase Chain Reaction (RT-PCR) for mRNA markers, and restriction enzyme digestion, cloning and Sanger sequencing for DNA markers. Before the dawn of high-throughput technologies these methods played important roles in marker identification and yielded significant discoveries in various diseases such as systemic lupus erythematosus, rheumatoid arthritis and breast cancer [32,52,107], which greatly enhanced the diagnostic efficiency in these diseases.

During the past two decades, high-throughput technologies emerged and have displayed great potential in large-scale studies for marker discovery. These technologies include protein microarrays [42,119], mass spectrometry for large-scale shotgun studies [116], and, more recently, high-throughput parallel sequencing (including RNA-Sequencing) [9,63,67]. This article will

¹These authors contributed equally to this manuscript.

*Corresponding author: Michael Snyder, 300 Pasteur Drive, Alway M-344, Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. Tel.: +1 650 736 8099; Fax: +1 650 725 1534; E-mail: mpsnyder@stanford.edu.

review disease marker discoveries using these systems biology approaches, with a focus on high-density protein microarray technologies. We will also briefly review current progress in disease marker identification using parallel sequencing and mass spectrometry technologies.

2. Marker discovery using high-density protein microarrays

Currently, high-density protein microarrays contain hundreds to thousands of proteins that are arrayed on coated glass microscope slides (e.g. nitrocellulose-coated slides) in an addressable format [23]. These arrays are usually probed with fluorescently labeled molecules and the signals are then acquired with a confocal laser scanner. A number of surface chemistries have been employed for their ability to bind proteins efficiently although there is often a trade-off between retention and decreased protein function or improper folding. There are several broad categories of protein microarrays: arrays composed of cell or tissue lysates or protein fractions isolated from crude lysates [33,74], antibody or analytical arrays that contain types of antibodies directed specific analytes [13] as well as so-called functional protein arrays [86,118]. Functional protein arrays contain full length proteins with intact catalytic function and proper epitope folding, which are often generated by arraying purified proteins produced individually prior to printing [118] or proteins produced *in situ* by *in vitro* transcription and translation of DNA that is printed directly on the surface of the array [86]. Our group has been developing the last type of protein microarrays by arraying purified proteins on nitrocellulose-coated glass slides. This type of functional protein microarray has clear advantages over the other alternatives: the ability to specifically identify individual proteins compared with cell lysate arrays and the ability to ensure the quality of each arrayed protein compared to the *in situ* transcription-translation arrays. After development of the first protein microarray that contains 5800 full-length proteins of the budding yeast *Saccharomyces Cerevisiae* [118], our group produced a number of protein arrays including the yeast N-terminal and C-terminal arrays, a 500 protein Arabidopsis array, and a coronavirus array [119]. In conjunction with Protometrix Corporation (now part of Invitrogen Corporation) we also collaborated in the development of a human protein array that currently holds more than 9,000 proteins expressed individually using

a Baculovirus/sf9 expression systems. These various arrays were used for a variety of applications including assaying for protein-protein, protein-lipid and protein-nucleic acid interactions as well as probing for substrates of protein kinases [41,45,118]. We also developed algorithms for positive signal calling and large dataset processing [50]. Recently, we have applied this technology in a novel proteomics-based approach to screen for human antibodies that react with foreign and self-antigens [64,65,79]. Particularly notable in this review is our use of these protein microarrays to analyze the immune response to coronaviruses [119] as well as our screening projects to analyze ovarian cancer [43], myeloma, multiple sclerosis and asthma. In this section we will review disease marker identification in several fields using high-density protein microarrays.

2.1. Antibodies as markers of viral infections

Currently, tests for the detection of microbial infections are the only clinical tests that rely on measuring antibody responses. ELISA-based detection methods are often used to detect a patient's antibody titer to epitopes of the microorganisms for diagnosis of the infection.

In late 2003, an outbreak of a novel coronavirus (CoV), the Severe Acute Respiratory Syndrome (SARS) virus, resulted in that killed over 900 deaths. Novel diagnostic tests were required to identify and monitor this disease, and ELISA, immunofluorescence and nucleic acids tests were employed. It was shown that protein array-based methods proved to be more accurate than any of the existing antibody based methods [119]. Our lab developed a coronavirus protein-microarray that contained 82 coronavirus proteins including all SARS-CoV proteins and proteins from five additional coronaviruses. The microarray was used to probe sera obtained from 399 Canadians and 203 Chinese during the SARS outbreak, including samples from confirmed SARS-CoV cases, other respiratory disease patients, and healthcare professionals. After detection with Cy3-labeled anti-Human IgG antibodies, the bound reactive antibodies to coronavirus-encoding proteins from sera were visualized and quantified. The reactivity results of the different proteins were analyzed using a variety of computational methods, and we developed computer algorithms based on the reactivity results to predict which patients were infected with SARS [119].

The protein microarray platform displayed a very high sensitivity, and reliably detected SARS-CoV reac-

tive antibodies even when serum was diluted at 16,000 fold. The assay showed good reproducibility with less than 10% variance in signal intensity between duplicate slides. Importantly, the method requires less than one microliter of serum for detection, which is desirable when serum samples are limited. Moreover, probing of the coronavirus protein microarrays with SARS infected serum samples also shows high specificity to SARS-CoV-specific proteins, as very little crossreactivity with proteins of other coronaviruses has been observed in SARS infected serum samples. To determine the best classifiers to distinguish SARS-positive from SARS-negative sera, we used one unsupervised clustering method and two supervised methods: k -nearest neighbor (k -NN) and logistic regression (LR). Both supervised models showed high sensitivity (90% for k -NN and 89% for LR) and specificity (93% for k -NN and 94% for LR) with a panel of 5 (k -NN) or 4 (LR) best classifiers, and these numbers were greater than 97% when the assay was performed in triplicate. The prediction methods were then tested on 56 sera from Chinese fever patients for SARS-infection prediction and were determined to have 100% sensitivity and 95% specificity, which are superior to two ELISA-based detection methods that were used during the SARS outbreak.

Our study in SARS-CoV infection diagnosis via detection of SARS-CoV-specific antibodies by protein microarrays demonstrated that this approach is sensitive (50-fold more sensitive than ELISAs), specific (little crossreactivity with other coronavirus proteins), and rapid (performed in a few hours). Nevertheless it should be noted that tests based on immune responses to foreign antigens are more likely to achieve higher accuracy than those based on autoantibody-autoantigen responses since there is no self-tolerance to the foreign antigens and the presence of pathogen associated molecular patterns (PAMPs) significantly increases the immune response. Overall, this study demonstrated for the first time that protein microarrays could be used to diagnose and monitor human antibodies as protein markers that are generated during the course of a disease. Moreover, it demonstrated the power of using a panel of multiple classifiers for diagnostics.

2.2. *Marker discovery for early detection and prognosis of cancer*

While protein microarray technology can efficiently and accurately detect antibodies generated against foreign antigens from infectious organisms, perhaps the

most intriguing application of this technology is in the discovery of novel protein markers for the early detection of various cancers. The identification of disease markers holds the promise of increasing the effectiveness of clinic therapies and marker-based routine screening programs and can potentially enable diagnosis at the earliest stages of the disease, before the development of clinically recognizable cancers that are usually at advanced stages. For instance, in heavy smokers autoantibodies recognizing mutant forms of the tumor suppressor p53 have been detected prior to the diagnosis of lung cancer [103]. Early detection and treatment would result in markedly improved survival rates, especially for patients whose cancers do not present symptoms during early stages such as pancreatic and ovarian cancer [25,29,91].

Oncoproteomics is a rapidly expanding field aimed at applying high-throughput proteomics approaches to understanding the mechanisms involved in cancer. Proteomic approaches to discover cancer markers have been an area of strong interest in recent years. In the past, these projects often involved serum screening with phage expression libraries prepared from cancer tissues, or SEREX (serological analysis of cDNA expression libraries), or by immunoblotting cancer cell lysates after two dimensional polyacrylamide gel electrophoresis (2DE-PAGE). These approaches have yielded some promising candidate markers but suffer from particular issues, such as the fact that phage expression libraries often contain out of frame and truncated protein targets and protein candidates discovered by 2DE-PAGE are difficult to identify since the proteins are unknown. Mass spectrometry is often required in order to identify the candidate autoantigens [20,46,53,90]. An additional problem with these approaches is that the samples are usually limited in amount and are difficult to reproduce. Protein microarrays overcome those difficulties as all of the spotted proteins are derived from known, well-characterized clones. Additionally, even a small amount of purified protein is sufficient to print hundreds of arrays for patient screening [4,7,16,47,81].

As most traditional disease markers are proteins that have become over- or under-expressed during the course of disease, there is much interest in the potential use of autoantibodies as a novel class of disease markers. Recently, detection in serum of circulating autoantibodies targeting Tumor-Associated Antigens (TAAs) has emerged as an effective approach for identifying cancer early detection markers (e.g. breast, lung and ductal pancreatic cancer [5,80,102]). This approach is based on the fact that the immune system produces an-

antibodies against abnormal/mutated proteins generated from apoptotic/necrotic cancer cells. These autoantibodies can then be detected with immunosorbent assays like ELISA. Because the levels/stability of autoantibodies are potentially much greater than those of the original autoantigens, they would be more easily detected. By comparing autoantibody profiles between different groups (cancer patients versus controls), it is possible to identify markers that are significantly differentially expressed. This method is expected to be superior to DNA array-based methods since changes in RNA expression levels do not necessarily correlate with protein expression.

The area of research in autoantibody marker discovery using protein microarrays has rapidly expanded over the last several years as the protein array platform continues to mature. The recent availability of high content protein microarrays allows for global profiling of autoantibodies to cancer antigens in both high-throughput (thousands of protein candidates) and high sensitivity (≥ 10 fg of protein) [43,118]. Improvements in printing techniques and increases in protein spot quantity have made these arrays promising vehicles for exploring the repertoire of autoantibodies in human disease. This approach has been applied, by various groups, for the discovery of autoantibody markers in breast cancer [5], lung cancer [80] and ovarian cancer [43], as well as a smaller study in pancreatic cancer [74]. Here we will review past and ongoing research in immune response profiling using protein microarrays relating to a number of disease states.

While self-tolerance usually abrogates the antibody response to self-proteins it is possible to elicit an autoimmune response under certain conditions present in cases of disease. The antigenicity of self proteins may result from overexpression of normal proteins such as in the case of Her-2 in breast cancer subtypes and prostate specific antigen (PSA) in prostate cancer, from aberrant post-translational modification such as different Mucin-1 glycoforms in breast cancer, or from mutations in the proteins as has been found to be the case with the tumor suppressor p53 in multiple cancer types [2,14,17,55]. Additionally, proteins that are usually restricted to expression in germ line cells or are expressed only in the early stages of development may be aberrantly expressed in cancer. This is the case with the testis antigen NY-ESO-1 and carcinoembryonic antigen (CEA) respectively. Because many of the proteins mentioned above are detected only at very low levels in serum, even in late stages of disease, they would be of little utility for screening purposes. However, even

slight increases in the expression of those antigens can lead to detectable increases in the corresponding autoantibody. Generally, we find the existence of a basal autoantibody level to many self-antigens, however, this response has been shown to be markedly increased in cases of diseases such as those mentioned above [2,26,87].

CA-125 is currently the only clinically approved marker for ovarian cancer screening. Unfortunately, although CA-125 serum levels are significantly elevated in advanced stages of the disease, its positive predictive value for the detection of early stage ovarian cancer is less than 10% [66]. For this reason the identification of new markers for this disease is of critical importance. Scientists, such as the group led by Gil Mor at Yale University, recruited proteomics-based approaches using antibody-based protein microarrays to identify new serum biomarkers, which, in combination with CA-125, may enhance the early detection of ovarian cancer [48,66,110]. Our group also launched a pilot study to profile ovarian cancer-associated autoantibodies with protein microarrays containing 5,005 full-length human proteins [43]. We compared the autoantibody profiles in 30 cases of epithelial ovarian cancer patients and 30 healthy controls, and after statistical analysis, identified 90 proteins to have significantly different immune reactivity in the patient group versus the control group. The results were validated by immunohistochemistry (IHC) and demonstrated high sensitivity (95%) and specificity (97.5%) when the top two markers (Lamin A/C and SSRP1) were combined. However, further validation is required before the candidate markers can be adopted in a clinical setting. Therefore we carried the top ranking candidates through to the validation phase in which a much larger set of samples will be tested to evaluate the performance of the potential markers. We have generated focused protein microarrays containing these candidates as well as control proteins such as CA-125 (Fig. 1). These arrays are printed in twelve blocks per slide allowing as many as twelve samples to be screened per array. This approach will allow hundreds to thousands of samples to be screened in order to determine which markers or combination of markers demonstrate the best receiver operator characteristic (ROC) performance [34,43].

Autoantigens in breast cancer subtypes such as Her-2/neu positive tumors have been shown to correlate with increased autoantibody responses in patients. Her-2/neu autoantibodies in those patients demonstrate approximately 18% sensitivity and 94% specificity. Other Groups have adapted similar approaches to multiplex-

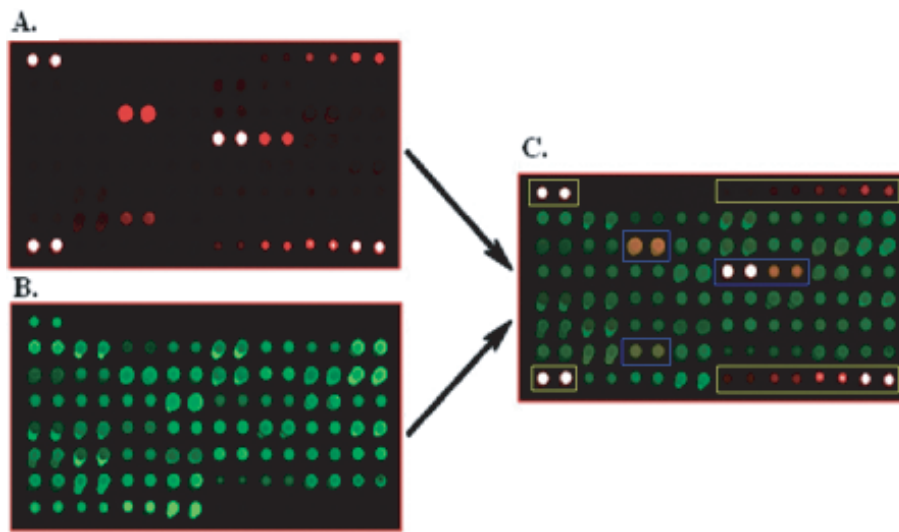


Fig. 1. A “focused” protein microarray composed of high ranking ovarian cancer autoantigen marker candidates discovered by Hudson, et al., and probed with serum diluted 1:50. Panel A shows the anti-human IgG channel. B shows anti-GST reactivity to GST-tagged autoantigens and controls. C is the merged image which demonstrates the utility of using the ratio data to analyze autoantibody reactivity relative to protein content. Blue boxes indicate autoantigens that were serum reactive while yellow boxes show control spots.

ing immune responses using high-density protein microarrays in breast cancer [15,117]. Among these studies, Anderson et al. used nucleic acid programmable protein arrays (NAPPA) for sera screening in breast cancer [5]. These arrays were generated by printing individual genes as plasmid DNA along with capture antibodies to GST tags on the fusion proteins. The arrays were then incubated with *in vitro* transcription and translation coupled cell-free lysates to produce the proteins and anchor them to the array surface. Each array consisted of 1700 cancer associated candidate proteins including p53 as well as the Epstein bar nuclear antigen (EBNA) as a control. Sera from four breast cancer patients and four healthy controls were used to probe the arrays and they found anti-p53 autoantibodies in cancer patients. Because the proteins are not produced until the arrays are ready to be probed they do not suffer from degradation during periods of storage. However, the protein quantity is more variable between spots compared to directly printed arrays. We have found that spot to spot variation in protein amount may be overcome by evaluating the autoantibody response relative to the protein amount (i.e. the ratio of autoantibody signal to the signal from an epitope tag on the autoantigens). In our protein array immune response profiling studies we have found that this approach results in decreased signal variance between replicate spots (unpublished data). To date, no studies that attempt to identify novel breast cancer markers have been performed using high-density protein microarrays. Pre-

vious high-throughput serum screens in breast cancer have relied on SEREX and 2DE-PAGE and involved relatively small sample sets [53,90].

A more interesting immune response profiling study may be the autoantibody responses to self-antigens in cancers of the blood and lymph, such as multiple myeloma. Our lab has been involved in a protein microarray based screening project aimed at elucidating the autoantigen repertoire in multiple myeloma. Multiple myeloma is a cancer of the bone marrow system resulting from the uncontrolled proliferation of monoclonal plasma cells (precursors of the B-cells responsible for the production of antibodies) [60, 83]. Monoclonal gammopathies of unspecified significance (MGUS) is a precursor disease to multiple myeloma characterized by bone marrow plasmacytosis and increased M-protein levels in the blood [84]. We have probed high-density protein arrays using plasma from dozens of cases of MGUS, multiple myeloma, and healthy controls. By comparing IgG responses to individual antigens on the arrays between the healthy and diseased groups we have identified multiple autoantigens that are significantly differentially targeted by IgG autoantibodies in early stage disease. This study was unique as it employed the highest density protein arrays for multiple myeloma immune response profiling to date and the patient samples were from a prospective collection. Because samples were drawn prior to disease onset there are no artifacts resulting from medical treatment of the patients. By querying such early stage

samples we have a better chance of identifying markers that will be effective for diagnosing patients at early stages when they are more treatable and will experience better outcomes. The markers identified in this study may yield insight into the biological processes and mutational events that contribute to the development of aggressive forms of multiple myeloma.

In addition to early detection, cancer markers may also enable clinicians to offer personalized treatment. Recent advances using the anti-cancer drugs Herceptin and Iressa illustrate this point. These drugs target specific patient populations: Herceptin is effective against those tumors expressing the Her2 receptor and Iressa is effective for patients with specific mutations in the epidermal growth factor receptor [58,109]. These drugs offer limited benefits to patients with the same cancer types when these markers are not present. Thus, determining the marker profile of an individual's disease can enable the identification of distinct patient populations, allowing tailored and more effective treatment.

One issue that we would like to note is that no single autoantibody response to an autoantigen has been confirmed to have sufficient sensitivity and specificity for screening purposes in early stage disease. However, by evaluating the antibody responses involving a panel of autoantigens, accuracy has been markedly improved [15,43,66]. Thus, future protein microarray immune response screening tests will likely combine multiple autoantigens. Numerous approaches have been applied to combining disease markers. Some of the common methods for combining multiple autoantibody responses are linear regression, split-point analysis, and k-nearest neighbor (k-NN) [24,119]. Still there are currently no clinical screening or diagnostic tests that rely on a panel of protein markers, although multi-parameter DNA microarray tests are becoming commonplace in diseases such as breast cancer [106]. Clearly, more work is needed before autoantigen based microarray tests can be implemented in a clinical setting.

2.3. Marker discovery in autoimmune diseases

The use of protein microarray technology for biomarker discovery in autoimmune diseases seems a natural extension of the technique as autoantibody responses have already been shown to contribute to disease progression in diseases such as lupus [59]. While antinuclear antibody tests are sometimes used to confirm diagnosis of certain autoimmune diseases, they are not disease specific [59,78].

Multiple sclerosis (MS) is a debilitating disease of the central nervous system characterized by rounds of axonal demyelination and repair. It affects mostly younger people and is more common in women than men. The underlying cause remains unknown, though there is mounting evidence that antibody responses to self proteins play a role in both demyelination and repair [28]. Previous efforts have identified a number of myelin specific proteins that demonstrate increased autoantibody responses in MS. These autoantibodies have been detected in both serum and cerebrospinal fluid (CSF). Recently, protein microarray screening has been applied to evaluate autoantibody responses to subsets of myelin proteins that are known to be associated with MS and other neurodegenerative diseases [78,82,97]. The antigens that were shown to elicit autoantibody responses include classical MS antigens such as myelin-basic protein (MBP), myelin associated glycoprotein (MAG) and myelin oligodendrocyte glycoprotein (MOG) as well as proteins that have not been demonstrated to play a significant role in MS previously. These studies have suffered from the fact that the arrays were focused on a relatively small number of previously known candidate autoantigens. Our group is currently conducting larger-scale screenings based on high-density protein microarrays. With this platform, we have tentatively identified a number of novel candidate markers in multiple sclerosis. We are currently working to validate these markers in a larger sample set. This less biased approach may result in the identification of novel autoantigens leading to a better understanding of the underlying mechanisms in the etiology of MS and result in improved screening and diagnostic tests.

Asthma, a common disease with a prevalence of 11% for all ages [75], and 13% in children under 18 years of age in the United States [11], is another disease involving an autoimmune mechanism [88,121]. This heterogeneous inflammatory disease of the airways is marked by recurrent episodes of airway obstruction and wheezing [95,114], and is anatomically characterized by bronchoconstriction, inflammation and thickening of the airway walls [37]. Considering asthma as an aberrant chronic wound healing process [36], it would not be surprising that some of the released/leaked cellular contents from the airway epithelium due to damage and remodeling, similar to necrotic cancer cells, may elicit autoimmunity. In fact, aberrant autoantibodies have been detected in asthmatic sera by autologous serum skin tests as compared to normal controls [44]. Autoreactive antibodies have also been de-

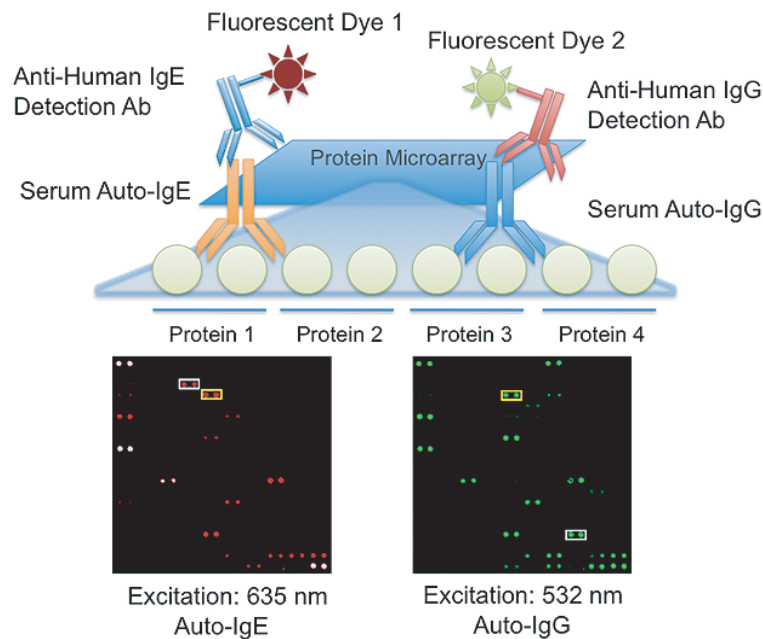


Fig. 2. Simultaneous detection of autoreactive IgE (auto-IgE) and autoreactive IgG (auto-IgG) in asthmatic serum samples. After probing asthmatic serum samples (1:500 dilution) on protein microarray slides, human auto-IgE and auto-IgG were detected with goat anti-human IgE and goat anti-human IgG secondary antibodies labeled with different fluorescent dyes (in this case Fluorescent Dye 1 = Alexa Fluor 647 and Fluorescent Dye 2 = Alex Fluor 555). After washing and drying, the slides were then scanned with the GenePix 4200A Scanner. Yellow boxes, proteins bound to both auto-IgE and auto-IgG; white boxes, proteins bound to only one subtype of autoantibodies.

ected in the asthmatic sera against the high-affinity IgE receptor $Fc\epsilon RI$ [100,101]. A few specific autoantigens have been identified in the serum of asthmatic patients, including the autoIgG-reactive β -adrenergic receptor [35,108], cytokeratin 18 [68], DFS70 [105], and α -enolase [54]. Moreover, studies on atopic dermatitis, which often occurs with asthma, revealed autoreactive IgE antibodies against Hom s 1–5 (Hom s 1 = SART1; Hom s 2 = α -NAC; Hom s 3 = BCL7B; Hom s 4 = a protein with calcium-binding motif; Hom s 5 = a Type II Cytokeratin) [70,104,105] and DFS70 [105]. However, these studies focused on small patient groups with no cross validation, as well as limited number of potential targets investigated. Since both autoreactive IgG and IgE may be involved in the pathogenesis of asthma, we conducted a large-scale screening for asthma-associated auto-IgG and IgE reactive autoantigens using protein microarrays with more than 8,000 protein candidates (unpublished data). This is the first large-scale study to profile asthma-associated autoantigens, and the results will greatly improve our understanding of the role of autoimmunity in the etiology of asthma. One unique feature of this study is that, in order to maintain uniform probing conditions, we multiplexed the detection for both autoreactive IgG and

autoreactive IgE in the serum samples simultaneously on the same array, with a mixture of anti-human IgG and IgE secondary detection antibodies labeled with distinct fluorescent dyes. Our result suggested that the protein array is capable of detecting both IgG and IgE reactive signals in distinct emission channels with high specificity and no/detectable signal bleeding across the channels (Fig. 2).

Similar applications of protein arrays have been performed in studies of other autoimmune diseases. Song et al. discovered 3 novel autoantigens, namely RPS20, Alba-like and dUTPase for autoimmune hepatitis (AIH) using a protein microarray containing 5011 nonredundant proteins [98]. Horn et al. profiled the repertoire of IgG autoantibodies in plasma samples from Dilated Cardiomyopathy (DCM) patients with a redundant protein microarray containing 37,200 total proteins and identified 26 autoreactive proteins to IgG (with 6 of them reactive specifically to the IgG3 subclass) [39]. Autoantigens were also identified for the chronic disease alopecia areata by protein microarray technology [61]. These examples demonstrate the great potential of protein microarray technology in the application of autoantigen marker identification in autoimmune diseases.

2.4. Limitations of protein microarray technology for disease marker screening

Although protein microarray technology provides a high-throughput method with high sensitivity and specificity for protein marker discovery, there are particular limitations that investigators need to be aware of before applying the technology to their research.

First of all, as probing protein microarrays for autoantibodies are *in vitro* studies with all the protein targets arrayed in a 2-D platform, one has to take into consideration off-target binding. Therefore findings from protein microarray screenings should be validated in larger sample sets, and the autoantigens have to be confirmed by direct detection methods such as Western Blotting, ELISA, or IHC in patient samples before an autoantigen can be confidently associated with the disease. Secondly, microarrays that contain full-length, folded proteins may not be recognized by autoantibodies that are directed against misfolded or degraded proteins expressed in disease cases, contributing to false negative detections. Patwa et al. developed a method to chemically digest the proteins with CNBr before printing them on the arrays, which may help to overcome this problem [74], however, as the digestion rate is hard to control, the final complex mixture of digested proteins at different levels may complicate normalization efforts and experimental control. Normalization of the array data is another important consideration. As we have already discussed, normalization of protein spot morphology and quantity can be achieved by probing with a labeled antibody directed against an epitope-tag appearing on all of the arrayed proteins (e.g. GST). Various software based methods have been adopted to adjust for regional defects and background that has traditionally been an issue for protein and DNA based microarrays [120]. Nevertheless, analysis and interpretation of the acquired large-scale data is still a challenge to both biologists and statisticians, therefore the development of improved algorithms is an ongoing effort. Under real probing conditions, uncontrollable events such as scratches on the slides, deposition of salt and non-homogeneous local concentrations can further complicate the analysis of the array data, although internal controls are often included to help overcome these defects and improved array surface chemistries have significantly decreased local and regional background defects.

3. Marker discovery using parallel sequencing technologies

In recent years it has become feasible to sequence entire genomes and transcriptomes using massively parallel sequencing platforms such as the 454 and Solexa Genome Analyzer. These platforms use a highly sensitive light sensor (such as CMOS sensors or CCD cameras) to capture fluorescent signals emitted from each deoxynucleotide as they are added to the DNA chain simultaneously in up to millions of parallel reactions in a flowcell [38], thus performing sequencing in a high-throughput manner to obtain short sequences (vary from 30 to 450 bp depending on the platform) from one or both ends. Currently, related platforms and products are available through Illumina IG, Applied Biosystems SOLiD, Roche 454 Life Science, and the Helicos Biosciences tSMS [112]. Another company, Pacific Biosciences, will manufacture a new sequencer that will perform single molecule sequencing by the end of 2010 [22].

A typical parallel sequencing procedure consists of the following steps: DNA/RNA isolation, fragmentation and DNA/cDNA library construction, high-throughput sequencing and read assembly and mapping. This method has many advantages compared to the traditional tiling microarray hybridization-based methods, or the more traditional RT-PCR and Sanger sequencing method. These new platforms achieve single-base resolution in a high-throughput manner, have low background, no cross-hybridization noise, low dependence on the availability of existing genomic sequence, high reproducibility and low cost per base, and there is no upper limit for quantification [112]. In this section we will briefly review recent efforts in genetic marker discovery with next-generation parallel sequencing.

3.1. Whole genome sequencing

This new generation of sequencing technology is shaping a new paradigm in disease marker research, in which massive amounts of sequence information from genomic DNA and expression libraries are screened for linkages and associations of genetic and genomic markers to specific diseases by comparing disease patients and healthy individuals [1,21,31,62]. Genomic DNA sequencing provides rich information on genetic variations (such as Single Nucleotide Polymorphisms, insertions and deletions) and structural variations (such as copy number variations, transposition and transloca-

tion) of the investigated genomes and is a powerful tool to reveal novel disease-associated markers. Genome sequencing can also detect integrated viral sequences which may help address studies of virus-associated diseases. Whole genome sequencing has already been applied in organisms with small genomes, such as *Acinetobacter baumannii* [96], *Toxoplasma gondii* [12], and *Drosophila melanogaster* [77], however, due to the large size of human genome and the high cost of parallel sequencing, human whole genome sequencing is still in its infancy. Ley et al. were the first to sequence the entire genome of one type of cancerous tissue, the acute myeloid leukemia (AML) cells (32.7X haploid coverage), as well as corresponding normal tissue, the patient's skin tissue (13.9X haploid coverage) [57]. Due to the unbiased nature of the sequencing methods, they were able to use read frequency to establish how rates of mutations vary within the cancer tissue. This concept is important for future works as we seek to understand the progression of mutational events that lead to the development of diseases like cancer. The researchers found that 59,209 single nucleotide variations were unique in the cancer tissue sample. These mutations resulted in changes to the coding regions in ten genes, two of which were previously implicated in cancer. Nonetheless, as sequencing costs continue to decrease with the maturation of the platforms, whole genome sequencing of larger sample sets is shedding light on new venues of genetic and genomic marker identification in various diseases. Both biologists and clinicians are preparing for this coming revolution, and projects have already been conceived such as ClinSeq, a pilot project led by Green et al. which currently enrolls about 1000 participants for whole genome sequencing [10].

3.2. Transcriptome sequencing

Whole transcriptome sequencing by RNA Sequencing (RNA-Seq) is another promising application of parallel sequencing technology and has already been applied to marker discovery in various cancers. Gene expression profiling has been shown to predict the outcome of breast and other types of cancer [40,76]. Shah et al. used paired end RNA-Seq to canvas the transcriptomes of four granulose-cell tumors (GCT) from ovarian cancer patients [93]. They found a common missense mutation in the FOXL2 gene (C402G) in those tumors that was not present in 11 other ovarian cancer transcriptomes that they also sequenced. Additionally, this mutation was confirmed to be present in 97% of other GCT cancers that they tested. Leven et al. performed

Illumina sequencing on tiling-array-hybridization enriched transcripts representing 467 cancer associated genes from the K-562 chronic myeloid leukemia cell line and detected a wide range of DNA and RNA sequence alterations in the targeted transcripts [56]. These alterations included fusion transcripts such as BCR-ABL1 and NUP214-XKR3, as well as SNPs within and splice isoforms of these transcripts. While whole genome sequencing is still a relatively expensive proposition, transcriptome sequencing can be performed at much lower cost, facilitating the discovery of any mutations in the transcriptome that may contribute to the development of disease. In addition, RNA sequencing also provides information that would not be obtained through whole genome sequencing, such as information on the expression level of each transcript, alternative splicing and RNA editing, as well as trans-splicing events. We should expect to see genetic and genomic biomarkers identified as causative or contributing factors in human disease in the near future thanks to the utility of RNA sequencing.

One other promising application for transcriptome sequencing is to identify viruses in the host sample. In a study carried out by Sorber et al., the authors successfully detected Hepatitis B Virus (HBV) sequences in the serum sample from a patient with HBV infection [99]. Similarly, Palacios et al. used RNA-Seq and identified a novel Old World arenavirus in RNA samples extracted from the liver and kidney of three deceased patients who had transplantation-related infection [72]. Nakamura et al. obtained 20–460 reads of influenza virus sequence in nasopharyngeal aspirates and 484–15,260 reads of norovirus sequence in fecal specimens from patients suffering from influenza or norovirus infections [69]. Meanwhile, Rwahni et al. sequenced the RNA from a grapevine with the Roche 454 system and revealed infection of 32 plant viruses as well as one novel virus in the diseased grapevine [3]. Our group has also carried out an RNA-Seq experiment to detect West Nile virus sequences in infected macrophages (unpublished data). We obtained 4700 reads (0.06% of the total mapped reads) mapped to the West Nile virus genome from the infected cells and very few reads (30, ~0.0003% of the total mapped reads) mapped to viral sequences in RNA isolated from mock control cells. After analysis recheck of the few reads that did map to the virus genome in the control cells, we found that they were redundant with sequences in the human genome. These studies proved that RNA-Seq can achieve high sensitivity and specificity to identify and profile infecting viruses, or the “virome”, us-

ing RNA samples isolated from the host. Moreover, RNA-Seq will also provide information on virus-host interactions by monitoring expression changes of the host's genes.

3.3. *ChIP-Seq and Sono-Seq*

While mutations in protein-coding sequence are well known to contribute to multiple diseases, RNA-Seq is limited to only those actively transcribed sequences of the genome. This bias results in overlooking variations in non-transcribed regions of the genome that can be important contributors to disease as mutations in these regions can result in aberrant gene regulation. Besides whole genome sequencing, ChIP-Seq, or Chromatin Immunoprecipitation-Sequencing, is another approach to address mutations in functional non-transcribed regions of the genome. This technology sequences the genomic regions bound by transcription factors or other DNA-binding proteins (such as histones), and provides information on the position of these binding sites as well as possible mutations in these sites. The binding site profiles of multiple transcription factors, as well as the identified sequence variations within, may act as new markers for diseases such as leukemia [8, 73]. Sono-Seq is a related technology developed in our lab, which parallel sequences sonicated formaldehyde cross-linked chromatin DNA via Illumina sequencing, and identifies the chromatin regions that are open and accessible (nucleosome-free therefore susceptible to sonication) [6]. With this technology we identified multiple highly accessible chromatin regions including actively transcribed promoter regions as well as the CTCF insulator protein binding sites. This technology is similar to another open chromatin finding technology, termed FAIRE (formaldehyde-assisted isolation of regulatory elements), which selects open chromatin regions for DNA microarray hybridization by phenol-chloroform extraction of sonicated cross-linked samples [27]. When interrogated in the background of a disease compared with healthy controls, the identified profiles of these nucleosome-free regions may provide a new type of disease marker for future studies.

3.4. *Limitations*

While next-generation sequencing holds great promise for the discovery of novel disease markers, there are issues with the current technology, such as artifacts due to sample preparation (both reverse transcription and polymerase chain reaction can generate biases) and data

processing (assembly of short reads can result in errors, especially in regions of repetitive sequence). Newer technologies from companies like Helicos Biosciences and Pacific Biosciences are on the horizon that could overcome these issues by direct RNA sequencing and long single molecule sequencing, fulfilling the promise of personalized medicine in the post genome era [22, 71].

4. **Marker discovery using mass spectrometry**

Mass spectrometry technology (MS) has been growing rapidly in the past several decades. Since John Bennett Fenn and Koichi Tanaka developed new soft desorption methods that made mass spectrometric analyses of biological macromolecules possible, this technology has been widely used in proteomic studies [19,49]. The ability to identify and quantify target molecules (e.g. peptides) makes mass spectrometry methods a popular tool for disease marker discovery.

Disease marker discovery with mass spectrometry is usually combined with varied sample separation methods such as 2DE (2-Dimensional Electrophoresis) and 2D-DIGE (2-Dimensional Differential In-Gel Electrophoresis) [18]. In a typical procedure, mixed proteins from pooled disease samples and pooled controls are separated with 1D or 2D electrophoresis, and individual protein bands or spots are visualized and differential bands or spots are then excised followed by enzyme digestion (e.g. trypsin). The digested peptides are then subjected to mass spectrometry analysis for protein identification. With this method, Shen et al. identified 40 potential markers for pancreatic adenocarcinoma, and the spectrum of these markers covered antioxidant proteins, chaperones, calcium-binding proteins, catalytic enzymes, signal transduction proteins and extracellular matrix proteins [94]. Similarly, Wang et al. identified 52 differentially expressed proteins (including 8 novel markers) associated with oral squamous cell carcinoma, and validated one of the eight markers named RACK1 using immunostaining and gene silencing studies [113]. Even though 2D electrophoresis can improve protein separation and assist in further identification by mass spectrometry, the discovery of low-abundance disease markers has been greatly limited by the poor resolution and sensitivity of 2DE or 2D-DIGE methods. Furthermore, the pooled samples will not only lose important information such as personal variation of the disease markers among different

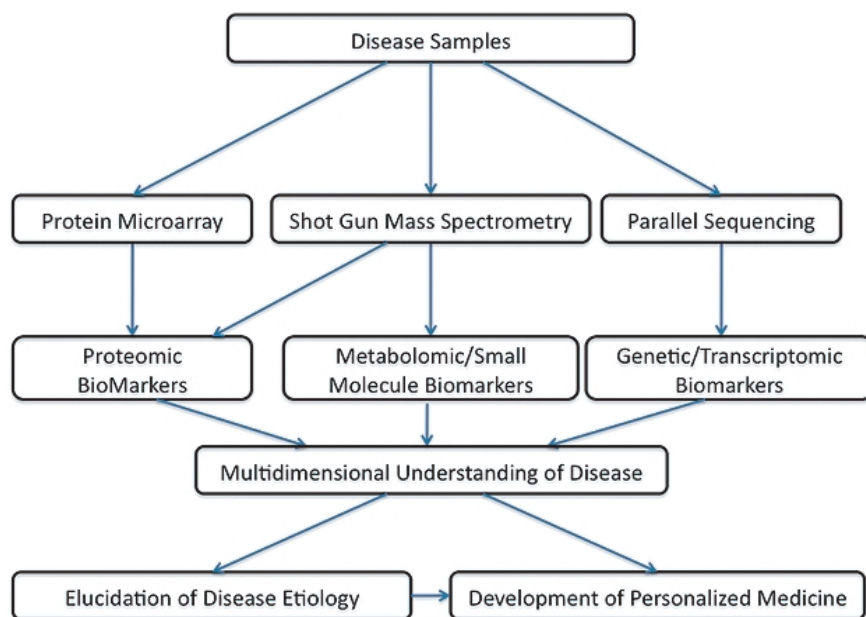


Fig. 3. Systems biology approaches in biomarker discovery contribute to comprehensive understanding of diseases.

individuals, but also miss protein that are only present in a subset of the sample population.

Recent improvement of mass spectrometry techniques as well as data analysis algorithms has enabled analysis of complex protein samples [116]. Currently, liquid chromatography (LC)-MS/MS using electrospray ionization (ESI) is one of the commonly used methods for large-scale shotgun proteomic studies. Gel-free methods, such as MudPIT (multidimensional protein identification technology) has been more and more popular and greatly enhanced detection limits [115], and the improved resolution and accuracy of mass spectrometers makes this technology more useful in disease marker discovery [111]. With 2D LC-MS/MS, Ralhan et al. identified 811 nonredundant proteins in head-and-neck squamous cell carcinoma from 15 individual cancer samples compared with one pooled normal control, and the panel of the three best performing markers achieved a sensitivity of 92% and specificity of 91% in cancer classification [85]. A most recent study to identify ovarian cancer biomarkers from patient ascites samples with 2D LC-MS/MS also yielded a panel of 25 known and 52 novel protein markers [51]. These studies demonstrated that the improved MS techniques not only enabled researchers to search for biomarkers in unpooled samples, but also could lead to the identification of a larger number of potential markers due to improved sensitivity.

In addition to protein marker identification, mass spectrometry is also capable of identifying metabolom-

ic markers [92]. Metabolomics is a novel field that studies the global profiles of all metabolites in a given sample. Since diseases such as cancer usually have unique metabolomes [30], the over- or under-presented metabolites could serve as potential markers of the disease. Moreover, certain metabolites in the cells will also influence the activity of larger biomolecules such as kinases (unpublished data), therefore identification of the cancer-specific metabolomes would be of great value. Currently, multiple metabolites have been associated with various tumors, such as Alanine, saturated lipids, CCMs, Glycine, lactate, myo-inositol, nucleotides, PUFAs and Taurine [30], and it will not be surprising if this list will grow dramatically in the coming years.

Although mass spectrometry is a powerful tool in molecular marker identification, the clinical application of this technology for diagnostic purposes is still limited. Mass spectrometry may one day become the platform of choice for detection of disease-associated markers, however, the high cost of mass spectrometers as well as lack of standardized methods are preventing it from being adopted by clinicians as a diagnostic tool. Moreover, the high level of molecular complexity of biological samples is still a large obstacle in both marker identification and application. There remains great space for improvement before mass spectrometry realizes its ultimate potential in the clinic.

5. Closing remarks

Disease markers are important for the efficient diagnosis, prognosis and treatment of a disease, therefore identifying these markers is crucial especially for diseases with high mortality rates such as cancer. Each technology reviewed in this article has a unique niche in disease marker discovery. Protein microarray technology excels in finding protein markers, especially antibody markers; next-generation parallel sequencing is designed for RNA and genetic/genomic marker discovery; and mass spectrometry specializes in the identification of protein markers as well as metabolomic markers. Each technology has its unique advantages and limitations, and the “-omics” information obtained with the help of these technologies may complement each other and leading to a comprehensive view of disease (Fig. 3). This information will greatly expand our understanding of the etiology and course of human diseases, resulting in more efficient diagnosis and treatment of disease.

Moreover, by adopting a comprehensive “-omics” view of human diseases, it will be interesting to discover the extent to which these systems interact with each other. Will we find that a disease associated with certain genetic markers also develops specific autoantibodies? Or a certain autoantigen response is actually due to the dysregulation of a specific metabolite or trans-spliced mRNA? Is it possible that diseases such as multiple sclerosis and asthma actually are caused by the coordinating effect of genetic susceptibility and, say, viral infection? While systems biology approaches to disease marker discovery are still in their infancy they have already led to the discovery of many promising genetic and protein markers in various diseases. Additionally, new dimensions in the development and application of these approaches may further revolutionize this field by interrogating additional “-omes”, such as the methyl-genome, the “kinome” and the “virome”. These systems may be as important as the systems reviewed above to establish a complete understanding of, and efficient treatments for, many diseases.

Acknowledgements

We cordially thank our collaborators, Drs. Gil Mor (ovarian cancer studies), Geoffrey Chupp (asthma studies) and Ruth Montgomery (West Nile Virus studies) at Yale University, and Dr. Jonas Bergqvist at Uppsala University, Sweden (multiple sclerosis study) for their contributions to the projects outlined in this review.

References

- [1] Comprehensive genomic characterization defines human glioblastoma genes and core pathways, *Nature* **455** (2008), 1061–1068.
- [2] J.K. Adam, B. Odhav and K.D. Bhoola, Immune responses in cancer, *Pharmacol Ther* **99** (2003), 113–132.
- [3] M. Al Rwahnih, S. Daubert, D. Golino and A. Rowhani, Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus, *Virology* (2009).
- [4] M.S. Alhamdani, C. Schroder and J.D. Hoheisel, Oncoproteomic profiling with antibody microarrays, *Genome Med* **1** (2009), 68.
- [5] K.S. Anderson, N. Ramachandran, J. Wong, J.V. Raphael, E. Hainsworth, G. Demirkan, D. Cramer, D. Aronson, F.S. Hodi, L. Harris, T. Logvinenko and J. LaBaer, Application of protein microarrays for multiplexed detection of antibodies to tumor antigens in breast cancer, *J Proteome Res* **7** (2008), 1490–1499.
- [6] R.K. Auerbach, G. Euskirchen, J. Rozowsky, N. Lamarre-Vincent, Z. Moqtaderi, P. Lefrancois, K. Struhl, M. Gerstein and M. Snyder, Mapping accessible chromatin regions using Sono-Seq, *Proc Natl Acad Sci USA* **106** (2009), 14926–14931.
- [7] I. Babel, R. Barderas, R. Diaz-Uriarte, J.L. Martinez-Torrecuadrada, M. Sanchez-Carbayo and J.I. Casal, Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays, *Mol Cell Proteomics* **8** (2009), 2382–2395.
- [8] A. Barski and K. Zhao, Genomic location analysis by ChIP-Seq, *J Cell Biochem* **107** (2009), 11–18.
- [9] S. Bennett, Solexa Ltd, *Pharmacogenomics* **5** (2004), 433–438.
- [10] L.G. Biesecker, J.C. Mullikin, F.M. Facio, C. Turner, P.F. Cherkuri, R.W. Blakesley, G.G. Bouffard, P.S. Chines, P. Cruz, N.F. Hansen, J.K. Teer, B. Maskeri, A.C. Young, T.A. Manolio, A.F. Wilson, T. Finkel, P. Hwang, A. Arai, A.T. Remaley, V. Sachdev, R. Shamburek, R.O. Cannon and E.D. Green, The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine, *Genome Res* **19** (2009), 1665–1674.
- [11] B. Bloom, R.A. Cohen and G. Freeman, Summary health statistics for U.S. children: National Health Interview Survey, 2007, *Vital Health Stat* **10** (2009), 1–80.
- [12] I.L. Bontell, N. Hall, K.E. Ashelford, J.P. Dubey, J.P. Boyle, J. Lindh and J.E. Smith, Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants, *Genome Biol* **10** (2009), R53.
- [13] A. Carlsson, C. Wingren, J. Ingvarsson, P. Ellmark, B. Baldertorp, M. Ferno, H. Olsson and C.A. Borrebaeck, Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays, *Eur J Cancer* **44** (2008), 472–480.
- [14] W.P. Carney, R. Neumann, A. Lipton, K. Leitzel, S. Ali and C.P. Price, Potential clinical utility of serum HER-2/neu oncoprotein concentrations in patients with breast cancer, *Clin Chem* **49** (2003), 1579–1598.
- [15] C. Chapman, A. Murray, J. Chakrabarti, A. Thorpe, C. Woolston, U. Sahin, A. Barnes and J. Robertson, Autoantibodies in breast cancer: their use as an aid to early diagnosis, *Ann Oncol* **18** (2007), 868–873.

- [16] M. Chatterjee, J. Wojciechowski and M.A. Tainsky, Discovery of antibody biomarkers using protein microarrays of tumor antigens cloned in high throughput, *Methods Mol Biol* **520** (2009), 21–38.
- [17] W.C. Cho, Contribution of oncoproteomics to cancer biomarker discovery, *Mol Cancer* **6** (2007), 25.
- [18] M.L. Cowan and J. Vera, Proteomics: advances in biomarker discovery, *Expert Rev Proteomics* **5** (2008), 21–23.
- [19] B.F. Cravatt, G.M. Simon and J.R. Yates, 3rd, The biological impact of mass-spectrometry-based proteomics, *Nature* **450** (2007), 991–1000.
- [20] R. Di Niro, S. D'Angelo, P. Secco, R. Marzari, C. Santoro and D. Sblattero, Profiling the autoantibody repertoire by screening phage-displayed human cDNA libraries, *Methods Mol Biol* **570** (2009), 353–369.
- [21] L. Ding, G. Getz, D.A. Wheeler, E.R. Mardis, M.D. McLellan, K. Cibulskis, C. Sougnez, H. Greulich, D.M. Muzny, M.B. Morgan, L. Fulton, R.S. Fulton, Q. Zhang, M.C. Wendl, M.S. Lawrence, D.E. Larson, K. Chen, D.J. Dooling, A. Sabo, A.C. Hawes, H. Shen, S.N. Jhangiani, L.R. Lewis, O. Hall, Y. Zhu, T. Mathew, Y. Ren, J. Yao, S.E. Scherer, K. Clerc, G.A. Metcalf, B. Ng, A. Milosavljevic, M.L. Gonzalez-Garay, J.R. Osborne, R. Meyer, X. Shi, Y. Tang, D.C. Koboldt, L. Lin, R. Abbott, T.L. Miner, C. Pohl, G. Fewell, C. Haipek, H. Schmidt, B.H. Dunford-Shore, A. Kraja, S.D. Crosby, C.S. Sawyer, T. Vickery, S. Sander, J. Robinson, W. Winckler, J. Baldwin, L.R. Chirieac, A. Dutt, T. Fennell, M. Hanna, B.E. Johnson, R.C. Onofrio, R.K. Thomas, G. Tonon, B.A. Weir, X. Zhao, L. Ziaugra, M.C. Zody, T. Giordano, M.B. Orringer, J.A. Roth, M.R. Spitz, Wistuba, II, B. Ozenberger, P.J. Good, A.C. Chang, D.G. Beer, M.A. Watson, M. Ladanyi, S. Broderick, A. Yoshizawa, W.D. Travis, W. Pao, M.A. Province, G.M. Weinstock, H.E. Varmus, S.B. Gabriel, E.S. Lander, R.A. Gibbs, M. Meyerson and R.K. Wilson, Somatic mutations affect key pathways in lung adenocarcinoma, *Nature* **455** (2008), 1069–1075.
- [22] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach and S. Turner, Real-time DNA sequencing from single polymerase molecules, *Science* **323** (2009), 133–138.
- [23] J. Fasolo and M. Snyder, Protein microarrays, *Methods Mol Biol* **548** (2009), 209–222.
- [24] Z. Feng and Y. Yasui, Statistical considerations in combining biomarkers for disease classification, *Dis Markers* **20** (2004), 45–51.
- [25] L.C. Fry, K. Mönkemüller and P. Malfertheiner, Molecular markers of pancreatic cancer: development and clinical relevance, *Langenbeck's archives of surgery/Deutsche Gesellschaft für Chirurgie* **393** (2008), 883–890.
- [26] M. Gion, R. Mione, A.E. Leon and R. Dittadi, Comparison of the diagnostic accuracy of CA27.29 and CA15.3 in primary breast cancer, *Clin Chem* **45** (1999), 630–637.
- [27] P.G. Giresi, J. Kim, R.M. McDaniell, V.R. Iyer and J.D. Lieb, FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin, *Genome Res* **17** (2007), 877–885.
- [28] J.J. Graber and S. Dhib-Jalbut, Protective autoimmunity in the nervous system, *Pharmacol Ther* **121** (2009), 147–159.
- [29] R.T. Greenlee, M.B. Hill-Harmon, T. Murray and M. Thun, Cancer statistics, 2001, *CA Cancer J Clin* **51** (2001), 15–36.
- [30] J.L. Griffin and J.P. Shockcor, Metabolic profiles of cancer cells, *Nat Rev Cancer* **4** (2004), 551–561.
- [31] W.C. Hahn, I.F. Dunn, S.Y. Kim, A.C. Schinzel, R. Firestein, I. Guney and J.S. Boehm, Integrative genomic approaches to understanding cancer, *Biochim Biophys Acta* **1790** (2009), 478–484.
- [32] S.P. Halbert, J. Karsh and M. Anken, Studies on autoantibodies to deoxyribonucleic acid and deoxyribonucleoprotein with enzyme-immunoassay (ELISA), *J Lab Clin Med* **97** (1981), 97–111.
- [33] D.A. Hall, J. Ptacek and M. Snyder, Protein microarray technology, *Mech Ageing Dev* **128** (2007), 161–167.
- [34] J.A. Hanley and B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology* **143** (1982), 29–36.
- [35] L.C. Harrison, J. Callaghan, J.C. Venter, C.M. Fraser and M.L. Kaliner, Atopy, autonomic function and beta-adrenergic receptor autoantibodies, *Ciba Found Symp* (1982), 248–262.
- [36] S.T. Holgate, Epithelium dysfunction in asthma, *J Allergy Clin Immunol* **120** (2007), 1233–1244; quiz 1245–1236.
- [37] S.T. Holgate, Pathogenesis of asthma, *Clin Exp Allergy* **38** (2008), 872–897.
- [38] R.A. Holt and S.J. Jones, The new paradigm of flow cell sequencing, *Genome Res* **18** (2008), 839–846.
- [39] S. Horn, A. Lueking, D. Murphy, A. Staudt, C. Gutjahr, K. Schulte, A. König, M. Landsberger, H. Lehrach, S.B. Felix and D.J. Cahill, Profiling humoral autoimmune repertoire of dilated cardiomyopathy (DCM) patients and development of a disease-associated protein chip, *Proteomics* **6** (2006), 605–613.
- [40] E. Huang, S.H. Cheng, H. Dressman, J. Pittman, M.H. Tsou, C.F. Horng, A. Bild, E.S. Iversen, M. Liao, C.M. Chen, M. West, J.R. Nevins and A.T. Huang, Gene expression predictors of breast cancer outcomes, *Lancet* **361** (2003), 1590–1596.
- [41] J. Huang, H. Zhu, S.J. Haggarty, D.R. Spring, H. Hwang, F. Jin, M. Snyder and S.L. Schreiber, Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips, *Proc Natl Acad Sci USA* **101** (2004), 16594–16599.
- [42] M.E. Hudson and M. Snyder, High-throughput methods of regulatory element discovery, *Biotechniques* **41** (2006), 673, 675, 677 passim.
- [43] M.E. Hudson, I. Pozdnyakova, K. Haines, G. Mor and M. Snyder, Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays, *Proc Natl Acad Sci USA* **104** (2007), 17494–17499.
- [44] A.S. Jang, J.S. Park, J.H. Lee, S.W. Park, D.J. Kim and C.S. Park, Autologous serum skin test for autoantibodies is associated with airway hyperresponsiveness in patients with asthma, *Respiration* **74** (2007), 293–296.
- [45] G. Jona and M. Snyder, Recent developments in analytical and functional protein microarrays, *Curr Opin Mol Ther* **5** (2003), 271–277.
- [46] Z. Kalnina, K. Silina, I. Meistere, P. Zayakin, A. Rivosh, A. Abols, M. Leja, O. Minenkova, D. Schadendorf and A. Line, Evaluation of T7 and lambda phage display systems for survey of autoantibody profiles in cancer patients, *J Immunol Methods* **334** (2008), 37–50.

- [47] G. Kijanka and D. Murphy, Protein arrays as tools for serum autoantibody marker discovery in cancer, *J Proteomics* **72** (2009), 936–944.
- [48] K. Kim, I. Visintin, A.B. Alvero and G. Mor, Development and validation of a protein-based signature for the detection of ovarian cancer, *Clin Lab Med* **29** (2009), 47–55.
- [49] T. Kocher and G. Superti-Furga, Mass spectrometry-based functional proteomics: from molecular machines to protein networks, *Nat Methods* **4** (2007), 807–815.
- [50] J. Kononen, L. Bubendorf, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M.J. Mihatsch, G. Sauter and O.P. Kallioniemi, Tissue microarrays for high-throughput molecular profiling of tumor specimens, *Nat Med* **4** (1998), 844–847.
- [51] C. Kuk, V. Kulasingam, C.G. Gunawardana, C.R. Smith, I. Batruch and E.P. Diamandis, Mining the ovarian cancer ascites proteome for potential ovarian cancer biomarkers, *Mol Cell Proteomics* **8** (2009), 661–669.
- [52] P. Lalle, M. De Latour, P. Rio and Y.J. Bignon, Detection of allelic losses on 17q12-q21 chromosomal region in benign lesions and malignant tumors occurring in a familial context, *Oncogene* **9** (1994), 437–442.
- [53] F. Le Naour, D.E. Misek, M.C. Krause, L. Deneux, T.J. Giordano, S. Scholl and S.M. Hanash, Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer, *Clin Cancer Res* **7** (2001), 3328–3335.
- [54] H.A. Lee, B. Kwon, G.Y. Hur, S.J. Choi, D.H. Nahm and H.S. Park, Isotype and IgG subclass distribution of autoantibody response to alpha-enolase protein in adult patients with severe asthma, *Yonsei Med J* **49** (2008), 923–930.
- [55] V.V. Levenson, Biomarkers for early detection of breast cancer: what, when, and where, *Biochim Biophys Acta* **1770** (2007), 847–856.
- [56] J.Z. Levin, M.F. Berger, X. Adiconis, P. Rogov, A. Melnikov, T. Fennell, C. Nusbaum, L.A. Garraway and A. Gnirke, Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts, *Genome Biol* **10** (2009), R115.
- [57] T.J. Ley, E.R. Mardis, L. Ding, B. Fulton, M.D. McLellan, K. Chen, D. Dooling, B.H. Dunford-Shore, S. McGrath, M. Hickenbotham, L. Cook, R. Abbott, D.E. Larson, D.C. Koboldt, C. Pohl, S. Smith, A. Hawkins, S. Abbott, D. Locke, L.W. Hillier, T. Miner, L. Fulton, V. Magrini, T. Wylie, J. Glasscock, J. Conyers, N. Sander, X. Shi, J.R. Osborne, P. Minx, D. Gordon, A. Chinwalla, Y. Zhao, R.E. Ries, J.E. Payton, P. Westervelt, M.H. Tomasson, M. Watson, J. Baty, J. Ivanovich, S. Heath, W.D. Shannon, R. Nagarajan, M.J. Walter, D.C. Link, T.A. Graubert, J.F. DiPersio and R.K. Wilson, DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome, *Nature* **456** (2008), 66–72.
- [58] B. Leyland-Jones, Trastuzumab: hopes and realities, *Lancet Oncol* **3** (2002), 137–144.
- [59] Q.Z. Li, J. Zhou, A.E. Wandstrat, F. Carr-Johnson, V. Branch, D.R. Karp, C. Mohan, E.K. Wakeland and N.J. Olsen, Protein array autoantibody profiles for insights into systemic lupus erythematosus and incomplete lupus syndromes, *Clin Exp Immunol* **147** (2007), 60–70.
- [60] P. Lin, Plasma cell myeloma, *Hematol Oncol Clin North Am* **23** (2009), 709–727.
- [61] A. Lueking, O. Huber, C. Wirths, K. Schulte, K.M. Stieler, U. Blume-Peytavi, A. Kowald, K. Hensel-Wiegel, R. Tauber, H. Lehrach, H.E. Meyer and D.J. Cahill, Profiling of alopecia areata autoantigens based on protein microarray technology, *Mol Cell Proteomics* **4** (2005), 1382–1390.
- [62] E.R. Mardis and R.K. Wilson, Cancer genome sequencing: a review, *Hum Mol Genet* **18** (2009), R163–168.
- [63] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Badger, L.A. Bembgen, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley and J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, *Nature* **437** (2005), 376–380.
- [64] G.A. Michaud, M. Salcius, F. Zhou, R. Bangham, J. Bonin, H. Guo, M. Snyder, P.F. Predki and B.I. Schweitzer, Analyzing antibody specificity with whole proteome microarrays, *Nat Biotechnol* **21** (2003), 1509–1512.
- [65] G.A. Michaud, M. Salcius, F. Zhou, V.V. Papov, J. Merkel, M. Murtha, P. Predki and B. Schweitzer, Applications of protein arrays for small molecule drug discovery and characterization, *Biotechnol Genet Eng Rev* **22** (2006), 197–211.
- [66] G. Mor, I. Visintin, Y. Lai, H. Zhao, P. Schwartz, T. Rutherford, L. Yue, P. Bray-Ward and D.C. Ward, Serum protein markers for early detection of ovarian cancer, *Proc Natl Acad Sci USA* **102** (2005), 7677–7682.
- [67] U. Nagalakshmi, Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein and M. Snyder, The transcriptional landscape of the yeast genome defined by RNA sequencing, *Science* **320** (2008), 1344–1349.
- [68] D.H. Nahm, Y.E. Lee, E.J. Yim, H.S. Park, H. Yim, Y. Kang and J.K. Kim, Identification of cytokeratin 18 as a bronchial epithelial autoantigen associated with nonallergic asthma, *Am J Respir Crit Care Med* **165** (2002), 1536–1539.
- [69] S. Nakamura, C.S. Yang, N. Sakon, M. Ueda, T. Tougan, A. Yamashita, N. Goto, K. Takahashi, T. Yasunaga, K. Ikuta, T. Mizutani, Y. Okamoto, M. Tagami, R. Morita, N. Maeda, J. Kawai, Y. Hayashizaki, Y. Nagai, T. Horii, T. Iida and T. Nakaya, Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach, *PLoS ONE* **4** (2009), e4219.
- [70] S. Natter, S. Seiberler, P. Hufnagl, B.R. Binder, A.M. Hirschl, J. Ring, D. Abeck, T. Schmidt, P. Valent and R. Valent, Isolation of cDNA clones coding for IgE autoantigens with serum IgE from atopic dermatitis patients, *FASEB J* **12** (1998), 1559–1569.
- [71] F. Ozsolak, A.R. Platt, D.R. Jones, J.G. Reifengerger, L.E. Sass, P. McInerney, J.F. Thompson, J. Bowers, M. Jarosz and P.M. Milos, Direct RNA sequencing, *Nature* **461** (2009), 814–818.
- [72] G. Palacios, J. Druce, L. Du, T. Tran, C. Birch, T. Briese, S. Conlan, P.L. Quan, J. Hui, J. Marshall, J.F. Simons, M. Egholm, C.D. Paddock, W.J. Shieh, C.S. Goldsmith, S.R. Zaki, M. Catton and W.I. Lipkin, A new arenavirus in a cluster of fatal transplant-associated diseases, *N Engl J Med* **358** (2008), 991–998.
- [73] T. Palomero and A.A. Ferrando, Genomic tools for dissecting oncogenic transcriptional networks in human leukemia, *Leukemia* **23** (2009), 1236–1242.
- [74] T.H. Patwa, Y. Wang, D.M. Simeone and D.M. Lubman, Enhanced detection of autoantibodies on protein microarrays using a modified protein digestion technique, *J Proteome Res*

- 7 (2008), 2553–2561.
- [75] S.P. Peters, G. Ferguson, Y. Deniz and C. Reisner, Uncontrolled asthma: a review of the prevalence, disease burden and options for treatment, *Respir Med* **100** (2006), 1139–1151.
- [76] J. Pittman, E. Huang, H. Dressman, C.F. Horng, S.H. Cheng, M.H. Tsou, C.M. Chen, A. Bild, E.S. Iversen, A.T. Huang, J.R. Nevins and M. West, Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes, *Proc Natl Acad Sci USA* **101** (2004), 8431–8436.
- [77] A.E. Platts, S.J. Land, L. Chen, G.P. Page, P. Rasouli, L. Wang, X. Lu and D.M. Ruden, Massively parallel resequencing of the isogenic *Drosophila melanogaster* strain w(1118); iso-2; iso-3 identifies hotspots for mutations in sensory perception genes, *Fly (Austin)* **3** (2009), 192–203.
- [78] M. Plebani, M. Pittoni, M. Celadin, D. Bernardi and M.M. Mion, Recent advances in diagnostic technologies for autoimmune diseases, *Autoimmun Rev* **8** (2009), 238–243.
- [79] J. Ptacek, G. Devgan, G. Michaud, H. Zhu, X. Zhu, J. Fasolo, H. Guo, G. Jona, A. Breitkreutz, R. Sopko, R.R. McCartney, M.C. Schmidt, N. Rachidi, S.J. Lee, A.S. Mah, L. Meng, M.J. Stark, D.F. Stern, C. De Virgilio, M. Tyers, B. Andrews, M. Gerstein, B. Schweitzer, P.F. Predki and M. Snyder, Global analysis of protein phosphorylation in yeast, *Nature* **438** (2005), 679–684.
- [80] J. Qiu, G. Choi, L. Li, H. Wang, S.J. Pitteri, S.R. Pereira-Faca, A.L. Krasnoselsky, T.W. Randolph, G.S. Omenn, C. Edelstein, M.J. Barnett, M.D. Thornquist, G.E. Goodman, D.E. Brenner, Z. Feng and S.M. Hanash, Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in pre-diagnostic lung cancer sera, *J Clin Oncol* **26** (2008), 5060–5066.
- [81] J. Qiu and S. Hanash, Autoantibody profiling for cancer detection, *Clin Lab Med* **29** (2009), 31–46.
- [82] F.J. Quintana, M.F. Farez, V. Viglietta, A.H. Iglesias, Y. Merbl, G. Izquierdo, M. Lucas, A.S. Basso, S.J. Khoury, C.F. Lucchinetti, I.R. Cohen and H.L. Weiner, Antigen microarrays identify unique serum autoantibody signatures in clinical and pathologic subtypes of multiple sclerosis, *Proc Natl Acad Sci USA* **105** (2008), 18889–18894.
- [83] M.S. Raab, K. Podar, I. Breitkreutz, P.G. Richardson and K.C. Anderson, Multiple myeloma, *Lancet* **374** (2009), 324–339.
- [84] S.V. Rajkumar, A. Dispenzieri and R.A. Kyle, Monoclonal gammopathy of undetermined significance, Waldenstrom macroglobulinemia, AL amyloidosis, and related plasma cell disorders: diagnosis and treatment, *Mayo Clin Proc* **81** (2006), 693–703.
- [85] R. Ralhan, L.V. Desouza, A. Matta, S. Chandra Tripathi, S. Ghanny, S. Datta Gupta, S. Bahadur and K.W. Siu, Discovery and verification of head-and-neck cancer biomarkers by differential protein expression analysis using iTRAQ labeling, multidimensional liquid chromatography, and tandem mass spectrometry, *Mol Cell Proteomics* **7** (2008), 1162–1173.
- [86] N. Ramachandran, J.V. Raphael, E. Hainsworth, G. Demirkan, M.G. Fuentes, A. Rolfs, Y. Hu and J. LaBaer, Next-generation high-density self-assembling functional protein arrays, *Nat Methods* **5** (2008), 535–538.
- [87] M. Reuschenbach, M. von Knebel Doeberitz and N. Wentzensen, A systematic review of humoral immune responses against tumor antigens, *Cancer Immunol Immunother* **58** (2009), 1535–1544.
- [88] M. Rottem and Y. Shoenfeld, Asthma as a paradigm for autoimmune disease, *Int Arch Allergy Immunol* **132** (2003), 210–214.
- [89] A.H. Rowe, Insulin Treatment of Diabetes Mellitus, *Cal State J Med* **21** (1923), 204–208.
- [90] M.J. Scanlan, I. Gout, C.M. Gordon, B. Williamson, E. Stockert, A.O. Gure, D. Jager, Y.T. Chen, A. Mackay, M.J. O’Hare and L.J. Old, Humoral immunity to human breast cancer: antigen definition and quantitative analysis of mRNA expression, *Cancer Immun* **1** (2001), 4.
- [91] P.E. Schwartz, Current diagnosis and treatment modalities for ovarian cancer, *Cancer Treat Res* **107** (2002), 99–118.
- [92] N.J. Serkova and K. Glunde, Metabolomics of cancer, *Methods Mol Biol* **520** (2009), 273–295.
- [93] S.P. Shah, M. Kobel, J. Senz, R.D. Morin, B.A. Clarke, K.C. Wiegand, G. Leung, A. Zayed, E. Mehl, S.E. Kalloger, M. Sun, R. Giuliany, E. Yorida, S. Jones, R. Varhol, K.D. Swenerton, D. Miller, P.B. Clement, C. Crane, J. Madore, D. Provencher, P. Leung, A. DeFazio, J. Khattra, G. Turashvili, Y. Zhao, T. Zeng, J.N. Glover, B. Vanderhyden, C. Zhao, C.A. Parkinson, M. Jimenez-Linan, D.D. Bowtell, A.M. Mes-Masson, J.D. Brenton, S.A. Aparicio, N. Boyd, M. Hirst, C.B. Gilks, M. Marra and D.G. Huntsman, Mutation of FOXL2 in granulosa-cell tumors of the ovary, *N Engl J Med* **360** (2009), 2719–2729.
- [94] J. Shen, M.D. Person, J. Zhu, J.L. Abbruzzese and D. Li, Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry, *Cancer Res* **64** (2004), 9018–9026.
- [95] B.O. Shum, M.S. Rolph and W.A. Sewell, Mechanisms in allergic airway inflammation – lessons from studies in the mouse, *Expert Rev Mol Med* **10** (2008), e15.
- [96] M.G. Smith, T.A. Gianoulis, S. Pukatzki, J.J. Mekalanos, L.N. Ornston, M. Gerstein and M. Snyder, New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis, *Genes Dev* **21** (2007), 601–614.
- [97] K. Somers, C. Govarts, P. Stinissen and V. Somers, Multiplexing approaches for autoantibody profiling in multiple sclerosis, *Autoimmun Rev* **8** (2009), 573–579.
- [98] Q. Song, G. Liu, S. Hu, Y. Zhang, Y. Tao, Y. Han, H. Zeng, W. Huang, F. Li, P. Chen, J. Zhu, C. Hu, S. Zhang, Y. Li, H. Zhu and L. Wu, Novel Autoimmune Hepatitis-Specific Autoantigens Identified Using Protein Microarray Technology, *J Proteome Res* (2009).
- [99] K. Sorber, C. Chiu, D. Webster, M. Dimon, J.G. Ruby, A. Hekele and J.L. DeRisi, The long march: a sample preparation technique that enhances contig length and coverage by high-throughput short-read sequencing, *PLoS ONE* **3** (2008), e3495.
- [100] R.S. Sun, X.H. Chen, R.Q. Liu, T.M. Cheng, X.Z. Ran and T. Yang, Autoantibodies to the high-affinity IgE receptor in patients with asthma, *Asian Pac J Allergy Immunol* **26** (2008), 19–22.
- [101] R.S. Sun, X.H. Chen, J.F. Sui, R.Q. Liu, T.M. Cheng, X.Z. Ran and T. Yang, Detecting anti-FcεpsilonRI autoantibodies in patients with asthma by flow cytometry, *J Int Med Res* **36** (2008), 1214–1219.
- [102] B. Tomaino, P. Cappello, M. Capello, C. Fredolini, A. Ponzetto, A. Novarino, L. Ciuffreda, O. Bertetto, C. De Angelis, E. Gaia, P. Salacone, M. Milella, P. Nisticò, M. Alessio, R. Chiarle, M.G. Giuffrida, M. Giovarelli and F. Novelli, Autoantibody signature in human ductal pancreatic adenocarci-

- noma, *J Proteome Res* **6** (2007), 4025–4031.
- [103] G.E. Trivers, V.M. De Benedetti, H.L. Cawley, G. Caron, A.M. Harrington, W.P. Bennett, J.R. Jett, T.V. Colby, H. Tazelaar, P. Pairolero, R.D. Miller and C.C. Harris, Anti-p53 antibodies in sera from patients with chronic obstructive pulmonary disease can predate a diagnosis of cancer, *Clin Cancer Res* **2** (1996), 1767–1775.
- [104] R. Valenta, S. Natter, S. Seiberler, S. Wichlas, D. Maurer, M. Hess, M. Pavelka, M. Grote, F. Ferreira, Z. Szepefalusi, P. Valent and G. Stingl, Molecular characterization of an autoallergen, Hom s 1, identified by serum IgE from atopic dermatitis patients, *J Invest Dermatol* **111** (1998), 1178–1183.
- [105] R. Valenta, S. Seiberler, S. Natter, V. Mahler, R. Mossabeh, J. Ring and G. Stingl, Autoallergy: a pathogenetic factor in atopic dermatitis?, *J Allergy Clin Immunol* **105** (2000), 432–437.
- [106] L.J. van 't Veer, H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards and S.H. Friend, Gene expression profiling predicts clinical outcome of breast cancer, *Nature* **415** (2002), 530–536.
- [107] M. Vejtorp, M. Hoier-Madsen and P. Halberg, Enzyme-linked immunosorbent assay for determination of IgM rheumatoid factor, *Scand J Rheumatol* **8** (1979), 65–70.
- [108] J.C. Venter, C.M. Fraser and L.C. Harrison, Autoantibodies to beta 2-adrenergic receptors: a possible cause of adrenergic hyporesponsiveness in allergic rhinitis and asthma, *Science* **207** (1980), 1361–1363.
- [109] C. Vicentini, C. Festuccia, G.L. Gravina, A. Angelucci, A. Marronaro and M. Bologna, Prostate cancer cell proliferation is strongly reduced by the epidermal growth factor receptor tyrosine kinase inhibitor ZD1839 in vitro on human cell lines and primary cultures, *J Cancer Res Clin Oncol* **129** (2003), 165–174.
- [110] I. Visintin, Z. Feng, G. Longton, D.C. Ward, A.B. Alvero, Y. Lai, J. Tenthorey, A. Leiser, R. Flores-Saaib, H. Yu, M. Azori, T. Rutherford, P.E. Schwartz and G. Mor, Diagnostic markers for early detection of ovarian cancer, *Clin Cancer Res* **14** (2008), 1065–1072.
- [111] P. Wang, J.R. Whiteaker and A.G. Paulovich, The evolving role of mass spectrometry in cancer biomarker discovery, *Cancer Biol Ther* **8** (2009), 1083–1094.
- [112] Z. Wang, M. Gerstein and M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nat Rev Genet* (2008).
- [113] Z. Wang, L. Jiang, C. Huang, Z. Li, L. Chen, L. Gou, P. Chen, A. Tong, M. Tang, F. Gao, J. Shen, Y. Zhang, J. Bai, M. Zhou, D. Miao and Q. Chen, Comparative proteomics approach to screening of potential diagnostic and therapeutic targets for oral squamous cell carcinoma, *Mol Cell Proteomics* **7** (2008), 1639–1650.
- [114] M.R. Warrier and G.K. Hershey, Asthma genetics: personalizing medicine, *J Asthma* **45** (2008), 257–264.
- [115] M.P. Washburn, D. Wolters and J.R. Yates, 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat Biotechnol* **19** (2001), 242–247.
- [116] L. Wu and D.K. Han, Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics, *Expert Rev Proteomics* **3** (2006), 611–619.
- [117] L. Zhong, K. Ge, J.C. Zu, L.H. Zhao, W.K. Shen, J.F. Wang, X.G. Zhang, X. Gao, W. Hu, Y. Yen and K.H. Kernstine, Autoantibodies as potential biomarkers for breast cancer, *Breast Cancer Res* **10** (2008), R40.
- [118] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R.A. Dean, M. Gerstein and M. Snyder, Global analysis of protein activities using proteome chips, *Science* **293** (2001), 2101–2105.
- [119] H. Zhu, S. Hu, G. Jona, X. Zhu, N. Kreiswirth, B.M. Willey, T. Mazzulli, G. Liu, Q. Song, P. Chen, M. Cameron, A. Tyler, J. Wang, J. Wen, W. Chen, S. Compton and M. Snyder, Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray, *Proc Natl Acad Sci USA* **103** (2006), 4011–4016.
- [120] X. Zhu, M. Gerstein and M. Snyder, ProCAT: a data analysis approach for protein microarrays, *Genome Biol* **7** (2006), R110.
- [121] B. Zweiman, Asthma and autoimmunity: is there a connection? *Curr Allergy Asthma Rep* **7** (2007), 157–158.