

# Antibody-based Proteomics for Human Tissue Profiling

Mathias Uhlen<sup>‡§</sup> and Fredrik Ponten<sup>¶</sup>

**Here, we describe the use of antibody-based proteomics involving the generation of protein-specific antibodies to functionally explore the human proteome. The antibodies can be used for analysis of corresponding proteins in a wide range of assay platforms, including i) immunohistochemistry for detailed tissue profiling, ii) specific affinity reagents for various functional protein assays, and iii) capture (“pull-down”) reagents for purification of specific proteins and their associated complexes for structural and biochemical analyses. In this review, the use of antibodies for such analysis will be discussed with focus on the possibility to create a descriptive and comprehensive protein atlas for tissue distribution and subcellular localization of human proteins in both normal and disease tissues. *Molecular & Cellular Proteomics* 4:384–393, 2005.**

A great need exists for the generation of protein-specific antibodies to explore the human proteome (1, 2). These efforts, usually denoted antibody proteomics, is here defined as the systematic generation and use of protein-specific antibodies to functionally explore the proteome. Affinity reagents can be utilized in numerous assay platforms (Fig. 1), including *in vivo* and *in vitro* protein profiling and for pull-out experiments to enable identification of interaction partners or post-translational modifications (3, 4). In addition, affinity reagents can be spotted onto microarrays for the detection of specific biomolecules in complex biosamples (5) or used for structural genomics efforts or studies of the biochemical function of specific proteins (6, 7). In this review, we will discuss the issues and strategic alternatives for the generation of antibodies, including choice of antigen and advantages and disadvantages with different types of antibodies. The use of tissue microarrays to allow high-throughput tissue profiling will also be discussed, as well as the value of a comprehensive protein atlas for clinical and biomedical research.

## GENERATION OF ANTIGENS

One of the main difficulties of high-throughput antibody generation is the production of sufficiently large amounts of

protein targets for the generation of antibodies. There are three principally different types of protein antigen that can be generated, excluding genetic immunization, and those are: i) full-length proteins, ii) synthetic peptides, or iii) recombinant protein fragments. The full-length protein route usually involves the expression of the recombinant protein in various eukaryotic hosts to mimic the folding and posttranslational modification of the native protein (6). Alternatively, the native protein is purified from an appropriate tissue by a series of biochemical unit operations. In some cases, but far from always, purified protein with native fold can be obtained, although the amount of protein is usually low as compared with the other strategies described below. It is interesting to note that the advantage of correctly folded antigens during immunization is not yet thoroughly investigated. It is not unlikely that, in many cases, native proteins are denatured during the immunization procedure, *i.e.* the mixing with mineral oil for the preparation of Freund's adjuvance (8, 9). The importance of generating a correctly folded protein, which is subsequently denatured immediately prior to the immunization, could therefore be questioned.

The use of synthetic peptides as antigen represents a completely different strategy. The synthetic route is technically simple and large amounts of pure peptide can readily be made in a short time. The peptides are usually less than 40 amino acids and it is therefore not likely that native folded structures are obtained. This means that a large fraction of the obtained antibodies recognize a peptide in a form not present on the folded target protein. In addition, the linear epitopes exposed on small peptides might not be exposed on the surface of a corresponding protein and thus lead to nonfunctional antibodies in the subsequent assays. However, many examples of successful peptide antibodies have been generated in both the academic and the commercial sector (10).

A third way to produce antigen is to generate recombinant protein fragments (11). The use of protein fragments makes the cloning procedures much easier than for full-length proteins, because relatively short PCR fragments can be used. In addition, affinity tags can be used for the purification, which makes the purification procedure easy to scale-up and reliable for the generation of pure antigen. Since the introduction of affinity tags for purification of recombinant proteins more than 20 years ago (12), a large number of tag systems are now available and the choice of system depends on factors such as the need for solubility and ease of purification (13). The

From the <sup>‡§</sup>Department of Biotechnology, AlbaNova University Center, Royal Institute of Technology, SE-10691 Stockholm, Sweden; and <sup>¶</sup>Department of Genetics and Pathology, Rudbeck Laboratory, SE-751 85 Uppsala University, Uppsala, Sweden

Received, February 5, 2005

Published, MCP Papers in Press, February 5, 2005, DOI 10.1074/mcp.R500009-MCP200

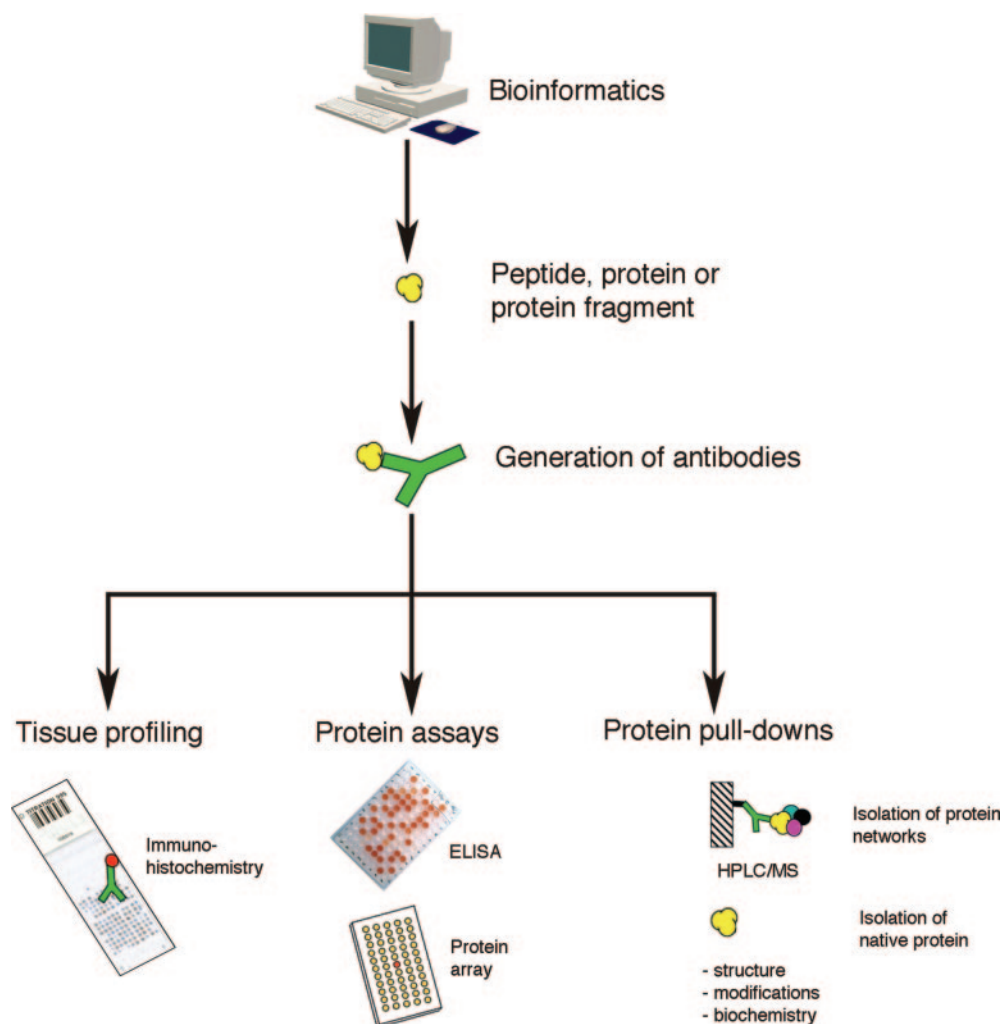


FIG. 1. **The principle of antibody proteomics, defined as the systematic generation and use of protein-specific antibodies to functionally explore the proteome.** The antibodies can be used for analysis of corresponding proteins in a wide range of assay platforms, including i) tissue profiling, ii) protein assays, and iii) as capture (“pull-down”) reagents for purification of specific proteins and their associated complexes for structural and biochemical analyses.

histidine tag system used for purification with immobilized metals (14), such as Co or Ni, is the most frequently used system, partly because the affinity purification can be performed in buffers containing chaotropic reagents, such as urea, which allows precipitated proteins (inclusion bodies) to be recovered (15). The fact that only part of the protein is expressed makes it unlikely that fully folded structures are obtained, although careful selection of a fragment corresponding to a protein domain of structural integrity could increase the probability of generating a native configuration. Both the synthetic peptide and the protein fragment routes allow for selection of peptides or fragments with low homology to other proteins of the human proteome, enabling the use of the information from the human genome sequence to design epitope-specific antibodies with low risk of cross-reactivity (16). Agaton *et al.* (11) showed that the selection of such unique regions of the protein called Protein Epitope

Signatures Tags (PrESTs)<sup>1</sup> made it possible to design and produce recombinant protein fragments suitable for antibody generation, as exemplified by a pilot study using the genes of human chromosome 21.

In conclusion, there are three principally different ways to produce antigens for the generation of antibodies. Full-length proteins are the most likely to have a correct fold, although this strategy is often cumbersome with relatively low success rates for the combined expression and purification steps. The synthetic peptide strategy is rapid and convenient, but is generally limited to the generation of linear epitopes. The

<sup>1</sup> The abbreviations used are: PrEST, protein epitope signatures tags; pAb, polyclonal antibody; mAb, monoclonal antibody; recAb, recombinant antibody; msAb, mono-specific antibodies; TMA, tissue microarrays; IgG, immunoglobulin G; PTM, posttranslational modification.

strategy with recombinant fragments combines ease of use with the potential to create conformational epitopes, but relatively often yields insoluble inclusion bodies in the recombinant host. The two latter approaches have the advantage of generating relatively large amounts of antigen, which allows for affinity purification of antibodies using the antigen as ligand. Availability of large amounts of antigen also facilitates subsequent quality assurance of antibodies, e.g. adsorption and binding assays to test for specificity and selectivity.

GENERATION OF ANTIBODIES

Currently, there are four major types of antibodies used for antibody-based analysis: i) polyclonal antibodies (pAb) made by immunization of an animal, such as rabbit, hen, goat, or rat; ii) monoclonal antibodies (mAb) made by immunization and screening of B cells from antibody-producing mouse (17, 18) or rabbit (19) cells; iii) recombinant antibodies (recAb) made through *in vitro* selection principles (20, 21); and iv) mono-specific antibodies (msAb) generated from pAbs using antigen-specific purification (22).

pAbs have the inherent advantage of multi-epitope binding, which makes them suitable for cross-platform assays involving proteins both in a native and denatured form. However, the specific antibodies in a pAb reagent (antiserum) usually constitute less than 1% of the total amount of antibodies. Thus, pAb produced by standard routes often show cross-reactivity and lack of reproducibility upon re-immunizations with the same antigen, which makes these reagents less attractive as a renewable resource.

mAbs have a defined single binding epitope, which makes them attractive for diagnostic and therapeutic applications. mAbs are currently the most commonly used immunoreagents in diagnostic applications (23). However, because monoclonal reagents recognize a single epitope of a protein, it is difficult to generate antibodies that can be used as single and defined reagents across whole platforms of functional and detection assays, where the proteins often are denatured in different ways, such as with detergents or formalin. In addition, time-consuming screening is needed to generate mAbs, which render them less suitable for large-scale efforts.

recAbs are usually selected by utilizing phage display technologies (21), or other selection principles, such as ribosomal display (20). Much of the *in vitro* selection technology has been based on recAb fragments, especially scFvs and Fabs (24), but alternative protein scaffolds have also been used (25). Nonproteinaceous binding ligands, such as the *in vitro* selected aptamers (26), have also been successfully utilized for specific recognition of target molecules. These affinity reagents can be generated without the use of animals and recently attempts have been made toward proteome-wide production of recAbs by phage display (27).

msAbs are generated from a polyclonal antiserum using the antigen as ligand (22). A disadvantage is that the purification procedure is technically more challenging as compared with

TABLE I  
Comparison of the use of pAbs, mAbs, recAbs, and msAbs

Parameter	pAb	mAb	recAb	msAb
Uniform epitope specificity	-	+	+	+
Renewable resource	+/-	+	+	+/-
Need to use animals	Yes	Yes	No	Yes
Suitability for diagnostics	+	++	++	++
Suitability for therapeutics	- <sup>a</sup>	+	+/-	-
Suitability for "pull-down"	+	+/-	+/-	+
Suitability for protein arrays	+	++	++	++
Detection of denatured proteins	++	+/-	+/-	++
Need for screening	No	Yes	Yes	No

<sup>a</sup> pAbs have been used for many decades as vaccines (gamma-globulin).

use of an antiserum consisting of nonpurified pAbs. However, the elimination of more than 99% of the antibodies directed to other, unknown epitopes results in an increased probability of obtaining a selective and specific antibody reagent. The affinity purification of pAbs on protein fragments with potentially many different epitopes often renders antibodies that are highly functional across a platform of different assays. An advantage as compared with mAbs and recAbs is that no screening is needed, which facilitates attempts to scale-up to whole-proteome applications.

In conclusion, each type of antibody has its advantages and disadvantages. The choice of antibody therefore depends on factors, such as the need for a renewable resource, ease of selection and screening, the need for use across several analysis platforms, and the possibility to use the antibody for routine diagnostics and therapy in patients. A summary of some of considerations related to choice of antibody class is shown in Table I with some comments on potential advantages and disadvantages for each class of antibodies.

ANALYSIS USING TISSUE MICROARRAYS (TMA)

Immunohistochemistry provides an important tool for *in situ* visualization of protein expression patterns in tissues and cells (28, 29). The possibility to detect and localize defined proteins at tissue, cellular, and subcellular levels present a deeper insight into normal cellular functions and pathogenic mechanisms leading to different types of disease. The use of multiple biopsies or curettages in a single paraffin block has been utilized in surgical pathology procedures for several decades. Although several articles have been published describing improvements of methodology and usefulness in immunohistochemistry (30, 31), multi-tissue blocks did not become a powerful tool in research until designated instruments were developed and the tissue array technology was standardized (32). TMA technology now provides an automated array-based high-throughput technique, in which up to 1,000 of paraffin-embedded tissue samples can be assembled into one paraffin block in an array format (33, 34). The diameter of the tissue sample cores is generally 0.6–1.5 mm in diameter.

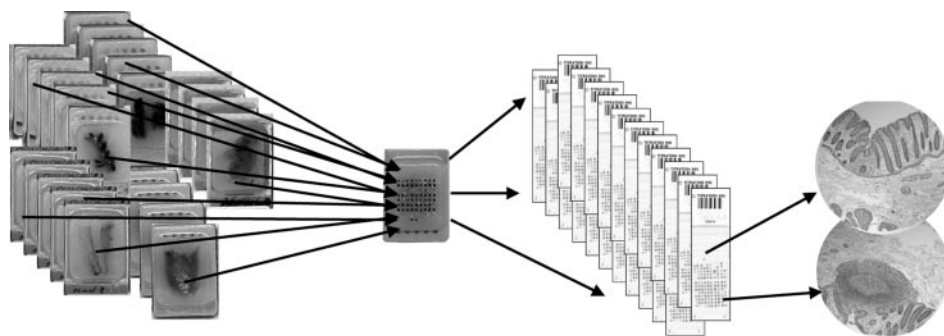


FIG. 2. **The principle of the generation of TMA.** A large number of paraffin blocks (*left*) containing tissues are collected and used as template for the designed TMA block (*center-left*). From each created recipient TMA-block (*center-right*) several hundred tissue sections (3–4  $\mu\text{m}$  thick) can be cut. Data on protein expression patterns can be mined from the immunohistochemically stained spots of tissue (*right*).

From each TMA block, thin tissue sections can be cut and mounted onto glass slides, producing several hundred nearly identical slides (Fig. 2).

The TMA technology allows investigators to use a single slide to conduct controlled studies on large cohorts of tissues using only small amounts of reagent. The source of tissue is only restricted by its availability in paraffin, and the small amount of tissue that is needed minimizes loss of unique and sparse tissues. As only limited amount of tissue is used in the TMA, individual heterogeneity is an issue that needs consideration. Several studies have compared the use of sections from whole tissue blocks with different number and size of tissue cores in the TMA format. The use of two to four cores, depending on design and content of TMA results in >95% accuracy and appears as a reasonable standard in TMA design (35). The combination of immunohistochemistry and TMA technology therefore presents an attractive strategy for high-throughput, antibody-based tissue proteomics (36, 37).

In a recent study, it was shown that high-throughput analysis of protein expression could be performed using a standardized set of TMAs containing both normal human tissues and various types of cancer (38). For representation of normal tissues, a set of 48 different tissues was selected from the human body. Tissues were selected to comprise the wide variety of phenotypes found within the human body, and all major organs and tissues were included. The selection was partly based on availability of surgical material, *i.e.* highly specialized tissues, *e.g.* eye, ear, spinal chord were not included due to lack of tissue resources. One tissue core from three different individuals was assembled for each of the selected 48 normal tissue types (Fig. 3). When applicable, the triplicate samples represented different sex and age. In this manner, protein expression patterns were thus analyzed in tissue spots from 144 different individuals for each tested antibody.

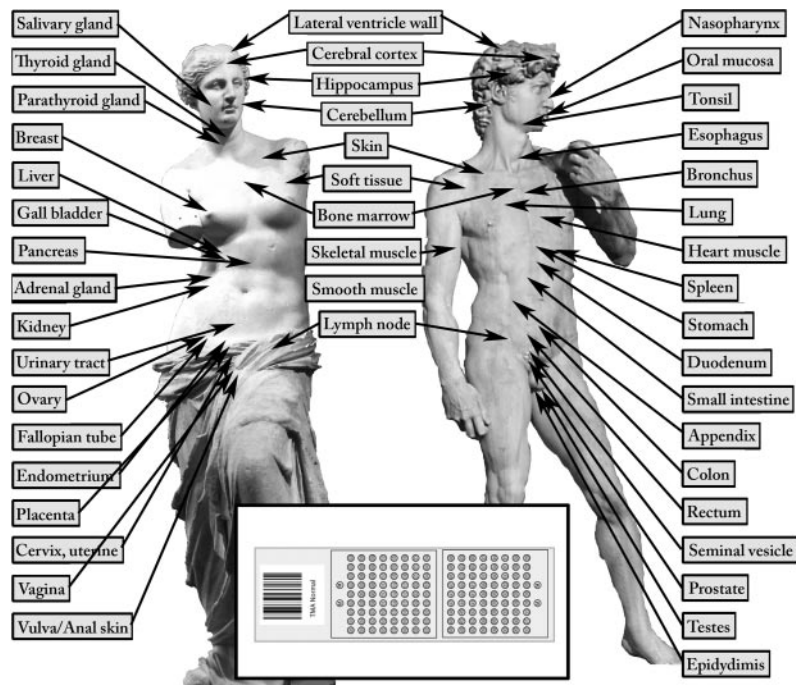
A similar approach was also used to analyze protein expression patterns in human cancer (38). Malignant tumors are classified into different cancer types, mainly based on localization of primary tumor and phenotypical characteristics. New molecular biological tools have recently altered the classical diagnostics resulting in a modified classification, based

on mutation analysis and protein expression patterns, for certain types of cancer (most evident for hematological malignancies). The TMA format is ideal for assembling samples from numerous individual tumors within each separate cancer type. Cancer TMAs including a reasonable amount of individual tumors representing the most common types of cancer allow for a primary screening of protein expression patterns of potential importance in cancer development. A schematic overview showing the 216 unique tumors that were analyzed for each antibody is shown in Fig. 4. Screening for tumor-associated protein expression profiles provides a tool for basic cancer research as well as a foundation for more in-depth studies with the aim of finding new and improved diagnostic and therapeutic agents. The presented approach is possible to scale-up and can be used for whole-proteome efforts, provided that an adequate number of specific antibodies are available.

As an example of the use of cancer TMAs, some representative images obtained for an antibody generated to a member of the ALEX family of proteins (ALEX3 protein) is shown in Fig. 5. The ALEX proteins may play a role in tumor suppression, and the encoded protein contains a potential N-terminal transmembrane domain and a single Armadillo (arm) repeat (39). Other proteins containing the arm repeat are involved in development, maintenance of tissue integrity, and carcinogenesis. This gene is localized together with other family members on the X chromosome. Three transcript variants encoding the same protein have been identified for this gene. When staining cancer TMAs including 216 different individual tumors with an antibody generated toward this protein, a subset of cancer types showed a strong positive staining (Fig. 5). Interestingly, this antibody stained a majority of the hepatocellular carcinomas, a subset of ovarian carcinomas and only rare other tumors, *e.g.* 1/12 malignant melanomas. In the TMA containing normal tissues, this antibody showed exclusive positive staining in liver, testes, brain tissues, and rare cells in the GI tract (data not shown). The example shows that staining patterns in normal and disease tissues can be rapidly screened for every protein in which a specific antibody is available. Potential biomarkers for disease can



FIG. 3. A schematic figure showing an example of a TMA representing normal tissues. Surgical specimens, representing a large variety of organs and tissues listed in the figure were selected by Kampf *et al.* (38) to generate the TMA blocks containing normal tissues (triplicate samples of each tissue type). In total 144 different tissue cores representing 48 different tissue types were assembled. Schematic illustration of TMA sections from normal TMAs are shown at the bottom of the figure.



thus be identified in a systematic manner and immunoassays can be designed for further validation in relevant patient cohorts.

It is also possible to generate arrays to study protein expression from *in vitro* cultured cells or cell samples from patients and sections from such a cell-TMA block can be used for immunohistochemistry (40, 41). The advantage of analyzing immunoreactivity in a single layer of dispersed cells, without the complexity of different cell populations and matrix background, includes a possibility to employ automated image analysis systems. Standardized image analysis algorithms would allow for more objective measures of qualitative as well as quantitative parameters. Furthermore, utilization of cell lines in TMAs facilitate a comparative analysis of transcription and protein expression using immunohistochemistry combined with image analysis and cDNA-array technologies. Cell-TMAs are also advantageous to study hematological malignancies, for which there are readily available cell aspirate samples as well as a multitude of established cell lines. A considerable amount of data including genotype, transcription profiles, protein expression, and phenotypic characteristics have been published for available and well-characterized cell lines, representing different stages of lymphoid and myeloid differentiation.

An issue for the analysis of tissue profiles is the choice between fluorescence- or enzyme-based immunohistochemistry. The enzyme-based method is sensitive, easy to automate, and allows a rudimentary analysis of intracellular localization. However, the fluorescent-based methods can be used for detailed subcellular analysis and three-dimensional reconstruction of the cell can be done using confocal micros-

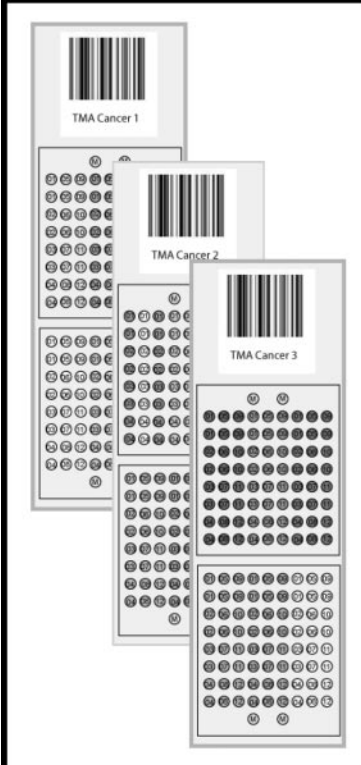
copy (42). It is thus possible to discriminate not only between nuclear, cytoplasmic, and membrane-bound localization, but also to determine proteins expressed in mitochondria, nucleoli, nuclear membrane, etc. In addition, fluorescent methods facilitate comparative localization studies because multiple dyes, which enable several proteins to be analyzed simultaneously, can be used.

A major benefit of using TMA technology for expression profiling of tissues is the ease of automation. Collection of tissues including production of TMAs and immunohistochemistry is fairly labor intensive, whereas slide scanning with subsequent image processing can be automated to a large degree (43). Various commercial software and hardware employing different technologies for scanning tissue sections on glass slides have recently been developed, yet comprehensive publications are sparse. One major advantage is that cumbersome and time-consuming handling of glass slides and evaluation in the microscope is avoided through the generation of digital data. Digital images allow for high-throughput analysis including rapid distribution over networks. Digital images can be viewed and scored using conventional web-based tools, *i.e.* virtual microscopy coupled to a database, facilitating easy access to both image and data.

#### A PROTEIN ATLAS FOR TISSUE PROFILES

One important objective for clinical proteomics is to generate a protein atlas displaying expression and localization patterns of proteins in all or most human tissues and organs (37, 44). Such a protein atlas would function as a knowledgebase with regard to the structural and temporal expression of the

FIG. 4. A schematic figure showing an example of a tissue microarray representing cancer tissues. In total, cores from 216 individual tumors (in duplicate) were selected by Kampf *et al.* (38) representing 20 different types of cancers. On the right side, the number of individual tumor used as templates for each cancer type is shown. An illustration of three TMA slides with sections from the cancer TMA blocks is shown (left).



<b>Breast cancer</b>	<b>12</b>
<b>Carcinoid</b>	<b>4</b>
<b>Cervical cancer</b>	<b>12</b>
<b>Colo-rectal cancer</b>	<b>12</b>
<b>Endometrial cancer</b>	<b>12</b>
<b>Glioma</b>	<b>12</b>
<b>Head-Neck cancer</b>	<b>4</b>
<b>Kidney cancer</b>	<b>12</b>
<b>Liver cancer</b>	<b>12</b>
<b>Lung cancer</b>	<b>12</b>
<b>Lymphoma</b>	<b>12</b>
<b>Melanoma</b>	<b>12</b>
<b>Ovarian cancer</b>	<b>12</b>
<b>Pancreatic cancer</b>	<b>12</b>
<b>Prostate cancer</b>	<b>12</b>
<b>Skin cancer</b>	<b>12</b>
<b>Stomach cancer</b>	<b>12</b>
<b>Testis cancer</b>	<b>12</b>
<b>Thyroid cancer</b>	<b>4</b>
<b>Urothelial cancer</b>	<b>12</b>

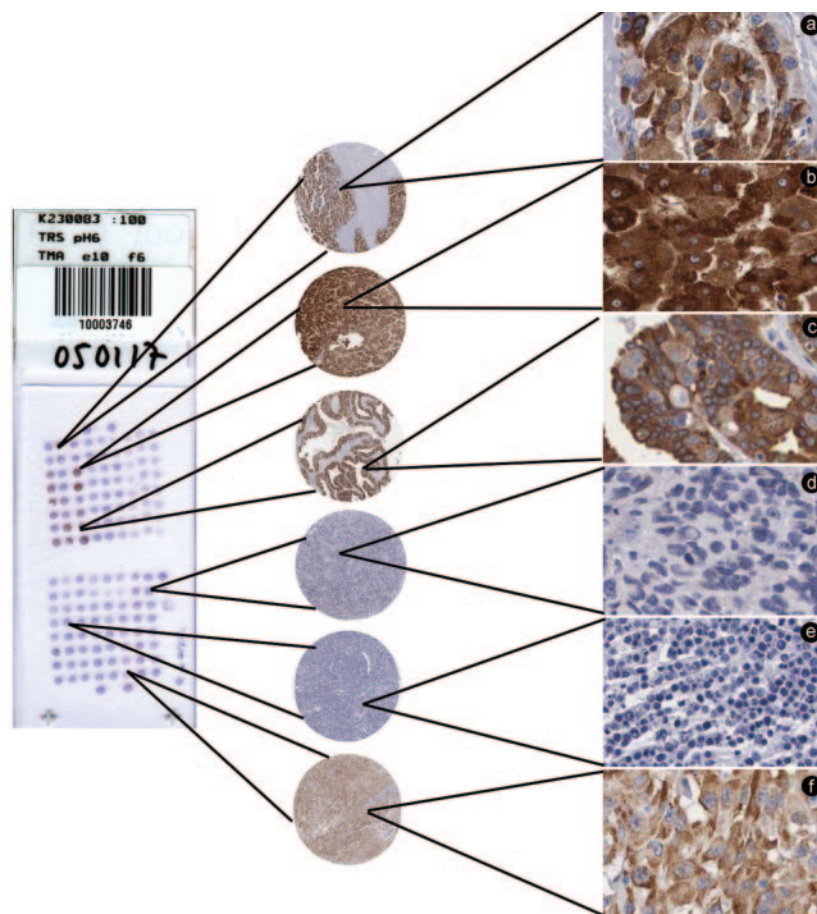
human proteins in various cells and tissues. Using the TMA approach described above, it is possible to generate a large number of digital high-resolution images corresponding to normal and cancer tissues and then display them using a conventional web-based database. In this manner, a protein atlas displaying tissue expression and localization can be developed for each protein with an available specific antibody. However, the file size of high-resolution images representing whole tissue spots is a challenge and it is probably necessary in most cases to compress the large image sizes to enable a web-based interface. For example, an original TIFF image representing a 1-mm core can be reduced from 50 Mbytes to 1 Mbyte using JPEG compression, without detectable impairment of image quality as viewed on an ordinary screen (Ponten and Uhlen, unpublished). Due to their smaller size, the JPEG files are suitable for web-based applications, while the original TIFF files can be saved as primary data to allow future analysis of the image contents using sophisticated image analysis algorithms.

To ensure a high quality of analysis, it is most likely necessary to use a trained pathologist for evaluation of immunoreactivity and annotation of the protein expression (45). One challenge is the lack of positive controls when analyzing patterns of immunoreactivity generated by antibodies directed toward proteins of unknown function. Basic requirements for such staining include sufficient signal-to-noise ratios and low background staining. Although there are several approaches to deal with issues concerning sensitivity and specificity, the

experience of a trained pathologist for distinguishing “true” staining patterns should not be underestimated. The inherent complexity of tissue structure often includes several different cell types in addition to the compound of biomolecules that form the matrix in which cells grow and differentiate. The composite cell populations act in concert to execute normal organ function. A tissue section from any normal organ will thus contain cells with distinct and different phenotypes, including differences in immunoreactivity for various antibodies. In certain tissues only a minority of present cells determine a specific function of a defined tissue, e.g. islet cells in pancreatic tissue or glomerulus cells in kidney tissue. Other tissues contain clandestine populations of cells with important functions that supplement main organ function, e.g. antigen-presenting Langerhan’s cells in epidermis (skin).

For a coherent assessment of protein expression profiles, the level of detail for a sufficient analysis can be debated and is at large determined by the time required for each annotation. For high-throughput tissue profiling, time for annotation is an important and limiting factor. A reasonable resolution for analysis most likely includes the major cell types that comprise a defined normal tissue, e.g. for cerebellum i) Purkinje cells, ii) cells of molecular layer, and iii) cells of granular layer and for uterus i) endometrial gland cells, ii) cytotogenic stroma cells, and iii) myometrial smooth muscle cells. Regardless of chosen strategy for annotation, a database containing high-resolution images of premium tissue sections stained with val-

FIG. 5. Images showing one example of immunohistochemically stained TMA slides from cancer tissues. The whole slide is shown in low magnification (left) with six examples of spots (center) where immunoreactive patterns is shown in corresponding tissues (right). The figure shows cancer tissues stained with PrEST antibody K230085 recognizing a fragment of the ALEX3 protein. This antibody shows strong cytoplasmic immunoreactivity mainly in hepatocellular carcinomas (a and b). Distinct positivity is also found in a subset of other tumor types such as seropapillary ovarian carcinoma (c) and malignant melanoma (f). Most other cancer types are totally negative as exemplified in malignant lymphoma (d) and malignant glioma (e).



idated antibodies will enable refined analysis at the cellular and partly subcellular level. Such images would also provide excellent templates for future automated image analysis systems.

#### RELEVANCE FOR CLINICAL PROTEOMICS

Antibody-based tissue profiling allows a streamlined approach for generating expression data both for normal and disease tissue. It is also possible to generate data on many different individual patients to evaluate heterogeneity of tissue profiles. The correlation between genotype and phenotype is still poorly understood and presents a critical challenge to functionally explore mechanisms underlying normal development and disease (46). The phenotypical appearance as seen in a tissue section is evidently based on pattern of expressed proteins within a cell as well as composition of the surrounding micro-environment. Antibody-based tissue profiling is therefore an important complement to transcript profiling for analysis of clinical material. As large amounts of data is generated through use of micro-array technologies (47), there is an increasing demand to validate and compare transcription profiles with protein expression profiles in various normal as well as diseased tissues. The tissue profiling provides a basis for validation of up- and down-regulated genes and subclassification of different cancer types. Several examples exist

where differences in transcription profiles have suggested that cancer types as defined through conventional histology includes several subtypes with differences in behavior, prognosis, and sensitivity to therapy, e.g. breast cancer (48, 49). One could therefore envision the development of specific antibody panels as tools for diagnostic and prognostic information in given types of cancer. Immunohistochemistry allows for an enhanced understanding of morphology as it adds important information based on specific immunoreactivity in defined cell populations. Today in clinical medicine, analysis of immunohistochemical profiles as well as mutation analysis and levels of transcription can add important information regarding prognosis and choice of treatment for certain forms of cancer.

It is also possible to use the antibody-based proteomics to analyze other clinically important diseases where tissue and cell functions are impaired, provided that clinical material can be obtained in sufficient amounts. Degenerative, cardiovascular, metabolic, infectious, and neurological diseases as well as monitoring organ transplantation provide examples suitable for antibody-based tissue profiling. The possibility to use archival material consisting of paraffin-embedded biosamples is advantageous not only for screening for expressed profiles but also in studies where patient data coupled to specific



TABLE II  
Estimated number of proteins in the human proteome as a result of different classifications

Class of protein	Size	Description
Nonredundant proteins	20,000–25,000	A representative protein from every gene locus
Protein variants	50,000–500,000	Different proteins obtained by splicing or proteolysis
Combinatorial variants	>10,000,000	Different proteins that are generated by somatic DNA rearrangements
Protein species	>100,000	Proteins that differ in chemical composition due to PTMs
Protein alleles	75,000–150,000	Proteins that differ by genetic variation giving amino acid differences (coding single nucleotide polymorphisms)

tissues are included, *e.g.* scrutinizing biomarkers relevant for survival, response to therapy, etc.

#### HOW LARGE IS THE HUMAN PROTEOME, ANYWAY?

A relevant question for efforts to generate a comprehensive collection of antibodies is how large the human proteome is. The estimated number of proteins, protein variants, and isoforms in the human proteome can vary considerably with numbers ranging from tens of thousands to over several million proteins. To clarify this point, we have in Table II listed different ways to classify the proteins with accompanied estimates of the approximate number of human proteins within each class. As can be seen in this list, the number of proteins in the proteome is significantly related to the definition of the protein class.

First, the nonredundant set of proteins has here been defined as a single representative protein from every gene locus. At present, the human genome contains ~22,000–23,000 genes, although the estimated number is still changing on a monthly basis ([www.ensembl.org](http://www.ensembl.org)). Thus, the nonredundant set of proteins in the human proteome is probably between 20,000 and 25,000. On the other hand, if one includes all the protein variants generated by RNA splicing or specific proteolytic processing, the number of protein isoforms rapidly increases. Splicing is a common phenomenon and frequently gives rise to different forms of the same protein, often through events linked to a targeting of the protein variants to different compartments of the cell (50). Site-specific proteolysis is particularly common in the processing of preproteins of neuropeptides, but has also been shown to be involved in the maturation of proteins, such as the cleavage of the C-peptide from the proinsulin molecule to create functional insulin and C-peptide (51). The estimated number of protein variants represented by splicing and proteolysis is still unknown, but the number might be between 50,000 and 500,000.

Another form of proteolysis is the degradation of proteins, such as the ubiquitin-mediated degradation of proteins (52). The degradation of functional proteins into amino acids yields, most likely, many millions of short-lived, intermediate products. These are, however, in most cases not functional and will be disregarded here. Another form of variation in the human proteome is the combinatorial variants created by somatic rearrangement in cells involved in the immune system. A well-known example is the immunoglobulin G (IgG)

molecule, which is created by rearrangement of the two separate genes coding for the heavy and light chains. The number of different IgG molecules in a human individual is probably more than 10 million molecules all with different complementarity-determining regions and thus different binding properties. The T cell receptor represents another protein class with similar combinatorial variation. Thus, all human individuals carry several tens of million combinatorial protein variants.

A different type of protein isoforms is the protein species represented by posttranslation modifications (PTM), such as glycosylation, acetylation, phosphorylations, etc. (3, 4). Glycosylation is common for the fraction of the proteome, which is secreted to the endoplasmic reticulum. The phosphorylation of proteins is a biologically important process and in many cases determines if a protein is active or not. The number of protein species represented by different PTMs is not easy to estimate, partly because glycosylation is a stepwise process and it is a matter of definition if also the intermediate forms of the proteins should be included in the size estimate. Finally, genetic differences between different individuals give rise to different forms of the protein. Recent studies have estimated that each gene has in average four coding single nucleotide polymorphisms defined as genetic variation that give rise to amino acid differences in more than 1% of the population (53). If the estimate of the number of genes is close to 25,000, the estimated number of variants, here called protein alleles, should thus be around 100,000.

In summary, the structural space of the proteome including combinatorial variants and various protein isoforms is many tens of millions molecules. For the study of the proteome, it could be argued that the primary objective of an antibody-based proteomics effort should be to generate a uniform set of antibodies for the nonredundant set of proteins, which is probably between 20,000 and 25,000 molecules. If these antibodies are generated to parts of the protein shared by the different isoforms, they could be used to subsequently study the various isoforms using antibody-specific “pull-downs” followed by biochemical studies, in particular with MS. These antibodies should enable all isoforms to be affinity purified for further analysis with a single antibody reagent.

#### CONCLUSIONS AND PERSPECTIVES

The possibility to map and understand the proteome of different organisms has increased greatly with the sequencing



of a large number of genomes, including the human. The knowledge of the genetic code offers new opportunities to explore function and communication within the proteome. In this review, we have focused on an antibody-based proteomics strategy that can be used to systematically explore clinical samples using a standardized set of tissue microarrays. The approach appears possible to scale up involving tissue profiles representing tens of thousands of antibodies. If specific antibodies on a proteomic scale can be generated within the framework of various international efforts, a comprehensive protein atlas for a large part of the human proteome is within reach.

*Acknowledgments*—We would like to thank Per-Åke Nygren, Sophia Hober, and Anna Asplund for critical comments and suggestions and the Knut and Alice Wallenberg Foundation for funding the Human Proteome Resource (www.hpr.se) program.

§ To whom correspondence should be addressed: Department of Biotechnology, AlbaNova University Center, Royal Institute of Technology, 10691 Stockholm, Sweden. E-mail: mathias.uhlen@biotech.kth.se.

REFERENCES

1. Agaton, C., Uhlen, M., and Hober, S. (2004) Genome-based proteomics. *Electrophoresis* **25**, 1280–1288
2. Hanash, S. (2004) HUPO initiatives relevant to clinical proteomics. *Mol. Cell. Proteomics* **3**, 298–301
3. Tyers, M., and Mann, M. (2003) From genomics to proteomics. *Nature* **422**, 193–197
4. Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M., and Fields, S. (2003) Protein analysis on a proteomic scale. *Nature* **422**, 208–215
5. Espina, V., Woodhouse, E. C., Wulfskuhle, J., Asmussen, H. D., Petricoin, E. F., 3rd, and Liotta, L. A. (2004) Protein microarray detection strategies: Focus on direct detection technologies. *J. Immunol. Methods* **290**, 121–133
6. Miroux, B., and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298
7. Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., and Hard, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* **11**, 313–321
8. Lise, L. D., and Audibert, F. (1989) Immunoadjuvants and analogs of immunomodulatory bacterial structures. *Curr. Opin. Immunol.* **2**, 269–274
9. Larsson, M., Graslund, S., Yuan, L., Brundell, E., Uhlen, M., Hoog, C., and Stahl, S. (2000) High-throughput protein expression of cDNA products as a tool in functional genomics. *J. Biotechnol.* **80**, 143–157
10. Meloan, R. H., Puijk, W. C., Langeveld, J. P., Langedijk, J. P., and Timmerman, P. (2003) Design of synthetic peptides for diagnostics. *Curr. Protein Pept. Sci.* **4**, 253–260
11. Agaton, C., Galli, J., Hoiden Guthenberg, I., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., Wurtz, D., Hoog, C., Lundeberg, J., Stahl, S., Ponten, F., and Uhlen, M. (2003) Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol. Cell. Proteomics* **2**, 405–414
12. Uhlen, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S., and Philipson, L. (1983) Gene fusion vectors based on the gene for staphylococcal protein A. *Gene* **23**, 369–378
13. Terpe, K. (2003) Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**, 523–533
14. Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598–599
15. Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R. A., Nordheim, A., andStunnenberg, H. G. (1991) Rapid and efficient purification of native

- histidine-tagged protein expressed by recombinant vaccinia virus. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 8972–8976
16. Lindskog, M., Rockberg, J., Uhlen, M., and Sterky, F. (2005) Selection of protein epitopes for antibody production. *BioTechniques*, in press.
17. Kohler, G., and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497
18. Milstein, C. (2000) With the benefit of hindsight. *Immunol. Today* **21**, 359–364
19. Spiekier-Polet, H., Sethupathi, P., Yam, P. C., and Knight, K. L. (1995) Rabbit monoclonal antibodies: Generating a fusion partner to produce rabbit-rabbit hybridomas. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 9348–9352
20. Lipovsek, D., and Pluckthun, A. (2004) In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290**, 51–67
21. Bradbury, A. R., and Marks, J. D. (2004) Antibodies from phage antibody libraries. *J. Immunol. Methods* **290**, 29–49
22. Agaton, C., Falk, R., Hoiden Guthenberg, I., Gostring, L., Uhlen, M., and Hober, S. (2004) Selective enrichment of monospecific polyclonal antibodies for antibody-based proteomics efforts. *J. Chromatogr. A* **1043**, 33–40
23. Borrebaeck, C. A. (2000) Antibodies in diagnostics—From immunoassays to protein chips. *Immunol. Today* **21**, 379–382
24. Liu, B., and Marks, J. D. (2000) Applying phage antibodies to proteomics: Selecting single chain Fv antibodies to antigens blotted on nitrocellulose. *Anal. Biochem.* **286**, 119–128
25. Nygren, P. A., and Skerra, A. (2004) Binding proteins from alternative scaffolds. *J. Immunol. Methods* **290**, 3–28
26. Brody, E. N., and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J. Biotechnol.* **74**, 5–13
27. Liu, B., Huang, L., Burlingame, A., and Marks, J. D. (2002) Towards proteome-wide production of monoclonal antibody by phage display. *J. Mol. Biol.* **1**, 1063–1073
28. Spicer, S. S. (1993) Advantages of histochemistry for the study of cell biology. *Histochem. J.* **25**, 531–547
29. Gown, A. (2002) Genogenic immunohistochemistry: A new era in diagnostic immunohistochemistry. *Curr. Diagn. Pathol.* **8**, 193–200
30. Battifora, H. (1986) The multitumor (sausage) tissue block: Novel method for immunohistochemical antibody testing. *Lab. Invest.* **55**, 244–248
31. Wan, W. H., Fortuna, M. B., and Furmanski, P. (1987) A rapid and efficient method for testing immunohistochemical reactivity of monoclonal antibodies against multiple tissue samples simultaneously. *J. Immunol. Methods* **103**, 121–129
32. Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G., and Kallioniemi, O. P. (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* **4**, 844–847
33. Hewitt, S. M. (2004) Design, construction, and use of tissue microarrays. *Methods Mol. Biol.* **264**, 61–72
34. Braunschweig, T., Chung, J. Y., and Hewitt, S. M. (2004) Perspectives in tissue microarrays. *Comb. Chem. High Throughput Screen* **7**, 575–585
35. Rimm, D. L., Camp, R. L., Charette, L. A., Costa, J., Olsen, D. A., and Reiss, M. (2001) Tissue microarray: A new technology for amplification of tissue resources. *Cancer J.* **7**, 24–31
36. Kallioniemi, O. P., Wagner, U., Kononen, J., and Sauter, G. (2001) Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum. Mol. Genet.* **10**, 657–662
37. Warford, A., Howat, W., and McCafferty, J. (2004) Expression profiling by high-throughput immunohistochemistry. *J. Immunol. Methods* **290**, 81–92
38. Kampf, C., Andersson, A. C., Wester, K., Bjorling, E., Uhlen, M., and Ponten, F. (2005) Antibody-based tissue profiling as a tool for clinical proteomics. *Clin. Proteomics*, in press.
39. Kurochkin, I. V., Yonemitsu, N., Funahashi, S. I., and Nomura, H. (2001) ALEX1, a novel human armadillo repeat protein that is expressed differentially in normal tissues and carcinomas. *Biochem. Biophys. Res. Commun.* **280**, 340–347
40. Hoos, A., and Cordon-Cardo, C. (2001) Tissue microarray profiling of cancer specimens and cell lines: Opportunities and limitations. *Lab. Invest.* **81**, 1331–1338
41. Wester, K., Andersson, A. C., Ranefall, P., Bengtsson, E., Malmstrom, P. U., and Busch, C. (2000) Cultured human fibroblasts in agarose gel as a multi-functional control for immunohistochemistry. Standardization of

- Ki67 (MIB1) assessment in routinely processed urinary bladder carcinoma tissue. *J. Pathol.* **190**, 503–511
42. Amos, W. B., and White, J. G. (2003) How the confocal laser scanning microscope entered biological research. *Biol. Cell* **95**, 335–342
43. Vrolijk, H., Sloos, W., Mesker, W., Franken, P., Fodde, R., Morreau, H., and Tanke, H. (2003) Automated acquisition of stained tissue microarrays for high-throughput evaluation of molecular targets. *J. Mol. Diagn.* **5**, 160–167
44. Marengo, E., Robotti, E., Antonucci, F., Cecconi, D., Camprostrini, N., and Righetti, P. G. (2005) Numerical approaches for quantitative analysis of two-dimensional maps: A review of commercial software and home-made systems. *Proteomics* **5**, 654–666
45. Becich, M. J. (2000) The role of the pathologist as tissue refiner and data miner: The impact of functional genomics on the modern pathology laboratory and the critical roles of pathology informatics and bioinformatics. *Mol. Diagn.* **5**, 287–299
46. Celis, J. E., Moreira, J. M., Gromova, I., Cabezon, T., Ralfkiaer, U., Guldborg, P., Straten, P. T., Mouridsen, H., Friis, E., Holm, D., Rank, F., and Gromov, P. (2005) Towards discovery-driven translational research in breast cancer. *FEBS J.* **272**, 2–15
47. Quackenbush, J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.* **2**, 418–427
48. Carr, K. M., Rosenblatt, K., Petricoin, E. F., and Liotta, L. A. (2004) Genomic and proteomic approaches for studying human cancer: Prospects for true patient-tailored therapy. *Hum. Genomics* **1**, 134–140
49. Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L., and Liu, E. T. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10393–10398
50. Garcia-Blanco, M. A., Baraniak, A. P., and Lasda, E. L. (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.* **22**, 535–546
51. Rigler, R., Pramanik, A., Jonasson, P., Kratz, G., Jansson, O. T., Nygren, P., Stahl, S., Ekberg, K., Johansson, B., Uhlen, S., Uhlen, M., Jornvall, H., and Wahren, J. (1999) Specific binding of proinsulin C-peptide to human cell membranes. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13318–13323
52. Hamel, F. G., Fawcett, J., Bennett, R. G., and Duckworth, W. C. (2004) Control of proteolysis: Hormones, nutrients, and the changing role of the proteasome. *Curr. Opin. Clin. Nutr. Metab. Care* **7**, 255–258
53. Fakhrai-Rad, H., Zheng, J., Willis, T. D., Wong, K., Suyenaga, K., Moorhead, M., Eberle, J., Thorstenson, Y. R., Jones, T., Davis, R. W., Namsaraev, E., and Faham, M. (2004) SNP discovery in pooled samples with mismatch repair detection. *Genome Res.* **14**, 1404–1412