

antibodies, aptamers, and protein-based aptamers. In one embodiment the capture molecule is an antibody.

A “biomarker” is a molecule of interest that is to be detected and/or analyzed, e.g., a peptide, or a protein. Typically a biomarker is associated with a particular physical condition, e.g., a disease or disease state, e.g., late stage breast cancer.

A biomarker that “binds” to a capture molecule is a term well understood in the art, and methods to determine such specific or preferential binding are also well known in the art. A molecule is said to exhibit “binding” if it reacts or associates more frequently, more rapidly, with greater duration and/or with greater affinity with a particular target than it does with alternative substances. A capture molecule “binds” to a target if it attaches with greater affinity, avidity, more readily, and/or with greater duration than it attaches to other substances. For example, a capture molecule that specifically or preferentially binds to a target is an antibody that binds this target with greater affinity, avidity, more readily, and/or with greater duration than it binds to other substances. It is also understood by reading this definition that, for example, a capture molecule that specifically or preferentially binds to a first target may or may not specifically or preferentially bind to a second target. As such, “binding” does not necessarily require (although it can include) exclusive binding. Generally, but not necessarily, reference to “binding” means preferential binding. The concept of “binding” also is understood by those of skill in the art to include the concept of specificity. Specific binding can be biochemically characterized as being saturable, and binding for specific binding sites can be biochemically shown to be competed.

An “antibody” is an immunoglobulin molecule capable of specific binding to a target, such as a carbohydrate, polynucleotide, lipid, polypeptide, etc., through at least one antigen recognition site, located in the variable region of the immunoglobulin molecule.

The terms “polypeptide,” “oligopeptide,” “peptide,” and “protein” are used interchangeably herein to refer to polymers of amino acids of any length. The polymer can be linear or branched, it can comprise modified amino acids, and it can be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component. Also included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), as well as other modifications known in the art.

An “array” format is a known or predetermined and ordered spatial arrangement of one or more capture molecules on a solid support. A “multiplexed array” format is an ordered spatial arrangement of two or more capture molecules on a solid support. In one embodiment, row and column arrangements are used due to the relative simplicity in making and assessing such arrangements. The spatial arrangement can, however, be essentially any form selected by the user, and preferably, but need not be, in a pattern. Array formats are characterized by the use of spatial location within the array to identify the feature present at that location.

“Detect” refers to identifying the presence, absence and/or amount of protein to be detected. Detection can be done visually or using a device, e.g., a scanner and detector.

The term “mammal” as used herein includes both humans and non-humans and include but is not limited to humans, non-human primates, canines, felines, murines, bovines, equines, and porcines.

“Solid support” refers to a material or group of materials having a rigid or semi-rigid surface or surfaces. In some aspects, at least one surface of the solid support will be substantially flat, although in some aspects it can be desirable to physically separate regions for different molecules with, for example, wells, raised regions, pins, etched trenches, or the like.

To “analyze” includes determining a set of values associated with a sample by measurement of constituent expression levels in the sample and comparing the levels against constituent levels in a sample or set of samples from the same subject or other subject(s).

A “predictive model” is a mathematical construct developed using an algorithm or algorithms for grouping sets of data to allow discrimination of the grouped data. As will be apparent to one of ordinary skill in the art, a predictive model can be developed using e.g., principal component analysis (PCA), and linear discriminant analysis (LDA).

A “score” is a value or set of values selected or used to discriminate a subject’s condition based on, for example, a measured amount of sample constituent from the subject.

The term percent “identity,” in the context of two or more nucleic acid or polypeptide sequences, refers to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent “identity” can exist over a region of the sequence being compared, e.g., over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat’l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., *infra*).

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information, most conveniently at its website.

Abbreviations used include: human estrogen receptor-2 (Her-2), matrix metalloproteinase-2 (MMP-2), cancer antigen 15-3 (CA 15-3), osteopontin (OPN), tumor protein 53 (p53), vascular endothelial growth factor (VEGF), cancer antigen 125 (CA 125), Serum Estrogen Receptor (SER)