PROTEIN ARRAYS AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority of U.S. Provisional Application No. 61/245,852, filed September 25, 2009, which is hereby incorporated by reference in its entirety.

FIELD

Embodiments herein disclosed relate to the field of protein arrays, methods for making protein arrays, and uses of the arrays for research and to diagnose disease, among other things.

I

Protein arrays allow many protein-based assays to carried out in parallel. They provide greatly increased throughput compared to individual assays. They typically use less reagent per assay and require less time per assay as well. As a result, arrays generally provide greatly reduced costs per assay, even when the cost of fabricating the arrays is taken into account. In addition, simultaneous performance of multiple assays in arrays provides for redundancy of individual assays and the ability to assay the same parameter in multiple ways, leading to improved precision and accuracy of results, compared to individual assays. In addition, full genomic protein arrays offer possibilities for detecting proteome wide protein-molecule interactions. Such genome wide surveys will be powerful tools for understanding protein-protein interactions, decoding antibody binding specificities and cross-reactions, and for identifying biomarkers for diagnosis and patient stratification, to name a few salient applications.

Several major formats of protein arrays have been described. In forward phase protein micro-arrays capture proteins with well defined

specificities for particular targets are immobilized at defined locations in an array and target compounds are identified and quantified by the positions and intensities of binding of a sample to the array. The primary use of forward phase arrays is to interrogate individual samples to determine the presence and amount of a large number of different components simultaneously. In one typical type of forward phase array, the array is made up of a panoply of antibodies specific for particular antigens and the array is used to measure the presence and amounts of these antigens in a sample.

In reverse phase arrays, a panoply of samples are arrayed and then probed with an identifying reagent, typically with a mono-specific reagent, such as an antibody specific for a particular antigen. The primary use of reverse phase arrays is to characterize a large number of samples for the presence and amount of one – or at most a few – components. An illustrative use of reverse phase arrays is to screen a series of mono-specific reagents, such as antibodies specific for particular antigens, against a collection of different cell types.

A number of reviews on protein arrays have been published, which describe types, uses, advantages and disadvantages of current protein arrays technologies, including those of Joos and Bachmann (2009); "Protein microarrays: potentials and limitations," *Frontiers in Biosciences* 14: 4376-4385; Chan *et al.* (2004); "Protein microarrays for multiplex analysis of signal transduction pathways," *Nature Medicine* 10(12): 1390-1396; Hartmann *et al.* (2009); "Protein microarrays for diagnostic assays," *Anal. Bioanal. Chem.* 393: 1407-1416, and Caron *et al.* (2007); "Cancer Immunomics Using Autoantibody Signature for Biomarker Discovery," *Molecular & Cellular Proteomocis* 6.7: 1115-1122. Additional references are provided under VII, below.

Previous protein arrays typically either comprised relatively small numbers of purified proteins or were made by reverse transfection of large numbers of previously arrayed DNAs into a host cell which, after growth and expression of the transfected DNAs, were lysed *in situ*. Arrays in the former category have been limited by the number of proteins that can be practically

obtained; that is, by the difficulties of protein purification that must be overcome for each protein in the array. Arrays in the latter category have been limited by heterogeneity in transfection and expression results and by the limited protein density that can be obtained from confluent cells lysed on a surface *in situ*. For these reasons, and a variety of others, proteins arrays presently available suffer from a variety of limitations and disadvantages, and there is a need for improved protein arrays and for arrays that provide functionality not available with present technology.

II

The following numbered paragraphs are provided by way of illustration and describe a few of the many aspects and embodiments of the inventions herein disclosed. Many others are described herein and will be readily apparent to those skilled in the arts to which they pertain. The use of the phrase "any of the foregoing or the following" in the numbered paragraphs indicates that the various elements set forth in the numbered paragraphs can be combined in any way, and it is used to provide explicit support for any such combination. Applicant reserves the right to set out explicitly and/or claim any one or more of the combinations thus abbreviated, in whole or part by amendment in this or any successor or related application.

1.01. A method for making a protein array, comprising applying a plurality of cell lysates, comprising a corresponding plurality of proteins to a corresponding plurality of positions on a support, wherein said plurality of proteins is expressed in said corresponding plurality of cells via a corresponding plurality of exogenous DNAs,

1.02. A method for making a protein array, comprising applying lysates L_1 through L_n comprising proteins P_1 through P_n to positions S_1 through S_n on a support,

wherein each lysate L_x is of cells C_x, comprising protein P_x expressed therein via exogenous DNA D_x and is applied to position S_x,

wherein

P_1 through P_n are all different from one another,

S_1 through S_n are all different from one another,

n is an integer greater than 1 and

x is an integer from 1 to n.

In embodiments, as set forth herein, x is a fraction of the genes, loci, or protein coding regions in a genome, particularly as set forth elsewhere herein. In embodiments, as set forth herein below, x is a set number of genes, loci or protein coding genes of an organism.

1.03. A method for making a protein array, comprising applying proteins P_1 through P_n to positions S_1 through S_n on a support,

wherein each protein P_x is expressed in cells C_x via exogenous DNA D_x and is applied to position S_x,

wherein

P_1 through P_n are all different from one another,

S_1 through S_n are all different from one another,

n is an integer greater than 1 and

x is an integer from 1 to n.

1.04. A method for making an array of proteins, comprising:

applying a plurality of lysates comprising a corresponding plurality of proteins to a corresponding plurality of positions on a support, thereby producing an array of said plurality of proteins,

wherein said lysates are produced by a method comprising expressing a plurality of proteins in a corresponding plurality of cell colonies or cultures via a corresponding plurality of exogenous DNAs in cells of said colonies or cultures, and lysing each of said plurality of cell colonies or cultures thereby to produce a corresponding plurality of lysates comprising said corresponding plurality of proteins.

1.05. A method for making an array of protein, comprising:

expressing a plurality of two or more proteins in a corresponding plurality of cell colonies or cultures via a corresponding plurality of exogenous DNAs in cells of said colonies or cultures;

lysing each of said plurality of cell colonies or cultures thereby to produce a corresponding plurality of lysates comprising said corresponding plurality of proteins;

applying said plurality of lysates comprising said corresponding plurality of proteins to a corresponding plurality of positions on a support;

thereby producing an array of said plurality of proteins.

2.01. A method according to any of the foregoing or the following, wherein said proteins are over expressed via said DNAs in said cells.

2.02. A method according to any of the foregoing or the following, wherein said proteins are present at high concentrations in said lysates.

2.03. A method according to any of the foregoing or the following, wherein at least any of 50, 60, 75, 80, 85, 90, 95, 99 or 100% of said proteins, other than controls, is at least any of 0.01, 0.025, 0.05, .0.10, 0.15, 0.25, 0.35, 0.50, .075, 1.00, 1.25, 1.50. 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.50, 4.00, 4.50, 5.00, 5.50, 6.00, 7.50, 10.0, 15.0 or 20 per cent of the total protein in said cells.

2.04. A method according to any of the foregoing or the following, wherein at least any of 50, 60, 75, 80, 85, 90, 95, 99 or 100% of said proteins, other than controls, is expressed in said cells via said exogenous DNA in an amount that is substantially more than any endogenous expression of said protein in said cells.

2.05. A method according to any of the foregoing or the following, wherein at least any of 50, 60, 75, 80, 85, 90, 95, 99 or 100% of said proteins, other than controls, is expressed via said DNAs in said cells in an amount that is at least any of 1.5, 2.0, 2.5, 3.0, 4.0 5.0, 7.5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 750, 1,000, 1,500 or more times any endogenous expression of said protein in said cells.

2.06. A method according to any of the foregoing or the following, wherein at least any of 50, 60, 75, 80, 85, 90, 95, 99 or 100% of said proteins, other than controls, is at least any of 0.01, 0.025, 0.05, .0.10, 0.15, 0.25, 0.35, 0.50, .075, 1.00, 1.25, 1.50. 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.50, 4.00, 4.50,

5.00, 5.50, 6.00, 7.50, 10.0, 15.0 or 20 per cent of the total protein in said lysates.

2.07. A method according to any of the foregoing or the following, wherein in at least any of 50, 60, 75, 80, 85, 90, 95, 99 or 100% of said lysates, other than controls, the concentration of each protein in each lysate is at least any of 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 ug/ml or at least any of 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 mg/ml.

3.01. A method according to any of the foregoing or the following, wherein the proteins collectively comprise, at least any of 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 98 per cent of the proteins encoded by a genome of an organism. In embodiments the organism is a mammal. In embodiments the organism is any one of a mouse, rat, sheep, goat, dog or primate.

3.02. A method according to any of the foregoing or the following, wherein the proteins collectively comprise at least any of 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 98 per cent of the proteins encoded by a human genome.

3.03. A method according to any of the foregoing or the following, wherein the array comprises proteins of at least any of 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21, 000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or 30,000 different loci of an organism.

3.04. A method according to any of the foregoing or the following, wherein the array comprises at least any of 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21, 000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or 30,000 different proteins.

3.05. A method according to any of the foregoing or the following, wherein the array comprises at least any of 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21, 000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or 30,000 positions to which said proteins have been applied.

3.06. A method according to any of the foregoing or the following, wherein there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 positions per cm^2.

3.07. A method according to any of the foregoing or the following, wherein there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 lysates applied per cm^2.

3.08. A method according to any of the foregoing or the following, wherein there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 of said proteins applied per cm^2.

3.09. A method according to any of the foregoing or the following, wherein there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 positions per cm^2 with said proteins applied to at least any of 50, 60, 70, 80, 90 or 95% of said positions.

3.10. A method according to any of the foregoing or the following, wherein the spots are any of 10-50, 25-75, 50-100, 75-150 100-200, 150-250, 200-300, 250-350, 300-400, 400-500, 500-750, 400-800, 750-1,000 um in diameter.

3.11. A method according to any of the foregoing or the following, wherein the area of the features are any of 10-50, 25-75, 50-100, 75-150 100-200, 150-250, 200-300, 250-350, 300-400, 400-500, 500-750, 400-800, 750-1,250, 1,000-2,000, 1,500-3,000, 2,500-5,000 μm^2

3.12. A method according to any of the foregoing or the following, wherein the center to center spacing of the features (spots) is any of 5-15, 10-20, 15-25, 20-40, 25-50, 25-75, 50-100, 75-150, 100-150, 125-175, 150-225, 200-250, 225-275, 250-350, 300-400 or 400-500 um.

4.01. A method according to any of the foregoing or the following, wherein the concentration of proteins in each lysate is at least any of 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 micrograms/mil or at least any of 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 mg/ml.

4.02. A method according to any of the foregoing or the following, wherein for at least any of 50, 60, 75, 80, 85, 90, 95, 99 or 100% of said lysates, other than controls, the amount of lysate protein is the amount of total protein in at least any of 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000 , 1,500, 2,000 or 2,500 of the cells of the lysate.

4.03 A method according to any of the foregoing or the following, wherein for at least any of 50, 60, 75, 80, 85, 90, 95, 99 or 100% of said proteins, other than controls, the amount of said protein expressed via an exogenous DNA is at least any of the amount of said protein expressed via said exogenous DNA in 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,500, 2,000 or 2,500 cells in which the protein was expressed.

5.01. A method according to any of the foregoing or the following, wherein said cells are eukaryotic cells.

5.02. A method according to any of the foregoing or the following, where said cells are prokaryotic cells.

5.03. A method according to any of the foregoing or the following, wherein the cells are any one or more of HEK293, COS, CV1, BHK, CHO, HeLa, LTK, or NIH 3T3 cells.

5.04. A method according to any of the foregoing or the following, wherein the cells are HEK293T cells.

6.01. A method according to any of the foregoing or the following, wherein one or more of said exogenous DNAs each encodes one of said proteins.

6.02. A method according to any of the foregoing or the following, wherein one or more of said exogenous DNAs each encodes one of said proteins and each such protein in each such exogenous DNAs is encoded by a cDNA, a genomic DNA, or a synthetic DNA.

6.03. A method according to any of the foregoing for the following, wherein one or more of said exogenous DNAs is an expression construct comprising *cis*-acting elements effective for transcription in said cells operably linked to DNAs encoding one of said proteins. In embodiments the *cis*-acting elements include a promoter.

6.04. A method according to any of the foregoing for the following, wherein one or more of said exogenous DNAs is an expression construct comprising a promoter (and, optionally, other *cis*-acting genetic elements) effective for transcription in said cells operably linked to DNAs encoding one of said proteins, wherein said DNA encoding said protein in each of said one or more exogenous DNAs is a cDNA, a genomic DNA or a synthetic DNA.

6.05. A method according to any of the foregoing or the following, wherein said promoter is any one or more of a CMV, SV40 or MMTV promoter.

6.06 A method according to any of the foregoing or the following, wherein one or more of said exogenous DNAs encodes a chimeric protein comprising substantially the amino acid sequence of a protein for said array fused in correct reading frame to a tag sequence effective for any one or more of attachment, immobilization, capture, purification, detection and/or quantification.

6.07. A method according to any of the foregoing or the following, wherein said tag is any one or more of a GST, HA, V5, HIS, DDK (or FLAG) or myc tag.

6.08. A method according to any of the foregoing or the following, wherein said tag is a myc/FLAG tag.

6.09. A method according to any of the foregoing or the following, wherein said expression construct is a pCMV6-entry expression vector.

6.10. A method according to any of the foregoing or the following, wherein said exogenous DNA comprises a construct for non-homologous recombinatorial activation of expression of an endogenous gene encoding a protein for said protein array.

6.11. A method according to any of the foregoing or the following, wherein said exogenous DNA comprises a construct for homologous recombinatorial activation of expression of an endogenous gene encoding a protein for said array.

7.01. A method according to any of the foregoing or the following wherein the support is any support described or listed elsewhere herein.

7.02. A method according to any of the foregoing or the following, wherein the support comprises *nitrocellulose*.

7.03. A method according to any of the foregoing or the following, wherein the support comprises a nitrocellulose coated glass slide.

8.01 A method according to any of the foregoing or the following, wherein the proteins are purified from the lysates prior to application to the support.

8.02 A method according to any of the foregoing or the following, wherein the proteins are purified prior to application to the support to at least any of 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% pure.

8.03. A method according to any of the foregoing or the following, wherein the proteins are purified prior to application to the support so as to at least any of 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% of the total protein applied to the support.

8.04. A method according to any of the foregoing or the following, wherein the proteins are purified by affinity chromatography prior to application to the support .

8.05. A method according to any of the foregoing or the following, wherein the proteins comprise an affinity tag and are purified prior to

application to the support by affinity chromatography specific for the affinity tag.

8.06. A method according to any of the foregoing or the following, wherein the proteins comprise a peptide affinity tag and are purified prior to application to the support by affinity chromatography specific for the peptide affinity tag.

8.07. A method according to any of the foregoing or the following, wherein, the proteins comprises a DDK affinity tag and are purified prior to application to the support by immunoaffinity chromatography using an antibody specific for the DDK tag.

9.01 A method according to any of the foregoing or the following, wherein the proteins are purified from the lysates following application to the support.

9.02 A method according to any of the foregoing or the following, wherein the proteins are purified following application to the support to at least any of 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%,. 98%, 99% or more of homogeneously pure.

9.03. A method according to any of the foregoing or the following, wherein the proteins are purified following application to the support so as to be at least any 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%,. 98%, 99% or the total protein immobilized on the support.

9.04. A method according to any of the foregoing or the following, wherein the proteins are purified following application to the support by binding to an affinity moiety on the support specific to the proteins expressed via the exogenous DNA, and removing unbound material.

9.05. A method according to any of the foregoing or the following, wherein the proteins comprise an affinity tag and are purified following application to the support by binding to an affinity moiety on the support specific for the tag, and removing unbound material.

9.06. A method according to any of the foregoing or the following, wherein the proteins comprise a peptide affinity tag and are purified following

application to the support by binding to an affinity moiety specific for the peptide tag and removing unbound material.

9.07. A method according to any of the foregoing or the following, wherein, the proteins comprises a DDK affinity tag and are purified following application to the support by binding to an affinity moiety specific for the DDK affinity tag and removing unbound material. In embodiments the DDK specific affinity tag is a DDK-specific antibody.

10.01. A protein array according to any of the foregoing or the following, made by any of the foregoing methods.

10.02. A protein array according to any of the foregoing or the following comprising at least any of 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21, 000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or 30,000 different proteins.

10.03. A protein array according to any of the foregoing or the following, comprising at least any of 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21, 000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or 30,000 different loci of a genome of an organism.

10.04. A protein array according to any of the foregoing or the following, comprising at least any of 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 98 per cent of the proteins encoded by a genome of an organism.

10.05. A protein array according to any of the foregoing or the following, wherein there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 lysates applied per cm^2.

10.06. A protein array according to any of the foregoing or the following, wherein there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 of said proteins expressed via said exogenous DNAs per cm^2.

10.07. A protein array according to any of the foregoing or the following, wherein at least any of 25, 50, 60, 75, 80, 85, 90, 95, 99 or 100 % of the positions to which said proteins are applied the amount of said protein expressed via said exogenous DNA is at least any of the amount of said protein expressed via said exogenous DNA in 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,500, 2,000 or 2,500 cells in which each said protein was expressed.

10.08. A protein array according to any of the foregoing or the following, comprising alignment markers.

11.01. A method of determining the anti-body specificity and/or cross reaction of one or more antibodies to proteins, comprising contacting an antibody with a protein array according to any of the foregoing or the following and determining binding of the antibody thereto. In embodiments the protein array comprises a substantial fraction of all the proteins encoded by a genome in accordance with any of the foregoing or the following.

11.02. A method for determining one or more specificities and/or one or more cross reactivities of an antibody preparation, comprising contacting an antibody preparation with a protein arrays in accordance with any of the foregoing or the following and determining binding of antibodies in the preparation thereto. In embodiments the antibody preparation is a whole cell anti-serum. In embodiments the protein array comprises a substantial fraction of the proteins encoded by a genome in accordance with any of the foregoing or the following.

11.03. A method for determining the binding specificity of an antibody or an antibody preparation comprising determining the binding of the antibody or antibody preparation to a protein array in accordance with any of the foregoing or the following and from the binding to the array thus determined identifying the proteins specifically bound thereby. In embodiments the protein array comprises a substantial fraction of the proteins encoded by a genome in accordance with any of the foregoing or the following.

11.04. A method for determining protein biomarkers of disease, comprising determining binding of samples from one or more healthy individuals and from one or more diseased individuals suffering from a disease to a protein array in accordance with any of the foregoing or the following, and from differences in the binding of the samples from the healthy and diseased individuals determining protein biomarkers of the disease. In embodiments the protein array comprises a substantial fraction of the proteins encoded by a genome in accordance with any of the foregoing or the following.

11.05. A method for determining biomarkers of an autoimmune disease, comprising determining binding to a protein array in accordance with any of the foregoing or the following of antibody-containing samples from one or more healthy subjects and from one or more subjects suffering from an autoimmune disease, and from differences in the binding of the antibodies in the samples from the healthy subjects and the subjects suffering from an autoimmune disease determining protein biomarkers of the autoimmune disease. In embodiments the protein array comprises a substantial fraction of the proteins encoded by a genome in accordance with any of the foregoing or the following.

11.06. A method for determining biomarkers of a disease characterized by the presence of antibodies not present in healthy individuals, comprising determining binding to a protein array in accordance with any of the foregoing or the following of samples from one or more healthy subjects and from one or more subjects suffering from a disease characterized by the presence of antibodies not present in healthy individuals, and from differences in the binding of the antibodies in the samples from the healthy subjects and the subjects suffering from the disease determining protein biomarkers of the disease. In embodiments the protein array comprises a substantial fraction of the proteins encoded by a genome in accordance with any of the foregoing or the following.

11.07. A method for diagnosing a disease characterized by the presence of antibodies not present in healthy individuals, comprising determining binding to a protein array in accordance with any of the foregoing

or the following of an antibody containing sample from a subject possibly suffering from the disease and from the binding of antibodies in the sample to the array determining the absence or the presence of the disease. In embodiments the protein array comprises a substantial fraction of the proteins encoded by a genome in accordance with any of the foregoing or the following.

11.08. A method for monitoring signaling transduction pathways, comprising determining binding to a protein array in accordance with any of the foregoing or the following of a sample comprising proteins of signal transduction pathway proteins, whereby said binding is indicative of the absence, the presence and/or the amount of said proteins. In embodiments the sample is a whole cell lysate. In examples the sample comprises cells in which protein expression via an exogenous DNA can alter the proteins of said signal transduction pathway. In embodiments changes in any one or more of abundance, post-translational modification, or stability of said proteins is monitored. In embodiments binding of the proteins is detected using one or more protein-specific antibodies. In embodiments binding to the proteins arrays is used to decode functional connections between proteins expressed via an exogenous DNA and endogenous proteins of one or more signal transduction pathways. In embodiments several determinations are made in succession and changes in the status of proteins in one more signal transduction pathways are monitored.

11.09. A method for determining interactions between small molecules and proteins, comprising determining the binding to protein arrays in accordance with any or the foregoing or the following of a sample comprising said small molecules. In embodiments the small molecular are any one or more of small organic molecules, fats, fatty acids, fatty acid esters, lipids, sugars, glycans, nucleic acids, polynucleotides, amino acids, peptides or polypeptides, or any other small molecules. In embodiments the small molecules are detectably labeled with a detectable label. In embodiment binding of the small molecules is detected using a secondary agent that binds to small molecules bound to the array.

III

Words, terms and phrases generally are used herein in accordance with their ordinary meanings to those skilled in the arts to which they pertain, except as may be defined otherwise herein. For clarity, illustrative explanations of certain terms and phrases are set forth below. These illustrative explanations are set out exclusively as an aid to understanding the inventions herein described, and they are not limitative of the invention, and should not be understood to unduly limit the invention in any way.

Lysate L_1 through L_n designates a plurality of n lysates, numbered consecutively 1 through n, where n is at least 2. Some of the lysates may be the same or they may all be different.

Cells (generally a population of cells) C_1 through C_n designates a plurality of n cells (generally n cell populations), numbered consecutively 1 through n, where n is at least 2. Some of the cells may be the same or they may all be different.

Proteins P_1 through P_n designates a plurality of n proteins, numbered consecutively 1 through n, where n is a least 2. Some of the proteins may be the same or they may all be different.

Positions S_1 through S_n designates a plurality of n positions (generally in an array) numbered consecutively 1 through n, where n is at least 2. Some of the proteins may be the same or they may all be different. The identities of some or all of the proteins may be know or unknown.

DNAs D_1 through D_n designates a plurality of n DNAs, numbered consecutively 1 through n, where n is at least 2. Some of the DNAs may be the same or they may all be different. The identities of some or all of the DNAs may be known or unknown.

The terms "respectively" (and "corresponding") used with these designations means a correspondence between them. For instance, lysates L_1 through L_n of cells C_1 through C_n expressing proteins P_1 through P_n, via DNAs D_1 through D_n, at positions S_1 through S_n respectively means lysate L_1 of

cells C_1 expressing protein P_1 via DNA D_1 at position S_1, lysate L_2 of cells C_2 expressing protein P_2 via DNA D_2 at position S_2, and so on through lysate L_n of cells C_n expressing protein P_n via DNA D_n at position S_n .

Antibody as used herein includes polyclonal and monoclonal antibodies and derivatives thereof, including but not limited to the following: $F(ab)_2$ and F(ab) fragments, including fragments of the following; hybrid (chimeric) antibody molecules, as described in for example Winter *et al.* (1991) Nature 349:293-299 and U.S. Pat. No. 4,816,567); Fv molecules (non-covalent heterodimers) as described in for example Inbar *et al.* (1972) Proc Natl Acad Sci USA 69:2659-2662 and Ehrlich *et al.* (1980) Biochem 19:4091-4096); single-chain Fv molecules (sFv) as described in for example Huston *et al.* (1988) Proc Natl Acad Sci USA 85:5879-5883; dimeric and trimeric antibody fragment constructs; minibodies, as described in for example Pack *et al.* (1992) Biochem 31:1579-1584 and Cumber *et al.* (1992) J. Immunology 149B:120-126; humanized antibody molecules, as described for example in Riechmann *et al.* (1988) Nature 332:323-327; Verhoeyan *et al.* (1988) Science 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published Sep. 21, 1994; and, any functional fragments obtained from such molecules, such as fragments that retain antigen binding properties.

Antigen is used herein broadly to indicate any agent which elicits an immune response in the body, typically by binding to an antibody T-cell receptor or other antigen binding an antibody, T-cell receptor or other antigen binding immune system molecule. An antigen typically has one or more epitopes.

Array as used herein generally refers to an ordered arrangement of discrete positions. A protein array typically is an ordered arrangement of proteins in discrete positions. Often proteins arrays comprise a set of discrete positions on a surface with proteins disposed at one or more of the positions. Typically, the positions, particularly those with proteins disposed therein, are at known locations in the array, and the positions typically have a spatial address, such a 2-dimensional denomination, akin to an x,y coordinate in a two

dimensional Cartesian coordinate system. Obviously, arrays can be made in any desired geometry and other addressing schemes can be employed to denote the unique locations of positions and/or of proteins in an array. In embodiments some or all of the proteins in an array are known proteins.

DNA is used herein to denote polydeoxyribonulcleotides, including modified forms of naturally occurring DNAs, such as DNAs with unusual bases, incorporating labels, or chemically modified DNAs. While many of the examples and illustrations herein are written in terms of DNA, other polynucleotides can be used in much the same way, such as RNAs. Moreover, when RNA is introduced into a host cell typically it is converted to DNA and the phrase expressed via an exogenous DNA thus includes expression resulting from the introduction of RNA.

DDK is used herein to denote a peptide tag, commercially known as FLAG. The terms DDK and FLAG are used interchangeably herein.

Epitopes are individual specific features (such as structural features) of an antigen that are recognized (bound) by an antibody. Antigens comprise one or more epitopes. Different antibodies may bind the same or different epitopes on a given antigen. The epitopes on a protein antigen may be defined by continuous or discontinuous portions of the amino acid sequence.

Recombinant protein is used herein to mean a protein produced using molecular cloning techniques, such as a protein expressed via an exogenous polynucleotide, such as an exogenous DNA. As discussed in greater detail elsewhere herein, expression of a protein via an exogenous polynucleotide can be engendered by introducing into a host cell a polynucleotide encoding the protein or by introducing into a host cell a polynucleotide that engenders increased expression of an endogenous gene, such as by promoter activation, or by other methods.

Specific binding-partner as used herein indicates an agent that binds specifically to a target. Specific binding indicates that the agent can distinguish a target, such as an antigen, or an epitope within an antigen, from other non-target substances. An antibody specific for an antigen and the

antigen are an example of specific binding partners. A specific binding partner is specific in the sense that it can be used to detect a target above background noise, typically a function of non-specific binding. For example, a specific binding partner of a protein can detect a specific feature such as a sequence or a topological conformation of the protein. A specific feature can be for instance a defined order of amino acids or a defined chemical moiety. For instance, an antibody that binds to a protein specifically may be specific for a short amino acid sequence of a protein, it may be specific for a specific amino acid modification, such as phosphorylation of tyrosine (phosphotyrosine), or it may be specific for a particular carbohydrate configuration (glycan structure) in the protein, among others.

Support is used broadly herein to mean a surface-providing structure to which proteins may be applied to form an array. Typically a support is solid and structurally stable to the manipulations required to make the array and to use it. Support can have one or more components, such as a glass slide for solidity and a nitrocellulose "pad" for immobilizing proteins in an array.

IV

BRIEF DESCRIPTIONS OF THE FIGURES AND TABLES

Table 1 is a schematic diagram showing a general method for making arrays in accordance with various embodiments of the inventions herein described.

FIGURE 1 is a schematic diagram showing a modular array layout, with an enlarged view of a subarray illustrating the layout of duplicate samples and controls.

FIGURE 2 shows a protein array of 3720 lysates spotted in duplicate (7500 spots in all) on a Schott nitrocellulose coated glass support slide. (A) shows the array after staining with colloidal gold to visualize total protein. (B) shows the array after immunostaining anti-FLAG antibody to visualize in each lysate the protein expressed via the exogenous DNA.

FIGURE 3 is a schematic diagram of a pCMV6-entry expression vector for expressing proteins in cells via an exogenous DNA. The diagram shows major functionalities of the vector, including the CMV promoter for strong transcription in eukaryotic cells, the SV40 origin for replication in eukaryotic cells, a DDK-myc tag encoding region, regions with multiple restriction sites for cloning, kanamycin / neomycin resistance genes to confer antibiotic resistance in prokaryotes and eukaryotes, respectively, polyadenylation signals for transcript polyadenylation in eukaryotes and an fi bacterial origin of replication for DNA replication in prokaryotes.

FIGURE 4 shows the specificity of binding of a characterized anit-p53 antibody to a protein array of 3720 lysates spotted in duplicate on a Schott nitrocellulose coated glass support slide. (A) shows the array after immunostaining with anti-FLAG antibody to visualize the protein expressed via the exogenous DNA in each lysate. (B) shows the array after immunostaining with the anti-p53 antibody. There is one positive lysate, highlighted by an arrow. No cross reaction was detected. (C) shows an enlarged image of the section of the array containing the spots binding to the p53 antibody. The upper panel in (C) shows anit-FLAG immunostaining in the enlarged area, including serial dilutions of control proteins. The lower panel in (C) shows the reaction of the duplicate spots binding the p53 antibody (highlighted by the arrow).

FIGURE 5 illustrates the use of a protein array to decode a monoclonal antibody generated by whole cell immunization. (A) shows immunostaining of the array with the monoclonal antibody. A positive signal is indicated by the dashed box and the arrow. The inset shows the positive area at high magnification, with the duplicate E-Cadhedrin I positive signal highlighted an arrow. (B) shows the results of a Western Blot analysis confirming specificity of the monoclonal antibody for E-Cadhedrin I.

FIGURE 6 illustrates the identification of breast cancer biomarkers using a protein lysate array.

The left panel shows the results of immunostaining the array with serum from a breast cancer patient. Positively reacting areas are set off in dashed boxes, highlighted by arrows, and shown enlarged in enlarged areas A, B and C.

The right panel shows control immunostaining with normal control serum. Enlarged areas A, B and C correspond to enlarged areas A, B and C in the left panel show. The control serum does not immunostain the positions immunostained by serum from the breast cancer patient; but, reaction of auto-antibodies in the control serum can be seen in C at different positions of the array.

TABLE 2 is a schematic diagram of an embodiment for making arrays using proteins purified from lysates. In embodiments the proteins are tagged with DDK epitopes and are purified by immunoaffinity using ant-DDK antibodies.

FIGURE 7 shows homogeneity by SDS-PAGE of ten myc-FLAG (DDK) tagged proteins purified from 10 randomly chosen whole cell lysates by high throughput immunoaffinity purification using an anti-DDK antibody.

TABLE 3 is a schematic diagram of an embodiment for making arrays using in-situ purification of proteins on the support.

FIGURE 8 is a schematic diagram of an embodiment for making arrays using a step of on-support purification, in which FLAG-tagged proteins are immobilized on an anti-FLAG coated support and other proteins are washed away, producing an array of purified proteins.

FIGURE 9 shows on support immunoaffinity purification of FLAG tagged proteins from lysates on an anti-FLAG antibody coated nitorcellulose support. (A) shows a small area of an array made on an anti-FLAG antibody coated support. (B) shows a small area of a matching array made without the anti-FLAG antibody coating. In both (A) and (B) the upper insets show immunostaining with anti-myc antibody to visualize myc spotted in this part of the array, and the lower insets show immunostaining with anti-beta actin

antibody indicative of (non-specific) binding of proteins that do not comprise the FLAG tag.

Table 1

Grow Host Cells

↓

Introduce Exogenous DNAs Into Cells

↓

Clonally Propagate Cells Expressing Proteins via the Exogenous DNAs
(Each Clonal Population Expressing One Protein via an Exogenous DNA)

↓

Prepare Lysates of Each Cell Population
(Each Lysate Containing One Protein Expressed via an Exogenous DNA)

↓

Spot Lysates onto Support to Make the Arrays.

Table 2

Grow Host Cells

↓

Introduce Exogenous DNAs Into Cells

↓

Clonally Propagate Cells Expressing Proteins via the Exogenous DNAs
(Each Clonal Population Expressing One Protein via an Exogenous DNA)

↓

Prepare Lysates of Each Cell Population
(Each Lysate Containing One Protein Expressed via an Exogenous DNA)

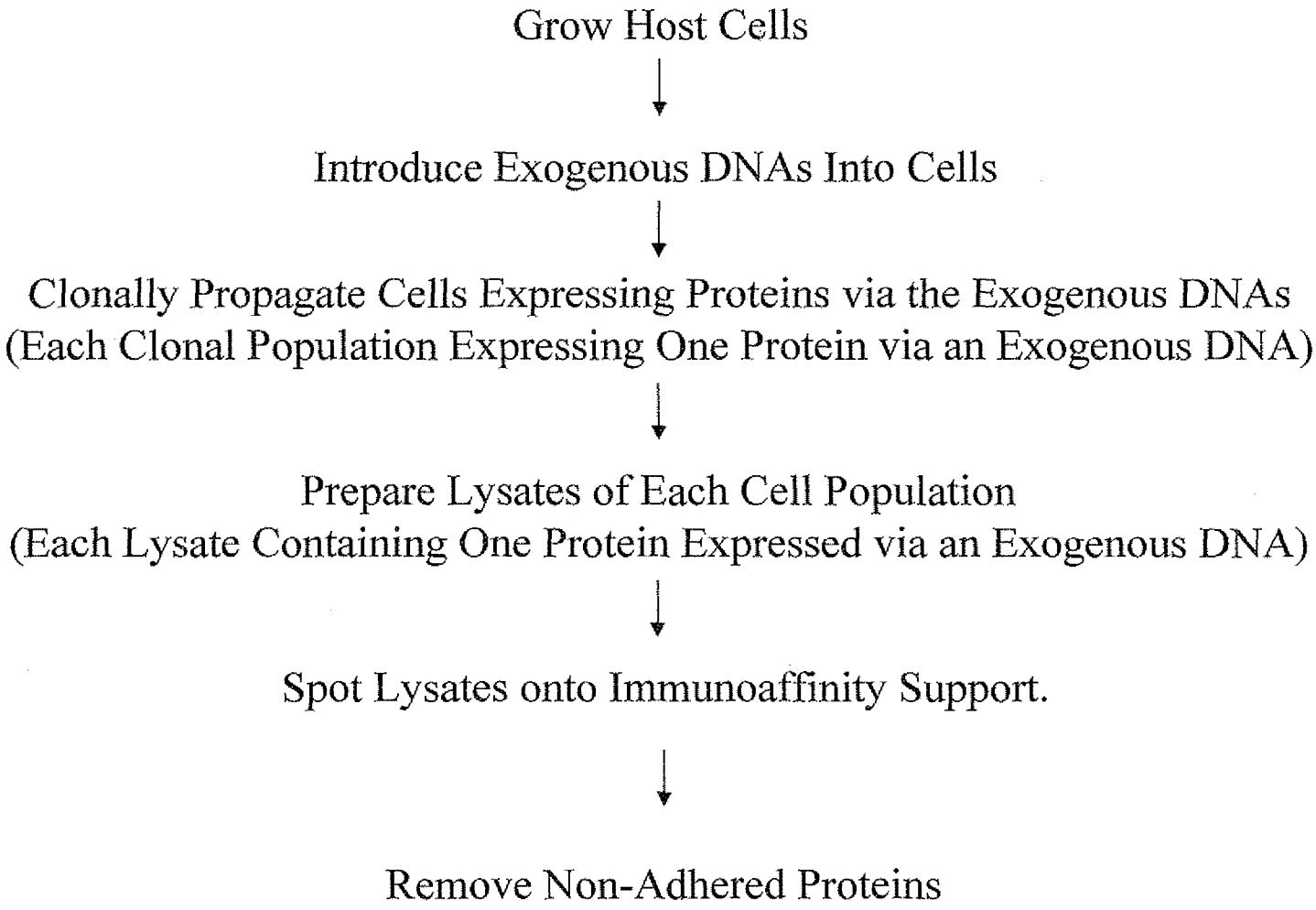
↓

Purify Proteins Expressed via the Exogenous DNA

↓

Spot Purified Proteins onto Support.

Table 3


Grow Host Cells
Introduce Exogenous DNAs Into Cells
Clonally Propagate Cells Expressing Proteins via the Exogenous DNAs
(Each Clonal Population Expressing One Protein via an Exogenous DNA)
Prepare Lysates of Each Cell Population
(Each Lysate Containing One Protein Expressed via an Exogenous DNA)
Spot Lysates onto Immunoaffinity Support.
Remove Non-Adhered Proteins

V

Embodiments of the invention herein described provide, among other things, protein arrays, methods for making protein arrays, methods for using protein arrays and devices that incorporate protein arrays. Certain embodiments provide methods to determine protein-protein interactions, verify antibody specificity and identify cross-reacting species, decode antibody specificities identify biomarkers, diagnose disease, and stratify patient populations, among other things.

As further illustrated herein in embodiments the proteins in the arrays are comprised in cell lysates. In embodiments, the lysates are made from cells in which the proteins are expressed via an exogenous DNA. In embodiments lysates are made from cells that over-express proteins via the exogenous DNA. In embodiments lysates comprising different over-expressed proteins are applied to specific positions in an array, such that the identity of the proteins is known by their positions in the array. In various embodiments the structure and/or the function of over-expressed proteins may or may not be known. In embodiments exogenous DNAs activate over-expression of an endogenous gene. In embodiments exogenous DNAs encode proteins. In embodiments exogenous DNAs comprise cDNAs and engender over production of the proteins the cDNAs encode. In embodiments the cells are mammalian cells. In embodiments the cells are human cells. In embodiments the proteins are mammalian proteins. In embodiments the proteins are human proteins. In embodiments arrays comprises a defined number of genes. In embodiments arrays comprises a defined fraction of the proteins encoded by a genome, such as the human genome.

In embodiments, proteins are purified from the lysates prior to immobilization. In embodiments the proteins are fusion proteins comprising an affinity tag and are purified by immunoaffinity purification and then immobilized in the array. In embodiments the proteins comprise a FLAG affinity tag and are purified by immunoaffinity using an anti-FLAG antibody.

In embodiments, proteins are purified from lysates *in situ* after application by binding to an affinity reagent coated support. In embodiments proteins are fusion proteins comprise an affinity tag that binds to an affinity reagent and the proteins are specifically bound to a support coated with the affinity reagent, by the interaction between the affinity tag and the affinity reagent, and unbound proteins in the lysate are removed from the support. In embodiments the affinity tag is a FLAG tag and the affinity reagent is an anti-FLAG antibody.

In embodiments protein arrays are used to determine protein-protein interactions. In embodiment arrays are used to identify, to determine and/or to quantify proteins to which a protein binds specifically and/or with which it cross-reacts.

In embodiments proteins arrays are used to determine the specificity and/or the cross-reactivity of antibodies, including, among others, polyclonal and monoclonal antibodies, and antibody derivatives. In embodiments arrays are used to decode antibodies, that is, to identify the proteins to which an antibody binds, such as, in particular, when that protein is not known.

In embodiments protein arrays are used to determine the proteins to which antibodies in autoimmune sera bind. In embodiments the autoimmune diseases are any one or more Lupus, RA or MS.

In embodiments protein arrays are used to identify autoimmune markers of health and/or disease. In embodiments, protein arrays are used to determine autoimmune markers of heath and/or disease. In embodiments the autoimmune markers are markers for autoimmune diseases or cancers. In embodiments the autoimmune diseases are any one or more Lupus, RA or MS.

In embodiments protein arrays are used to determine the binding of non-protein substances to proteins. In embodiments protein arrays are used to identify the protein binding partners of non-protein substances.

Array making methods

In embodiments the arrays are formed by, for each protein of a population of proteins, preparing a cell lysate comprising the protein and applying the lysate to a position on a support, wherein the lysate for each protein is applied to a different position, the application of the lysates forms an array of the proteins on the support, and the protein is expressed via an exogenous DNA in the cells from which the lysates are made. Figure 1 shows a general scheme for making arrays from cell lysates in accordance with embodiments of the invention. Figure 1 shows a two dimensional protein array made in accordance with the general method of embodiments set forth in Table 1. Table 2 shows a method for making arrays in accordance with an embodiment wherein proteins are purified from lysates prior to forming the array. Table 3 shows a method for making arrays in accordance with an embodiment wherein proteins are purified after forming the array by an *in situ* affinity method.

In aspects of the inventions herein described, embodiments relate to arrays that encompass a substantial fraction of the proteins expressed by genes in a genome of an organism. Certain embodiments moreover relate to arrays in which the proteins are over-expressed via an exogenous DNA in a host cell prior to application to the support. Certain further embodiments relate to arrays in which the concentration of proteins in the array is higher than in host cells in which it is expressed, and in embodiments to arrays in which the concentration of proteins in the array is higher than it is in host cells in which it is expressed via an exogenous DNA.

Methods for making arrays in embodiments, as illustrated herein, can employ any suitable methods for expressing proteins in cells via exogenous DNAs (or other polynucleotides), lysing the cells and applying the lysates (or proteins purified there from) to a support to form an array. Methods for expressing proteins, making the lysates, and applying the lysates (or purified proteins) to supports to form arrays in accordance with various illustrative embodiments are described in greater detail below, and further illustrated in the Examples.

Proteins for arrays

Proteins for arrays in embodiments of the invention herein described are obtained via exogenous polynucleotides in host cells, often DNAs. In embodiment the host cells are eukaryotic cells. In embodiments the cells are mammalian. In embodiments they are human cells, as further discussed elsewhere herein. The polynucleotides, such as DNAs for expressing the proteins in the host cells in embodiments are members of a library. In this regard the term library means a collection or set of polynucleotides, such as DNAs. In embodiments libraries may comprise a defined number of unique loci, genes, protein coding genes or regions, open reading frames or the like of a genome, such as a mammalian genome, such as a mouse, rat, goat, sheep, pig, cow, horse, monkey, gorilla or human genome.

In this regard, a "locus" or "gene locus" refers to a distinct position on a chromosome. A gene locus is precisely mapped by nucleotide sequence to a defined chromosomal region within the genome that includes all possible exons that can be spliced together. More than one transcript can originate from a single genomic locus because of alternative exon usage and/or differential splicing. Thus, each unique gene locus can be represented by multiple expression clones, each containing a polynucleotide for a different transcript originating from the same unique gene locus. By proteins comprising loci as used herein is meant that the proteins correspond to loci, that they are encoded there.

The total number of genes, protein-coding genes, unique loci and the like in the human genome is a matter of on-going research. For instance, see Nature 431, 931 - 945 (21 October 2004) and other articles in that issue which describe the number of human genes. The NCBI maintains a comprehensive, integrated, non-redundant set of nucleotide sequences from the human genome referred to herein as the Reference Sequence Collection ("RefSeq"). The collection, which is meticulously curated and continually updated, is described in, for instance, Pruitt KD, Katz KS, Sicotte H, Maglott DR , Trends Genet. 2000 Jan;16(1):44-47; Pruitt KD, Maglott DR, Nucleic Acids Res 2001

Jan 1;29(1):137-140; The NCBI handbook [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information, 2002 Oct. Chapter 17, The Reference Sequence (RefSeq) Project (available via http://ncbi.nlm.nih.gov/entrez). Sequences of polynucleotides for expressing proteins, as well as numbers of loci and genes, can be located in RefSeq and in other data bases such as GenBank, SwissProt, GenSeq, EMBL, UniProt, ASD, IMGT, IPD, IPI.

In embodiments such libraries comprise a substantial portion of the unique loci in the genome, such as at least any of 5, 10, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 97, 99% (and values in between) of the unique loci in a genome, such as the human genome. In embodiments such libraries may comprise a substantial portion of genes in the genome, such as at least any of 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 97, 99% or more (and values in between) of the genes in a genome, such as the human genome. In embodiments such libraries may comprise a substantial portion of genes in the genome, such as at least any of 5, 10, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 97, 99% or more (and values in between) of the protein-coding genes in genome, such as the human genome.

In embodiments such libraries comprise polynucleotides, such as DNAs, for a specified number of unique loci, genes, protein coding genes, open reading frames or the like, such as at least any of 5,000, 6,000, 7,000, 8,000, 9,.000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21,000, 22,000, 24,000, 26,000, 28,000, 30,000, 35,000, 40,000 loci, genes. protein coding genes, open reading frames and the like.

Subarrays

In embodiments protein arrays are organized into subarrays based on particular, general proteins features, such as functional or structural features, when a protein is expressed during the cell cycle or during the development, or where it is expressed in an organism, or where it is located in cells, its involvement in a given metabolic pathway, its relationship to a disease, *etc.* In

embodiments arrays and/or subarrays may group proteins that are related by involvement in a specific disease. In embodiments arrays and/or subarrays may group proteins such as transmembrane (plasma membrane); G-protein coupled receptors; G-protein coupled receptors, non-olfactory; G-protein coupled receptors, olfactory; hormone receptors; steroid hormone receptors; neurotransmitter receptors; enzymes; kinases; cytoplasmic; organellar, nuclear; nuclear membrane; endoplasmic reticulum; mitochondrial; lysosomal; cytoskeleton; immune system; tissue type (e.g., breast, prostate, brain, heart, etc); ion channels; nuclear hormone receptors, cytochrome P450; phosphatases; proteases; phosphodiesterases; protein trafficking; ATP-binding cassette (ANC); cytokines; homeobox and HOX genes; integrins; transporters; DexH/D protein family (RNA metabolism), etc.

Protein production via exogenous DNAs

In embodiments proteins are expressed in cells from which the lysates are made via an exogenous DNA, which is to say that the amount of the protein in the cells – and thereby in the lysates – is engendered substantially by the exogenous DNA. In embodiments the protein is over-expressed in the cells via the exogenous DNA, by which is meant that the protein is produced in the cells in excess of the amount the cells would produce were it not for the presence and action of the exogenous DNA. In embodiments the protein is produced endogenously in cells; but, it is produced at distinguishably higher levels via the cells via the exogenous DNA. In embodiments the protein is over-produced via the exogenous DNA in amounts that substantially exceed the amount produced in its absence. In embodiments the protein is produced in amounts that are at least any of 1.2, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0, 10, 25, 50, 75, 100, 150, 200, 250, 300, 400, 500, 750, 1,000, 2,000, 3,000, 5,000, 7,500, 10,000 or more times as much as the amounts produced in the cells without the exogenous DNA.

Exogenous DNAs for Expressing Proteins

By exogenous DNA is meant a DNA that is not a naturally occurring DNA in its natural setting in a genome; that is, that it is not an unaltered

endogenous gene in its unaltered endogenous setting. Typically, exogenous DNAs are DNAs introduced into cells via well known recombinant DNA techniques. Often the exogenous DNA encodes the protein to be expressed, either in its natural form, as a mutein and/or as a fusion protein. In embodiments in this regard exogenous DNAs are introduced into cells in the form or expression vectors or constructs, as discussed in greater detail below. Exogenous DNA also may be an activator DNA that does not encode the protein to be expressed, such as a RAGE construct that acts by non-homologous recombination. It may be an activator DNA that comprises only a portion of the gene for the protein to be expressed, such as a construct for gene activation by homologous recombination. It may be a construct that encodes the protein, such as an expression construct, in which case the coding region maybe uninterrupted for interrupted. And it may be other exogenous DNA that engenders the production in the cells of desired amounts of one or more proteins for an array.

Expression vectors

In embodiments a protein for an array is expressed in cells via an exogenous DNA that is an expression vector (also referred to as an expression construct) that encodes the protein and in which the coding sequence for the protein is operably linked to expression control sequences (also referred to as *cis*-acting control sequences) that provide for the desired transcription and, ultimately, production of the protein in a host cell. In embodiments expression vectors replicate autonomously, such as those that persist as episomal elements in cell. In embodiments expression vectors integrate into host cell DNA, such as those that replicate with the host cell DNA. Any suitable expression control sequences can be used to produce proteins in cells. Such expression control sequences include but are not limited to promoters, enhancers, ribosome interaction sites, such as ribosome binding sites, polyadenylation sites, transcription splice sequences, transcription termination sequences, sequences that stabilize mRNA, and other sequences that engender, regulate, facilitate, increase and/or achieve a desired effect on

production of proteins via exogenous DNA in a host cell. Such control sequences can be selected for host compatibility, inducible expression, high mRNA copy number, and other desirable effects. In embodiments promoters useful in this regard include trp, lac, tac, or T7 promoters for bacterial hosts; alpha factor, alcohol oxidase, or PGH promoters for yeast, and MMTV; SV40;.CMV, and RSV "promoters" for eukaryotic cells, such as mammalian cells.

An illustrative expression vector useful in certain embodiments of the invention, pCMV6-entry, is shown schematically in Figure 3.

Introduction of exogenous DNA into cells

Any suitable system or method can be used to introduce DNAs or other polynucleotides into cells for expression of proteins for making arrays in embodiments of the invention. There are many well known methods for introducing DNAs and other polynucleotides that can be used in this regard, such those described in the references listed further below. Among suitable methods described therein and elsewhere that can be used in embodiments of the invention herein described are calcium phosphate precipitation, electroporation, injection, DEAE-Dextran-mediated transfection, fusion with liposomes, association with agents which enhance its uptake into cells, and viral transduction. Methods can be used for introducing DNAs and other polynucleotides into cells in which, after entry into the cell, the DNA (or other polynucleotide) persists extra-chromosomally or integrated into a chromosome(s) of the host cell. The DNA, or other polynucleotide, can be transiently, constitutively and/or inducibly expressed, in accordance with well know methods. Where the polynucleotide introduced into the cell is not DNA, it will often be the case that it is copied into DNA, which DNA ultimately is the template for expressing the protein of interest.

Cells

As noted above, cells that express proteins of interest can be made by introducing exogenous DNAs (or other polynucleotides) into host cells, selecting the cells that have taken up the DNA, clonally propagating the cells,

confirming that they express the protein of interest, and then storing, and/or further expanding the cells to produce a sufficient population of cells to make lysates sufficient for making desired arrays. Suitable methods are well known and routine in the art, such as for instance the methods set forth in the references on molecular cloning listed further below.

In embodiments proteins for making arrays can be made by in any suitable cell type, such as, without limitation prokaryotic cells or eukaryotic cells, including bacterial, plant or animal cells, yeast or mammalian cells, and human cells, such as COS, CV1, BHK, CHO, HeLa, LTK, NIH 3T3, 293, and HEK293 cells, such as HEK293T cells.

Lysates

Lysates can be made from cells using any suitable method. In embodiments the lysate methods preserve desired structural and/or functional features of the proteins. Many such methods are well known to those of skill. For instance, lysates can be made using detergentless buffers and buffers with detergents, such as including RIPA buffer, lysis buffer containing SDS, hypotonic lysis buffer and the like. Methods for lysing cells are well known in the art and include but are not limited to detergent lysis, sonication lysis, and lysis under pressure (French Press) and the like.

In embodiments the concentration of proteins in each lysate in at least 20, 30, 40, 50, 60, 70, 80, 90, 95 or 100% of the lysates in the array is least any of 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 micrograms/ml, 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 mg/ml, or 1, 2, 3, 4, or 5 gm/ml.

In embodiments the concentration of proteins expressed via the recombinant DNAs in at least 20, 30, 40, 50, 60, 70, 80, 90, 95 or 100% of the lysates in the array is at least any of 0.01, 0.05, 0.10, 0.20, 0.50, 0.75, 1.00, 2.00, 3.00, 4.00, 5.00, 10.0, 15.0, 20.0 per cent of the total protein in the lysate.

In embodiments, lysate concentrations are 0.2 – 4 mg/ml and the protein expressed via an exogenous DNA is between 0.1 and 2% of the total protein.

In embodiments wherein recombinant proteins are purified before application to the array, the concentration of the recombinant protein in the application buffer for at least 20, 30, 40, 50, 60, 70, 80, 90, 95 or 100% of the proteins applied to the array is least any of 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 micrograms/ml, 1, 2, 3, 5, 10, 15, 20, 25, 35, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 or 900 mg/ml, or 1, 2, 3, 4, or 5 gm/ml.

Supports

Arrays can be made on any suitable support, whether in one part or several. In embodiments the solid support can be any material that is an insoluble matrix and can have a rigid or semi-rigid surface. In embodiments the support is a membrane, such as nitrocellulose, nylon, and the like, among other membrane materials suitable to act as supports for applying proteins to make arrays. Such membrane supports may be free standing or may be themselves supported, such as nitrocellulose membrane material on a glass slide. In embodiments supports may be glass, such as glass slides, silicon, including the surfaces of elements in integrated circuits and MEMs devices, and plastics, including plastic plates, such as microtiter plates, including, for instance, 96-well, 384-well microtiter plates, as well as those of other capacities.

Exemplary solid supports include, but are not limited to, substrates such as nitrocellulose (e.g., in membrane on a glass slide or in microtiter well form); polyvinylchloride (e.g., sheets, on glass or in microtiter wells); polystyrene latex (e.g., bead, on glass or in microtiter plates); polyvinylidine fluoride; diazotized paper; nylon membranes; activated beads, magnetically responsive beads, etc. Particular supports include plates, pellets, disks, capillaries, hollow fibers, needles, pins, solid fibers, cellulose beads, pore-glass beads, silica gels, polystyrene beads optionally cross-linked with divinylbenzene, grafted co-poly

beads, polyacrylamide beads, latex beads, dimethylacrylamide beads optionally crosslinked with N-N'-bis-acryloylethylenediamine, and glass particles coated with a hydrophobic polymer.

In embodiments proteins are attached to supports via covalent and/or non-covalent bonding. In embodiments proteins can be attached in unmodified form or they can be modified to facilitate attachment or removal after attachment or both. In embodiments proteins can be modified to facilitate or enable attachment to glass, polylysine, polystyrene, polyacrylate, polyimide, polyacrylamide, polyethylene, polyvinyl, polydiacetylene, polyphenylene-vinylene, polypeptide, polysaccharide, polysulfone, polypyrrole, polyimidazole, polythiophene, polyether, epoxies, silica glass, silica gel, siloxane, polyphosphate, hydrogel, agarose, cellulose, and/or other supports, coatings or films.

Among these are glass slides coated with nitrocellulose, such as Schott Nexterion nitrocellulose slides. .

Application of Lysates to Supports

Proteins, such as those in lysates expressed via an exogenous gene, can be applied to arrays in a variety of ways. In embodiments they are applied using a microarray printer. Microarray printers can be differentiated into three groups by their printing tip architecture and mechanisms for spotting samples: quill pins (split pins), piezoelectric (ink jet) spotters, and solid pins (Barbulovic-Nad et al., 2006). The solid pin arrayer developed by Aushon is specially designed for printing complex mixtures, such as cell lysate, and it works well with viscous protein solutions to produce uniform spots on slides (Spurrier et al., 2008). Uniformity with this spotter is very good, as illustrated in Figure 2A, which shows an array stained with colloidal gold. Total protein in the spots is uniform across the array, and the concentration series in each sub-array show appropriate scaling, which likewise is uniform across the array.

Array Geometry and Spot Density

Arrays can be made in a wide variety of formats, sizes, modulaity and can be made with a wide variety of positions, proteins, features, feature sizes,

feature spaces, feature occupancy, controls, alignment markers and references among others.

In embodiments there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 positions per cm^2 in the arrays.

In embodiments there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 lysates applied to different positions on the array per cm^2.

In embodiments there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 proteins expressed via an exogenous DNA applied to different positions on the array per cm^2.

In embodiments there are at least any of 10, 25, 50, 75, 100, 150, 200, 250, 350, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,500, 10,000, 15,000, or 20,000 positions per cm^2 on the arrays with said proteins applied to at least any of 50, 60, 70, 80, 90 or 95% of said positions.

In embodiments the lysates and/or proteins are applied in spots (features) that are any of 10-50, 25-75, 50-100, 75-150 100-200, 150-250, 200-300, 250-350, 300-400, 400-500, 500-750, 400-800, 750-1,000 um in diameter.

In embodiments area of the features comprising lysates or proteins in the array are any of 10-50, 25-75, 50-100, 75-150 100-200, 150-250, 200-300, 250-350, 300-400, 400-500, 500-750, 400-800, 750-1,250, 1,000-2,000, 1,500-3,000, 2,500-5,000 μm^2

In embodiments the center to center spacing of the features (spots) in the array is any of 5-15, 10-20, 15-25, 20-40, 25-50, 25-75, 50-100, 75-150, 100-150, 125-175, 150-225, 200-250, 225-275, 250-350, 300-400 or 400-500 um.

In embodiments the protein spot size is 110 to 300um in diameter. In embodiments center to center spacing or positions and/or proteins on the array is 150 – 250 um.

Using Arrays to Detect Binding

Any suitable methodology can be used for detecting and/or measuring binding of agents to proteins in proteins arrays. In embodiments solid phase assays are used. In embodiments sandwich assays are used. In embodiments radiometric, colorimetric, chemiluminescence and/or fluorimetric based assays are used. In certain embodiments relating to antibody binding assays, for instance, any suitable immunoassay can be used, including for instance RIAs (radioimmunassays), ELISAs, (enzyme-linked-immunosorbent assays), EIAs (enzyme-immunoasays), immunofluorescence assays, and immunoprecipitation assays, and the like. In embodiments direct labeling methods are used, in which agents are directly labeled and binding to proteins in the array is determined by detecting and/or measuring the directly bound label. In embodiments indirect labeling methods are used in which binding of agents is detected by interaction with a detection moiety that is not part of the agent and not part of the protein on the array. For instance, in indirect ELISAs binding of an antibody to a protein in the array is detected by a labeled secondary antibody that binds to the first antibody. Colorimetric, radiometric and fluorimetric detectable markers that are useful in embodiments include but are not limited to rhodamine or rhodamine derivative, biotin, avidin, strepavidin, a fluorescent compound, such as Cy3, Cy5, Alexa-555, Alexa-647, Dylight-549 or Dylight-649a, chemiluminescent compound, such as dimethyl acridinium ester, and the like.

In embodiments enzyme-immuno assays can be used for detection. A variety of such assays are well known and routinely employed in the art that readily can be applied to protein arrays, such as those described in for example Voller, A., "The Enzyme Linked Immunosorbent Assay (ELISA)," 1978, Diagnostic Horizons 2, 1-7, Microbiological Associates Quarterly Publication, Walkersville, Md.; Voller, A. *et al.*, 1978, J. Clin. Pathol. 31, 507-520; Butler, J. E., 1981, Meth. Enzymol. 73, 482-523; and Maggio, E. (ed.), 1980, Enzyme Immunoassay, CRC Press, Boca Raton, Fla..

ELISAs utilize enzymatic reactions to produce colored (absorbent) or fluorescent products from colorigenic or fluorigenic substrates, or luminescence from chemiluminescent substrates. One or more enzymes may be employed. In a simple implementation an enzyme that acts on a chromogenic substrate is conjugaetd to an antibody. The conjugaete is incubated with, for instance, proteins immobilized in microtiter dish well. After incubation conjugate that has not bound to proteins in the array is washed away. Signal generating, such as a chromogenic substrate is added incubated in the wells for a period of time to allow any enzyme conjugaete bound to the protein in the microtiter plate well to generate the colored product. In the linear regime of the reaction, the amount of color produced is proportional to the amount of bound conjugate. For protein arrays, the product of the reaction generally will either precipitate onto or bind the surface, so that it does not diffuse away from locations where antibody is bound. The use of an enzymatic reaction greatly amplifies the signal from each binding event. ELISAs can employ two or more enzymes for additional amplification. A very wide variety of ELISAs are known to the art and readily can be adapted to use with protein arrays as herein described.

Many enzymes have been used successfully in ELISAs that can be employed in embodiments, including but not limited to: malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, .alpha.-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, .beta.-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase.

Any suitable substrate and label for these enzymes (and others) can be used in ELISAs, such as but not limited to (as mentioned above) colorigenic, fluorigenic, biolumingenic and chemilumigenic substrates, which give rise, respectively to colored, fluorescent and chemiluminescent products. Radiolabeling also can be used. Fluorescent labels useful in embodiments

include but are not limited to the following: fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine. Chemiluminescent labels useful in embodiments include but are not limited to luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester. Bioluminescent labels useful in embodiments include but are not limited to luciferin, luciferase and aequorin. Any suitable label can be employed, and the foregoing are merely some of the better known and more effective labels that have been developed and employed in ELISAs and other binding assays that can be effective in this regard.

Using Protein Arrays to Screen for Binding Partners

The interaction of any type of sample or substance (referred to herein as agents) that reacts with or binds to a polypeptide (protein) in an array can be identified, In embodiments arrays are used to detect binding to proteins in arrays, such as the binding of samples (and components thereof) or agents, such as candidate binding compounds or antibodies. Illustrative uses in this regard are described below.

In embodiments the agents are specific-binding partners of proteins in an array, such as antibodies, receptor ligands, aptamers, polypeptides, and other binding molecules. Agents can be enzymes or other substances that modify polypeptides, such as kinases (which phosphorylate proteins). Agents can comprise any substance or moiety that can bind a protein (polypeptide) including, but not limited to, chemical compounds; biomolecules, such as polypeptides (amino acids), lipids, nucleic acids (nucleotides and polynucleotides), and carbohydrates; inorganic molecules; organic molecules and the like, alone or combined.

Using Protein Arrays to Identify Proteins to Which Antibodies Bind

In embodiments protein arrays as described herein can be used to screen for and identify the proteins to which antibodies bind. Antibodies, because of their specificity, are widely employed for diagnostic and therapeutic

purposes. Because of limitations in current assays methods, the antigens with which these antibodies interact are not entirely known.

When antibodies are generated against antigens, the resulting antibodies are generally characterized as being specific for that antigen. In the case of protein antigens, the proteins typically comprise one or more epitopes to which individual anti-bodies bind specifically. Epitopes in proteins may be formed not only by continuous portions of the protein but also by discontinuous regions of the protein that are folded into proximity in the protein's three dimensional conformation. The binding specificity of an antibody – such as to a protein - can be complicated by cross-reaction to other proteins that contain the same or similar epitopes. Cross-reactivity can be a significant problem both when antibodies are used for analytical and therapeutic purposes.

Sometimes such cross-reaction arises because an amino acid sequence that defines an epitope occurs in different proteins. This can occur in differentially-spliced variants of the same primary transcript or because, simply, the sequence occurs in two proteins independently. Cross-reacting proteins may occur in the same cell and tissue types, and in different types, such as occurs when splice variation occurs in a tissue-specific manner.

In general, it is important to understand the specificity of an antibody for use in detection assay, such as a diagnostic assay, or in a therapeutic. For many purposes it is important to characterize the antibody's interaction with its specific target and its cross-reactivity. Global understanding of antibody cross reactivity has not been readily obtainable using current technology, because there has not been any way to determine the interaction of an antibody with the proteome in whole. Much the same is true for the interaction of other proteins (and non-protein agents) that bind to protein partners.

In embodiments herein described, protein arrays can be used to screen antibodies for cross-reaction to proteins other than the primary antigen and, in particular, to gain an understanding of interactions with a substantial fraction of the proteins encoded in a given genome. In embodiments in this regard

genome wide arrays as described herein are particularly useful. Much the same is true for other types of agents that bind proteins.

Using Protein Arrays to Identify Biomarkers of Disease

In certain embodiments protein arrays described herein can be used to detect antibodies produced by autoimmune diseases and by other diseases, such as cancers. In autoimmune conditions, subjects generate an immune response against self-antigens, and these antibodies often are useful markers for disease diagnosis and prognosis. Sometimes the cognate antigen for such auto-antibodies is not known, in which case, in embodiments protein arrays as described herein can be used to determine the proteins to which such protein-binding auto-antibodies bind. In other cases, cognate antigens are known, at least in part, and protein arrays as described herein can be used to determine, characterize and/or measure the auto-antibodies.

For antibodies that bind known or unknown proteins, in embodiments protein arrays as described herein can be used to determine the absence, the presence and/or the amount of such auto-antibodies. For instance, in accordance with certain embodiments of the invention in this regard, antisera, blood components, fluids, and/or cells (to name a few) from subjects, such as those at risk for or actually suffering from autoimmune conditions, can be applied to protein arrays as herein described, to determine the target antigen (proteins) to which they bind, and/or to determine the absence, presence and/or the amount of auto-antibodies in the samples that bind to particular proteins in the array, so as to characterize the auto-immune antibodies in the sample and thereby diagnosis health, risk or actual disease or the like in the subject.

Similarly, in accordance with certain embodiments of the invention in this regard, antisera, blood components, fluids, and/or cells (to name a few) from subjects, such as those at risk for or actually suffering from diseases that engender the production of antibodies not generally found in the absence of the disease, can be applied to protein arrays as herein described, to determine the target antigen (proteins) to which they bind, and/or to determine the

absence, presence and/or the amount of auto-antibodies in the samples that bind to particular proteins in the array, so as to characterize the antibodies in the sample and thereby diagnosis health, risk or actual disease or the like in the subject.

The foregoing description and the examples below illustrate various embodiments of the inventions herein disclosed. It is to be appreciated that a wide variety of additional aspects, features and embodiments will be apparent from reading the disclosure to those skilled in the arts pertaining thereto, which are all within the scope of the inventions herein disclosed.

V

The following examples describe illustrative embodiments of protein microarrays in accordance with various aspects of inventions herein described and a few illustrative applications of them. These examples are in no way limitative of the inventions herein described.

EXAMPLE 1: Protein Microarray with Approximately 50% Coverage of Human Protein Genes

Embodiments of the invention provide arrays with substantial fractions of all of the proteins coding genes in a genome. The International Human Genome Sequencing Consortium estimates that there are about 20,000-25,000 protein-coding genes in the human genome (Stein, 2004). The following examples illustrate the production and use of protein arrays with approximately 10,000 – 20,000 spots containing from approximately 3,500 – 10,000 individual human genes. The arrays were produced using OriGene, Inc. libraries of validated human cDNAs cloned into the mammalian expression vector pCMV6-entry as described below.

EXAMPLE 2: pCMV-entry Vector for Expressing the Human Proteins

Proteins for the arrays were expressed via a pCMV-entry expression vector, schematically depicted in Figure 3. The vector has several features that make it especially effective for overexpressing mammalian proteins in

mammalian host cells for making protein arrays. It comprises an origin of replication effective for efficient episomal replication in eukaryotic cells (SV40 *Ori*), and an origin for replication in bacteria. It comprises an expression cassette for convenient cloning and efficient expression in mammalian cells, comprising a CMV promoter, multicloning regions, and a polyadenylation signal. The expression cassette also includes myc and DDK epitopes just upstream of the polyadenylation signal for expressing C-terminal myc-DDK tagged proteins. The tags are effective and convenient moieties for detecting and purifying the recombinantly expressed proteins. The vector comprises a T7 promoter upstream of the multicloning regions for efficient transcription in bacterial hosts (and in vitro). It comprises a second expression cassette for expressing drug resistance markers for selection in mammalian and bacterial host cells (kanamycin and neomycein resistance genes, respectively). And it comprises C-terminal myc and DDK tag sequences in 3' region of the CMV expression of tagged fusion proteins. It comprises an origin of replication for propagation in bacterial cells as well. (FLAG is a proprietary name for DDK.)

EXAMPLE 3: Over-Expression Lysates

More than 12,000 over-expression lysates have been made using human cDNAs cloned into the pCMV-entry vector and over-expressed in HEK293T cells, and they have been validated by anti-Flag immunoblot analyses. Expression profile analyses also showed that proper posttranslational modifications occur in this expression system. The expression profiles for all the lysates were examined and annotated individually. Figure 2 shows expression profiles for 8 randomly chosen lysates.

Expression levels for most of the recombinant proteins are at least 100 times higher than its endogenous counter-partner, illustrated in Figure 2. This level of over-expression provides an extremely high signal to noise ratio relative to the background from the host cells themselves.

The overall success rate for anti-Flag immunoblot is around 95%.

EXAMPLE 4: Printing Over-expression Lysates on Nitrocellulose Slides

Overexpression lysates were printed on a Schott nitrocellulose slide. Figure 1 shows an overall layout and subarray specifications. As shown in the figure, each slide was divided into subarrays, typically 40 subarrays per slide. The enlarged area in the figure shows the layout of each subarray. As indicated in the figure each subarray contained the following controls and markers: purified BSA-cy3 and BSA-cy5 orientation markers; purified mouse and rabbit IgG (positive controls); lysates of HEK293T cells transfected with empty pCMV-entry vector DNA (negative control), reference dilution series of purified GST-myc-Flag fusion proteins (to establish reference concentration curves for quantifying exogenous recombinant protein expression in the subarray lysates). The signal from the GST-myc-FLAG concentration series served to establish a standard curve of signal intensity vs. concentration for determining mys-FLAG tagged protein expression.

Arrays were made on Schott nitrocellulose standard microarray slides, using pin spotters, in particular an Aushon 2470 array spotter. 9.000 spots (features), 200-300 pico liters each, were printed on 21mm x 51mm nitrocellulose pads on the standard slides using 110 um pins. 16,000 spots were printed with 85 um pins. As many as 22,000 150-200 pico liter spots were printed on the slides using an 85 um pin on a somewhat larger nitrocellulose pad (21mm x 60mm). In keeping with ambient analyte theory, detection sensitivity increased with decreasing spot size (Ekins, 1989). Uniformity of signal was greater for 85 um spots than 110 um spots, while signal intensity was about the same. Details of array fabrication are set out below.

Slide type:	*Schott NC slide*
Aushon arrayer pin size	*85um*
Slide NC pad dimension:	*21mm x 60mm*
Total subarrays:	*48 (4 columns x 12 rows)*
Subarray Size:	*4200 um x 4400 um*
Subarray Dimensions:	*21 columns (Horizontal) x 22 rows (Vertical)*
Median Spot Diameter:	*~150um*

Spot Center to Center Spacing: *200um*
Distances between Subarrays: *200um*
Replicates per Sample: *2*

Slide made to these specifications can comprise 22,176 features (spots), such as duplicates of 10,464 unique proteins spots (such as lysates) and 1,248 control features.

Lysates in RIPA buffer containing 1% NP-40 were spotted onto the nitrocellulose slides using a solid pin spotter. Other solutions can be used for printing such as those that contain other detergents or chaotropic reagents, such as those described in Chan *et al.*, 2004 and Nishizuka *et al.*, 2003. Lysates were spotted directly from source plates and all spotting was carried out in a controlled environment at 70% relative humidity to minimize evaporative effects on sample concentrations. This worked well for printing arrays of approximately 9,000 spots.

For larger arrays, the fabrication time can be kept the same by spotting from several source plates, instead of one, serially or in parallel. To further minimize evaporative effects on concentration during array fabrication the lysates can be distributed into several source plates and spotted in parallel, so that none of the samples is exposed so long that it is adversely affected. Alternatively or additionally, 4% glycerol (or other stabilizing and anti-evaporative agents) can be added to the spotting buffer.

EXAMPLE 5: Evaluation of Chip Quality

Quality of the protein arrays was evaluated by staining with colloidal gold to evaluate total protein and by anti-Flag immunostaining to evaluate recombinant protein Colloid gold staining shows high uniformity of spot morphology for total protein across the arrays, illustrated in Figure 3A. Immunostaining with anti-FLAG antibody showed binding of FLAG-myc fusion proteins across the array, seen in figures 3B and 5A. Variability in amount of anti-FLAG binding reflects differences in fusion protein expression in the host cells and consequent differences in concentrations in the lysates. Purified GST-myc-Flag fusion protein was applied directed to the arrays to serve as a

reference standard for determining concentrations of fusion proteins in lysates. The dilution series is graphically illustrated in the enlarged area of the array in Figure 1, and uniformity of the dilution series can be seen in the upper right panel in Figure 4C

EXAMPLE 6: Identifying Antibody Specificities Using Lysate Arrays

Antibody specificity often is critical for molecular biology research and therapeutic antibody development, as well a crucial feature for the diagnostic and therapeutic use of antibodies. Cross-reactivities can cause false positive for biological research and side effects for therapeutic antibody treatment as described in Tabrizi et al., (2009) for instance. Embodiments of inventions herein disclosed include the use of overexpression lysate microarrays for identifying and validating antibody specificity, including identifying and/or characterizing primary specificities and cross-reactivities of antibodies. The arrays furthermore can be used to investigate, identify and validate the binding specificities and cross-reactivities of other proteins, as well as other types of molecules.

By way of illustration in this regard, a previously characterized polyclonal antibody for p53 was screened against a lysate array comprising over 3700 overexpression lysates comprising 3700 distinct myc-FLAG human fusion proteins. The p53 antibody reacted both with the p53 expressing lysate and with endogenous p53 of the HEK293T host cells. The signal from the overexpression lysate was more than 10 times greater than the signal from background HEK293T expression.

In addition, the antibody bound to several other lysates at levels notably above background, but less than the binding to the p53 lysate. Further examination showed that the higher p53 signals in several of these lysates was due to stimulation of p53 expression by the exogenous protein rather than cross-reactivity of the exogenous protein to the anti-p53 antibody. Similar results were obtained using a mouse monoclonal anti-p53 antibody.

The results, illustrated in Figure 4 show that over-expression lysate arrays can be used to study protein interaction specificity and cross-reactivity

and to determine effects of overexpression of a large number of proteins (individually and/or in concert with one another) on expression of other proteins, particularly, for instance, endogenous proteins.

EXAMPLE 7: Decoding Monoclonal Antibodies Generated by Whole Cell Immunization

Whole cell immunization can be used to generate monoclonal antibodies, such as highly specific monoclonal antibodies for biomarker assays and cancer therapy. However, wider use of whole cell immunization techniques is hampered by the difficulty of determining the targets of the monoclonal antibodies that are initially obtained. Often this task is very expensive and takes years to carry out.

In embodiments overexpression lysates microarray chips can be used to quickly determine the protein-binding targets of protein binding agents, such as the cellular protein binding specificities of monoclonal antibodies generated by whole cell immunization techniques. Determining the binding specificity or specificities of binding partners, such as antibodies, is referred to herein as decoding.

Embodiments in this regard are illustrated by identification of the targets of a commercially available anti-E-cadherin antibody that was initially obtained by immunizing mouse with MCF-7 mammary carcinoma cells (Shimoyama et al., 1989). Immunostaining data show that the target stands out clearly from over 3700 different genes (Figure 5A). The conclusion is further supported by western blot analysis (Figure 5B).

EXAMPLE 8: Tumor Biomarker Discovery

A variety of proteins serve as disease indicators and surrogate end points for developing therapeutics. Auto-antibodies produced by patients with cancer against tunors represent a class of proteins that could prove valuable diagnostic and prognostic indicators of disease. The possibility such proteins represent has not been realized, in part because of the difficulty of characterizing auto-antibodies in human sera.

A variety of methods have been developed in hopes of overcoming this difficulty, such as SEREX (Serological Identification of Antigens by Recombinant Expression Cloning) and SERPA (Serological Proteome Analysis) (Gunawardana and Diamandis, 2007); but, they all have substantial disadvantages.

SEREX can provide wide breath of coverage with clear annotation for each clone; but, it is based on the prokaryotic cDNA expression library screening. As a result the recombinant proteins used for screening do not have any posttranslational modifications. Moreover, the technology makes it difficult to study large number of patient serum samples at discovery stage.

SERPA identifies auto-antibody targets by immunoblotting and MS. In essence, sera containing auto-antibodies is used as a probe to detect cognate antigens in human tissue lysates subjected to 2-D IEF/SDS PAGE. Protein antigens in the gel that bind auto-antibodies in the sera then are identified by mass spectroscopy. The technique subjects proteins to harshly denaturing conditions, and suffers from detection insensitivity and irreproducibility. In addition, it can be difficult to identify antigens from the limited data that the technique provides – IEP, size and a mass spectrum, typically contaminated by other proteins involved in carrying out the western blot and detection steps.

Embodiments herein described overcome the limitations of methods such as SEREX and SERPA for identifying and characterizing protein antigens of auto-antibodies. In embodiments, the proteins arrays have a clear annotation for each gene, so that the proteins are known at each location in the array. Moreover, in embodiments the proteins in the arrays are produced and well processed post-translationally in HEK293T expression system.

Such embodiments are illustrated by the identification of auto-antibodies in breast cancer patients. Microarray slides were incubated with sera from breast cancer patients or from age matched healthy patients and then immunostained to visualize where auto-antibodies in the patient sera bound proteins in the microarrays. The results reveal a distinct autoantibody immunoreactive pattern for different human serum, illustrated in Figure 6.

EXAMPLE 9: Purified Protein Arrays – One Step Immunopurification

Previous research has shown that anti-Flag immunoaffinity purification technology can be used to isolate Flag-tagged multiple subunit protein complexes from overexpression lysates under native conditions (Chiang et al., 1993; Gloeckner et al., 2009). We applied this approach to isolate FLAG-myc tagged proteins expressed using the pCMV6-entry vector in HEK2093T cells. The general approach is depicted schematically in Tables 2 and 3 and results for10 randomly chosen lysates are shown in Figure 7. The approach can be used with any epitope tags, including, but not limited to His, myc, FLAG, V5, GST, T7, HSV, VSV-g, Glu-Glu, HA, E-tag and others.

EXAMPLE 10: On Chip Purification

On-chip purification can be used to produce microarrays as described herein. A general scheme for making protein arrays using on-chip purification is depicted in Table 3 and Figure 8.

This example describes the production of a protein array with 10,464 purified human recombinant proteins using a DDK epitope and anit-DDK antibody (FLAG epitope and anit-FLAG antibody). Flag is a highly immunogenic peptide. The interaction between Flag epitope tag and anti-Flag antibody is exceptionally strong and specific (Chiang and Roeder, 1993). High quality anti-Flag antibodies have been produced in different species, including mouse, rabbit, goal and even chicken. Such antibodies are available commercially, such as those from OriGene, Inc., which offers high quality anti-Flag mouse monoclonal and rabbit polyclonal antibodies, often used for immunoprecipitation analysis.

Efficacy of the on-chip purification is depicted in Figure 9. HEK293T cell lysates comprising FLAG-myc tagged proteins expressed in the HEK293T cells via the pCMV6-entry vector or lysates of empty pCMV6-entry vector transformed cells (negative controls) were spotted onto uncoated nitrocellulose slides (negative control) and nitrocellulose slides coated with anti-FLAG antibody. Slides were then probed with anit-myc antibodies to visualize

immobilized FLAG-myc tagged proteins or with anti-beta-actin antibodies to visualize actin, representing untagged cellular protein.

As illustrated in Figure 9, anti-myc antibody binding revealed that myc-FLAG tagged proteins bound to the anti-FLAG coated nitrocellulose slides in a tight, relatively uniform, densely staining spot, whereas it bound to the uncoated slides in a broader, more diffuse and less dense spot. There was no staining of the negative controls by anti-myc antibody on either type of slide. There was no anti-beta actin immunostaining of the spots on the anti-FLAG coated slide, showing that blocking step was effective to prevent non-specific binding and that bulk cellular protein was efficiently washed away after spotting the lysate on the slide. Anti-myc antibody binding to the uncoated slides showed binding of bulk cellular protein in a diffuse spot with a dark annulus.

The on-chip one step immunoaffinity purification can be used with any tagged proteins and thus can be applied broadly to proteins produced via an exogenous DNA using a vector that expresses tagged fusion proteins. A variety of tags can be used in the same way as DDK (FLAG) for this purpose.

VII

The following references and those cited elsewhere herein are expressly incorporated herein in their entireties, particularly as to the specific subject for which they are referenced herein.

Information on polynucleotides, expression of exogenous genes, expression vectors, protein production in transformed cells which may be useful n carrying out embodiments of the invention is well known and widely available in the arts to which embodiments pertain. Such information may be found in, for instance, in the following references.

Hames *et al.*, Polynucleotide Hybridization, IRL Press, 1985.

Davis *et al.*, Basic Methods in Molecular Biology, Elsevir Sciences Publishing, Inc., New York, 1986.

Sambrook *et al.*, Molecular Cloning, 3rd Ed. CSH Press, 2001.

Howe, Gene Cloning and Manipulation, Cambridge University Press, 1995.

Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons, Inc., 1994 to the present

Additional information, noted in citations herein, may be found in the following references.

Barbulovic-Nad, I., Lucente, M., Sun, Y., Zhang, M., Wheeler, A. R., and Bussmann, M. (2006). Bio-microarray fabrication techniques--a review. Crit Rev Biotechnol *26*, 237-259.

Chan, S. M., Ermann, J., Su, L., Fathman, C. G., and Utz, P. J. (2004). Protein microarrays for multiplex analysis of signal transduction pathways. Nat Med *10*, 1390-1396.

Cheadle, C., Vawter, M. P., Freed, W. J., and Becker, K. G. (2003). Analysis of microarray data using Z score transformation. J Mol Diagn *5*, 73-81.

Chiang, C. M., Ge, H., Wang, Z., Hoffmann, A., and Roeder, R. G. (1993). Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerases II and III. Embo J *12*, 2749-2762.

Chiang, C. M., and Roeder, R. G. (1993). Expression and purification of general transcription factors by FLAG epitope-tagging and peptide elution. Pept Res *6*, 62-64.

Ekins, R. P. (1989). Multi-analyte immunoassay. J Pharm Biomed Anal *7*, 155-168.

Gloeckner, C. J., Boldt, K., Schumacher, A., and Ueffing, M. (2009). Tandem Affinity Purification of Protein Complexes from Mammalian Cells by the Strep/FLAG (SF)-TAP Tag. Methods Mol Biol *564*, 359-372.

Goshima, N., Kawamura, Y., Fukumoto, A., Miura, A., Honma, R., Satoh, R., Wakamatsu, A., Yamamoto, J., Kimura, K., Nishikawa, T., *et al.* (2008). Human protein factory for converting the transcriptome into an in vitro-expressed proteome. Nat Methods *5*, 1011-1017.

Guilleaume, B., Buness, A., Schmidt, C., Klimek, F., Moldenhauer, G., Huber, W., Arlt, D., Korf, U., Wiemann, S., and Poustka, A. (2005). Systematic comparison of surface coatings for protein microarrays. Proteomics *5*, 4705-4712.

Gunawardana, C. G., and Diamandis, E. P. (2007). High throughput proteomic strategies for identifying tumour-associated antigens. Cancer Lett *249*, 110-119.

Haab, B. B. (2006). Applications of antibody array platforms. Curr Opin Biotechnol *17*, 415-421.

He, M., Stoevesandt, O., Palmer, E. A., Khan, F., Ericsson, O., and Taussig, M. J. (2008). Printing protein arrays from DNA arrays. Nat Methods *5*, 175-177.

Hultschig, C., Kreutzberger, J., Seitz, H., Konthur, Z., Bussow, K., and Lehrach, H. (2006). Recent advances of protein microarrays. Curr Opin Chem Biol *10*, 4-10.

Husi, H., and Grant, S. G. (2001). Isolation of 2000-kDa complexes of N-methyl-D-aspartate receptor and postsynaptic density 95 from mouse brain. J Neurochem *77*, 281-291.

Ikura, T., Ogryzko, V. V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell *102*, 463-473.

LaBaer, J., and Ramachandran, N. (2005). Protein microarrays as tools for functional proteomics. Curr Opin Chem Biol *9*, 14-19.

Li, A. G., Piluso, L. G., Cai, X., Gadd, B. J., Ladurner, A. G., and Liu, X. (2007). An acetylation switch in p53 mediates holo-TFIID recruitment. Mol Cell *28*, 408-421.

MacBeath, G., and Schreiber, S. L. (2000). Printing proteins as microarrays for high-throughput function determination. Science *289*, 1760-1763.

Monneret, C. (2005). Histone deacetylase inhibitors. Eur J Med Chem *40*, 1-13.

Nabholtz, J. M., Reese, D. M., Lindsay, M. A., and Riva, A. (2002). HER2-positive breast cancer: update on Breast Cancer International Research Group trials. Clin Breast Cancer *3 Suppl 2*, S75-79.

Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W. C., Waltham, M., Kouros-Mehr, H., Bussey, K. J., Lee, J. K., Espina, V., *et al.* (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. Proc Natl Acad Sci U S A *100*, 14229-14234.

Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., and Dunker, A. K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics *9 Suppl 1*, S1.

Payne, M. E., Fong, Y. L., Ono, T., Colbran, R. J., Kemp, B. E., Soderling, T. R., and Means, A. R. (1988). Calcium/calmodulin-dependent protein kinase II. Characterization of distinct calmodulin binding and inhibitory domains. J Biol Chem *263*, 7190-7195.

Pelham, H. R. (1990). The retention signal for soluble proteins of the endoplasmic reticulum. Trends Biochem Sci *15*, 483-486.

Ramachandran, N., Raphael, J. V., Hainsworth, E., Demirkan, G., Fuentes, M. G., Rolfs, A., Hu, Y., and LaBaer, J. (2008). Next-generation high-density self-assembling functional protein arrays. Nat Methods *5*, 535-538.

Schnack, C., Hengerer, B., and Gillardon, F. (2008). Identification of novel substrates for Cdk5 and new targets for Cdk5 inhibitors using high-density protein microarrays. Proteomics *8*, 1980-1986.

Sheng, Y., Saridakis, V., Sarkari, F., Duan, S., Wu, T., Arrowsmith, C. H., and Frappier, L. (2006). Molecular recognition of p53 and MDM2 by USP7/HAUSP. Nat Struct Mol Biol *13*, 285-291.

Shimoyama, Y., Hirohashi, S., Hirano, S., Noguchi, M., Shimosato, Y., Takeichi, M., and Abe, O. (1989). Cadherin cell-adhesion molecules in human epithelial tissues and carcinomas. Cancer Res *49*, 2128-2133.

Spurrier, B., Honkanen, P., Holway, A., Kumamoto, K., Terashima, M., Takenoshita, S., Wakabayashi, G., Austin, J., and Nishizuka, S. (2008). Protein and lysate array technologies in cancer research. Biotechnol Adv *26*, 361-369.

Spurrier, B., Washburn, F. L., Asin, S., Ramalingam, S., and Nishizuka, S. (2007). Antibody screening database for protein kinetic modeling. Proteomics *7*, 3259-3263.

Stein, L. D. (2004). Human genome: end of the beginning. Nature *431*, 915-916.

Stornaiuolo, M., Lotti, L. V., Borgese, N., Torrisi, M. R., Mottola, G., Martire, G., and Bonatti, S. (2003). KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and Golgi complex. Mol Biol Cell *14*, 889-902.

Tabrizi, M. A., Bornstein, G. G., Klakamp, S. L., Drake, A., Knight, R., and Roskos, L. (2009). Translational strategies for development of monoclonal antibodies from discovery to the clinic. Drug Discov Today *14*, 298-305.

VanMeter, A., Signore, M., Pierobon, M., Espina, V., Liotta, L. A., and Petricoin, E. F., 3rd (2007). Reverse-phase protein microarrays: application to biomarker discovery and translational medicine. Expert Rev Mol Diagn *7*, 625-633.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., *et al.* (2001). Global analysis of protein activities using proteome chips. Science *293*, 2101-2105.

What is claimed is:

1. A method for making a protein array, comprising applying lysates L_1 through L_n comprising proteins P_1 through P_n to positions S_1 through S_n on a support,

wherein each lysate L_x is of cells C_x, comprising protein P_x expressed therein via exogenous DNA D_x and is applied to position S_x,

wherein

P_1 through P_n are all different from one another,

S_1 through S_n are all different from one another,

n is an integer greater than 1 and

x is an integer from 1 to n.

2. A method for making a protein array, comprising applying proteins P_1 through P_n to positions S_1 through S_n on a support,

wherein each protein P_x is expressed in cells C_x via exogenous DNA D_x and is applied to position S_x,

wherein

P_1 through P_n are all different from one another,

S_1 through S_n are all different from one another,

n is an integer greater than 1 and

x is an integer from 1 to n.

3. A method according to claim 1 or 2, wherein the proteins comprise at least 1,000 different loci of an organism.

4. A method according to claim any of claims 1 through 3, wherein the proteins collectively comprise at least 20 per cent of the proteins encoded by the genome of an organism.

5. A method according to any of claims 1 through 4, wherein the genome is a human genome

6. A method according to any of claims 1 through 5, wherein for at least 50 per cent of said lysates, the amount of lysate protein applied to each position in the array is the amount of total protein in at least 100 cells of the lysate.

7. A method according to any of claims 1 through 6, wherein the proteins or the lysates are applied to nitrocellulose on a glass slide.

8. A method according to any of claims 1 through 7, wherein the support is coated with a capture reagent specific for an affinity tag and the proteins comprise the tag and are purified by binding to the capture reagent.

9. A protein array, comprising lysates L_1 through L_n comprising proteins P_1 through P_n at positions S_1 through S_n on a support,

wherein each lysate L_x is of cells C_x, comprising protein P_x expressed therein via exogenous DNA D_x and applied to position S_x,

wherein

P_1 through P_n are all different from one another,

S_1 through S_n are all different from one another,

n is an integer greater than 1 and

x is an integer from 1 to n.

10. A protein array, comprising proteins P_1 through P_n at positions S_1 through S_n on a support,

wherein each protein P_x is expressed in cells C_x via exogenous DNA D_x and applied to position S_x,

wherein

P_1 through P_n are all different from one another,

S_1 through S_n are all different from one another,

n is an integer greater than 1 and

x is an integer from 1 to n.

11. A protein array according to claim 9 or 10, wherein the proteins comprise at least 1,000 different loci of an organism.

12. A protein array according to claim any of claims 9 through 11, wherein the proteins collectively comprise at least 20 per cent of the proteins encoded by the genome of an organism.

13. A protein array according to any of claims 9 through 12, wherein the genome is a human genome

14. A protein array according to any of claims 9 through 13, wherein for at least 50 per cent of said lysates, the amount of lysate protein applied to each position in the array is the amount of total protein in at least 100 cells of the lysate.

15. A protein array according to any of claims 9 through 14, wherein the proteins or the lysates are applied to nitrocellulose on a glass slide.

16. A protein array according to any of claims 9 through 15, wherein the proteins comprise an affinity tag and are bound to the support by a capture reagent specific for the affinity tag immobilized therein.

17. A method for determining the binding specificity of an antibody or antibody preparation comprising determining the binding of the antibody or antibody preparation to a protein array according to any of claims 9 through 16 and from the determination identifying the binding specificity of the antibody or antibody preparation for proteins in the array.

18. A method for determining protein biomarkers of disease, comprising determining binding of samples from one or more healthy individuals and from one or more diseased individuals suffering from a disease to a protein array according to any of claims 9 through 16, and from differences in the binding of the samples from the healthy and diseased individuals determining protein biomarkers of the disease.

19. A method for determining biomarkers of an autoimmune disease, comprising determining binding of antibody-containing samples from one or more healthy subjects and from one or more subjects suffering from an autoimmune disease to a protein array in according to any of claims 9 through 16, and from differences in the binding of the antibodies in the samples from the healthy subjects and the subjects suffering from an autoimmune disease determining protein biomarkers of the autoimmune disease.

20. A method for determining biomarkers of a disease characterized by the presence of antibodies not present in healthy individuals, comprising determining bindingof samples from one or more healthy subjects and from one or more subjects suffering from a disease characterized by the presence

of antibodies not present in healthy individuals to a protein array according to any of claims 9 through 16, and from differences in the binding of the antibodies in the samples from the healthy subjects and the subjects suffering from the disease determining protein biomarkers of the disease.

21. A method for diagnosing a disease characterized by the presence of antibodies not present in healthy individuals, comprising determining binding of an antibody containing sample from a subject possibly suffering from the disease to a protein array in according to any of claims 9 through 16 and from the binding of antibodies in the sample to the array determining the absence or the presence of the disease.

22 A method for monitoring signaling transduction pathways, comprising determining binding of a sample comprising proteins of signal transduction pathway proteins to a protein array according to any of claims 9 through 16, whereby binding to proteins in the array is indicative of the absence, the presence and/or the amount of said proteins.

23. A method for determining interactions between small molecules and proteins, comprising determining the binding of a sample comprising said small molecules to protein arrays according to any of claims 9 through 16.