

A Gene-centric Human Protein Atlas for Expression Profiles Based on Antibodies*

Lisa Berglund‡§, Erik Björling‡§, Per Oksvold‡, Linn Fagerberg‡, Anna Asplund¶, Cristina Al-Khalili Szigarto‡, Anja Persson‡, Jenny Ottosson‡, Henrik Wernérus‡, Peter Nilsson‡, Emma Lundberg‡, Åsa Sivertsson‡, Sanjay Navani||, Kenneth Wester¶, Caroline Kampff¶, Sophia Hober‡, Fredrik Pontén¶, and Mathias Uhlen‡**

An attractive path forward in proteomics is to experimentally annotate the human protein complement of the genome in a gene-centric manner. Using antibodies, it might be possible to design protein-specific probes for a representative protein from every protein-coding gene and to subsequently use the antibodies for systematic analysis of cellular distribution and subcellular localization of proteins in normal and disease tissues. A new version (4.0) of the Human Protein Atlas has been developed in a gene-centric manner with the inclusion of all human genes and splice variants predicted from genome efforts together with a visualization of each protein with characteristics such as predicted membrane regions, signal peptide, and protein domains and new plots showing the uniqueness (sequence similarity) of every fraction of each protein toward all other human proteins. The new version is based on tissue profiles generated from 6120 antibodies with more than five million immunohistochemistry-based images covering 5067 human genes, corresponding to ~25% of the human genome. Version 4.0 includes a putative list of members in various protein classes, both functional classes, such as kinases, transcription factors, G-protein-coupled receptors, etc., and project-related classes, such as candidate genes for cancer or cardiovascular diseases. The exact antigen sequence for the internally generated antibodies has also been released together with a visualization of the application-specific validation performed for each antibody, including a protein array assay, Western blot analysis, immunohistochemistry, and, for a large fraction, immunofluorescence-based confocal microscopy. New search functionalities have been added to allow complex queries regarding protein expression profiles, protein classes, and chromosome location. The new version of the protein atlas thus is a resource for many areas of biomedical research, including protein science and biomarker discovery. *Molecular & Cellular Proteomics* 7:2019–2027, 2008.

From the ‡Department of Proteomics, School of Biotechnology, AlbaNova University Center, Royal Institute of Technology (KTH), SE-106 91 Stockholm, Sweden, ¶Rudbeck Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden, and ||Lab Surgpath, 204 Bombay Market, 400 034 Mumbai, India

Received, July 30, 2008

Published, MCP Papers in Press, July 31, 2008, DOI 10.1074/mcp.R800013-MCP200

One of the largest challenges in the postgenome era is to explore and understand the instructions embedded in the human genome. The sequencing of the human genome has revealed ~20,500 protein-encoding genes (1), and it is now possible to envision a gene-centric proteomics approach to map the protein-based molecular architecture of the human body (2). The generation of “probes” in the form of protein-specific antibodies (3) allows for studies of the corresponding proteins and protein isoforms using a wide range of assays, including analysis of tissue and cell protein profiles.

The major repository for protein information is the Universal Protein Resource (UniProt) (4), a collaborative effort between the Swiss Institute of Bioinformatics, the European Bioinformatics Institute, and the Protein Information Resource. An alternative portal was recently described, the Human Proteinpedia (5), in which researchers themselves can submit annotated protein data from a large number of technology platforms, such as mass spectrometry, Western blots, and immunohistochemistry, with accompanying experimental evidence. In the current version of UniProtKB (release 14), there are 20,069 reviewed human protein entries. Only approximately half (11,371) of these entries have evidence at the protein level. This emphasizes the need for protein-specific antibodies to extend the number of verified proteins with the aim to create a complete list of all human proteins forming a basis for protein-related research in human biology and medicine.

Here we report on a new version (4.0) of the publicly available Human Protein Atlas with new features and a new structure based on a gene-centric and genome-complete view of the protein complement of the human genome. The protein atlas portal is the result of a large scale effort to create a resource for human protein expression profiles in normal tissues, cancer cells, and cell lines with validated antibodies as reagents. The results show a path forward to create a comprehensive resource covering all human proteins.

THE PROGRESS OF THE HUMAN PROTEIN ATLAS

The first version (1.0) of the Human Protein Atlas portal was released in August 2005 and contained 275 internally generated antibodies and 443 external antibodies obtained from various commercial antibody vendors (Table I). The antibodies

TABLE I
The progress of the Human Protein Atlas

Version	Year	No. of genes	No. of external Abs ^a	No. of internal Abs	Total no. of Abs	New features
1.0	2005	650	443	275	718	Protein profiles for 48 normal tissues (IHC); protein profiles for 20 cancer types (IHC)
2.0	2006	1344	759	755	1514	Protein profiles for 59 cell lines and cells (IHC)
3.0	2007	2600	1241	1774	3015	Subcellular localization in three cell lines (IF) ^b ; validation results (Western blot/protein array); search function (tissues and cancers)
4.0	2008	5067	2251	3869	6120	Gene-centric structure with all predicted genes; advanced searches (protein classes and cells); visualization of antigen

^a Abs, antibodies.

^b Immunofluorescence.

were used to stain tissue microarrays (6) with a comprehensive set of normal and cancer tissues (7). For each antibody, 576 high resolution images representing different tissues and individuals were generated and subsequently annotated and curated by certified pathologists (8). Tissue profiles corresponding to 650 human genes were released as part of version 1.0, and more than 400,000 high resolution images were available through the database portal. A year later, the next version of the atlas was released, more than doubling the number of antibodies to 1514, corresponding to 1344 genes (Table I). In addition, 59 human cells and cell lines were stained in duplicates for each antibody, and the 118 images were analyzed using automated image analysis software (TMAx) (9). The number of immunohistochemistry (IHC)¹ images available in this release was ~1 million. The third version of the atlas was released in 2007, and again the number of antibodies and genes doubled as compared with the previous version (Table I). Altogether 1774 “in-house” generated monospecific antibodies and 1241 external antibodies were analyzed in this manner generating more than two million annotated images, all publicly available on the Human Protein Atlas portal. A new search function was released allowing relatively complex queries involving tissue profiles in normal and cancer tissues (10). Some of the antibodies were also used for analysis of subcellular localization of the target protein in three human cell lines using fluorescence microscopy (11). All confocal images were generated with three additional probes to stain the nucleus, the endoplasmic reticulum, and the cytoskeleton, and images were generated to allow visualization of these four probes simultaneously or separately. The new version 4.0 of the protein atlas again doubles the content of antibodies to 6120 (Table I) with more than five million images and with the majority (63%) of the antibodies obtained through the internal effort. Altogether 5067 genes have been analyzed, corresponding to ~25% of all the predicted protein-coding genes of the human genome (1).

¹ The abbreviations used are: IHC, immunohistochemistry; HPA, Human Protein Atlas; CAB, collaborator antibody; GI, gastrointestinal; ID, identifier.

A GENE-CENTRIC PROTEIN ATLAS

The first three versions of the protein atlas were “antibody-centric” with a database structure based on the results from each antibody. The new version 4.0 has been changed to a “gene-centric” structure in which all genes predicted by Ensembl (12) are included as part of the database. Altogether the 21,528 predicted protein-coding genes of Ensembl release 50.36 are contained in the database at present; this is slightly higher than the 20,500 genes predicted by Clamp *et al.* (1). A search in the portal will result in a list of protein-coding genes, and this is illustrated in Fig. 1 in which a partial list of all the genes on chromosome X is displayed. The *Antibody ID column* shows whether an antibody has been used to generate tissue profiles as is the case for most of the genes shown here. However, the protein product of the first gene (*ATP1B4*) and two other genes further down (*ATXN3L* and *AVPR2*) do not yet have approved antibodies, and consequently there are no tissue profiles in the database at present as shown by the lack of an Antibody ID and no data in the *Validation column*. Other genes have several antibodies to the same protein target as exemplified by the second gene (*ATP2B3*), which has both an internally generated antibody (HPA) and an external commercial antibody (CAB), and the last gene (*BCAP51*), which has three antibodies directed toward its protein product, one internally generated (HPA) and two external antibodies (CAB), in this case from two different antibody vendors. Consequently three different antibodies from three different origins are available to this gene product, and this allows for comparative studies and better validation of the staining results.

VALIDATION OF INTERNALLY GENERATED ANTIBODIES

The specificity of the internally generated antibodies was validated by protein microarrays (8, 13) in which 384 human recombinant protein fragments were spotted on a microarray, and the antibody specificity was determined using a fluorescence-based analysis. The antibodies were further validated by Western blotting using tissue extracts from liver and tonsil, pooled human plasma depleted of IgG and albumin, and cell extracts from two human cell lines (8). In addition, an IHC

#	Gene name	Description	Chr	Links	Class	Antibody ID	Validation
110	ATP1B4	Xipotassium-transporting ATPase subunit beta-m (X,K-ATPase subunit beta-m).	X:q24				<input type="checkbox"/>
111	ATP2B3	Plasma membrane calcium-transporting ATPase 3 (EC 3.6.3.8) (PMCA3) (Plasma membrane calcium pump isoform 3) (Plasma membrane calcium ATPase isoform 3).	X:q28			HPA001583	<input checked="" type="checkbox"/> PA <input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH <input checked="" type="checkbox"/> IF
						CAB005607	<input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH
112	ATP6AP1	Vacuolar ATP synthase subunit S1 precursor (EC 3.6.3.14) (V-ATPase subunit S1) (V-ATPase S1 accessory protein) (V-ATPase Ac45 subunit) (Protein XAP-3).	X:q28			CAB015218	<input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH
113	ATP6AP2	Renin receptor precursor (Renin/prorenin receptor) (ATPase H(+)-transporting lysosomal accessory protein 2) (ATPase H(+)-transporting lysosomal-interacting protein 2) (Vacuolar ATP synthase membrane sector-associated protein M8-9) (V-ATPase M8.9 subunit)	X:p11.4			HPA003156	<input checked="" type="checkbox"/> PA <input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH <input checked="" type="checkbox"/> IF
114	ATP7A	Copper-transporting ATPase 1 (EC 3.6.3.4) (Copper pump 1) (Menkes disease-associated protein).	X:q21.1			HPA012887	<input checked="" type="checkbox"/> PA <input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH
115	ATRX	Transcriptional regulator ATRX (EC 3.6.1.-) (ATP-dependent helicase ATRX) (X-linked helicase II) (X-linked nuclear protein) (XNP) (Znf-HX).	X:q21.1			HPA001906	<input checked="" type="checkbox"/> PA <input checked="" type="checkbox"/> IH
						CAB009372	<input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH
116	ATXN3L	Putative ataxin-3-like protein (EC 3.4.22.-) (Machado-Joseph disease protein 1-like).	X:p22.2				<input type="checkbox"/>
117	AVPR2	Vasopressin V2 receptor (Renal-type arginine vasopressin receptor) (Antidiuretic hormone receptor) (AVPR V2).	X:q28				<input type="checkbox"/>
118	BCAP31	B-cell receptor-associated protein 31 (BCR-associated protein Bap31) (p28 Bap31) (Protein CDM) (BC6-AG tumor-associated antigen).	X:q28			HPA003906	<input checked="" type="checkbox"/> PA <input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH <input checked="" type="checkbox"/> IF
						CAB015350	<input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH
						CAB015424	<input checked="" type="checkbox"/> WB <input checked="" type="checkbox"/> IH

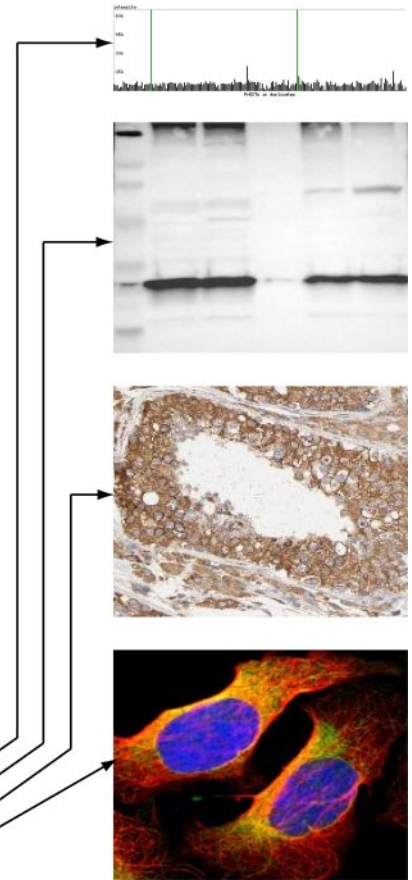


FIG. 1. Gene list resulting from a search on chromosome X in the Human Protein Atlas. 884 protein-coding genes for chromosome X are reported of which nine examples of gene entries are displayed here. Gene name, a short description of the gene, and the chromosomal localization of the gene (*Chr*) are given together with links to external databases containing information about this gene. Abbreviations of protein classes to which the gene belongs are displayed under “Class” and link to a protein class page. If the atlas contains expression profiles for proteins encoded by the gene, the antibody ID is shown and will link to the expression profile and antibody information pages. The results for different assays (protein array (PA), Western blotting (WB), immunohistochemistry (IH), and immunofluorescence (IF)) are color-coded (green, supportive; yellow, uncertain; red, non-supportive).

validation was performed using a tissue microarray containing 13 normal tissue types, four tumor types, and eight cell lines in duplicate or triplicate with 42 microarray cores altogether. The results of the Western blots and the immunohistochemistry were annotated based on known literature and bioinformatics data (presence of signal peptide, transmembrane regions, or other localization signals).

The results of the validation of more than 9000 internally generated antibodies are shown in Table II. Approximately half (49%) of the antibodies fail because of a staining pattern that is not consistent with literature or bioinformatics data. Antibodies with supportive validation constituted 22% of the reagents, and of these 7% were paired antibodies with similar staining pattern, consistent with experimental and/or bioinformatics data (score 1). 29% of the antibodies were categorized as uncertain with no literature data available or with staining pattern only partly consistent with earlier experimental data and/or bioinformatics. The

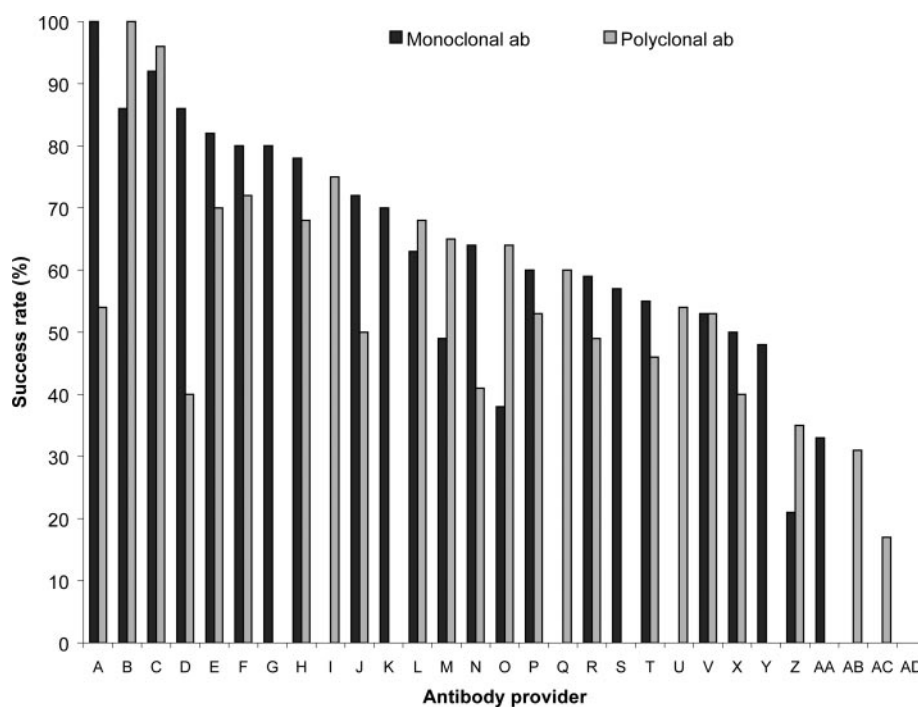
results show that approximately half (51%) of the internally generated antibodies are scored 1–4, and these are subsequently made public through the Human Protein Atlas.

It is noteworthy that the validation of the antibodies (immunohistochemistry and Western blotting) depends on subjective decisions based on a comparison of experimental results with information obtained via bioinformatics prediction methods and literature, which in many cases are inconclusive. As an example, the lack of developmental tissues in the microarray setup excludes the validation of antibodies to protein targets that are only expressed during developmental stages. Observed discordant IHC patterns can be explained by the presence of protein isoforms, such as splice variants or post-translationally modified proteins, but cross-reactivity to other proteins cannot be excluded. The availability of validation data on a public database thus facilitates input from the scientific community regarding the possible cross-reactivity

TABLE II
Validation of 9358 internally generated antibodies

Validation score	Validation category	Criteria	Number of antibodies	Fraction
1. High	Supportive	Two independent antibodies with similar staining pattern consistent with experimental and/or bioinformatics data	687	7
2. Medium	Supportive	Staining pattern consistent with experimental and/or bioinformatics data	1414	15
3. Low	Uncertain	Staining pattern partly consistent with experimental and/or bioinformatics data	1853	20
4. Very low	Uncertain	No literature or bioinformatics data available	861	9
5. Failed	Failed	Staining pattern not consistent with experimental and/or bioinformatics data	4543	49

FIG. 2. Success rates in validation of antibodies from external providers. 5399 antibodies from external providers were validated by the standardized approach using Western blotting and tissue microarrays. The figure displays the fraction of approved antibodies (grouped by monoclonal or polyclonal antibodies) per external provider (2374 monoclonal antibodies and 3025 polyclonal antibodies from 29 antibody providers). Only external providers that supplied at least five monoclonal/polyclonal antibodies are represented in the figure. *ab*, antibody.



and specificity of each antibody (14). In this context, we have decided to make the supporting validation results available with the supporting images available for review (Fig. 1). Our aim is to encourage open access to the validation results to facilitate international efforts to compare antibodies generated by different programs.

VALIDATION OF EXTERNAL ANTIBODIES

A call for antibodies toward human proteins was sent to a large number of commercial vendors. 5436 external antibodies from 51 different antibody providers were obtained and subsequently validated. Of these, 1410 monoclonal antibodies and 1316 polyclonal antibodies were approved by a standardized validation using Western blotting and IHC on tissue microarrays as described above. The success rates stratified by the different providers (Fig. 2) showed large differences, ranging from 0 to 100% of the antibodies with an average success rate of 49%. It

is important to point out that many of these antibodies have not been approved by the antibody providers for immunohistochemistry; this might explain the low success rate in our hands. In addition, some of the antibodies might give supportive data if a more specialized assay was designed for each antibody making sure that the target protein is present in the assay. Anyhow the results enforce the need for publicly available validation results for antibodies because half of the commercially available antibodies did not pass our quality assurance.

INFORMATION ABOUT THE ANTIGEN

We have earlier argued (13) that a long term objective for a human antibody initiative could be to generate two independent antibodies toward all non-redundant proteins, allowing the results and validation of one antibody to be used to validate the other. The objective to generate paired antibodies with non-overlapping epitopes opens up the need to

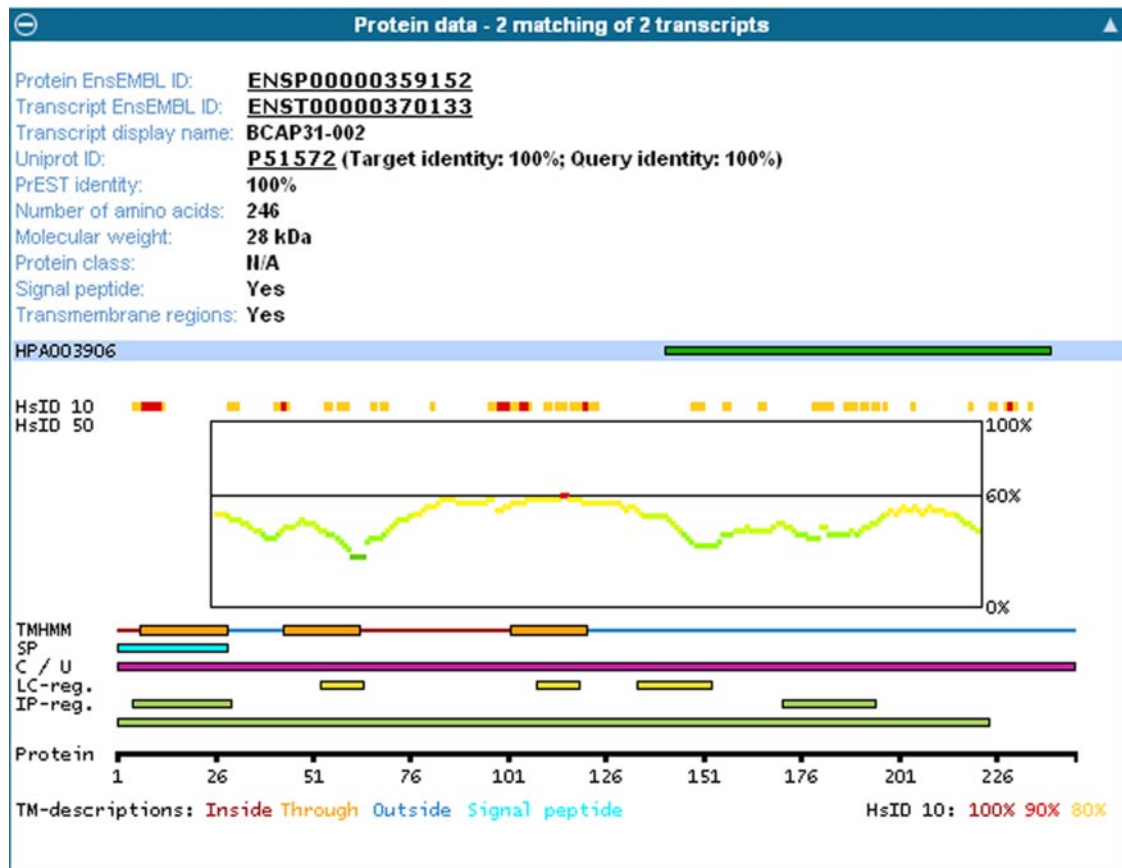


FIG. 3. **Protein information in the Human Protein Atlas.** The gene/protein information page for a gene displays gene and protein information from different external sources as well as from in-house-generated data. The protein view displayed here visualizes protein features such as sequence identity of different fractions of the protein to other human proteins (HsID), predicted transmembrane regions (TMHMM), predicted signal peptide (SP), regions common to all splice variants of the gene or exclusive to this particular splice variant (C/U), low complexity regions (LC-reg.), and protein domains (IP-reg.). The position of the antigen HPA003906 used for generation of the antibody is shown in the *upper right corner* of the protein view.

gain precise knowledge about the binding region of each antibody either by epitope mapping of each antibody, such as the use of overlapping synthetic peptides, or if a synthetic peptide or a recombinant fragment has been used to gain access to the sequence used to generate the antibody. All the internally generated antibodies have been produced via a protocol for polyclonal (monospecific) antibodies immunized and affinity-purified using a recombinant protein epitope signature tag (PrEST) as antigen (15). With the new version of the Human Protein Atlas (4.0), all antigen sequences for the internally generated antibodies (63% of all the antibodies) are disclosed as part of the antibody information page to facilitate such development and to aid in international efforts to generate paired antibodies. The position of the antigen in the target protein is given in a new protein view (16) (Fig. 3) together with a visualization of the various characteristics of a particular protein, such as plots showing the sequence identity of different fractions of the protein to other proteins (17, 18), predicted transmembrane regions (19), predicted signal peptide (20), InterPro regions (21), etc.

PROTEIN CLASSES

The new version of the atlas portal also includes the possibility to stratify the proteome according to various protein classes. This can be done in three different ways: (i) functional classes, such as kinases, transcription factors, or G-protein-coupled receptors, (ii) a list of interesting proteins, such as plasma proteins, candidate markers of cardiovascular disease, or candidate cancer markers, and/or (iii) chromosomal location of the corresponding gene. Some examples of protein classes included in the protein atlas are shown in Table III with information about the number of genes in each class. The table also shows the fraction of proteins with tissue profiles (validated antibodies) for each protein class ranging from 26% for the transcription factors to 65% for the candidate markers of cardiovascular disease (Table III).

ADVANCED SEARCHES

The search function described by Björling *et al.* (10) has been extended to allow complex queries, including combined searches based on protein classes, chromosomal location,

TABLE III
Protein classes

The number of genes in each protein class is given by "Predicted genes" (as determined by Ensembl version 50.36 (12)). "Tissue profiles" gives the number of genes for each class analyzed by antibodies; "Fraction" shows the percentage of analyzed proteins. "Source" is the source from which the original gene or protein identifiers for each class have been retrieved. TCDB, Transporter Classification Database; DBD, DNA-binding domain.

Protein class	Predicted genes	Tissue profiles	Fraction	Source
			%	
Kinases	503	273	54	UniProt (4)
Transporters	472	162	34	TCDB (30)
Transcription factors	1492	391	26	DBD (31)
Enzymes	2492	980	39	ENZYME (32)
G-protein-coupled receptors ^a	392	119	30	UniProt (4)
Plasma proteins	687	384	56	Plasma Proteome Inst. (33)
Candidate markers of cardiovascular disease	166	108	65	Plasma Proteome Inst. (34)
Candidate cancer biomarkers	1067	668	63	Plasma Proteome Inst. (22)
Chromosome 22	504	176	35	Ensembl (12)
Chromosome X	884	321	36	Ensembl (12)

^a Excluding olfactory and gustatory receptors.

and/or tissue profiles. It is also possible to include searches based on the expression in the cell lines in addition to tissue profiles in the normal and/or cancer tissues. In Fig. 4, two examples of combined searches are presented to illustrate the new search function. The first example (Fig. 4A) is aimed to identify proteins from a certain protein class encoded from a particular chromosome and expressed in a defined cell type. A search was made for transcription factors located on the human chromosome X with high expression in ovary stromal cells, and four proteins fit these criteria. The first protein is the erythroblast transformation specific domain-containing protein Elk-1, a transcription factor part of the ternary complex factor subfamily. This protein is the nuclear target for the Ras-Raf-MAPK (mitogen-activated protein kinase) signaling cascade, and the atlas summary shows strong expression in ovary and placenta but also in parts of the brain (cerebellum). The second protein is the methyl-CpG-binding protein 2 (MeCP2), which is a nuclear protein that binds to methylated DNA. This protein can suppress transcription by binding to CpG-containing promoters, and the protein is essential for embryonic development. The gene is subject to X chromosome inactivation and is the major cause of female mental retardation. The tissue profiling summary shows high expression in almost all tissues with an expected nuclear localization confirmed by the confocal images. The third protein is the multifunctional transcription factor YY2 (yin and yang 2), suggested to antagonize YY1 (also known as δ transcription factor NF-E1), which shows high sequence similarity to YY2. The protein YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins, and it is involved in repressing and activating a diverse number of promoters. The tissue profiles of the atlas show an expected nuclear localization and expression in most cells except the brain. Finally the query found zinc finger protein 75, which is also known as zinc finger protein 82. The function of

this protein is unknown. The tissue profile summary shows expression in all cells. Again the immunohistochemistry shows a nuclear staining, and this is supported by the confocal microscopy images showing staining of nuclei in all three cell lines.

The second query is aimed to identify proteins overexpressed in cancer as compared with its normal cellular counterpart. Based on the list of candidate cancer proteins identified by Polanski and Anderson (22), a search was made to investigate whether any proteins on this list were expressed at high levels in at least six of 12 colorectal cancer patients, whereas the expression in normal glandular cells in the colon is low (Fig. 4B). Again four proteins fit these criteria. The first protein is the serine/threonine-protein kinase 12 (Aurora-B), also known as Aurora-related kinase 2. The Aurora kinases associate with microtubules during chromosome movement and segregation, and this protein localizes to microtubules near kinetochores, specifically to the specialized microtubules called K-fibers. The tissue profiles of the Human Protein Atlas show strong expression in six of 12 colon cancers, and the expression is, as expected, in the nuclei. The protein is also expressed highly in a subset of lung cancer and cervical cancer patients. The second protein is the G₂/mitotic-specific cyclin-B1. This protein is a regulatory protein involved in mitosis, and the gene product complexes with p34 (Cdc2) to form the maturation-promoting factor. The tissue profiles show low expression in the normal colon (and the GI tract), whereas six of 12 colon cancer patients express this protein abundantly. The protein is expressed highly in a few breast cancers and malignant lymphomas. The third protein is the macrophage migration-inhibitory factor, also known as the glycosylation-inhibiting factor, which is a lymphokine involved in cell-mediated immunity, immunoregulation, and inflammation. It plays a role in the regulation of macrophage function in host defense through the suppression of anti-

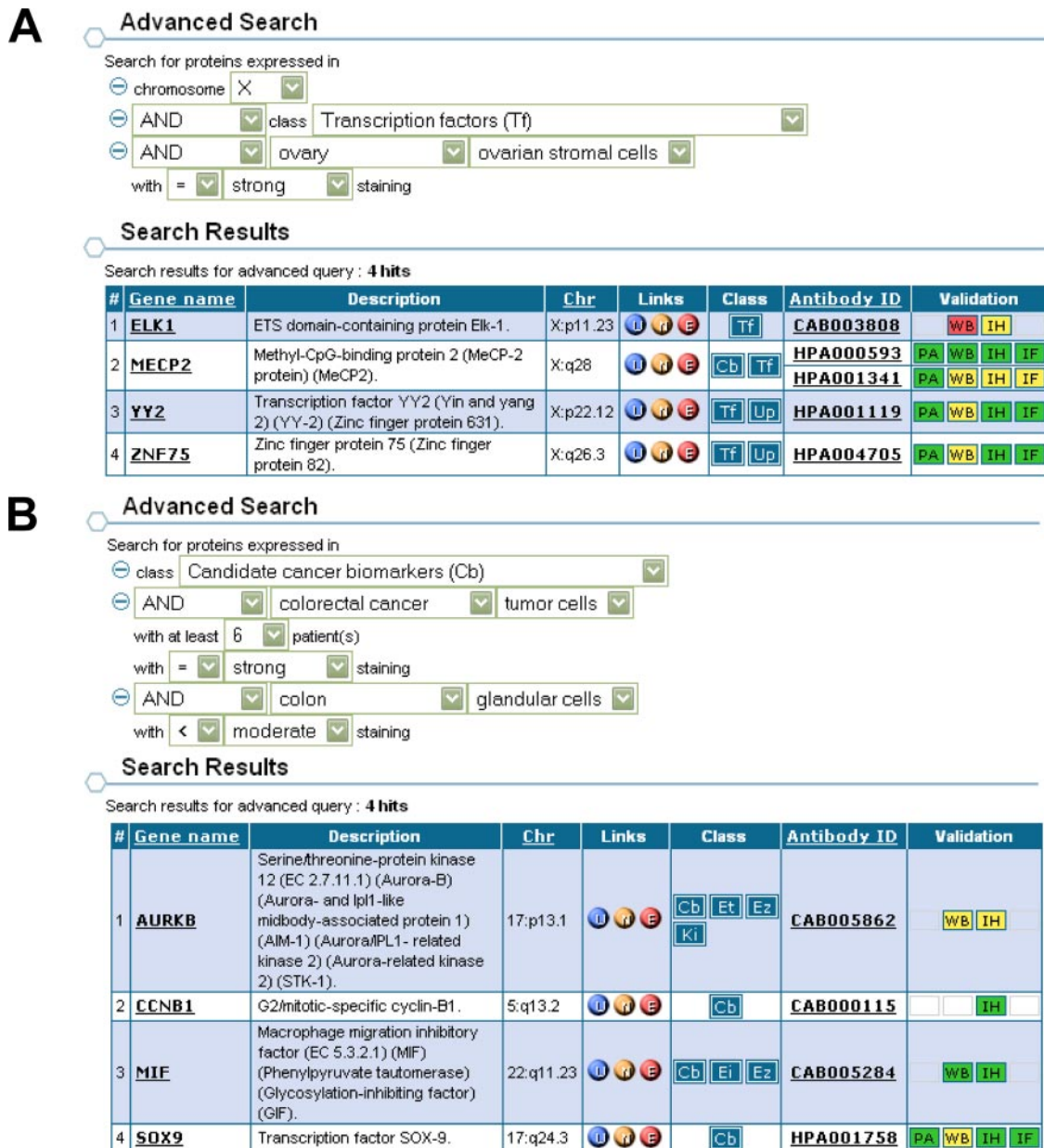


FIG. 4. Lists of genes resulting from advanced searches. The new advanced search allows for complex queries involving expression profiles, protein classes, and chromosomal localization of genes (*Chr*). A, a search for genes coding for transcription factors on chromosome X with strong staining in ovarian stromal cells results in a list of four candidate genes. B, a search for high levels of expression of candidate cancer biomarkers in colorectal cancer with low levels in normal colon results in a list of four candidate genes. PA, protein array; WB, Western blotting; IH, immunohistochemistry; IF, immunofluorescence.

inflammatory effects of glucocorticoids. This protein is expressed in low amounts in normal colon and the GI tract, whereas six colon cancer patients show high expression. Finally the search identified the human transcription factor SOX-9. This protein recognizes the sequence CCTTGAG along with other members of the high mobility group box class DNA-binding proteins. It acts during chondrocyte differentiation and regulates transcription of the anti-Muellerian hormone gene. Deficiencies result in the skeletal malformation syndrome camptomelic dysplasia. The tissue

profiles show a distinct nuclear staining, which is confirmed by the analysis using confocal microscopy. Eight of 12 colon cancers express high amounts of this protein in contrast to the normal GI tract. SOX-9 is also highly expressed in other cancers, in particular skin cancer.

CONCLUDING REMARKS

Here we describe a first attempt to experimentally annotate the human protein complement of the genome in a gene-centric manner. A new version (4.0) of the Human Protein Atlas

has been developed with the inclusion of all human genes and splice variants from the Ensembl (12) database. The new version is based on tissue profiles generated from 6120 antibodies with more than five million immunohistochemistry-based images covering 5067 human genes.

An important objective of a gene-centric antibody-based proteomics project is to generate paired antibodies with non-overlapping epitopes (8). In this context, an attractive path forward might be to use the results from the antibody pairs with separate and distinct epitopes to subtract staining that is only observed using one antibody. In this manner, the staining observed for both antibodies can be scored as specific, although caution has to be taken not to exclude true differences between the antibodies due to the presence of splice variants or protein modifications. A non-redundant set of the human proteome corresponds to ~21,000 proteins (1), and the objective of an initial human antibody-based proteome project would then, according to this strategy, be to generate 42,000 validated affinity reagents directed to two non-overlapping epitopes of each human protein. Ideally the binder pairs should consist of different types of affinity reagents, such as monoclonal antibodies (23), *in vitro* selected antibody fragments (24, 25), affibodies (26), or aptamers (27). If one of the affinity reagents could be generated by the strategy used here involving the generation of affinity-purified monospecific antibodies, the other affinity reagent should thus be generated with an alternative technology platform, preferably generating a renewable resource of reagents available to the scientific community. This emphasizes the need for international efforts to coordinate such antibody-based efforts as previously discussed by Haab *et al.* (28) and Taussig *et al.* (29).

The antibody-based proteomics strategy described opens up the possibility to perform antibody-based functional studies of the complete proteome with an almost endless set of analyses to investigate the corresponding proteins. Such studies should be complemented with alternative proteomics techniques, such as green fluorescent protein tagging, interaction analysis, and high resolution mass spectrometry analysis, to further validate the results from protein probe-based assays. Such studies could include detailed analysis of protein isoforms and global spatial and temporal protein profiles in both normal and disease tissues with the objective to create a gene-centric human proteome map as a basis for biomedical and biological research.

Acknowledgments—We are grateful to Claes Wilhelmsson and Kenneth Nilsson for useful comments and advices. The entire staffs of the Human Proteome Resource centers in Stockholm and Uppsala, Sweden and in Mumbai, India are acknowledged for tremendous efforts.

* This work was supported by grants from the Knut and Alice Wallenberg Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Joint first authors.

** To whom correspondence should be addressed. Tel.: 46-8-5537-8325; Fax: 46-8-5537-8482; E-mail: mathias@biotech.kth.se.

REFERENCES

- Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K., and Lander, E. S. (2007) Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19428–19433
- Uhlén, M. (2007) Mapping the human proteome using antibodies. *Mol. Cell. Proteomics* **6**, 1455–1456
- Uhlén, M., and Pontén, F. (2005) Antibody-based proteomics for human tissue profiling. *Mol. Cell. Proteomics* **4**, 384–393
- The UniProt Consortium (2008) The universal protein resource (UniProt). *Nucleic Acids Res.* **36**, D190–D195
- Mathivanan, S., Ahmed, M., Ahn, N. G., Alexandre, H., Amanchy, R., Andrews, P. C., Bader, J. S., Balgley, B. M., Bantscheff, M., Bennett, K. L., Björling, E., Blagoev, B., Bose, R., Brahmachari, S. K., Burlingame, A. S., Bustelo, X. R., Cagney, G., Cantin, G. T., Cardasis, H. L., Celis, J. E., Chaerkady, R., Chu, F., Cole, P. A., Costello, C. E., Cotter, R. J., Crockett, D., DeLany, J. P., De Marzo, A. M., DeSouza, L. V., Deutsch, E. W., Dransfield, E., Drewes, G., Droit, A., Dunn, M. J., Elenitoba-Johnson, K., Ewing, R. M., Van Eyk, J., Faca, V., Falkner, J., Fang, X., Fenselau, C., Figeys, D., Gagne, P., Gelfi, C., Gevaert, K., Gimble, J. M., Gnad, F., Goel, R., Gromov, P., Hanash, S. M., Hancock, W. S., Harsha, H. C., Hart, G., Hays, F., He, F., Hebbar, P., Helsen, K., Hermeking, H., Hide, W., Hjerno, K., Hochstrasser, D. F., Hofmann, O., Horn, D. M., Hruban, R. H., Ibarrola, N., James, P., Jensen, O. N., Jensen, P. H., Jung, P., Kandasamy, K., Kheterpal, I., Kikuno, R. F., Korf, U., Korner, R., Kuster, B., Kwon, M. S., Lee, H. J., Lee, Y. J., Lefevre, M., Lehvaslaiho, M., Lescuyer, P., Levander, F., Lim, M. S., Lobke, C., Loo, J. A., Mann, M., Martens, L., Martinez-Heredia, J., McComb, M., McRedmond, J., Mehrle, A., Menon, R., Miller, C. A., Mischak, H., Mohan, S. S., Mohmood, R., Molina, H., Moran, M. F., Morgan, J. D., Moritz, R., Morzel, M., Muddiman, D. C., Nalli, A., Navarro, J. D., Neubert, T. A., Ohara, O., Oliva, R., Omenn, G. S., Oyama, M., Paik, Y. K., Pennington, K., Pepperkok, R., Periaswamy, B., Petricoin, E. F., Poirier, G. G., Prasad, T. S., Purvine, S. O., Rahiman, B. A., Ramachandran, P., Ramachandra, Y. L., Rice, R. H., Rick, J., Ronnholm, R. H., Salonen, J., Sanchez, J. C., Sayd, T., Seshi, B., Shankari, K., Sheng, S. J., Shetty, V., Shivakumar, K., Simpson, R. J., Sirdeshmukh, R., Siu, K. W., Smith, J. C., Smith, R. D., States, D. J., Sugano, S., Sullivan, M., Superti-Furga, G., Takatalo, M., Thongboonkerd, V., Trinidad, J. C., Uhlén, M., Vandekerckhove, J., Vasilescu, J., Veenstra, T. D., Vidal-Taboada, J. M., Vihinen, M., Wait, R., Wang, X., Wiemann, S., Wu, B., Xu, T., Yates, J. R., Zhong, J., Zhou, M., Zhu, Y., Zurbig, P., and Pandey, A. (2008) Human Proteinpedia enables sharing of human protein data. *Nat. Biotechnol.* **26**, 164–167
- 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
- Kampf, C., Andersson, A. C., Wester, K., Björling, E., Uhlén, M., and Pontén, F. (2004) Antibody-based tissue profiling as a tool in clinical proteomics. *Clin. Proteomics* **1**, 285–300
- Uhlén, M., Björling, E., Agaton, C., Szizyarto, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Ödling, J., Oksvold, P., Olsson, I., Öster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Skölleremo, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Pontén, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4**, 1920–1932
- Strömberg, S., Björklund, M. G., Asplund, C., Skölleremo, A., Persson, A., Wester, K., Kampf, C., Nilsson, P., Andersson, A. C., Uhlén, M., Kononen, J., Pontén, F., and Asplund, A. (2007) A high-throughput

- strategy for protein profiling in cell microarrays using automated image analysis. *Proteomics* **7**, 2142–2150
10. Björling, E., Lindskog, C., Oksvold, P., Linné, J., Kampf, C., Hober, S., Uhlén, M., and Pontén, F. (2008) A web-based tool for in silico biomarker discovery based on tissue-specific protein profiles in normal and cancer tissues. *Mol. Cell. Proteomics* **7**, 825–844
 11. Barbe, L., Lundberg, E., Oksvold, P., Stenius, A., Lewin, E., Björling, E., Asplund, A., Pontén, F., Brismar, H., Uhlén, M., and Andersson-Svahn, H. (2008) Toward a confocal subcellular atlas of the human proteome. *Mol. Cell. Proteomics* **7**, 499–508
 12. Flicek, P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Eyre, T., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K. L., Howe, K., Johnson, N., Jenkinson, A., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A. J., Vogel, J., White, S., Wood, M., Birney, E., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Hubbard, T. J., Kasprzyk, A., Proctor, G., Smith, J., Ureta-Vidal, A., and Searle, S. (2008) Ensembl 2008. *Nucleic Acids Res.* **36**, D707–D714
 13. Nilsson, P., Paavilainen, L., Larsson, K., Ödling, J., Sundberg, M., Andersson, A. C., Kampf, C., Persson, A., Al-Khalili Szigartyo, C., Ottosson, J., Björling, E., Hober, S., Wernérus, H., Wester, K., Pontén, F., and Uhlén, M. (2005) Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **5**, 4327–4337
 14. Service, R. F. (2005) Proteomics. New database to track protein locations. *Science* **309**, 1310
 15. Agaton, C., Galli, J., Höiden Guthenberg, I., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., Wurtz, D., Höög, C., Lundberg, J., Ståhl, S., Pontén, F., and Uhlén, M. (2003) Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol. Cell. Proteomics* **2**, 405–414
 16. Berglund, L., Björling, E., Jonasson, K., Rockberg, J., Fagerberg, L., Al-Khalili Szigartyo, C., Sivertsson, A., and Uhlén, M. (2008) A whole-genome bioinformatics approach to selection of antigens for systematic antibody generation. *Proteomics* **8**, 2832–2839
 17. Andrade, J., Berglund, L., Uhlén, M., and Odeberg, J. (2006) Using Grid technology for computationally intensive applied bioinformatics analyses. *In Silico Biol.* **6**, 495–504
 18. Berglund, L., Andrade, J., Odeberg, J., and Uhlén, M. (2008) The epitope space of the human proteome. *Protein Sci.* **17**, 606–613
 19. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580
 20. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795
 21. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M. D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N. J., Oinn, T. M., Pagni, M., Servant, F., Sigrist, C. J., and Zdobnov, E. M. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40
 22. Polanski, M., and Anderson, L. (2006) A list of candidate cancer biomarkers for targeted proteomics. *Biomark. Insights* **2**, 1–48
 23. Köhler, G., and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497
 24. Bradbury, A. R., and Marks, J. D. (2004) Antibodies from phage antibody libraries. *J. Immunol. Methods* **290**, 29–49
 25. Lipovsek, D., and Pluckthun, A. (2004) In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290**, 51–67
 26. Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M., and Nygren, P. Å. (1997) Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat. Biotechnol.* **15**, 772–777
 27. Brody, E. N., and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J. Biotechnol.* **74**, 5–13
 28. Haab, B. B., Paulovich, A. G., Anderson, N. L., Clark, A. M., Downing, G. J., Hermjakob, H., Lobaer, J., and Uhlén, M. (2006) A reagent resource to identify proteins and peptides of interest for the cancer community: a workshop report. *Mol. Cell. Proteomics* **5**, 1996–2007
 29. Taussig, M. J., Stoevesandt, O., Borrebaeck, C. A., Bradbury, A. R., Cahill, D., Cambillau, C., de Daruvar, A., Dubel, S., Eichler, J., Frank, R., Gibson, T. J., Gloriam, D., Gold, L., Herberg, F. W., Hermjakob, H., Hoheisel, J. D., Joos, T. O., Kallioniemi, O., Koegl, M., Konthur, Z., Korn, B., Kremmer, E., Krobitsch, S., Landegren, U., van der Maarel, S., McCafferty, J., Muyldermans, S., Nygren, P. A., Palcy, S., Pluckthun, A., Polic, B., Przybylski, M., Saviranta, P., Sawyer, A., Sherman, D. J., Skerra, A., Templin, M., Ueffing, M., and Uhlén, M. (2007) ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods* **4**, 13–17
 30. Saier, M. H., Jr., Tran, C. V., and Barabote, R. D. (2006) TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.* **34**, D181–D186
 31. Wilson, D., Charoensawan, V., Kummerfeld, S. K., and Teichmann, S. A. (2008) DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.* **36**, D88–D92
 32. Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.* **28**, 304–305
 33. Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326
 34. Anderson, L. (2005) Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J. Physiol.* **563**, 23–60