

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date
4 August 2005 (04.08.2005)

PCT

(10) International Publication Number
WO 2005/071419 A2

(51) International Patent Classification⁷: **G01N 33/68**

de Lattre de Tassigny, Super Valmont, La Garde, F-13013 Marseille (FR).

(21) International Application Number:

PCT/IB2005/000261

(74) Agents: **BREESÉ, Pierre et al.**; Breesé Derambure M- jerowicz, 3, avenue de l'Opéra, F-75001 Paris (FR).

(22) International Filing Date: 17 January 2005 (17.01.2005)

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/537,412 16 January 2004 (16.01.2004) US
11/036,298 14 January 2005 (14.01.2005) US

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (for all designated States except US): **IPSO-GEN** [FR/FR]; 163 avenue de Luminy, F-13009 Marseille (FR). **INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)** [FR/FR]; 101, rue de Tolbiac, F-75654 Paris Cedex 13 (FR). **INSTITUT PAOLI-CALMETTES** [FR/FR]; 232 boulevard Sainte-Marguerite, F-13009 Marseille (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **JACQUEMIER, Jocelyne** [FR/FR]; 15, boulevard Notre-Dame, F-13011 Marseille (FR). **BERTUCCI, François** [FR/FR]; Immeuble Le Sully, Parc de la Cadenelle, 122 rue du Commandant Rolland, F-13008 Marseille (FR). **BIRNBAUM, Daniel** [FR/FR]; 9, rue Baldacchini, La Croix du Sud, F-13009 Marseille (FR). **DEBONO, Stéphane** [FR/FR]; Résidence La Palmeraie-Borely, Bât D22, 70 avenue d'Haïfa, F-13008 Marseille (FR). **TAGETT, Rebecca** [GB/FR]; 430 avenue

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/071419 A2

(54) Title: PROTEIN EXPRESSION PROFILING AND BREAST CANCER PROGNOSIS

(57) Abstract: A method for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of proteins in breast tissues or cells, said pool comprising all or part, for example one, two, three or more of a protein set comprising: Afadin, Aurora A, a-Catenin, b-Catenin, BCL2, Cyclin D1, Cyclin E, Cytokeratin 5/6, Cytokeratin 8/18, E-Cadherin, EGFR, ERBB2, ERBB3, ERBB4, Estrogen receptor, FGFR1, FHIT, GATA3, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3, Cytokeratin 6, Cytokeratin 18, Ang1, AuroraB, BCRP1, CathepsinD, CD10, CD44, CK14, Cox2, FGF2, GATA4, Hifla, MMP9, MTA1, NM23, NRG1a, NRG1beta, P27, Parkin, PLAU, S100, SCRIBBLE, Smooth Muscle Actin, THBS1, TIMP1.

**Protein expression profiling and breast cancer
prognosis**

I - Field of the invention

5 The present invention relates to protein analysis and, in particular, to protein expression profiling of breast tumors and cancers.

II - Background

10 Adjuvant systemic therapy has a favorable impact on survival in patients with early breast cancer.^{1, 2} The decision to give or withhold such therapy is based upon a series of histoclinical 15 prognostic criteria reviewed in consensus conferences (i.e. National Institute Health NIH and St-Gallen).^{3, 4} However, despite the establishment of standardized criteria, the heterogeneity of breast tumors remains poorly understood. For example, 20 clinical treatment decisions on whether to treat patients with node-negative breast cancer by surgery and radiotherapy alone, or in combination with adjuvant chemotherapy are currently being made with scant information on patient risk for metastatic 25 relapse. Additionally, identifying among the patients who receive chemotherapy those who will benefit and those who will not benefit from standard anthracyclin-based protocols remains elusive. However, the relatively limited efficacy of current 30 protocols (~30-40% of failure rate) and the increasing availability of new therapies make this issue clinically important. Furthermore, the development of molecularly-targeted drugs such as trastuzumab (Herceptin™), a monoclonal antibody

against the ERBB2 tyrosine kinase receptor, is needed.⁵ With few exceptions, such as estrogen receptor and ERBB2 receptor, the available molecular markers are of limited value in clinical practice.

5 High-throughput molecular technologies such as DNA arrays, have recently significantly contributed to enhance understanding of the molecular complexity of breast cancer.⁶ Several studies have demonstrated the potential clinical utility of gene expression signatures defined by the combined RNA expression of a few tens of genes. These signatures have lead to the development of a new molecular taxonomy of disease, including the identification of previously indistinguishable prognostic subclasses.⁷⁻¹⁵ The 10 clinical impact of these tests on disease management must be subsequently evaluated in large retrospective and prospective studies of adequate statistical power on fully annotated patient samples, followed by the development of gene 15 expression-based diagnostics adapted to the clinical setting.

20 Unfortunately, the cost, technical complexity, and interpretation of DNA microarray technology still complicate investigation with cancer specimens 25 and are currently unsuitable for routine use in the standard clinical setting. Issues that must be addressed prior to validation and integration of this technology to clinical pathology laboratories include the requirement for high-quality RNA 30 extracted from unfixed tissues, intra-tumoral heterogeneity of excised patient samples, and bias resulting from the asymmetry of variables with a number of hybridized samples greatly inferior to the 35 number of genes being tested leading to non-trivial statistical problems. Finally, the sensitivity,

specificity, reproducibility and technical feasibility outside large academic centers will have to be addressed, and experimental conditions will have to be standardized and data compared in multi-center clinical trials.

Additional opportunities to validate and/or identify prognostic expression signatures are provided by alternative high-throughput approaches, which may be used either separately or in combination with DNA microarrays. One of these is the tissue microarray (TMA) technique,¹⁶⁻¹⁸ which allows for the simultaneous study of hundreds of tumor specimens at the DNA, RNA or protein level. Immunohistochemistry (IHC) is applicable to paraffin-embedded samples that constitute the bulk of pathology archives, avoiding the requirement for high-quality RNA extracted from frozen specimens. IHC is relatively inexpensive, straightforward and well established in standard clinical pathology laboratories. Thus, IHC on TMA may be a practical approach both in validation studies and in routine testing. However, analytical classification methods to efficiently process and interpret multiple target IHC data have not been previously developed.

Recent studies have shown the reliability of hierarchical clustering for classifying cancers when applied to IHC TMA data of a significant range of markers.¹⁹⁻²⁴ However none addressed the prognostic issue.

The aim of the present invention is to provide means capable of analyzing histopathologic features of breast disease, in particular of classifying breast cancers into prognostically relevant subclasses. After exhaustive testing on a retrospective panel of 552 early breast cancer

samples we have found that this classification was possible by analyzing a consistent set of proteins. Classification of samples, based on this multidimensional protein data set, was first done 5 using classical unsupervised hierarchical clustering. We then developed a supervised bioinformatic method that further improved the classification as compared with usual prognostic factors.

10

III - Summary of the invention

The present invention provides a protein expression signature identified by protein expression profiling and which may be used for analysing histopathologic features of breast disease as well as methods for carrying out such analysis. In particular, protein expression profiling may be a clinically useful approach to assess breast cancer heterogeneity and prognosis in patients with stage 15 I, II, or III disease. It may be used both for breast tumor management in clinical settings and as 20 a research tool in academic laboratories

25

The invention provides in one aspect a method for analyzing differential protein expression associated with histopathologic features of breast disease, in particular breast tumours, e.g., breast carcinomas, comprising the detection of the overexpression or underexpression of a pool of 30 proteins in breast tissues or cells, said pool comprising all or part of a protein set comprising

Afadin, Aurora A, α -Catenin, β -Catenin, BCL2, Cyclin D1, Cyclin E, Cytokeratin 5/6, Cytokeratin 35 8/18, E-Cadherin, EGFR, ERBB2, ERBB3, ERBB4,

Estrogen receptor, FGFR1, FHIT, GATA3, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3.

5 By "all or part" is meant 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51 or 52 proteins.

10 By "Cytokeratin 5/6" is meant Cytokeratin 5 and/or Cytokeratin 6. The same is applicable to "Cytokeratin 8/18".

15 The following table displays the proteins of the present invention and their corresponding amino-acid sequences (SEQ ID NO. 1 to 52). These proteins are identified by their common names (first column) in the methods, libraries, sets, pools etc. of the invention. Other names in the literature which designate the same proteins (alias, synonyms etc.) are covered as well, and are incorporated herein by reference.

20
25 The present invention may also define these proteins by their amino-acid (polypeptidic) sequences (SEQ ID NO.), or portions or modifications thereof in accordance with the definition of "protein" provided below.

Table 0

Protein Name	SEQ ID NO.
Afadin	1
Aurora A	2
a-Catenin	3
b-Catenin	4
BCL2	5

Cyclin D1	6
Cyclin E	7
Cytokeratin 5	8
Cytokeratin 8	9
E-Cadherin	10
EGFR	11
ERBB2	12
ERBB3	13
ERBB4	14
Estrogen receptor	15
FGFR1	16
FHIT	17
GATA3	18
Ki67	19
Mucin 1	20
P53	21
P-Cadherin	22
Progesterone receptor	23
TACC1	24
TACC2	25
TACC3	26
Cytokeratin 6	27
Cytokeratin 18	28
Ang1	29
AuroraB	30
BCRP1	31
CathepsinD	32
CD10	33
CD44	34
CK14	35
Cox2	36
FGF2	37
GATA4	38
Hif1a	39

MMP9	40
MTA1	41
NM23	42
NRG1a	43
NRG1beta	44
P27	45
Parkin	46
PLAU	47
S100	48
SCRIBBLE	49
Smooth Muscle Actin	50
THBS1	51
TIMP1	52

"Over or underexpression of a pool of protein" means that overexpression of certain proteins are detected simultaneously to the underexpression of others said proteins. "Simultaneously" means concurrent with or within a biologic or functionally relevant period of time during which the over expression of a protein may be followed by the under expression of another protein, or conversely, e.g., because both expressions are directly or indirectly correlated.

In a further aspect, the invention provides a method for analyzing for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of protein in breast tissues comprising a protein set comprising:

Aurora A, a-Catenin, b-Catenin, Cyclin D1, Cytokeratin 8/18, ERBB2, ERBB3, Estrogen receptor,

FGFR1, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC2.

In a further aspect, the invention provides a
5 method for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of protein in breast tissues comprising a protein set comprising:
10

Afadin, Aurora A, a-Catenin, BCL2, Cyclin D1, Cytokeratin 5/6, Cytokeratin 8/18, E-Cadherin, ERBB2, ERBB3, ERBB4, Estrogen receptor, FGFR1, FHIT, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC2, TACC3.
15

According to a preferred embodiment the pool of protein comprises a protein set comprising

Afadin, Aurora A, a-Catenin, b-Catenin, BCL2, Cyclin D1, Cyclin E, Cytokeratin 5/6, Cytokeratin 8/18, E-Cadherin, EGFR, ERBB2, ERBB3, ERBB4, Estrogen receptor, FGFR1, FHIT, GATA3, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3.
25

According to another embodiment the pool of protein comprises a protein set comprising all proteins of the Table 0 above.

The method further comprises at least one of the following embodiments :

- the detection of overexpression of at least one, preferably at least two, three or all of the following proteins :

EGFR, P53, Ki67, FGFR1, ERBB2, ERBB3, ERBB4, Cyclin D1, Cyclin E, Cytokeratin 5/6.

5 - the detection of overexpression of at least one, preferably at least two, three or all of the following proteins :

Estrogen Receptor, FHIT, GATA3, Mucin 1, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3, Afadin, Aurora A, α -Catenin, β -Catenin, BCL2, Cytokeratin 8/18, E-Cadherin.

10

A further object of the invention is to provide a protein library useful for the molecular characterization of histopathologic features of breast disease comprising or corresponding to a pool 15 of protein sequences, over or under expressed, in breast tissue or cells, said pool corresponding to the protein sets previously described.

20

Preferably, said protein librairies may be immobilized on a solid support which may be preferably selected from the group comprising nylon membrane, nitrocellulose membrane, polyvinylidene difluoride, glass slide, glass beads, polystyrene plates, membranes on glass support, silicon chip or gold chip.

30

In a further aspect, the present invention provides a method for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of protein in breast tissues comprising :

a) obtaining breast tissue cells from a patient, and

b) measuring in the tissue cells obtained in step (a) over or underexpression of proteins of a library as previously described.

5 Alternatively to breast tissue cells from a patient, the detection of over or under expression of the pool of protein may be carried out on breast tumor cell lines.

10 The proteins may be directly or indirectly labeled before reaction step (b) with a label which may be selected from the group comprising radioactive, colorimetric, enzymatic, molecular amplification, bioluminescent or fluorescent labels.
15 Advantageously, one or more specific label are used for each protein of the library according to the invention. A person skilled the art will be able to choose appropriate labels and labelling methods to carry out the invention. For example, one may use a
20 label selected in the group comprising and not limited to : biotine, digoxygenin.

25 The measuring of over or under expression of proteins may be carried out on cell or tissue, frozen or embedded in any appropriate material, e.g., paraffin, e.g. tissue microarray. Various known method of the prior art may be used as, e.g., ImmunoHistoChemistry (IHC) technologies. The measuring of over or under expression of proteins may be also be carried out by the use of, e.g., protein (micro)arrays, antibody (micro)arrays, antigen (micro)arrays or any other appropriate technology, e.g., by using the previously defined supports.

According to an advantageous embodiment, the method for analysing differential protein expression of the invention further comprises:

- a) obtaining a control sample
- 5 b) measuring in the control sample obtained in step (a) expression level of each protein corresponding to library according to the invention
- c) comparing expression level of each protein with the level of equivalent protein in breast tissue cells from a patient, or in cell lines.

The present invention is useful for detecting, diagnosing, staging, monitoring, predicting, preventing conditions associated with breast cancer. It is particularly useful for predicting clinical outcome of breast cancer and/or predicting occurrence of metastatic relapse and/or determining the stage or aggressiveness of a breast disease in at least 50%, e.g., at least 55%, e.g., at least 60%, e.g., at least 65%, e.g., at least 70%, e.g., at least 75%, e.g., at least 80%, e.g., at least 85%, e.g., at least 90%, e.g., at least 95%, e.g., 20 100% of the patients. The invention is also useful for selecting more appropriate doses and/or schedule of chemotherapeutics and/or biopharmaceuticals and/or radiation therapy to circumvent toxicities in 25 a patient.

In particular, the invention is also useful for 30 selecting appropriate doses and/or schedule of chemotherapeutics and/or (bio)pharmaceuticals, and/or targeted agents, among which one may cite Aromatase Inhibitors (e.g., Exomestane, Anastrazole, Letrozole), Anti-estrogens (e.g., Fluvestrant, 35 Tamoxifen), Taxanes (e.g., PacliTaxol, Docetaxel),

Antracyclines (e.g., Doxurubicin, Cyclophosphamide), CHOP (Doxurubicin, Cyclophosphamide, ocovorin, prednisone when taken in combination). Other drugs like Velcade™, 5-Fluorouracil, Vinblastine, 5 Gemcitabine, Methotrexate, Goserelin, Irinotecan, Thiotepla, Topotecan or Toremifene may be cited as well.

For targeted therapies, one may cite Iressa (gefitnib, ZD1839, anti-EGFR, PDGFR, c-kit, Astra-Zeneca); ABX-EGFR (anti-EGFR, Abgenix/Amgen); Zarnestra (FTI, J & J/Ortho-Biotech); Herceptin (anti-HER2/neu, Genentech); Avastin (bevancizumab, anti-VEGF antibody, Genentech); Tarceva (ertolinib, OSI-774, RTK inhibitor, Genentech-Roche); ZD66474 (anti-VEGFR, Astra-Zeneca); Erbitux (IMC-225, cetuximab, anti-EGFR, Imclone/BMS); Oncolar (anti-GRH, Novartis); PD-183805 (RTK inhibitor, Pfizer); EMD72000, (anti-EGFR/VEGF ab, MerckKgaA); CI-1033 (HER2/neu & EGF-R dual inhibitor, Pfizer); EGF10004; Herzyme (anti-HER2 ab, Medizyme Pharmaceuticals); Corixa (Microsphere delivery of HER2/neu vaccine, Medarex).

Further relevant anti-breast cancer agents are described by Awada et al. in "The pipeline of new anticancer agents for breast cancer treatment in 2003" Critical Reviews in Oncology/Hematology 48 (2003) 45-63, the content of which is incorporated herein by reference.

Advantageously, in a method according to the present invention, breast tissue cell may be obtained from a patient regardless of whether said patient has received or not a neo-adjuvant or

adjuvant, e.g., systemic, therapy. Similarly, treated or untreated cell lines may be used.

5 Advantageously, in a method according to the present invention, breast tissue cell may be obtained from a patient regardless of ER receptor expression.

10 In a further aspect, the present invention provides a method for treating a patient with a breast cancer comprising (i) the implementation of a method for analysing differential protein expression according to the invention on a sample from said patient, and (ii) determining a treatment for this 15 patient based on the analysis of differential protein expression profile obtained in step i).

20 In a further aspect, the present invention relates to a method for analyzing differential protein expression associated with histopathologic features of breast disease according to the invention wherein the detection of the overexpression or underexpression of said pool of protein in breast tissues comprises the detection of 25 the overexpression or underexpression of nucleic acids coding for said proteins.

30 The present invention further relates to a nucleic acids library useful for the molecular characterization of histopathologic features of breast disease comprising nucleic acids coding for the over or underexpressed proteins according to the invention, or equivalent thereof.

The sequences of the nucleic acids of the library according to the invention are easily available for a person skilled in the art that may, for example, use printed publications describing said sequences and/or public databases, e.g., the National Center for Biotechnological Information (NCBI) database, that provide such sequences as well. The content of the NCBI database may be available via internet at the following address
10 <http://www.ncbi.nlm.nih.gov/>.

Definitions

15 "aggressiveness of cancer" refers to cancer growth rate or potential to metastasise; a so-called "aggressive cancer" will grow or metastasise rapidly or significantly affect overall health status and quality of life

20 "adjuvant therapy" refers to treatment involving radiation, chemotherapy (drug treatment), biologic therapy (vaccines) or hormone therapy, or any combination given after primary treatment.

25 "antibody" is intended to include whole antibodies, e.g., of any isotype, and includes fragments thereof which are also specifically reactive with a vertebrate, e.g., mammalian, protein. Antibodies can be fragmented using conventional techniques and the
30 fragments screened for utility in the same manner as described above for whole antibodies. Thus, the term includes segments generated by proteolytic cleavage or prepared recombinant portions of an antibody molecule capable of selectively reacting with a certain protein. Non-limiting examples of such
35

proteolytic and/or recombinant fragments include Fab, F(ab')₂, Fab', Fv, and single chain antibodies (scFv) containing a V[L] and/or V[H] domain joined by a peptide linker. The scFv's may be covalently or non-covalently linked to form antibodies having two or more binding sites. Antibodies may include polyclonal, monoclonal, or other purified preparations of antibodies and recombinant antibodies.

10

"associated with" refers to a disease in a subject which is caused by, contributed to by, or causative of an abnormal level of expression of a protein.

15

"control" comprises for example proteins from a sample of the same patient or from a pool of different patients, or selected among reference proteins which may be already known to be over or under expressed. The expression level of said control can be an average or an absolute value of the expression of reference proteins. These values may be processed in order to accentuate the difference relative to the expression of the proteins according to the invention. The analysis of the over or under expression of proteins can be carried out on sample such as biological material derived from any mammalian cells, including cell lines, xenografts, human tissues preferably breast tissue, etc. The method according to the invention may be performed on sample from a, e.g., cell lines, healthy donors, patients or an animal (for example for veterinary application or preclinical studies).

35

"directly or indirectly labeled" include proteins the sub-constituants of which, i.e., amino acids or

amino acid groups or atoms, are themselves labeled (directly), as well as proteins labeled by the intermediate of any element able to recognize and bind to the targeted protein, e.g., an antibody.

5

"equivalent" includes nucleic acids encoding functionally equivalent proteins. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the nucleic acids of the invention because of the degeneracy of the genetic code.

10

"good-prognosis" and "poor-prognosis" respectively refer to favorable (e.g., remission) or unfavorable (e.g., metastasis, death) patient clinical outcome.

15

"histopathologic features of breast diseases" includes diseases, disorders or conditions known as, lethality or not, affecting breast cells and/or tissues, including but not limited to breast tumours, for example i) non cancerous breast diseases, for example, hyperplasias, metaplasias, fibroadenomas, fibrocystic disease, papillomas, sclerosing adenosis or preneoplastic, or ii) breast cancer. As "breast cancer" one may cite :

20

A) noninvasive breast cancers including i) ductal carcinoma in situ (also called intraductal carcinoma or DCIS), consisting of cancer cells in the lining of the duct ii) Lobular carcinoma in situ, or LCIS (also known as lobular neoplasia);

25

B) Invasive cancer occurring when cancer cells spread beyond the basement membrane which covers the

30

underlying connective tissue in the breast, and which include i) Infiltrating ductal carcinoma that penetrates the wall of a duct and ii) Infiltrating lobular carcinoma which spread through the wall of a lobule and may sometimes appear in both breasts, sometimes in several separate locations.

5 "ImmunoHistoChemistry (IHC)" refers to methods using histochemical localization of immunoreactive substances using antibodies as reagents on cells or tissues by technologies such as, but not limited to flow cytometry, ELISA, Western and Southwestern Blot Analysis, and frozen and paraffin-embedded samples.

10 15 "Nucleic acids" refers to polynucleotides, e.g., isolated, such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides. ESTs, chromosomes, cDNAs, mRNAs, and rRNAs are representative examples of molecules that may be 20 25 referred to as nucleic acids.

30 "over or underexpression" may comprise the detection of difference in the expression of the proteins according to the present invention in relation to at least one control.

"predicting clinical outcome" refers to the ability for a skilled artisan to classify patients into at least two classes "good prognosis" and "bad

prognosis" showing significantly different long-term Metastasis Free Survival (MFS)

5 "Protein" refers to a polypeptide with a primary, secondary, tertiary or quaternary structure, or any portion or modification, e.g., a mutant, or isoform thereof. A "portion" or "modification" of a protein retains at least one biological or antigenic characteristic of a native (wild-type) protein.

10 "Protein microarray" refers to a spatially defined and separated collection of individual proteins immobilised on a solid surface.

15 "Treating" as used herein is intended to encompass treating as well as ameliorating at least one symptom of the condition or disease.

20 **IV - Description of the figures**

Figure 1 represents hierarchical clustering analysis of global protein expression profiles in breast cancer as measured by IHC on TMA.

25 A/ Graphical representation of hierarchical clustering results based on expression profiles of 26 proteins in 552 early breast cancer samples. Each row represents a sample and each column represents a protein. Immunostaining results are depicted according to a color scale: red or brown for strong or moderate positive staining, respectively, green for negative staining, gray for missing data. Dendograms of samples (to the left of matrix) and proteins (above matrix) represent overall similarities in expression profiles. Three major
30
35

clusters of tumors (A1, A2 and B) are shown (A1 and A2 correspond to luminal cells; B corresponds to basal cells). Colored bars to the right and colored branches in the dendrogram indicate the locations of 5 3 sample clusters of interest zoomed in C. B/ Dendrogram of proteins. Two major clusters "P1" (basal/stem cells) and "P2" (luminal/glandular cells) are identified and further divided in 4 smaller clusters designated "proliferation", 10 "mitosis", "ER-related" and "adhesion" cluster, respectively. C/ Expanded view of selected sample clusters showing a partial grouping of tumors with similar histological type (LOB: lobular, DUC: ductal, OTH: other, MIX: mixed; blue bar) or ER 15 status (positive, red bar and negative, orange bar).

Figure 2 represents classification of 552 breast cancer samples based on the expression of the 20 21-protein discriminator set identified by supervised analysis.

A and B/ Correlations between the molecular grouping based on the combined expression of the 21 proteins and the occurrence of metastatic relapse in the learning (A) and the validation (B) set of samples. 25 C/ Supervised classification of all 552 samples using the 21-protein expression signature. Each row of the data matrix (left panel) represents a sample and each column represents a protein. Immunostaining results are depicted according to the color scale used in Figure 1. The 21 proteins, listed above the 30 matrix (ER*: means of three independent ER analyses), are ordered from left to right according to decreasing ΔP (ΔP is the difference between the probability of positive staining and the probability 35 of negative staining in non-metastatic samples).

Tumor samples are numbered from 1 to 552 and are ordered from top to bottom according to their increasing "Metastasis Score" (right panel). The orange dashed line indicates the threshold 0 that separates the two classes of samples, "poor-prognosis" (under the line) and "good-prognosis" (above the line). The middle panel indicates the occurrence (black square) or not (white square) of metastatic relapse for each patient.

10

Figure 3 represents Kaplan-Meier analysis of the metastasis-free survival of patients with breast cancer according to the molecular classification based on the 21-protein expression signature or the St-Gallen and the NIH consensus criteria.

Patients (pts) were classified in the "good-prognosis" class or the "poor-prognosis" class using the 21-protein signature identified by supervised analysis (*A, B, E and F*) or in the "low risk" class or the "high risk" class using the St-Gallen and the NIH consensus criteria (*C and D*). The P-values are calculated using the log-rank test. *A/* Survival of all 552 patients. *B/* Survival of 292 patients with node-negative cancer (N-) and 255 patients with node-positive cancer (N+). The difference of survival is significant between the "good-prognosis" class and the "poor-prognosis" class for the node-negative patients, as well as for the node-positive patients. In contrast, survival is not significantly different between the node-positive patients from the "good-prognosis class" and the node-negative patients from the "poor-prognosis class". *C/* Survival of 292 patients with node-negative cancer (N-) according to the St-Gallen criteria. *D/* Survival of 292 patients with node-negative cancer

(N-) according to the NIH criteria. E/ Survival of 186 patients without any adjuvant chemotherapy (CT) and hormone therapy (HT). F/ Survival of 133 patients who received adjuvant chemotherapy (CT) without hormone therapy (HT).

5

Figure 4 represents expression of proteins studied by IHC on tissue microarrays (TMA).

10

A/ Representative Hematoxylin-Eosin and Safran staining of a paraffin block section (25x30 mm²) from a TMA containing 552 early breast cancer cases with 0.6 mm tumor cores. B/ Immunohistochemical staining of a tumor core for the 21 proteins identified by supervised analysis (magnification x200). C/ Examples of IHC staining for 5 proteins with differential expression in cancer tissue (bottom) compared with normal tissue (top). 1, FHIT expression in cytoplasm in normal lobules, down-regulation in cancer sample (arrow); 2, Apical normal expression of MUC1, down-regulation and miss-localization in the cytoplasm of cancer sample (arrow); 3, Absence of ERBB2 expression in normal lobule (arrow), overexpression on the cytoplasmic membrane in positive cancer sample (arrow); 4, Absence of nuclear expression of Cyclin D1 in normal lobules (arrow), overexpression in nucleus of positive cancer sample (arrow); 5, Normal myoepithelial cells are immunostained by P Cadherin (arrow), overexpression in cancer sample (arrow). Magnification is x400.

15

20

25

30

V - Detailed description of the invention

5

We have combined IHC and TMA to measure the expression levels of selected proteins in a consecutive series of 552 patients with early stage breast cancer. Our aim was to determine protein combinations to refine tumor classification and improve the prognostic classification of disease.

10

V.1) Protein expression profiling identifies subclasses of breast cancer

15

20

25

30

35

Analysis and interpretation of the large amount of data generated (552 samples and 26 antibodies, ~14.000 data points) required the development of bioinformatic tools. As a first step, we applied pre-existing unsupervised hierarchical clustering algorithms as previously reported.¹⁹⁻²⁴ Two recent studies on breast cancer analyzed the expression of 15 proteins in 166 tumors,²² and 13 proteins on 107 samples,¹⁹ respectively. Several of these markers were included in the present work (BCL2, ER, PR, ERBB2, EGFR, Cyclins, Cytokeratins, MIB1, P53), allowing for direct comparison of results. In our analysis, clustering allowed the identification of four major coherent protein clusters designated according to the function of most included proteins: "ER-related cluster", "adhesion cluster", "mitosis cluster" and "proliferation cluster". Correlated expression of proteins may be due to different mechanisms such as coregulation (e.g., ER/BCL2³⁰), functional interaction (e.g., STK6/Taxins^{27, 28}), phenotypic association (e.g., ERBB2/P53³¹) or chromosomal location (e.g., FGFR1/TACC1 located on 8p11). Some co-expressed proteins were previously reported in RNA or protein expression profiling studies. For example, ER, PR, BCL2 and GATA3 clustered together.^{8-10, 13} This "ER-related cluster"

was negatively correlated with the "mitosis" and "proliferation" clusters, in agreement with the higher proliferation index in ER-negative tumors³² and the known proliferation-differentiation balance in carcinomas. The "ER-related cluster" was close to the "adhesion cluster" that included other markers that may correlate positively with ER expression such as FHIT,³³ CK8/18,^{19, 22} CCND1³⁴ and MUC1.⁸ Our "proliferation cluster" had some similarities to that identified by others with the common presence of P53, Ki67, CCNE, ERBB2 and CK5/6¹⁹ or CCNE, ERBB2, EGFR and CK5/6.²² Interestingly, this cluster also included CDH3/P-Cadherin, present in a "basal cluster" identified in gene expression analyses⁹ and previously shown to be overexpressed in a subgroup of breast carcinomas associated with higher proliferation rates and aggressive behavior.³⁵

Hierarchical clustering sorted tumors into three clusters that correlated with relevant histoclinical parameters, including histological type, SBR grade, ER status, ERBB2 status and the presence or absence of peritumoral vascular emboli. Correlations were found between the characteristics of these tumor clusters and their protein expression profiles. For example, the high number of grade III tumors in cluster B, as well as the high number of ERBB2-positive samples, agreed with the frequent strong expression of the "proliferation" cluster - which included ERBB2 - and the "mitosis" cluster in these tumors. Conversely, 99% of cluster A1 samples were ER-positive, and showed a frequent strong expression of the "ER-related" cluster and low expression of the "proliferation cluster".³²

Interestingly, the tumor clusters also correlated with a breast cancer classification

recently proposed in two series of analyses that provided a new conceptual framework of mammary oncogenesis. First, phenotypic analyses have established a three-cell phenotypic classification of breast cancer cells.^{22, 36, 37} These authors suggested that biomarkers such as intermediate filaments cytokeratins (CK), encoded by a large number of keratin genes, are able to distinguish between distinct cell subpopulations within the mammary gland epithelial compartment. It has been proposed that "basal" cells contain mammary gland progenitor cells able to give raise to both "luminal" and "myoepithelial"³⁸ cells.(³⁹ for review) Progenitor cells express type II keratins CK5 and 6. In contrast, differentiated "luminal" cells express type II keratin CK8 and type I keratin CK18, which are also observed in normal simple and glandular epithelia. Luminal cells also express ER.^{10, 11} Use of tissue microarray screening has confirmed this emerging theory.^{19, 22} Second, recent gene expression analyses using DNA microarrays have led to a similar identification of subclasses of breast tumors that corresponded to the phenotypic classification.⁹⁻¹¹ These experiments concurred to establish a distinction between several types of epithelial cells in the mammary gland. The origin of the breast malignant cell remains unknown. Two major types of breast cancer may derive from basal/progenitor or luminal cells, respectively. Alternatively, most tumors may originate from pluripotent stem cells and reach different stages of differentiation.⁴⁰ Our results support this new classification model. Tumor cluster A1 may be approximated to a cluster of luminal cell-like tumors, with frequent strong expression of ER and CK8/18. Cluster B may consist

of tumors with basal/progenitor, ER-negative characteristics, i.e. strong expression of CK5/6 and proliferation markers. A2 tumors, with an intermediate profile, may represent a transitory "baso-luminal" stage, or consist of tumors that have lost ER function. It can be expected that luminal A1 tumors, in which the bulk of cells are more differentiated and express ER-related cluster proteins, are of better prognosis, whereas more undifferentiated and proliferative basal B tumors are associated with poor prognosis. The significant differences in clinical outcome observed between the three defined tumor clusters in this study are consistent with this model and recent studies.^{9-11, 41}

In addition, we show that lobular carcinomas are luminal-like tumors, and consist of differentiated luminal cells that express CK8/18.

V.2) Protein expression profiling predicts
20 clinical outcome of breast cancer

Thus classical unsupervised hierarchical clustering applied to all tested proteins was able to identify biologically and clinically relevant classes of breast cancer. Recently, supervised methods have been successfully applied to gene expression data analysis in parallel with unsupervised approaches.⁴² In a second step, we thus developed a supervised method to identify the best combination within 26 proteins that would further improve the prognostic classification. To our knowledge, our study is the first application of such supervised methods to large-scale IHC data. We identified a 21-protein set which optimally classified patients into two classes ("good-prognosis" and "poor-prognosis class") with

significantly different long-term MFS. Initially identified in a random learning set of 368 patients, this prognostic signature was validated in an independent set of 184 patients, showing its robustness. Our discriminator set included 10 proteins coded by genes identified across recent gene expression studies,⁷⁻¹⁵ as well as other proteins with unclear role in disease progression and sensitivity to systemic therapy. The prognostic value of the signature was increasingly accurate with the addition of other proteins as evidenced by univariate and multivariate analyses, further highlighting the strength of large-scale molecular analyses for understanding tumor heterogeneity through the identification of expression signatures.

The classification based on the 21-protein predictor was associated with a highly significant difference in clinical outcome. The 5-year MFS was 90% for patients of the "good-prognosis class" and only 62% for patients of the "poor-prognosis class". When compared in multivariate analysis with classical prognostic factors and with each tested protein separately, our classification performed significantly better for predicting the occurrence of metastatic relapse. Such prognostic association persisted when applied to patients with lymph node-positive and lymph node-negative cancer. Interestingly, the MFS of node-negative patients from the "poor-prognosis class" was similar to that of node-positive patients from the "good-prognosis class". Notably, our molecular classification performed better than that defined by St-Gallen and NIH criteria for node-negative patients. This finding is of particular significance, since ~75% of node-negative patients candidate for adjuvant

chemotherapy based on the St. Gallen/NIH criteria are currently thought to be over-treated. In the present study, our 21-protein predictor assigned fewer node-negative patients to the "poor-prognosis class", and their clinical outcome was more frequently unfavorable than it was for patients assigned to the high-risk class defined by St-Gallen or NIH criteria. Our predictor also performed well in patients irrespective of ER status. The 5-year MFS was 90% for ER-positive patients from the "good-prognosis class", and 58% for ER-positive patients from the "poor-prognosis class", suggesting our 21-protein set may provide more accurate clinical information than ER status alone, possibly reflecting functional differences in the ER pathway.

Additionally, our molecular classification conserved its predictive impact for patients independent of adjuvant systemic therapy. Since distant metastasis may be influenced by adjuvant therapy, we separately analyzed the 186 patients who did not receive any chemo- and hormone therapy, as well as the 133 patients who exclusively received adjuvant chemotherapy with anthracyclin-based regimen in most cases. Interestingly, we found within the group of 186 untreated patients an odds ratio of 7.45 for metastatic relapse in the "poor-prognosis class" when compared with patients of the "good-prognosis class". Similar discrimination was observed within the 133 patients treated with chemotherapy alone with a corresponding odds ratio of 3. Thus, the 21-protein signature may facilitate the selection of appropriate treatment options in early breast cancer patients. It may be an important clinical tool to circumvent unnecessary, toxic and costly treatment of node-negative patients, and it may help for

selecting, among patients who need adjuvant chemotherapy, those who might benefit from standard protocol and those who would be candidates to other protocol or other form of systemic therapy.

5

VI - Materials and Methods

VI.1) Patients and histological samples

A consecutive series of 552 women with early (stage I, II or III) breast cancer treated at the Institut Paoli-Calmettes before December 1999 was studied using the TMA technology. The stage of disease was defined according to TNM classification (Union Internationale Contre le Cancer, UICC, TNM, 5th edition). Patients with locally advanced, inflammatory or metastatic disease, or with previous history of cancer were not included. Tumors were invasive adenocarcinomas including, according to the WHO histological typing, 388 ductal carcinomas (70%), 72 lobular (13%), 24 mixed (4%), 40 tubular (8%), 8 medullary (1%) and 20 other types (4%). Clinical annotation of each sample included patient age, axillary lymph node status, pathological tumor size, Scarff-Bloom-Richardson (SBR) grade, peritumoral vascular invasion, estrogen receptor (ER), progesterone receptor (PR) and ERBB2 status as evaluated by IHC with positivity cut-off values of 1% for hormone receptors and with 2 or 3+ score (HercepTest kit scoring guidelines) for ERBB2. The characteristics of patients are listed in Table 1 (see first column only).

Table 1. Histoclinical characteristics of 552 breast cancer patients, according to the membership to the

"good-prognosis" or the "poor-prognosis class" as defined using the expression of the 21-protein set.

Characteristics	All patients (N=552)	Good-prognosis class* (N=358)	Poor-prognosis class* (N=194)	P-value **
	no. of patients (% of evaluated cases)			
Age, years				
≤50	153 (28)	100 (28)	53 (27)	0.87
>50	399 (72)	258 (72)	141 (73)	
Lymph node metastasis				0.12
0	292 (53)	199 (56)	93 (49)	
1-3	158 (29)	103 (29)	55 (29)	
>3	97 (18)	55 (15)	42 (22)	
Pathological tumor size				0.69
pT1	245 (45)	171 (48)	74 (38)	
pT2	228 (42)	136 (38)	92 (48)	
pT3	75 (13)	48 (14)	27 (14)	
SBR grade				<0.0001
I	181 (33)	150 (42)	31 (16)	
II	229 (42)	153 (43)	76 (39)	
III	139 (25)	53 (15)	86 (45)	
Peritumoral vascular invasion				0.10
absent	345 (63)	233 (65)	112 (58)	
present	206 (37)	124 (35)	82 (42)	
ER status				<0.0001
negative	129 (23)	12 (4)	117 (60)	
positive	422 (77)	345 (96)	77 (40)	
PR status				<0.0001
negative	195 (35)	67 (19)	128 (66)	
positive	355 (65)	290 (81)	65 (34)	
ERBB2 status				<0.0001
negative	461 (87)	317 (92)	144 (77)	
positive	70 (13)	27 (8)	43 (23)	
Chemotherapy				0.0001
no	291 (53)	208 (58)	83 (43)	
yes	261 (47)	150 (42)	111 (57)	
Hormone therapy				<0.0001
no	286 (52)	161 (47)	125 (71)	
yes	233 (48)	181 (53)	52 (29)	
Follow-up***, months	57 (2, 182)	56 (3, 181)	58 (2, 182)	NS
median (range)				

5-year MFS % [95%CI]	80 [76.2 - 83.7]	90 [86.0 - 93.3]	62 [54.7 - 70.0]	<0.0001
-------------------------	------------------	------------------	------------------	---------

*, as defined using the 21-protein signature;
 **, P-values for the comparison of numbers of patients were calculated using the Chi-2 test, and P-values for
 5 the comparison of metastasis-free survival (MFS) were calculated using the log-rank test; NS, not significant;
 ***, calculated, for the 450 patients who did not experience metastatic relapse as a first event, from the date of diagnosis to the time of last follow-up;
 10 CI denotes confidence interval.

Patients were treated according to the following guidelines : all had primary surgery that included complete resection of breast tumor (modified radical mastectomy in 28% of cases and lumpectomy in 72%) and axillary lymph node dissection; 96% of patients (including 100% of those treated with breast-conservative surgery) received adjuvant local-regional radiotherapy; 47% were given adjuvant chemotherapy (anthracyclin-based regimen in most cases), and 42% received adjuvant hormone treatment (tamoxifen for most cases). After completion of local-regional treatment, patients were evaluated at least twice per year for the first 25 years and at least annually thereafter. The median follow-up was 57 months (range, 2 to 182) after diagnosis for the 450 patients who did not experience metastatic relapse as a first event, 37 months (range, 4 to 151) for the 102 patients with metastasis as first event, and 51 months (range, 2 to 182) for all patients. The 5-year MFS rate was 80% [95%CI 76.2 - 83.7].

VI.2) Tissue microarrays construction

TMA's were prepared as previously described²⁵ with slight modifications. For each tumor, three representative areas from the primary tumor were carefully selected from a hematoxylin-eosin stained section of a donor block. Core cylinders with a diameter of 0.6 mm each were punched from each of these areas and deposited into three separate recipient paraffin blocks using a specific arraying device (Beecher Instruments, Silver Spring, MD). The technique of TMA allows the analysis of tumors and controls under identical experimental conditions. In addition to tumor tissues, the recipient block also received 10 normal breast tissue samples from 10 healthy women that underwent reductive mammary surgery and pellets from nine mammary cell lines. Five- μm sections of the resulting TMA block were made and used for IHC analysis after transfer onto glass slides. We previously assessed the reliability of the method by comparison with the standard immunohistochemical method for the usual prognostic parameters; the value of the kappa test was 0.95.²⁵

VI.3) Selection of the 26 markers

The selection of the proteins was done according to the following criteria: known or potential importance in breast cancer and availability of a corresponding antibody that performed well in IHC on paraffin-embedded tissues. Twenty-six proteins were selected including hormone receptors (ER, PR), subclass markers (Cytokeratins), oncogenes and proliferation proteins (ERBB family members, BCL2, Cyclins, MIB1, FGFR1, Aurora A, Taxins), tumor suppressors (P53, FHIT), adhesion molecules (Cadherins, Catenins, Afadin), proteins from oncogenes of amplified genomic regions (ERBB2,

CCND1, STK6), and other potential prognostic markers identified in specific studies or previous DNA microarray experiments (CCNE, GATA3, MUC1). Twelve out of the 26 proteins were mentioned as potential significant genes in RNA expression profiling studies in breast cancer.⁶⁻¹⁵ The characteristics of the antibodies used are listed in Table 4. When available, several antibodies were studied for comparison, and only the reagents that gave the best quality data were kept for the global analysis.

Table 2. Proteins tested by immunohistochemistry on TMAs and characteristics of the corresponding antibodies.

Protein (acronym)	Antibody	Origin	Clone	Pretreatment	Dilution
1 Adhesion molecule Afadin (AF6)	Mmab	Transduction laboratories	35	DTRS (40 min, 98°C)	1/50
2 Aurora A kinase (STK6/STK15)	Mmab	C. Prigent, Rennes	/	DTRS (40 min, 98°C)	1/25
3 α -Catenin (CTNNA1)	Mmab	Zymed Laboratories	α CAT-7A4	Citrate buffer (40min, 98°C)	1/200
4 β -Catenin (CTNNB1)	Mmab	Transduction laboratories	14	Citrate buffer (40min, 98°C)	1/2500
5 Anti-apoptotic BCL2	Mmab	Dako Corporation	124	Citrate buffer (40min, 98°C)	1/100
6 Cyclin D1 (CCND1)	Mmab	Zymed laboratories	AM29	Citrate buffer (40min, 98°C)	1/200
7 Cyclin E (CCNE)	Mmab	Novocastria laboratories	13A3	Citrate buffer (40min, 98°C)	1/200
8 Cytokeratins 5 and 6 (CK5/6)	Mmab	Dako Corporation	D5/16B4	DTRS (40 min, 98°C)	1/50
9 Cytokeratins 8 and 18 (CK8/18)	Mmab	Zymed Laboratories	Zym5.2	DTRS (40 min, 98°C)	1/200
10 Adhesion molecule E-Cadherin (CDH1)	Mmab	Transduction laboratories	36	Citrate buffer (40min, 98°C)	1/2000
11 Epidermal growth factor receptor (EGFR)	Mmab	Zymed Laboratories	31G7	Pepsin (30 min, 37°C)	1/20
12 Tyrosine kinase	Mmab	Novocastria	CB 11	Citrate	1/500

	receptor ERBB2		Laboratories	buffer (40min, 98°C)	
13	Tyrosine kinase receptor ERBB3	Mmab	NeoMarkers	SGP1 None	1/40
14	Tyrosine kinase receptor ERBB4	Mmab	NeoMarkers	HFR-1 None	1/50
15	Estrogen receptor (ER)	Mmab	Novocastra Laboratories	6F11 (40min, 98°C)	Citrate buffer 1/60
16	Fibroblast growth factor receptor 1 (FGFR1)	Rpab	Santa Cruz Biotechnology	Sc-121 (40 min, 98°C)	DTRS 1/200
17	Fragile histidine triad (FHIT)	Rpab	Zymed Laboratories	ZR44 (40min, 98°C)	Citrate buffer 1/300
18	Transcription factor GATA3	Mmab	Santa Cruz Biotechnology	SC-268 (40min, 98°C)	Citrate buffer 1/100
19	MIB1/Ki67	Mmab	Dako Corporation	Ki-67 (40min, 98°C)	Citrate buffer 1/100
20	Mucin 1 (MUC1)	Mmab	Transgene	H23 (40min, 98°C)	None 1/1000
21	Tumor suppressor P53	Mmab	Immunotech	DO-1 (40min, 98°C)	Citrate buffer 1/4
22	Adhesion molecule P-Cadherin (CDH3)	Mmab	Transduction Laboratories	56 (40 min, 98°C)	DTRS 1/75
23	Progesterone receptor (PR)	Mmab	Dako Corporation	PgR 636 (40min, 98°C)	Citrate buffer 1/80
24	Transforming acidic	Rpab	Upstate	07-229 DTRS	1/200

	coiled-coil 1/Taxin 1 (TACC1)	Rpab	Biotechnology	(40 min, 98°C)	
25	Transforming acidic coiled-coil 2/Taxin 2 (TACC2)	Rpab	Upstate Biotechnology	07-228 (40 min, 98°C)	DTRS 1/40
26	Transforming acidic coiled-coil 3/Taxin 3 (TACC3)	Rpab	Upstate Biotechnology	07-233 (40 min, 98°C)	DTRS 1/100

Mmab: mouse monoclonal antibody; Rpab: rabbit polyclonal antibody; DTRS: Dako target retrieval solution.

VI.4) Immunohistochemical analysis

IHC was carried out on five- μm sections of tissue
5 fixed in alcohol formalin for 24 h and embedded in paraffin. Sections were deparaffinized in Histolemon (Carlo Erba Reagenti, Rodano, Italy) and rehydrated in graded alcohol. Antigen retrieval was accomplished by incubating the sections in pre-treatment solutions
10 depending on the antibody used. Pretreatment conditions are listed in Table 2. The reactions were carried out using an autoimmunostainer (Dako Autostainer). Staining was performed at room temperature as follows: rehydrated tissues were washed in phosphate buffer,
15 followed by quenching of endogenous peroxidase activity by treatment with 0.1% H₂O₂, slides, incubated with blocking serum (Dako) for 30 min., then with the affinity-purified antibody for one hour. After washes, slides were sequentially incubated with biotinylated
20 antibody against rabbit IgG for 20 min. followed by streptadivin-conjugated peroxidase (Dako LSAB^R2 kit), then visualized with Diaminobenzidine (3-amino-9-ethylcarbazole). Slides were counter-stained with hematoxylin, coverslipped using Aquatex (Merck,
25 Darmstadt, Germany) mounting solution, then evaluated under a light microscope by two pathologists. The results were expressed in terms of percentage (P) and intensity (I) of positive cells as previously described²⁵. For each sample, the mean of the score of a minimum
30 of two core biopsies was calculated. The results were then scored by the quick score (Q) (Q = P X I), except for ERBB2 status that was evaluated with the Dako scale (HercepTestTM kit scoring guidelines).

Quick score allowed separating tumors into two or three classes. Homogeneous classes were defined by grouping samples with an equivalent staining level according to the distribution curves as described.²⁵

5 Two classes (negative and positive) were defined for Afadin, α and β Catenins, BCL2, Cyclins D1 and E, Cytokeratins 5/6 and 8/18, EGFR, ERBB3, ERBB4, FGFR1, GATA3, MIB1, P53, P-Cadherin, PR and TACC3, with a positivity cut-off value of $Q = 1$, except for Cyclin D1 and MIB1 with a positivity cut-off value of 10 and 20, respectively.

10 Three classes were defined (negative, moderate and strong staining) for Aurora A, E-Cadherin, ER, FHIT, MUC1, TACC1, and TACC2, with negative ($Q = 0$), moderate ($0 < Q \leq 100$) or strong expression ($100 < Q \leq 300$).

15 For ERBB2, three classes (0/1+, 2+, 3+) were obtained with the Dako scale.

VI.5) Data analysis

A combination of exploratory unsupervised and supervised bioinformatic methods was used to analyze these immunohistochemical profiles. First, we applied unsupervised hierarchical clustering similar to that used in gene expression profiling studies. Data were reformatted using the following scoring system: -2 designated negative staining, 1 weakly positive staining, 2 strongly positive staining and missing data were left blank in the scored table. Hierarchical clustering investigates relationships between samples and between proteins, based on the similarity of sample immunoreactive scores. We used the Cluster program (average-linkage with Pearson correlation as similarity

metric) and results were displayed with the TreeView software.²⁶

We then performed supervised analysis to identify the protein-set that best distinguished between two classes of samples with different clinical outcome. To simplify the analyses, the IHC scores were recorded as negative (negative staining) or positive (weakly and strong positive staining). The classifier was derived through training on a subset of chosen samples (2/3 of population, learning set) and then validated on the remaining subset (1/3 of population, validation set). The assignment of samples to each set was random, but the ratio between tumors with and without metastatic relapse was preserved. An exhaustive testing comprising all combinations of 1 to 5 proteins, as well as the complementary combinations of 21 to 25 proteins was performed to assess their ability to classify tumors into 2 classes ("poor-prognosis" and "good-prognosis") in agreement with their clinical outcome.

Using the protein expression scores of each combination, we developed a "Metastasis Scoring" system that assigned to each tumor a probability to belong to the "poor-prognosis class" or the "good-prognosis class". Consider a combination of N proteins P_1, \dots, P_N (where N ranges from 1 to 5 and 21 to 26) and two predefined classes X, Y of tumors within the learning set: $X = \{X_1, \dots, X_K\}$ includes samples with metastatic relapse during the follow-up and $Y = \{Y_1, \dots, Y_M\}$ includes samples without any metastatic relapse. For each protein combination tested, one tumor is represented as a ternary vector (e.g. $X_1 = \{X_1(P_1), \dots, X_1(P_N)\}$) where each

component is scored 0 for missing data or +1/-1 for positive/negative IHC staining. Every tumor Z has a score $S(Z)$ defined as follows. For each protein P_i , we compute the frequencies of +1/-1 value in the X class (adjusted to avoid a 0 probability):

$$f_X^i(+1) = \frac{\text{card}\{k : X_k(P_i) = +1\} + 1}{\text{card}\{k : X_k(P_i) \neq 0\} + 2} \quad \text{and}$$

$$f_X^i(-1) = \frac{\text{card}\{k : X_k(P_i) = -1\} + 1}{\text{card}\{k : X_k(P_i) \neq 0\} + 2}$$

where, for instance, $\text{card}\{k : X_k(P_i) = +1\}$ is the number of X tumors with positive IHC staining for protein P_i . Similarly we compute the frequencies $f_Y^i(+1)$ and $f_Y^i(-1)$ in the Y class and we define $f_Y^i(0) = 1$. The Metastasis Score of tumor Z is the log ratio of the joint probabilities:

$$S(Z) = \sum_{i=1}^N \log(f_X^i(Z(P_i))) - \sum_{i=1}^N \log(f_Y^i(Z(P_i))).$$

Samples were then sorted according to their $S(Z)$ score. The natural threshold that divides the population in 2 classes is $S=0$: if $S(Z) > 0$ then Z is more similar to the class X and is predicted to belong to the "poor-prognosis class" and if $S(Z) < 0$ then Z is more similar to the class Y and is predicted to belong to the "good-prognosis class". The number of misclassifications (error rate) was defined as the number of X tumors classified in the "good-prognosis class" plus the number of Y tumors classified in the "poor-prognosis class". The best classifier protein-set was that with the minimal rate of misclassified tumors.

5

Once identified, the prognostic power of the classifier was tested on the validation set by classifying the remaining independent tumors using the same approach. Finally, it was assessed on the whole population. For each tumor set, the prognostic impact was further estimated by univariate analyses that compared the rate of metastatic relapses within the two molecularly defined classes of tumors (Fisher exact test).

10

VI.6) Statistical methods

15

20

25

Distributions of molecular markers and other categorical variables were compared using either the standard Chi-2 test or Fisher exact test. The follow-up was calculated from the date of diagnosis to the time of metastasis as first event or time of last follow-up for censored patients. The end point was the metastasis-free survival (MFS), calculated from the date of diagnosis, first metastasis being scored as an event. All other patients were censored at the time of the last follow-up, death, recurrence of local or regional disease, or development of a second primary cancer, including contralateral breast cancer. Survival curves were derived from Kaplan-Meier estimates and were compared by log-rank test. The influence of molecular grouping, adjusted for other factors including classical prognostic factors and significant IHC measurement, was assessed in multivariate analysis by the Cox proportional hazard models. Survival rates and odds ratios (OR) are presented with their 95% confidence intervals (95%CI). Statistical tests were two-sided at the 5% level of significance. All statistical tests were done using SAS Version 8.02.

VII - Results

VII.1) Expression protein profiling of breast cancers using tissue microarrays.

The expression of 26 proteins was studied by IHC on TMA containing 552 early stage breast tumor samples and controls (Figure 4A). As expected, staining for all antibodies was homogeneous among the 10 normal breast samples (data not shown), but much more heterogeneous for tumor samples. Sixteen proteins were underexpressed in 12% (for MUC1) to 60% (for Aurora A) of cases, and overexpressed for 10 proteins in 11% (for Ki67/MIB1) to 66% (for ERBB4) of cases in cancerous tissues compared to normal samples. Examples of IHC staining are shown in Figure 4 (panels B and C). Results are summarized in Table 3.

Table 3. Expression of proteins tested by immunohistochemistry in 552 early breast cancers deposited on TMA and Kaplan-Meier analysis of the metastasis-free survival (MFS).

Protein	Type of alteration in tumor samples*, frequency of alteration*, cell sublocalization	No. of patients	5-year MFS [95%CI]	P-value**
Afadin	negative positive	Downregulated, 14%, membrane and cytoplasm	48 300	0.13
Aurora A	negative positive	Downregulated, 60%, nucleus	267 177	0.25
α -Catenin	negative positive	Downregulated, 30%, membrane	105 267	66.9 [56.8 - 77.0] 84.9 [80.1 - 89.7] 0.0046
β -Catenin	negative positive	Downregulated, 40%, membrane	152 229	72.2 [64.2 - 80.1] 82.1 [76.9 - 88.8] 0.031
BCL2	negative positive	Downregulated, 21%, cytoplasm	88 324	57.6 [45.3 - 69.9] 83.9 [79.4 - 88.4] <0.0001
Cyclin D1	≤ 10 > 10	Upregulated, 21%, nucleus	380 101	0.82
Cyclin E	negative positive	Upregulated, 15%, nucleus	363 66	0.44
Cytokeratin 5/6	negative positive	Upregulated, 32%, membrane and cytoplasm	246 125	0.06

Cytokeratin 8/18	negative positive	Downregulated, 14%, membrane and cytoplasm	29 456		0.07
E-Cadherin	negative positive	Downregulated, 17%, membrane	61 424		0.41
EGFR	negative positive	Upregulated, 21%, membrane	349 92		0.45
ERBB2	0 - 1 2 - 3	Upregulated, 12 %, membrane	433 60	81.9 [77.8 - 86.0] 64.2 [48.8 - 79.6]	0.030
ERBB3	negative positive	Upregulated, 58%, cytoplasm and membrane	158 223		0.29
ERBB4	negative positive	Upregulated, 66%, cytoplasm and membrane	135 260		0.99
Estrogen receptor	negative positive	Downregulated, 24%, nucleus	133 408	67.0 [58.1 - 75.9] 85.2 [81.3 - 89.1]	<0.0001
FGFR1	negative positive	Upregulated, 45%, cytoplasm and membrane	193 233		0.92
FHIT	negative positive	Downregulated, 16%, cytoplasm	69 353		0.37
GATA3	negative positive	Downregulated, 45%, nucleus	170 268	69.7 [61.9 - 77.5] 85.1 [80.3 - 89.9]	0.0006
MIB1/Ki67	≤ 20 > 20	Upregulated, 11%, nucleus	406 53	83.4 [79.2 - 87.5] 56.0 [39.4 - 72.5]	<0.0001
Mucin 1	negative positive	Downregulated, 12%, cytoplasm and membrane	53 390		0.22

P53	negative positive	Upregulated, 26%, nucleus	383 132	82.2 [77.8 - 86.5] 71.2 [62.5 - 80.0]	0.003
P-Cadherin	negative positive	Downregulated, 55 %, membrane	248 207		0.28
Progesterone receptor	negative positive	Downregulated, 36%, nucleus	185 333	71.7 [64.4 - 79.0] 84.9 [80.5 - 89.3]	0.0007
TACC1	negative positive	Downregulated, 47%, cytoplasm	208 231		0.88
TACC2	negative positive	Downregulated, 27%, cytoplasm	107 288	72.8 [63.7 - 81.9] 80.3 [74.8 - 85.7]	0.048
TACC3	negative positive	Downregulated, 39%, cytoplasm	184 286		0.20

* , as compared to 10 normal breast samples.
 ** , p-values for the comparison of MFS were calculated using the log-rank test.
 CI denotes confidence interval.

5 VII.2) Unsupervised hierarchical classification
of 552 breast tumors upon protein expression
profiling.

VII.2.1) Hierarchical clustering

The overall expression patterns for the 552 samples were first analyzed with hierarchical clustering. Results are displayed in a color-coded matrix in Figure 1A. The clustering algorithm orders proteins on the horizontal axis and samples on the vertical axis on the basis of similarity of their expression profiles. This similarity is shown as a dendrogram where the length of branch between two elements reflects their degree of relatedness. Protein expression scores are represented according to a color scale: red for strong positive staining, brown for weak positive staining and green for negative staining. Despite significantly heterogeneous expression, such combinatorial analysis and color display highlighted groups of correlated proteins across correlated samples.

Figure 1B displays the dendrogram of related proteins. As expected, the three interpretations of ER staining made independently by two pathologists were highly correlated (R^2 between 0.87 and 0.96) (Figure 1C, middle and bottom panels). Furthermore, there was a high degree of concordance for expression of ER between IHC on full sections and on TMA ($p<0.0001$, Chi-2 test). Two major protein clusters - designated "P1" and "P2" - were identified (Figure 1B). These clusters were further divided into smaller sub-groups including a cluster (thereafter designated "ER-related cluster") of ER-associated proteins (PR, BCL2, GATA3) and an

"adhesion cluster" (E-Cadherin, β -Catenin, Afadin). We²⁷ have demonstrated that Aurora A (STK6) and Taxins (TACC1-3) are interacting partners and involved in cell division. This translated in the formation of a third cluster (thereafter designated "mitosis cluster"). The fourth cluster (thereafter designated "proliferation cluster") defined by the routinely used marker Ki67/MIB1, revealed that proteins such as EGFR, ERBB2, P53 and the G1 cyclin CCNE are preferentially overexpressed in tumors undergoing rapid growth.

The combined protein expression patterns defined two major clusters of tumors designated cluster A (462 cases) and cluster B (89 cases) in Figure 1 (1 case that clustered outside of the 2 clusters was excluded from further analysis). Cluster A could be further subdivided into two subclusters, A1 (393 cases) and A2 (89 cases). Globally, cluster A1 tumors displayed a strong expression of the "ER cluster" and the "adhesion cluster" and a low expression of the "proliferation cluster" in most of cases, whereas the "mitosis cluster" was strongly expressed in ~50% of samples. In general, cluster B tumors displayed overall a low expression of the "ER cluster" but a strong expression of the three other protein clusters. Cluster A2 included ER-positive and ER-negative tumors that displayed an intermediate profile characterized overall by strong expression of the "adhesion cluster" and a low expression of the "ER cluster", the "proliferation cluster" and the "mitosis cluster".

VII.2.2) Correlation with histoclinical parameters and survival

We identified correlations between tumor clusters and relevant biopathological parameters. In each cluster, the most frequent histological type was the ductal type; however in cluster A1, 19% of samples were of the lobular type compared with 12% in cluster A2 and only 7% in cluster B ($p=0.03$; Chi-2 test). Figure 1C (top panel) shows, within cluster A1, a subcluster of 24 tumors that includes 21 lobular or mixed (lobular/ductal) carcinomas with low expression of E-Cadherin, consistent with a previous report.²⁹ Correlation also existed with SBR grade; in cluster A1, 41% of cases were grade I and 15% were grade III compared with 23% and 35% in cluster A2, and 7% and 63% in cluster B ($p<0.0001$; Chi-2 test), respectively. In cluster B, samples were more likely to be ERBB2-positive (2+ or 3+ in IHC, 36% of cases) compared with 8% in cluster A1 and 12% in cluster A2 ($p<0.0001$, Chi-2 test). Conversely, cluster A1 samples were more likely to be ER-positive (99% of cases) compared with 35% in cluster A2 and 10% in cluster B ($p<0.0001$, Chi-2 test). Finally, peritumoral vascular emboli were more frequent in A2 tumors (53% of cases) than in B (37%) and A1 (35%) tumors ($p=0.02$, Chi-2 test). Interestingly, no correlation was found with age of patients, pathological size of tumors, and axillary lymph node status.

Importantly, the tumor clusters correlated with clinical outcome. With a median follow-up of 57 months, the 5-year MFS was significantly different ($p<0.0001$, log-rank test) between cluster A1 (54 metastases, 86% MFS [95%CI 82.1 - 89.9]), cluster A2 (21 metastases, 68% MFS [95%CI 79.9 - 56.5]) and cluster B (26 metastases, 66% MFS [95%CI 54.3 - 77.6]) (data not shown).

VII.3) Supervised analysis and clinical outcome

5 We developed a supervised analysis method to search for smaller sets of discriminator proteins that might improve our prognostic classification. Analysis was conducted using two equivalent but independent tumor sets (learning and validation sets).

10 VII.3.1) Supervised analysis and classification of patients

The learning set of samples ($n=368$) allowed the identification of a combination of proteins (protein expression signature) that correlated with long-term MFS. The number of proteins in the "metastatic predictor" was optimized by iteratively testing all combinations of 1 to 5 proteins and the complementary combinations of 21 to 25 proteins and by assessing their ability for correct classification of samples using a "Metastatic Score". The optimal combination for these tumors contained 21 proteins (Figure 2C). Examples of IHC staining for these 21 proteins are shown in Figure 4B. Samples from the learning set were ordered using the "Metastatic Score". Two classes of samples ("poor-prognosis class", positive scores and "good-prognosis class", negative scores) were defined using a cut-off value of 0. As shown in Figure 2A, the classifier predicted rather successfully the actual clinical outcome of patients: 47 out of the 128 patients (37%) with positive score displayed metastatic relapse whereas only 21 out of the 240 (9%) with negative score experienced metastasis during follow-up (odds ratio, OR=6.1 [95%CI 3.3 - 11.3], $p<0.0001$, Fisher exact test).

We then shown the ability of this multiprotein signature to predict prognosis in an independent set of 184 patients (validation set). Using the same threshold for the "Metastatic Score" previously described, we identified two classes of patients that strongly correlated with clinical outcome. There were 24 metastatic relapses out of the 63 patients (38%) in the "poor-prognosis class" and only 10 out of the 121 (8%) in the "good-prognosis class" (odds ratio, OR=6.8 [95%CI 2.8 - 17.3], p<0.0001, Fisher exact test) (Figure 2B). These results confirmed and validated the predictive capacity and robustness of our 21-protein signature.

When all 552 cases (learning and validation cases) were analyzed together, the predictor correlated well with long-term MFS. Figure 2C shows the expression profiles of the 21 proteins in the 552 tumors in a color-coded matrix. Samples are ordered from top to bottom according to their increasing "Metastatic Score" and proteins from left to right according to decreasing ΔP (ΔP is the difference between the probability of positive staining and the probability of negative staining in non-metastatic samples). The orange dashed line indicates the threshold 0 that separates the two classes, "good-prognosis" (above the line) and "poor-prognosis" (under the line).

VII.3.2) Correlation of molecular classification with histoclinical parameters and survival

Table 1 (see the three last columns) shows the characteristics of patients in each class. The histoclinical parameters significantly associated with this classification were SBR grade (p<0.0001,

Chi-2 test), hormone receptor status ($p<0.0001$, Fisher exact test), ERBB2 status ($p<0.0001$, Fisher exact test), and whether patients received adjuvant chemotherapy ($p=0.001$, Fisher exact test) or hormone therapy ($p<0.0001$, Fisher exact test). There was no correlation with patient age, tumor size, and number of involved lymph nodes. In contrast, a strong correlation with clinical outcome was observed (Figure 2C): 65 of 194 patients (34%) assigned to the "poor-prognosis class" displayed metastatic relapse whereas only 37 of 358 (10%) assigned to the "good-prognosis class" experienced metastasis during follow-up (odds ratio, OR=4.4 [95%CI 2.7 - 7.0], $p<0.0001$, Fisher exact test). The 5-year MFS was 62% [95%CI 54.7 - 70.0] in the "poor-prognosis class", and 90% [95%CI 86.0 - 93.3] in the "good-prognosis class" ($p<0.0001$, log-rank test) (Figure 3A).

20 VII.3.3) Survival and lymph node status

Our protein expression signature also classified the 255 patients with node-positive disease into two classes that correlated with clinical outcome. In the "good-prognosis class", 28 out of 158 patients experienced metastatic relapse during follow-up as compared with 43 out of 97 in the "poor-prognosis class" (odds ratio, OR=3.7 [95%CI 2.0 - 6.8], $p<0.0001$, Fisher exact test) (Figure 3B).

30 The same was true for the 292 patients with node-negative breast cancer. In this group, the odds ratio for metastasis was 6.5 ([95%CI 2.7 - 16.8], $p<0.0001$, Fisher exact test) among the 93 women from the "poor-prognosis class", as compared with the 199 women from the "good-prognosis class" (Figure 3B).

As shown, there was no significant difference for MFS between the 158 node-positive patients from the "good-prognosis class" and the 93 node-negative patients from the "poor-prognosis class" ($p=0.142$, log-rank test).

We compared our prognostic classification of node-negative patients with those provided by the consensus criteria established during the St-Gallen and NIH conferences.^{3, 4} These criteria classified all 292 patients into two groups (low risk versus high risk) (Figures 3C and 3D). Our multiprotein signature classified many more patients into the "good-prognosis class" (199 vs 80 vs 43, respectively) and less patients in the "poor-prognosis class" (93 vs 209 vs 245) as compared with St-Gallen and NIH classifications, and interestingly, with a percentage of metastatic relapse similar in the classes with low risk (4.5% vs 5% vs 7%, respectively), but greater in the classes with high risk (24% vs 13% vs 11%, respectively). In fact, the low-risk group and the high-risk group defined according to consensual criteria could further be subdivided in prognostic subgroups when the 21-protein signature was applied (data not shown).

VII.3.4) Survival and estrogen receptor status.

The same analysis was separately applied to ER-positive and ER-negative tumors. In the ER-positive group ($n=422$), 35 of 345 patients from the "good-prognosis class" displayed metastatic relapse as compared with 29 of 77 from the "poor-prognosis class" (odds ratio, $OR=5.4$ [95%CI 2.8 - 9.9], $p=<0.0001$, Fisher exact test). The corresponding 5-year MFS were 90% [95%CI 85.9 - 93.3] and 58% [95%CI

45.4 - 70.6], respectively ($p<0.0001$, log-rank test) (data not shown). The same trend was observed, although not significant ($p=0.21$, log-rank test), for the 129 ER-negative tumors with 5-year MFS of 5 91% [95%CI 76.0 - 100.0] and 66% [95%CI 56.0 - 75.1], respectively.

VII.3.5) Survival and adjuvant systemic therapy

Since the occurrence of metastatic relapse may 10 be influenced by the delivery of adjuvant systemic therapy, the classification based on our 21-protein signature was applied to 186 women who received neither chemotherapy nor hormone therapy after local-regional treatment. Importantly, the 21- 15 protein signature successfully predicted prognosis in these patients: 6 metastatic relapses of 119 patients in the "good-prognosis class" and 19 of 67 in the "poor-prognosis class" (odds ratio, OR=7.4 [95%CI 2.6 - 23.9], $p<0.0001$, Fisher exact test) 20 (Figure 3E).

Similar results were observed when we focused 25 on the 133 patients who received adjuvant chemotherapy without hormone therapy. In the "good-prognosis class", 12 of the 58 patients displayed metastatic relapse whereas 33 of 75 experienced metastasis in the "poor-prognosis class" (odds ratio, OR=3 [95%CI 1.3 - 7.2], $p=0.006$ Fisher exact test) (Figure 3F).

30 VII.3.6) Uni- and multivariate prognostic analysis

We finally compared the prognostic ability of 35 our molecular grouping of tumors with classical histoclinical factors and individual protein markers. In univariate analysis, the histoclinical

5 factors that correlated with MFS ($p<0.05$, log-lank test) were pathological tumor size (≤ 20 mm, > 20), tumor grade (SBR I, II, III), number of positive axillary lymph nodes (0, 1-3, ≥ 4), and peritumoral vascular invasion (negative, positive). Proteins significantly correlated to MFS were BCL2 ($p<0.0001$), GATA3 ($p=0.0006$), MIB1 ($p<0.0001$), ER ($p<0.0001$), PR ($p=0.0007$), P53 ($p=0.003$) and α -Catenin ($p=0.005$) (Table 4).

10

Table 4. Cox proportional-hazards multivariate analyses in metastasis-free survival (n=552).

Variable	Hazard ratio [95%CI]	P-value
Molecular classification (21-protein set)		
"good-prognosis class"	1	<0.0001
"poor-prognosis class"	2.20 [1.25 - 3.89]	
Tumor size		
≤ 20 mm	1	
> 20 mm	3.17 [1.74 - 5.75]	0.0003
Axillary lymph node metastasis		
≤ 3	1	0.0018
> 3	2.48 [1.45 - 4.25]	
MIB1/Ki67 status		
negative	1	
positive	2.38 [1.30 - 4.33]	0.0030
Hormone therapy		
no	1	
yes	0.48 [0.27 - 0.87]	0.0137

15

CI denotes confidence interval.

The influence on the risk of distant metastasis of our multiprotein-based grouping, adjusted for other prognostic factors, was assessed in multivariate analysis by the Cox proportional hazards model. The parameters entered in the model were dichotomised and included the classification based on the discriminator 21-protein set ("good-prognosis class" and "poor-prognosis class"), age of patients (≤ 50 years, > 50 years), number of positive axillary lymph nodes (0, 1-3, ≥ 4), pathological tumor size (≤ 20 mm, > 20), tumor grade (SBR I, II, III), estrogen receptor status (negative, positive), progesterone receptor status (negative, positive), peritumoral vascular invasion (negative, positive), chemotherapy (delivery or not), hormone therapy (delivery or not) and each of the proteins (negative, positive) significantly associated with survival in univariate analyses. Results are shown in Table 4. Several independent factors predictive of distant metastasis as first event were evidenced including the prognosis signature based on the 21-protein combination, pathological size of tumors, axillary lymph node status (only when dichotomized ≤ 3 vs > 3), Ki67/MIB1 status and delivery of hormone therapy. However, the 21-protein signature was the strongest predictor with a hazard ratio of 2.2 for "poor-prognosis class" patients, compared to "good-prognosis class" patients ([95%CI 1.25 - 3.89], $p < 0.0001$).

References :

1. Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet* 1998; 351:1451-67.
2. Polychemotherapy for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet* 1998; 352:930-42.
3. Eifel P, Axelson JA, Costa J, et al. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 2001; 93:979-89.
4. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. *J Clin Oncol* 2001; 19:3817-27.
5. Leyland-Jones B. Trastuzumab: hopes and realities. *Lancet Oncol* 2002; 3:137-44.
6. Bertucci F, Viens P, Hingamp P, Nasser V, Houlgatte R, Birnbaum D. Breast cancer revisited using DNA array-based gene expression profiling. *Int J Cancer* 2003; 103:565-71.
7. Bertucci F, Houlgatte R, Benziane A, et al. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet* 2000; 9:2981-2991.
8. Bertucci F, Nasser V, Granjeaud S, et al. Gene expression profiles of poor-prognosis primary breast

- cancer correlate with survival. *Hum Mol Genet* 2002; 11:863-72.
9. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000; 406:747-52.
10. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003; 100:8418-23.
11. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003; 100:10393-8.
12. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347:1999-2009.
13. van 't Veer LJ, Dai H, van De Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415:530-6.
14. Huang E, Cheng SH, Dressman H, et al. Gene expression predictors of breast cancer outcomes. *Lancet* 2003; 361:1590-6.
15. Cheng Q, Lau WM, Tay SK, Chew SH, Ho TH, Hui KM. Identification and characterization of genes involved in the carcinogenesis of human squamous cell cervical carcinoma. *Int J Cancer* 2002; 98:419-26.
16. Hoos A, Cordon-Cardo C. Tissue microarray profiling of cancer specimens and cell lines: opportunities and limitations. *Lab Invest* 2001; 81:1331-8.

17. Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998; 4:844-7.
- 5 18. Richter J, Wagner U, Kononen J, et al. High-throughput tissue microarray analysis of cyclin E gene amplification and overexpression in urinary bladder cancer. *Am J Pathol* 2000; 157:787-94.
- 10 19. Callagy G, Cattaneo E, Daigo Y, et al. Molecular classification of breast carcinomas using tissue microarrays. *Diagn Mol Pathol* 2003; 12:27-34.
20. Hsu FD, Nielsen TO, Alkushi A, et al. Tissue microarrays are an effective quality assurance tool for diagnostic immunohistochemistry. *Mod Pathol* 2002; 15:1374-80.
- 15 21. Liu CL, Prapong W, Natkunam Y, et al. Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. *Am J Pathol* 2002; 161:1557-65.
- 20 22. Korschning E, Packeisen J, Agelopoulos K, et al. Cytogenetic alterations and cytokeratin expression patterns in breast cancer: integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis. *Lab Invest* 2002; 82:1525-33.
- 25 23. Alkushi A, Irving J, Hsu F, et al. Immunoprofile of cervical and endometrial adenocarcinomas using a tissue microarray. *Virchows Arch* 2003; 442:271-7.
- 30 24. Nielsen TO, Hsu FD, O'Connell JX, et al. Tissue Microarray Validation of Epidermal Growth Factor Receptor and SALL2 in Synovial Sarcoma with Comparison to Tumors of Similar Histology. *Am J Pathol* 2003; 163:1449-56.

25. Ginestier C, Charaffe-Jauffret E, Bertucci F, et al. Interest and limitations of tissue-microarrays for validation of breast tumor markers selected upon cDNA array analysis. Am J Pathol 2002; 161:1223-1233.
- 5
26. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998; 95:14863-8.
- 10
27. Conte N, Delaval B, Ginestier C, et al. The TACC1-chTOG-Aurora A protein complex in breast cancer. Oncogene in press.
- 15
28. Giet R, McLean D, Descamps S, et al. Drosophila Aurora A kinase is required to localize D-TACC to centrosomes and to regulate astral microtubules. J Cell Biol 2002; 156:437-51.
- 20
29. Droufakou S, Deshmane V, Roylance R, Hanby A, Tomlinson I, Hart IR. Multiple ways of silencing E-cadherin gene expression in lobular carcinoma of the breast. Int J Cancer 2001; 92:404-8.
- 25
30. Teixeira C, Reed JC, Pratt MA. Estrogen promotes chemotherapeutic drug resistance by a mechanism involving Bcl-2 proto-oncogene expression in human breast cancer cells. Cancer Res 1995; 55:3902-7.
31. Menard S, Fortis S, Castiglioni F, Agresti R, Balsari A. HER2 as a prognostic factor in breast cancer. Oncology 2001; 61:67-72.
32. Fisher ER, Osborne CK, McGuire WL, et al. Correlation of primary breast cancer histopathology and estrogen receptor content. Breast Cancer Res Treat 1981; 1:37-41.
- 30
33. Ginestier C, Bardou VJ, Popovici C, et al. Loss of FHIT protein expression is a marker of

- adverse evolution in good prognosis localized breast cancer. *Int J Cancer* 2003; 107:854-62.
34. Hui R, Cornish AL, McClelland RA, et al. Cyclin D1 and estrogen receptor messenger RNA levels are positively correlated in primary breast cancer. *Clin Cancer Res* 1996; 2:923-8.
- 5 35. Paredes J, Milanezi F, Reis-Filho JS, Leitao D, Athanazio D, Schmitt F. Aberrant P-cadherin expression: is it associated with estrogen-independent growth in breast cancer? *Pathol Res Pract* 2002; 198:795-801.
- 10 36. Lakhani SR, Chaggar R, Davies S, et al. Genetic alterations in 'normal' luminal and myoepithelial cells of the breast. *J Pathol* 1999; 189:496-503.
- 15 37. Dontu G, Al-Hajj M, Abdallah WM, Clarke MF, Wicha MS. Stem cells in normal breast development and breast cancer. *Cell Prolif* 2003; 36 Suppl 1:59-72.
- 20 38. Lakhani SR, O'Hare MJ. The mammary myoepithelial cell--Cinderella or ugly sister? *Breast Cancer Res* 2001; 3:1-4.
- 25 39. Boecker W, Buerger H. Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. *Cell Prolif* 2003; 36 Suppl 1:73-84.
- 30 40. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 2003; 100:3983-8.
41. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a

group of breast carcinomas with poor clinical outcome. Am J Pathol 2002; 161:1991-6.

42. Brazma A, Vilo J. Gene expression data analysis. FEBS Lett 2000; 480:17-24.

Claims

1) A method for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of proteins in breast tissues or cells, said pool comprising all or part, for example one, two, three or more of a protein set comprising:

Afadin, Aurora A, a-Catenin, b-Catenin, BCL2, Cyclin D1, Cyclin E, Cytokeratin 5/6, Cytokeratin 8/18, E-Cadherin, EGFR, ERBB2, ERBB3, ERBB4, Estrogen receptor, FGFR1, FHIT, GATA3, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3, Cytokeratin 6, Cytokeratin 18, Ang1, AuroraB, BCRP1, CathepsinD, CD10, CD44, CK14, Cox2, FGF2, GATA4, Hif1a, MMP9, MTA1, NM23, NRG1a, NRG1beta, P27, Parkin, PLAU, S100, SCRIBBLE, Smooth Muscle Actin, THBS1, TIMP1.

2) A method for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of proteins in breast tissues or cells, said pool comprising all or part, for example one, two, three or more of a protein set comprising:

Afadin, Aurora A, a-Catenin, b-Catenin, BCL2, Cyclin D1, Cyclin E, Cytokeratin 5/6, Cytokeratin 8/18, E-Cadherin, EGFR, ERBB2, ERBB3, ERBB4, Estrogen receptor, FGFR1, FHIT, GATA3, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3.

5

3) A method for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of protein in breast tissues comprising a protein set comprising:

10

Afadin, Aurora A, a-Catenin, BCL2, Cyclin D1, Cytokeratin 5/6, Cytokeratin 8/18, E-Cadherin, ERBB2, ERBB3, ERBB4, Estrogen receptor, FGFR1, FHIT, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC2, TACC3.

15

4) The method according to claims 1 to 3 wherein the pool comprises a protein set comprising:

20

Afadin, Aurora A, a-Catenin, b-Catenin, BCL2, Cyclin D1, Cyclin E, Cytokeratin 5/6, Cytokeratin 8/18, E-Cadherin, EGFR, ERBB2, ERBB3, ERBB4, Estrogen receptor, FGFR1, FHIT, GATA3, Ki67, Mucin 1, P53, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3.

25

5) The method according to claim 1 to 4 comprising the detection of overexpression of the following proteins :

EGFR, P53, Ki67, FGFR1, ERBB2, ERBB3, ERBB4, Cyclin D1, Cyclin E, Cytokeratin 5/6.

30

6) The method according to claim 1 to 5 comprising the detection of underexpression of the following proteins :

35

Estrogen Receptor, FHIT, GATA3, Mucin 1, P-Cadherin, Progesterone receptor, TACC1, TACC2, TACC3, Afadin, Aurora A, a-Catenin, b-Catenin, BCL2, Cytokeratin 8/18, E-Cadherin.

7) A protein library useful for the molecular characterization of histopathologic features of breast disease comprising or corresponding to a pool of protein sequences, over or under expressed, in breast tissue or cells, said pool corresponding to the protein defined in any of claims 1 to 6.

10 8) A protein library according to Claim 7 immobilized on a solid support.

15 9) A protein library according to claim 7 or 8 wherein the support is selected from the group comprising nylon membrane, nitrocellulose membrane, polyvinylidene difluoride, glass slide, glass beads, polyustyrene plates, membranes on glass support, silicon chip or gold chip.

20 10) A method for analyzing differential protein expression associated with histopathologic features of breast disease comprising the detection of the overexpression or underexpression of a pool of protein in breast tissues comprising :

25 a) obtaining breast tissue cells from a patient, and
b) measuring in the tissue cells obtained in step (a) over or underexpression of proteins of a library according to any of Claims 7 to 9.

30 11) The method according to Claim 10 wherein said proteins are directly or indirectly labeled before reaction step (b).

5

12) The method according to claim 11 wherein the label is selected from the group consisting of radioactive, colorimetric, enzymatic, molecular amplification, bioluminescent or fluorescent labels.

10

13) The method according to claim 12 wherein one or more specific label are used for each protein of a library according to any of Claims 7 to 9.

15

14) The method according to any of claims 10 to 13, wherein said measuring of over or under expression of proteins is carried out on tissue microarray.

20

15) The method according to any of claims 10 to 14, wherein the measuring of over or under expression of protein is carried out by ImmunoHistoChemistry (IHC)technologies.

25

16) A method according to claim 10 wherein the detection of over or under expression of the pool of protein is alternatively carried out on breast tumor cell lines.

30

17) The method according to any of claims 10 to 16 further comprising
a) obtaining a control sample
b) measuring in the control sample obtained in step (a) expression level of each protein corresponding to library according to any of Claims 7 to 9

c) comparing expression level of each protein with the level of equivalent protein in a tissue sample according to claim 10.

5 18) A method according to any of claims 1 to 6 or 10 to 17, for detecting, diagnosing, staging, monitoring, predicting, preventing conditions associated with breast cancer.

10 19) The method according to any of claims 1 to 6 or 10 to 17 for predicting clinical outcome of breast cancer.

15 20) The method according to any of claims 1 to 6 or 10 to 17 for predicting occurrence of metastatic relapse.

20 21) The method according to claim 18 for determining the stage or aggressiveness of a breast cancer.

25 22) A method according to any of claims 1 to 6 or 10 to 21 wherein the breast tissue sample is obtained from a patient regardless of whether said patient has received a neo adjuvant or an adjuvant therapy.

30 23) The method according to claim 22 wherein the breast tissue sample is obtained from a patient who has received an adjuvant therapy.

24) The method according to claim 22 wherein the breast tissue sample is obtained from a patient who has not received an adjuvant therapy.

5

25) A method for treating a patient with a breast cancer comprising (i) the implementation of a method according to any of claims 1 to 6 or 10 to 24 on a sample from said patient, and (ii) determining a treatment for this patient based on the analysis of differential protein expression profile obtained with said method.

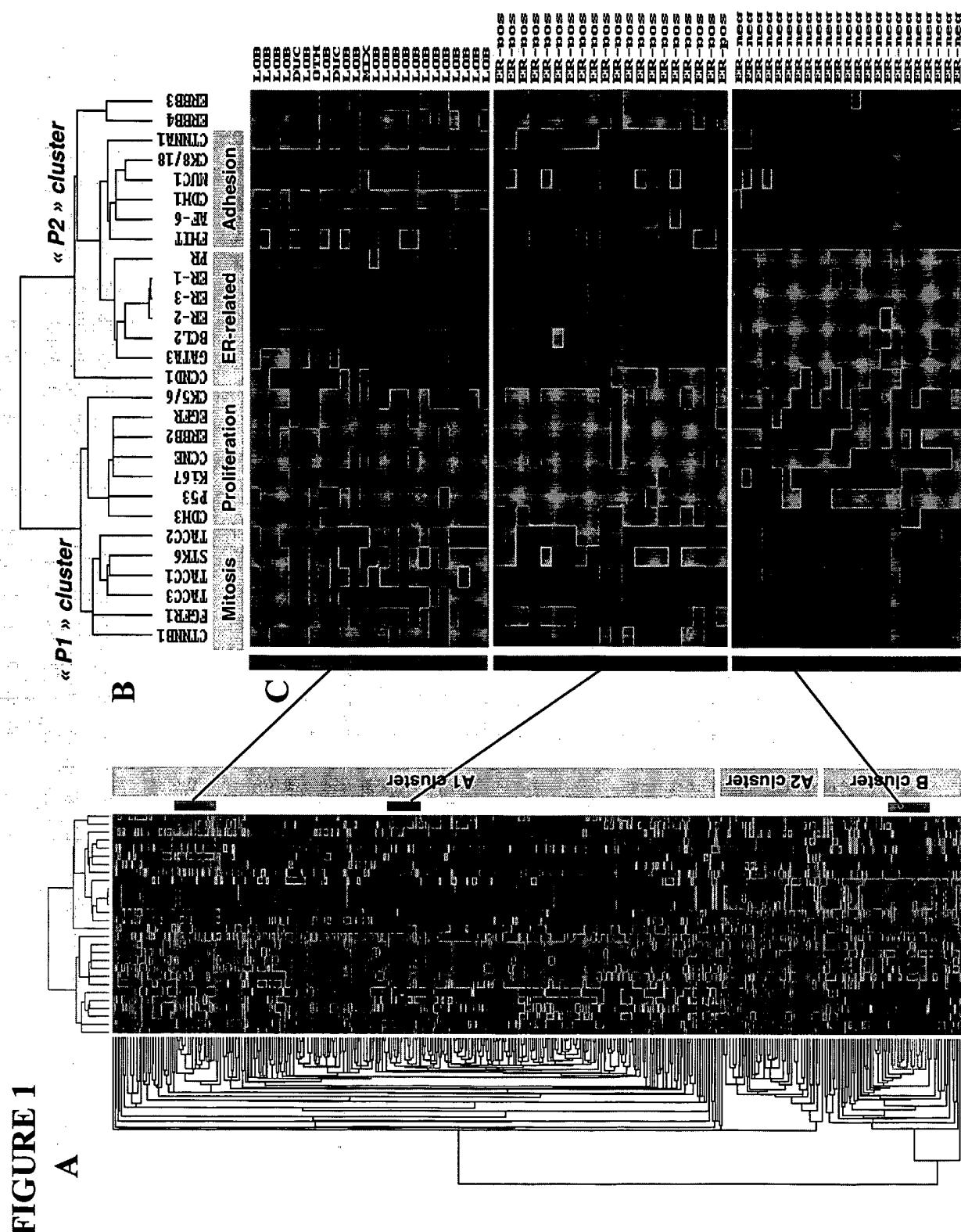
10

15

26) A method for analyzing differential protein expression associated with histopathologic features of breast disease according to claim 1 to 6 wherein the detection of the overexpression or underexpression of said pool of protein in breast tissues comprises the detection of the overexpression or underexpression of nucleic acids coding for said proteins.

20

27) A nucleic acids library useful for the molecular characterization of histopathologic features of breast disease comprising nucelic acids according to claim 26.



2/3

FIGURE 2

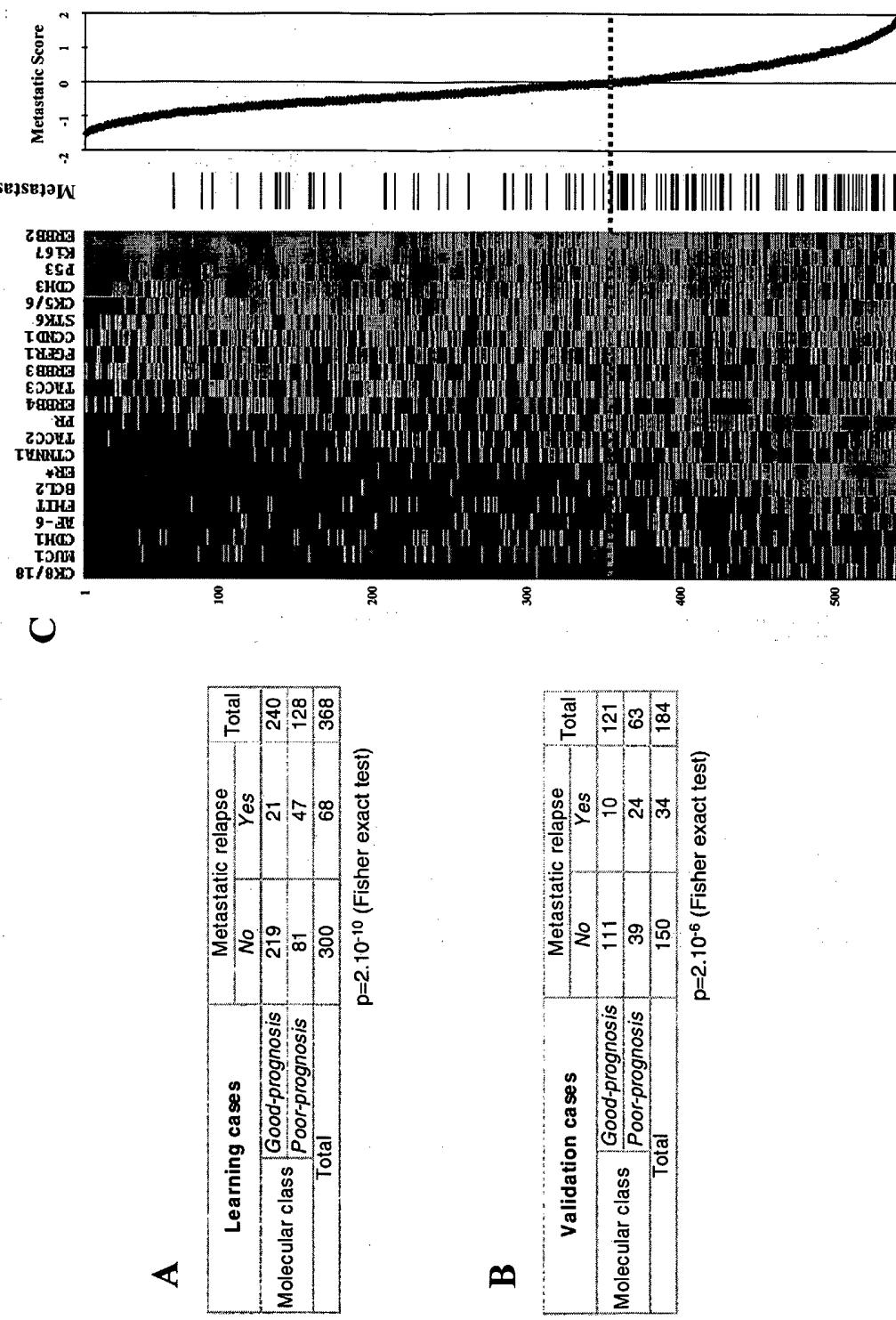
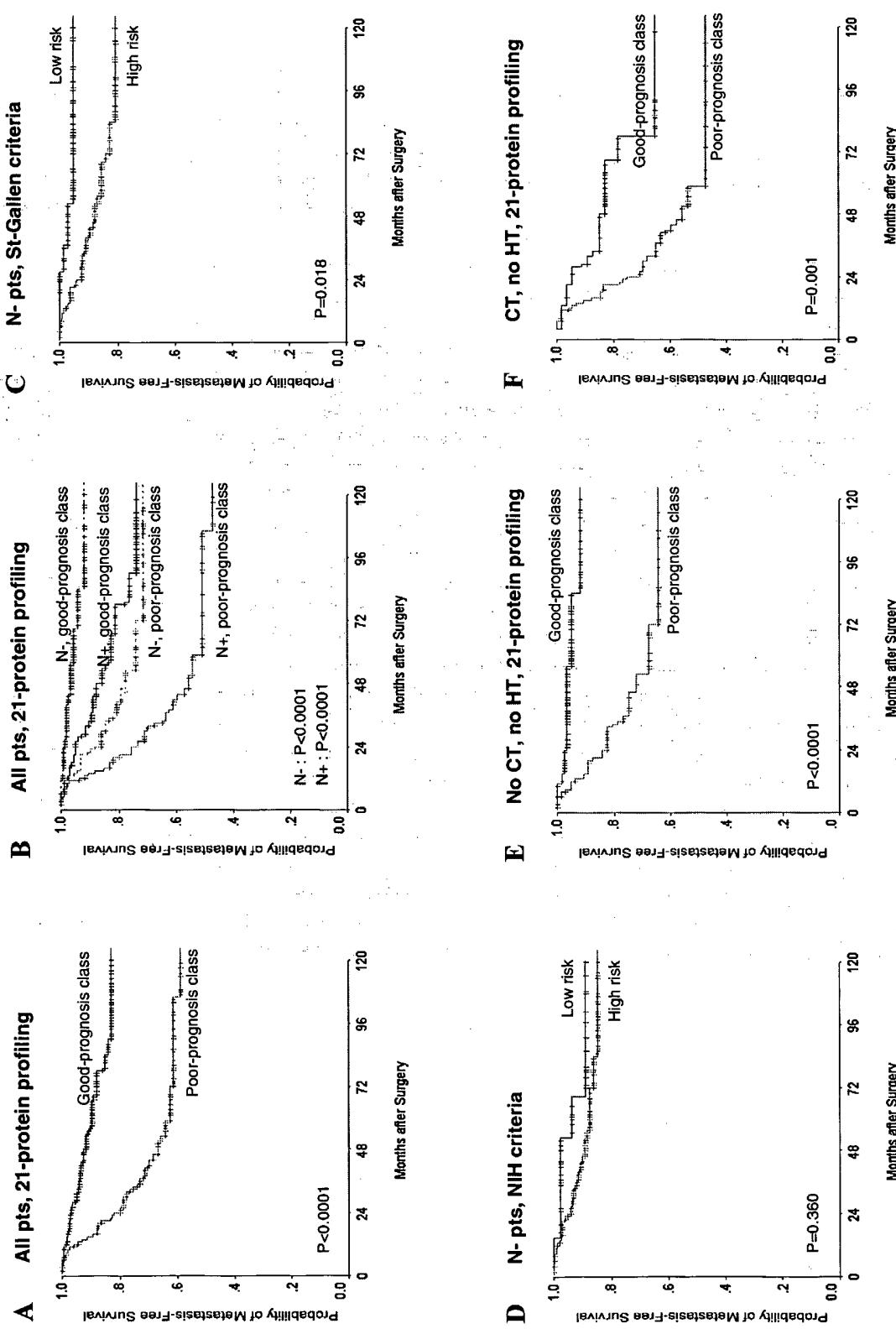


FIGURE 3

SEQUENCE LISTING

<110> IPSOGEN, INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MEDICALE
-INSERM -, INSTITUT PAOLI CALMETTES

<120> Protein expression profiling and breast cancer prognosis

<130> 35674/PCT

<140> PCT/IB05/xxxxx

<141> 2005-01-17

<150> US 11/xxxxxxx

<151> 2005-01-14

<150> US 60/537,412

<151> 2004-01-16

<160> 28

<210> 1

<211> 1612

<212> PRT

<213> Homo Sapiens

<400>

MSAGGRDEERRKLADIHHWNANRLDLFEISQPTEDLEFHGVMRFYFQDKAAGNFATKCIRVSSTATTQDVIELT
AEKFRPDMRMLSSPKYSLYEVHVGERRLDIDEKPLVVQLNWNKDDREGFVULKNNENDAIPPKAQSNGPEKQEKE
GVIQNFKRTLSKKEKEKKREKEALRQASDKDRPFQGEDVENSRALAAEVYKDMPETSFTRTISNPVVVMKRR
QQKLEKRMQEFRSSDGRPDGGTLRIYADSLKPNIPYKTISSSTDPADFAVAEALEKYGLEKENPKDYCIAVRM
LPPGAQHSDEKGAKEIILDDDECPLQIFREWPSDKGILVFLKRRPPDHIPKKTGHLEGKTPKGKERADGSVY
STLPPEKLPYLVELSPDGSDSRDKPKLYRLQLSVTEVGTEKLDNSIQLFGPGIQPHHCDLTNMDGVVTVP
DAETYVEGQRISETTMLQSGMKVQFGASHVFKFVDPSQDHALAKRSVDGGLMVKGPRHKPGIVQETTFDLGGDIH
SGTALPTSKSTTRLDSDRVSSASSTAERGMVKPMIRVEQQPDYRRQESRTQDASGPELILPASIEFRESSEDSFL
SAIINYNTNSSTVHFKLSPTYVLYMACRYVLSNQYRDISPSTERHKVIAAVVNKMVSMMEGVIQKQKNIAGALAFW
MANASELLNFIQDRDLSRITLDAQDVLAHLVQMAFKYLHVCLQSELNNYMPAFLDDPEENSLQRPKIDDVLHTL
TGAMSLRRRCRVNAALTIQLFSQLFHFINMWLFNRLVTDPSGLCSHYWGAIIRQQLGHIEAWAEKQGLEAADC
HLSRIVQATTLLMDKYAPDDIPNINSTCFKLNSLQLQALLQNYHCA
DGREVQLEEDPDLQLPFLPEDGYSCDVRNIPNGLQEFLDPLCQRGFCR
RENTELAQPLRKEPEIITVTLKKQNGMGLSIVA
IEAAMDRKSDSDMW
PPP
ESPQLPWA
EAYAPI
1 mdrskencis
2
3 gpvkatapvg
4 gpkrvlvtqq
5 ipcqnplpvn
6 sgqaqrvlcp
7 snssqrpvplq
8 61 aqklvsshkp
9 vqnqkqkqlq
10 atsvphpv
11 plnnntqkskq
12 plpsapenp
13 eeelaskqkn
14 121 eeskkrqwal
15 edfeigrplg
16 kgkfgnvyla
17 rekqskfila
18 lkvlfkaqle
19 kagvehqlrr
20 181 eveiqshlrh
21 pnirlgyyf
22 hdatrvylil
23 eyaplgtvyr
24 elqklskfde
25 qrtatyitel

<210> 2

<211> 403

<212> PRT

<213> Homo Sapiens

<400>

1 mdrskencis
2 gpvkatapvg
3 gpkrvlvtqq
4 ipcqnplpvn
5 sgqaqrvlcp
6 snssqrpvplq
7 61 aqklvsshkp
8 vqnqkqkqlq
9 atsvphpv
10 plnnntqkskq
11 plpsapenp
12 eeelaskqkn
13 121 eeskkrqwal
14 edfeigrplg
15 kgkfgnvyla
16 rekqskfila
17 lkvlfkaqle
18 kagvehqlrr
19 181 eveiqshlrh
20 pnirlgyyf
21 hdatrvylil
22 eyaplgtvyr
23 elqklskfde
24 qrtatyitel

241 analyschsk rvihrdikpe nlllgsagel kiadfgwsvh apssrrttlc gtldylppem
301 iegrmhdekv dlwslgvlcy eflvgkppfe antyqetykr isrveftfpd fvtegardli
361 srllkhnpsq rpmlrevleh pwitansskp sncqnkesas kqs

<210> 3
<211> 907
<212> PRT
<213> Homo Sapiens

<400>
1 mtavhagnin fkwdpkslei rtlaverlle plvtqvttlv ntnskgpsnk krgrskkahv
61 laasveqate nflekgdkia kesqflkeel vvavedvrkq gdlnmkaage faddpcssvk
121 rgnmvraara llsavtrlli ladmadvykl lvqlkvvedg ilklrnagne qdlgnqykal
181 kpevdklnim aakrqqelkd vghrdqmaaa rgilqsnvpi lytasqaclq hpdvaaykan
241 rdliykqlqq avtgisnaaq atasddasqh qgggggelay alnnfdkqii vdplsfseer
301 frpsleerle siisgaalma dssctrddrr erivaecnav rqacrtcvse ymgagnagrker
361 sdalnsaidk mtkktrdlrr qlrkavmdhv sdsfletnvp llvliaeakn gnekevkeya
421 qvfrehankl ievanlacsi snneegvklv rmsasqleag cpqvinaatw alapkqskl
481 agenmdlfke qwekqrvl davdditsid dflavsenhi ledvnkcvia lgekdvdgld
541 rtagairgra arvihvvtse mdnyepgvyt ekvleatkll sntvmpfrte qveaaveals
601 sdpaqpmden efidasrlvy dgirdirkav lmirtpeeld dsdfetedfd vrsetsqvte
661 ddqliaggqa raimaqlpqe qkakireqva sfgeekskld aevskwddsg ndiivlakqm
721 cmimmemtdf trkgkplknt sdvisaakki aeagsrmdkl grtirdhcpd sackqdllay
781 lqrialychq lnicskvkae vqnlggelvv sgvdsamsli qaaknlmnav vqtvkasyva
841 stkyqksqgm aslnlpavsm kmkapekkpl vkrekqdetq tkikrasqkk hvnpvqalse
901 fkamdsi

<210> 4
<211> 781
<212> PRT
<213> Homo Sapiens

<400>
MATQADLMELDMAMEPDRKAAVSHWQQQSYLDGIGATTTAPSLSGKGPNPEEDVDTSQVLYEWEQGFSQSFT
QEJVADIDGQYAMTRAQRVRAAMFPETLDEGMQIPSTQFDAAHPTNVQRLAEPQMLKHAVVNLIINYQDDAELAT
RAIPELTKLNNDEDQVVNKAAMVHQLSKKEASRHAIMRSPQMVAIVRTMQNTNDVETARCTAGTLHNLSHRR
EGLLAIFKSGGIPALVKMLGPVDSDLFYAITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAITTD
LQILAYGNQESKLIIIASGGPQAUVNIMRTTYEKLLWTTSRVLKVLSCVSSNKPAIVEAGGMQALGLHLDPSQ
RLVQNCLWTLRNLSDAATKQEGMEGLLGTLVQLLGSDDINVVTCAAGILSNLTCNNYKNKMMVCQVGGIEALVRT
VLRAGDREDITEPAICALRHLTSRHQEAEMAQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHA
PLREQGAIPRLVQLLVRAHQDTQRRTSMGGTQQQFVEGVRMEEIVEGCTGALHILARDVHNIRIVRLNTIPLFV
QLLYSPIENIQRVAAGVLCELAQDKEAAEAEIAEGATAPLTELLHSRNEGVAITYAAAVLFRMSEDKPQDYKKRSL
VELTSSLFRTEPMANETADLGLDIGAQGEPLGYRQDDPSYRSFHSGGYGQDALGMDPMMEHEMGHHHPGADYPV
DGLPDLGHAQDLMGDPGDSNQLAWFDTDL

<210> 5
<211> 239
<212> PRT
<213> Homo Sapiens

<400>
MAHAGRTGYD NREIVMKYIH YKLSQRGYEW DAGDVGAAPP GAAPAPGIFS SQPGHTPHPA
ASRDPVARTS PLQTPAAPGA AAGPALSPPV PVVHLTLRQA GDDFSRRYRR DFAEMSSQLH
LTPFTARGRF ATVVEELFRD GVNWGRIVAF FEFGGVMCVE SVNREMSPLV DNIALWMTEY
LNRHLHTWIQ DNGGWDAFVE LYGPSMRPLF DFSWLSLKTL LSLALVGACI TLGAYLGHK

<210> 6
<211> 295
<212> PRT
<213> Homo Sapiens

<400>

MEHQQLCCEV ETIRRAYPDA NLLNDRVLRA MLKAEETCAP SVSYFKCVQK EVLPSMRKIV 60
 ATWMLEVCEE QKCCEEVFPL AMNYLDRFLS LEPVKKSRLQ LLGATCMFVA SKMKETIPLT 120
 AEKLCIYTDN SIRPEELLQM ELLL VNKLKW NLAAMTPHDF IEHFLSKMPE AENKQIIRK 180
 HAQTFVALCA TDVKFISNPP SMVAAGSVVA AVQGLNLRSP NNFLSYYRLT RFLSRVIKCD 240
 PDCLRACQE_Q IEALLESSLR QAQQNMDPKA AEEEEEEEEE VDLACTPTDV RDVDI295

<210> 7

<211> 395

<212> PRT

<213> Homo Sapiens

<400>

5	10	15	20	25	30																									
1	M	K	E	D	G	G	A	E	F	S	A	R	S	R	K	R	K	A	N	V	T	V	F	L	Q	D	P	D	E	E
31	M	A	K	I	D	R	T	A	R	D	Q	C	G	S	Q	P	W	D	N	N	A	V	C	A	D	P	C	S	L	I
61	P	T	P	D	K	E	D	D	D	R	V	Y	P	N	S	T	C	K	P	R	I	I	A	P	S	R	G	S	P	L
91	P	V	L	S	W	A	N	R	E	E	V	W	K	I	M	L	N	K	E	K	T	Y	L	R	D	Q	H	F	L	E
121	Q	H	P	L	L	Q	P	K	M	R	A	I	L	L	D	W	L	M	E	V	C	E	V	Y	K	L	H	R	E	T
151	F	Y	L	A	Q	D	F	F	D	R	Y	M	A	T	Q	E	N	V	V	K	T	L	L	Q	L	I	G	I	S	
181	L	F	I	A	A	K	L	E	E	I	Y	P	P	K	L	H	Q	F	A	Y	V	T	D	G	A	C	S	G	D	E
211	I	L	T	M	E	L	M	I	M	K	A	L	K	W	R	L	S	P	L	T	I	V	S	W	L	N	V	Y	M	Q
241	V	A	Y	L	N	D	L	H	E	V	L	L	P	Q	Y	P	Q	Q	I	F	I	Q	I	A	E	L	L	D	L	C
271	V	L	D	V	D	C	L	E	F	P	Y	G	I	L	A	A	S	A	L	Y	H	F	S	S	S	E	L	M	Q	K
301	V	S	G	Y	Q	W	C	D	I	E	N	C	V	K	W	M	V	P	F	A	M	V	I	R	E	T	G	S	S	K
331	L	K	H	F	R	G	V	A	D	E	D	A	H	N	I	Q	T	H	R	D	S	L	D	L	D	K	A	R	A	
361	K	K	A	M	L	S	E	Q	N	R	A	S	P	L	P	S	G	L	L	T	P	P	Q	S	G	K	K	Q	S	S
391	G	P	E	M	A																									

<210> 8

<211> 590

<212> PRT

<213> Homo Sapiens

<400>

MSRQSSVSFRSGGSRSFSTASAITPSVSRSTSFTSVSRSGGGGGFGRVSLAGACGVGGYGSRSLYNLGGSKRIS
 ISTRGGSFRNRFGAGAGGGYGFGGGAGSGFGFGGGAGGGFGLGGGAGFGGGFPGFPVCPPGGIQEVTVNQSL
 TPLNLQIDPSIQRVRTEEREQIKTLNNKFASFIDKVRFLEQQNKVLDTKWTLLQQGTKTVRQNLEPLFEQYINN
 LRRQLDSIVGERGRLDSELRNMQDLVEDFKNKYEDEINKRTTAENEVFMLKKDVAAYMNKVELEAKVDALMDEI
 NFMKMFFDAELSQMOTHVSDTSVVLSDMDNNRNLDLDSIIAEVKAQYEEIANRSRTEAESWYQTKEELQQTAGRH
 GDDLRTNKHEITEMNRMIQQLRAEIDNVKKQCANLQNAIADAEGRLAKDARNKLAEEALQAKQDMARLL
 REYQELMNTKLALDVEIATYRKLEGEECRLSGEGVGPVNISVTSSVSSGYGSGSGYGGGLGGGLGGGLGGGLA
 GGSSGSYYSSSSGGVGLGGGLSVGGSGFSASSGRGLGVFGSGGGSSSSVKFVSTTSSSRKSFKS

<210> 9

<211> 483

<212> PRT

<213> Homo Sapiens

<400>

MSIRVTQKSYKVSTSGPRAFSRSYTSGPGSRISSSFSRVGSSNFRGGLGGGYGGASGMGGITAVTVNQSLSP
 LVLEVDPNIQAVRTQEKEQIKTLNNKFASFIDKVRFLEQQNKVLETKWSLLQQQKTARSNMDNMFESYINNLRQ
 LETLGQEKLKLEAELGNMQGLVEDFKNKYEDEINKRTEMENEFLIKKDVEAYMNKVELESRLEGLTDEINFLR
 QLYEEEIRELQSQISDTSVVLSDMDNSRSIDMDSIIAEVKAQYEDIANRSRAEAESMYQIKYEELQSLAGKHDDL
 RRTKTEISEMNRNISRLQAEIEGLKGQRASLEAAIAADEQRGELAIKDANAKLSELEAALQRAKQDMARQLREYQ
 ELMNVKLALDIEIATYRKLEGEESRLESGMQNMSIHTKTTSGYAGGLSSAYGGLTSPGLSYSLGSSFGSGAGSS
 SFSRTSSSSRAVVVKIETRDGKLVSESSDVLPK

<210> 10

<211> 882

<212> PRT
<213> Homo Sapiens

<400>

MGPWSRSLSALLLLOVSSWLCQEPEPCHPGFDAESYTFVPRRLERGRVLGRVNFEDECTGRQRTAYFSLDTRF
KVGTDGVITVKRPLRFHNPQIHFLVYAWDSTYRKFKTQLNTVGHHRPPHQASVSGIQAELLTFPNSSPGLR
RQKRDWVIPPSCPENEKGPFPKNLVQIKSNKDKEGKVFYSITGQGADTPPVGVFIIERETGWLKVTEPLDRERI
ATYTLFSHAVSSNGNAVEDPMELILITVTDQNDNKPEFTQEVFKG SVMEGALPGTSVMEVTATDADDVNNTYAAI
AYTILSQDPELPDKNMFTINRNTGVISVVTGLDRESFPYTILVVQAADLQGEGLSTTATAVITVTTDNPPIF
NPTTYKGQVPENEANVVITTLKVTDADAPNTPAWEAVYTILNDDGGQFVVTNPVNNDGILKTAKGLDFEAKQQY
ILHVAVTNVVPFEVSLLTSTATVTDVLDVNEAPIFVPPKEKREVESEDFGVGQEIITSYTAQEPDTFMEQKITYRI
WRDTANWLEINPDTGAISTRAELDREDFEHVKNSTYTALIIATDNGSPVATGTGTLLLILSDVNDNAPIPEPRTI
FFCERNPKPQVINIIDADLPPNTSPFTAELTHGASANWTIQYNDPTQESIILPKPKMALEVGDYKINLKLMDNQNKK
DQVTTLEVSVCDCEGAAGVCRKAQPVEAGLQIPAILGILGGILALLLILLLLLFLRRRAVVKEPLLPEDDTRD
NVYYYDEEGGGEEDQFDLSQLHRGLDARPEVTRNDVAPTLMSVPRYLPRPANPDEIGNFIDENLKAADTDPTAP
PYDSLVLVFDYEGSGSEAASLSSLNSSESVDQDYDYLNEWGNRFKKLADMYGGGEDD

<210> 11

<211> 1210

<212> PRT

<213> Homo Sapiens

<400>

MRPSGTAGAALLALLAALCPASRALEEKVCQGTSNKLQLGT FEDHFLSLQRMFNNCVV LGNLEITYVQRNYD
LSFLKTIQEVA GYV LIALNTVERIPL ENLQIIRGNMYYENS YALAVLSNYDANKTGLKELPMRN LQEI LHGA VRF
SNNPA CNVESI QWR DIVSS DFLS NMS MDF QNH LGSC QKCDP SCPNG SCWG AGEEN CQKLT KII CAQ QCS GR C RG
KSPSD CCH NQCA AGC TGP RES DCL VCR KFR DEAT C KDT CP PLM LYNP TT QMD VN PEG KYS FG AT CV KK C PR NYV
VTD HGSC VRAC GAD SYEME EDG VR CKK CE GP CRK VCN GIGIG EFK DLS IN ATNI HFKN CTS IS GDL HIL PVA
FRG DSFT HTPPL DPQ ELD I LKTV KEIT GFL LIQ AW PEN RT DLH AFEN LEI IRG RT KQHG QF SLAV VSL NI TSL GL
RSL KEIS DGD VI IS GNK NL CYANT IN WKL FGT SG QKTK II IS NRGEN SCK AT GQV CHAL CS PEG CWG PE PR DC VS
CRN VSRG REC VDK C KLL EGE PREF VEN SE CI QCH PE CLP QAMN IT CT GRG PD NC IQ CAHY IDG PHC V KTC PAG VM
GEN NT LVW KYADAGH VCH LCH P NCT YG CT GP GL EG CP T NG PKI P SI AT GMV G ALL LLV VAL GI LF MRR R HIV
KRT L RLL Q EREL V EPL TPS GE APN Q ALL R I LK ETE FKK I KV LG SG AFG T VY KGL WI PE GE KV KI PVA I KEL REA
TSP KAN KE IL DE AY VMA SVDN PHC RLL GIC LT STV QL I T QL MP FG C L LDY VRE HKD NI G SQ YL LN WC V QIA KGM
N YLED RR L V HR DLA ARN VL V KTP QHV KIT D FGL A K L GAE E KEY HAEG GK VPI KWM ALES I L H R I YTH Q SDV WSY
GVT VWE LM TFG SKP YDG I PASE I S SILE KGER LPQ PPI CT ID VY MIM V K C W M I D A S R P K F R E L I I E FSK MARD P
Q RYLV IQG DER MHL P SPT DS NF Y RAL M D E M D D V D A D E Y L I P Q Q G F S S P S T R P L L S L S A T S N N S T V A C I
DRN GL QSC PI KED SFL QRY SS DPT GALT E D S I D D TFL P VPE YIN Q SV P KRP AG S V Q N P V Y H N Q P L N P A P S RD P HY
QDP H STAV GN P EY L N T V Q P T C V N S T F D S PA HWA Q KG SH Q I S LD N P D Y Q Q D F F P K E A K P N G I F K G S T A E N A E Y L R V
APQSSEFIGA

<210> 12

<211> 1255

<212> PRT

<213> Homo Sapiens

<400>

MELAALCRWG LLL ALL PPGA AST QV C T GTD MKL RL PASPE TH LDML RH LY QGC QV V QGN L
EL TYL PTN AS LSFL QD I QEV QGYV LIAHN Q VRQ VPL QRL R I V RGT QL FED NYA LA VLD NG
DPL NN TTP VT GAS PG GL REL QL RSL TE I LK GG VLI Q RNP Q LCY QD TIL WK D IFH KNN QLA
LTL IDT NRS R ACH PC SPM CK GSRC WG ESS E DC QSL TRT VC AGG CAR CK GP LPT DCC HEQC
AAG CT GP KHS DCL AC LH FN H SG IC EL HC PA LVT YN TDT FE SMPN PEG RYT FG AS CV TACP
Y NYL STD VGS CTL VCP LHN Q EV TA E DGT QR CEK C SK PC AR VCY GLG MEHL REV RAV TSAN
I QE FAG C KKI FG SLA FL PES FD GD P AS NTA PL QP EQL QV F ET LEE IT GYL Y IS AWP D SLP
DLS V FQ NL QV IR GRIL HNG A YSL TL Q GLGI SWL GLR SL RE LG SGL ALI HH NTH LCF VHT V
PWD QL FRN PH Q ALL HTA NR P EDE CV GE GLA CH QL CARG HC WG PG PT QCVN CSQFL RG QEC
VEE CRV LQ GL PRE YVN ARHC LP CH PEC QP Q NGS VTC FG PE AD QCV AC AHY KDPP FC V AR C
PSG V KPD LSY MPI WKF DEE GAC QPC P INC TH SCV D LDD K GCP AE QRASP LTS II SAV VG
ILL VV VLG VV FG I LK R RQ Q KIR KYT M RRL LO ETEL VEPL T PSG A MPN QA QMR ILK ETEL
RK V KVL GSG A FG T VY KGI WI PD GEN V KIP V AI KVL RENTS PK AN KE IL DE AY VMA GVG SP
Y V S RLL GIC L TST VQ LVT QL MPY GCL LDH V REN RGR LGS Q DLL NW CM QIA KGM SY LED V R
L V HR DLA ARN VLV KSPN HVK IT DFG LAR LL DIDE TEY HAD GG K VPI KW MA LES I LRR RFT
HQ SDV WSY GV TV WEL MT FGA KPY DG I PARE IP DLE KGER LP QP PICT ID VY MIM V K C WM

IDSECRPRFR ELVSEFSRMA RDPQRFVVIQ NEDLGPASPL DSTFYRSLL DDDMGDLVDA
EEYLVPQQGF FCPDPAPGAG GMVHHRHRS STRSGGGDLT LGLEPSEEEA PRSPLAPSEG
AGSDVFDGL GMGAAKGLQS LPTHDPSPLO RYSEDPTVPL PSETDGYVAP LTCSPQPEYV
NQPDVRPQPP SPREGPLPAA RPAGATLERP KTLSPGKNGV VKDVFAGGA VENPEYLTPQ
GGAAPQPHPP PAFSPA FDNL YYWDQDP PER GAPPSTFKGT PTAENPEYLG LDVPV

<210> 13

<211> 1342

<212> PRT

<213> Homo Sapiens

<400>

MRANDALQVLGLLFSLARGSEVGNSQAVCPGTLNGLSVTGDARENQYQTLKYKLYERCEVMGNLEIVLTGHNA DLS
FLQWIREVTGYVLVAMNEFSTLPLPNLRVVRGTQVYDGKFAIFVMLNYNTNSSHALRQLRLTQLTEILSGGVYIE
KNDKLCHMDTIDWRDIVDRDRAEIVVKDNGRSCPPCHEVCKGRCWPGPSEDCQTLTKTICAPQCNGHCFGPNPQ
CCHDECAGGCSGPQDTDCFACRFNDSGACVPRCPQPLVYNKLTQLEPNPHTKYQYGGVCASC PHNFVVQDQTS
CVRACPPDKM EVDKNGLKMCEPCGGLCPKACEGTGSRSRFTQVDSSNIDGFVNCTKILGNLDFLITGLNGDPWHK
IPALDPEKLN VFRTVREITGYLNIQS WPPHMHNFSVFSNLTTIGGRSLYNRGFSLLIMKNLNVTSLGFRSLKEIS
AGRIYISANRQLCYHHSNWTKVLRGPTEERLDIHKHNRPRRDCVAEGKVC DPLCSSGGCWGP GPQCLSCRNYSR
GGVCVTHCNFLNGEPREF AHEAECSCHPECQPMGGTATCNGGSDTCAQCAHFRDGPCHVSSCPHGVLA KGP
YKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLI GKTHTLMALTVIAGLVVIFMMLGGTFLYWRGRRIQKR
AMRRYLERGESIEPLDPSEKANKVLA RIFKETELRKLKVLGSGVFGTVHKGVWIPEGESIKIPVCIKVIEDKSGR
QSFQAVTDHMLAIGSLDHAIHVRLLGLCPGSSLQLVTQYLP LGSLLD HVRQH RGALGPQ LNNWGVQIAKGMYYL
EEHGMVHRNLAARNVLLKSPSQVQADFGVADLLPPDKQLLYSEAKPIKWM ALESIHFGKYTHQSDVWSYGV
VWELMTFGAEPYAGLRLAEPDPLLEKGERLAQPQIC TIDVYVMVMKCWMIDENIRPTFKE LANE FTRMARDPPRY
LVIKRESGPGIAPGPEPHGLTNKKLEEVELEPELDL DLDLEAEEDNLATTLSA SLPVGTLNRP RGSQSLSP
SSGYMPMNQGNLGGSCQESAVGSSERCPRPVSLHPMPRGCLASESSEGHVTGSEAE LQEKVSMCRSRSRSRSPR
PRGDSAYHSQRHSLLTPVTPLSPPGLEEE DVNGYVMPDTHLKGT PSSREGTLSSVGLS VLGTEEDEDEEYEYM
NRRRRHSPHPPR PSSLEELGYEYMDVGSDLSASLGSTQSCP LHPVPIMPTAGTT PDEDYE YMNRQ RDGGPGGD
YAAMGACPASEQGYEEMRAFQGP GHQAPHVHYARLKTLSLEATDSA FDNP DYWH SRLFPKANA QRT

<210> 14

<211> 1308

<212> PRT

<213> Homo Sapiens

<400>

MKPATGLWWVSLLVAAGTVQPSDSQSVCAGTENK LSSLS DLEQQYRALRKYYENCEVVMGNLEITSIEHNRDLS
FLRSVREVTGYVLVALNQFRYLPLENL RIIRGT KLYEDRYALAI FLN YRKDG NGFLQ E GLK NLTEILNGGVYVD
QNKF LCYADTIHWQD IVRN PWP SNL TLV STNGSSG CRCH KSCT GR CWGP TENHC QT LRTV C AEQCDGRCYGP
VSDCCCHRECAGGCSGPKD TDCFAC MNF ND SGAC VTQ CPQT FV YNPTT QLE HNFNAK YTG AF CV KKCP HFVV D
SSSC VRAC PSK MEVEENG I K MCK P CT DICPK ACDG I GTG SLS MAQ TVD SSN IDK F IN CT KING N L I F L V TGI H
DPYNAIEAIDPEKLN VFRTVREITGFL NIQS WPPN MTD FS VFS NL V TIG R VLY SGL SLL L I K Q Q G I T S L Q F Q S L
KEISAGNIYI TD NSN L CYY HTI NWT LF STIN Q RIV I RD NRKA EN CTAEGMVCN HLCSS DGC WGP DQ CLS CRR
FSRGRICIESCN LYDGEFREFENG SIC VEC DP QCE K MED GLL TCH GPG PDN CT KCS HFK DGP NC V E K CP DGL QGA
NSFIF KYADPDRECHPCHPNCTQGCNGPTSHDCI Y PWTGH STLPQHARTPLIAAGVIGGL FIL VIVGLT FAVY V
RRKSIKKRALRRFLETELVEPLTPSGTAPNQAQLRILKETELKRVKVLGSGA FGTVYKG I WVP EGETV KIP VAI
KILNETTGP KAN VEF MDE ALIMASMDHPHLVRLLGVC LSP TIQLV TQLMP HGCL LEYVHEH KDN IGSQ LLL NWCV
QIAKG MYLEERRL VHD LAAR NVL V KSPN HV KITDFGLARLLEGDEKEYNADGGKMP IKWMALE CIHYRK FTH Q
SDVWSYGV TIWELMTFGGKPYDGI PTREI PDLLEKGERLPQPP I CTIDVYVMVMKCWMIDADSRPKF KELAAEFS
RMARDPQ RYLV I QGDDRMKLPSPNDSKFFQNLL DEEDLEDMMDAE EYLV P QAFNIPPIY TS RAR IDSNR SEIGH
SPPPAYTPMSGNQFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAE I FDDSCCNGTLRKP VAPHVQEDSST
QRYSADPTV FAPERS PRGEL DEEGYMTPMRD KPK QEYLN PVEENP F VSRRKNGDLQALDNPEYHN ASNGPPKAED
EVVNEPLYLNTFANTLGKA EYLKNN ILSMPEKAKFA FDNP DYWNH SLPPRSTLQHPDYLQ EY STK YFYK QNGRIR
PIVAENPEYLS EFSLKPGTVL PPPYRHRNTVV

<210> 15

<211> 595

<212> PRT

<213> Homo Sapiens

<400>

MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPavyNYPEGAAYEFNAAAAANAQVYQG
TGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLHPPQQLSPFLQPHQQVPPYLENEPSGYTVREAGPPAFY
RPNSDNRQGRERLASTNDKGSMAMESAKETRYCAVCNDYASGYHYGVWSCFGCKAFFKRSIQGHNDYMCPATN
QCTIDKNRKSCQACRLRKCYEVGMMKGGIRKDRRGGMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMIKR
SKKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQ
VHLLECAWLEILMIGLVWRSMEHPVKLLFAPNLLDRNQGKCVGMVEIFDMLLATSSRFRMMNLQGEFFVCLKS
IILLNSGVYTFLSSTLKSLEEKDHHRVLDKITDTLIELMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMELH
YSMKCKNVVPLYDLLMELDAHRLAPTSRGASVEETDQSHLATAGSTSSHSLQKYYITGEAEGFPATV

<210> 16

<211> 820

<212> PRT

<213> Homo Sapiens

<400>

MWSWKCLLFWAVLVTATLCTARPSTLPEQAQPWGAPVEVESFLVHPGDLLQLRCRLRDDVQSINWL RDGVQLAE
SNRTRITGEEVEVQDSVPADSGLYACVTSSPGSDTTYFSVNVSALPSSEDDDDDDDSSSEEKETDNTKPNPVA
PYWTSPEKMEKKLHVPAAKTVFKCPSSGTPNPTLRWLKNGKEFKPDHRIGGYKVRYATWSIIMDSVVPSDKGN
YTCIVENEYGSINHTYQLDVVERSPHRPILQAGLPANKTVALGSNVEFMCKVYSDPQPHIQWLKHIEVNGSKIGP
DNLPYVQILKTAGVNTTDKEMEVVLHRLNVSFEDAGEYTCLAGNSIGLSHHSAWLTVLEALEERPAVMTSPLYLEI
IIYCTGAFLISCMVGSVIVYKMKGTKSDFHSQMAVHKLAKSIPLRRQVTVSADSSASMNSGVLLVRPSRLSSS
GTPMLAGVSEYELPEDPRWELPRDRLVLGPLGEGCFGQVVLAEAIGLDKDKPNRVTKAVKMLKSDATEKDLSD
LISEMEMMMKIGKHKNIINLLGACTQDGPLYVIVEYASKGNLREYLQARRPPGLEYCYNPSPHNPEEQQLSSKDLVS
CAYQVARGMEYLASKCICHRDLAARNVLVTEDNVMKIADFGLARDIHIDYYKKTTNGRLPVKWMPEALFDRIY
THQSDVWSFGVLLWEITLGGSPYGPVVEELFKLLKEGHRMDKPSNCTNELYMMMRDCWHAVPSQRPTFKQLVE
DLDRIVALTSNQEYLDLSMP LDQYSPSFDPTRSCTSSGEDSVFSHEPLPEEPCLPRHAPQLANGGLKRR

<210> 17

<211> 147

<212> PRT

<213> Homo Sapiens

<400>

MSFRFGHQHLIKPSVVFLKTELSFALVNRPVVPGHVLVCPLRPVERFHDLRPDEVADLFQTTQRVGTVVEKFHG
TSLTFSMQDGPEAGQTVKHVVHVLPKAGDFHRNDSIYEELQKHDKEDFPASWRSEEMAAALRVYFQ

<210> 18

<211> 444

<212> PRT

<213> Homo Sapiens

<400>

MEVTADQPRWVSHHHPAVLNGQHPDTHHPGLSHSYMADAQYPLPEEVDFVNIDGQGNHV
PPYYGNSVRATVQRYPPTHHGSQVCRPPLLHGSPLWLDGGKALGSHHTASPWNLSFPSKT
SIHHGSPGPLSVYPPASSSSLSGGHASPHLFTFPPTPKDVSPDPSLSTPGSAGSARQDE
KECLKYQVPLPDMSMKLESSHSRGSMTA LGGASSSTHPI TTYPPYVPEYSSGLFPPSSL
GGSPTGFCKSRPKARSSTEGRECVNCGATSTPLWRRDGTGHYLCNACGLYHKMNGQNRP
LIKPKRRLSAARRAGTSCANCQTTTLWRRNANGDPVCNACGLYYKLHNINRPLTMKE
GIQTRNRKMSKSKKKVHDSLEDFPKNSSFNPAALSRHMSSLHISPFSHSSHMLTT
TPMHPPSSLSFGPHHPSSMVTAMG

<210> 19

<211> 3256

<212> PRT

<213> Homo Sapiens

<400>

MWPTRRLVTIKRSGVDGPHPLSLSTCLFGRGIECDIRIQLPVVKQHCKIEIHEQEAIL
HNFSSNTPTQVNGSVIDEPVRLKHGDVITIIDRSFRYENESLQNGRKSTEFPRKIREQEP
ARRVSRSSFSSDPDEKAQDSKAYS KITEGKVSGNPQVHIKNVKEDSTADDSDKSVAQGTT
NVHSSEAGRNGRNAADPISGDFKEISSVKLVSRYGELKSVPTTQCLDNSKKNESPFWKL

YESVKKELDVKSQKENVLQYCRKSGLQTDYATEKESADGLQGETQLLVSRSRKSRPKSGGSG
HAVAEPASPEQELDQNKGKGRDVESVQTPSKAVGASFPLYEPAKMTPVQYSQQQNSPQK
HKNKDLYTTGRRESVNLGKSEGFKAGDKTLTPRKLSTRNRTPAKVEDAADSATKPENLSS
KTRGSIPTDVEVLPTETEIHNEPFLTLWTQVERKIQKDSLSKPEKLGTTAGQMCSGLPG
LSSVDINNFGDSINESEGIPLKRRRVSGFGLRPELFDENLPNTPLKRGFAPTKRKSLV
MHTPPVLKKIIKEQPQPSGKQESGEIHVEVKAQSLVISPPAPSPRKTPVASDQRSSCK
TAPASSSSQTEVPKRGGERVATCLQKRVSIERSQHDILQMICSKRRSGASEANLIVAKS
WADVVVKLGAKQTQTKVIKHGPQRSMNKRQRRPATPKPVGEVHSQFSTGHANSPTIIIG
KAHTEKVHVPARPYRVLNFIISNQKMDFKEDLSGIAEMFKTPVKEQPQLTSTCHIAINS
ENLLGKQFQGTDSEEPPLLPTSESFGGNVFFSAQNAAKQPSDKCSASPPLRRQCIRENGN
VAKTPRNITYKMTSLETKTSDTEPESKTVSTVRSGRSTEFRNIQKLPVESKSEETNTEI
VECILKRGQKATLLQQRREGEMKEIERPFETYKENIELKENDEKMKAMKRSRTWGQKCAP
MSDLTDLKSLPDTELKDARGQNLQTDHAKAPKSEKGKITKMPQCQLQPEPINTPTH
TKQQLKASLGKVGVKEELLAVGKFTRTSGETTHTHREPADGDKSIRTFKESPKQILDPA
RVTGMKKWPRTPKEEAQSLEDLAGFKELFQTPGPSEESMTDEKTTKIACKSPPPESVDT
TSTKQWPKRSLRADVEEEFLALRKLTAGKAMTPKPGAGDEKDIKAFMGTPVQKLDL
AGLPGSKRQLQTPKEKAQALEDLAGFKELFQTPGHTEEELVAAGKTTKIPCDSPQSDPVD
TPTSTKQRPKRKSIRKADVEGELLACRNLMPSAGKAMHTPKPSVGEEKDIIIFVGTPVQKL
DLTENLTGSKRRPQTPKEEAQALEDLTGFKELFQTPGHTEEAVAAGKTTKMPCESSPPES
ADPTSTRRQPKTPLEKRDVQKELSALKLTQTSGETTHTDKVPGGEDKSINA
FRETAKQ
KLDPAASVTGSKRHPKTK
KEKAQPLEDLAGW
KELFQTPV
CTDKPTTHEKTTKIA
CRSQDP
VDTPTSSK
PQSKRSLR
KVDVEE
FFALRK
RTPSAGK
AMHTPK
PAVS
GEKNIYAF
MGTPVQ
KLDL
TENLTGS
KRRL
QTP
KEKA
QALE
DLAG
FKEL
FQTP
GHTEE
SMTNDK
TAKV
ACKS
SSP
PES
MDT
PTST
RRR
PKT
PLG
KRD
IVE
ELS
ALK
QLT
QT
SGET
THTE
PTGD
GKSM
AFMESP
KQIL
DSAAS
LTGS
KRQL
TPKG
SEVP
EDLAG
FIEL
FQTP
SHTE
SMTNE
KTTK
VSYRAS
QPD
LVDT
PTSS
KQPK
RSLR
KAD
TEE
EFLA
FRK
QTP
SAGK
AMHT
PKPAV
GEEK
DINTFLG
TPV
QKLD
QPGN
LPGSN
RRL
QTR
KEKA
QALE
ELTG
FREL
FQTP
CTDN
PTADE
KTTK
KILCK
SPQ
SDPA
DPT
NTK
QRP
KRS
LKK
AD
VEE
EFLA
FRK
LTP
SAGK
AMHT
PKAA
VGE
EKF
INTF
VGTP
VEK
L DLL
GNL
PGSKR
RQ
PQ
TP
KEKA
KALE
DLAG
FIEL
FQTP
GHTEE
SMT
DDK
ITE
VS
CKSP
QDP
PV
KTP
TSS
SK
QRL
KIS
LG
KVG
VKE
EL
PV
GK
LT
QTS
GK
TT
QTH
REAG
DG
KSIK
AF
KES
AK
QML
DP
ANY
GT
MER
WP
RTP
K
P
EE
AQ
SL
ED
LAG
FKEL
FQTP
C
D
HTE
EST
TDD
KTTK
I
ACK
S
P
P
E
S
MD
T
PT
ST
RRR
PK
T
PL
G
K
R
D
I
V
E
E
L
S
A
L
K
Q
L
T
Q
T
G
E
K
I
V
T
P
Q
K
L
L
G
N
L
P
G
S
K
R
Q
P
T
K
E
A
E
D
L
V
G
F
K
E
L
F
Q
T
P
G
H
T
E
E
S
M
T
D
D
K
I
T
K
I
P
C
K
S
P
P
E
L
D
T
A
T
S
T
K
R
C
P
K
T
R
P
K
E
V
K
E
E
L
V
A
G
K
L
T
Q
T
G
E
K
G
T
K
A
F
K
Q
P
A
K
R
N
D
A
E
V
I
G
S
R
R
P
R
A
P
K
E
K
A
Q
P
L
E
D
L
A
G
F
K
E
L
F
Q
T
P
G
H
T
E
E
L
A
N
G
A
A
D
S
F
T
S
A
P
K
Q
T
P
D
S
G
K
P
L
K
I
S
R
R
L
R
A
P
K
V
E
P
V
G
D
V
V
S
T
R
D
P
V
K
S
Q
K
S
N
T
S
L
P
L
P
F
K
R
G
G
GKD
GSV
GT
K
R
L
R
C
P
A
P
E
E
I
V
E
E
L
P
A
S
K
K
Q
R
V
A
P
R
A
R
G
K
S
S
E
P
V
V
I
M
K
R
S
L
R
T
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N
D
M
K
T
N
K
E
E
H
K
L
Q
D
S
V
P
E
N
K
G
I
S
L
R
S
R
R
Q
D
K
T
E
A
E
Q
Q
I
T
E
V
F
V
L
A
E
R
I
E
I
N
R
E
K
K
P
M
K
T
S
P
E
M
D
I
Q
N
P
D
D
G
A
R
K
P
I
P
R
D
K
V
T
E
N
K
R
C
L
R
S
A
R
Q
E
S
S
Q
P
K
V
A
E
E
S
G
G
Q
K
S
A
K
R
I
E
P
A
E
E
L
N
S
N<br

<210> 21
<211> 393
<212> PRT
<213> Homo Sapiens

<400>							
MEEPQSDPSV	EPPLSQETFS	DLWKLLPENN	VLSPLPSQAM	DDLMLSPDDI	EQWFTEDPGP	60	
DEAPRMPEAA	PPVAPAPAPA	TPAAPAPAPS	WPLSSSVPSQ	KTYQGSYGFR	LGFLHSGTAK	120	
SVTCTYSPAL	NKMFQLAKT	CPVQLWVDST	PPPGRTRVRAM	AIYKQSQHMT	EVVRRCPHHE	180	
RCSDSDGLAP	PQHLIRVEGN	LRVEYLDDRN	TFRHSVVVPY	EPPEVGSDCT	TIHYNYMCNS	240	
SCMGGMNRRP	ILTTITLEDS	SGNLLGRNSF	EVRCACPGR	DRRTEEENLR	KKGEPHHELP	300	
PGSTKRALPN	NTSSSPQPKK	KPLDGEYFTL	QIRGRERFEM	FRELNEALEL	KDAQAGKEPG	360	
GSRAHSSHLSK	SKKGQSTSRRH	KKLMFKTEGP	DSD393				

<210> 22
<211> 829
<212> PRT
<213> Homo Sapiens

<400>	MGLPRGPLAS	LLLLQVCWLQ	CAASEPCRAV	FREAEVTLEA	GGAEQECPQA	LGKVFMCPCG
QEPALFSTDN	DDFTVRNGET	VQERRSLKER	NPLKIFPSKR	ILRRHKRDWV	VAPISVPENG	
KGPFPQRNLQ	LKSNKDRDTK	IFYSITGPAGA	DSPPEGVFAV	EKETGWLLLN	KPLDREEIAK	
YELFGHAVSE	NGASVEDPMN	ISIIVTDQND	HKPKFQTQDTF	RGSVLEGVLP	GTSVMQVTAT	
DEDDAIYTYN	GVVAYSIHQS	EPKDPHDLMF	TIHRSTGTIS	VISSGLDREK	VPEYTLTIQA	
TDMGDGDSSTT	TAVAVVEILD	ANDNAPMFDP	QKYEAHVPE	AVGHEVQLT	VTLDLAPNSP	
AWRATYLIIMG	GDDGDHFIT	THPESNQGIL	TTTRKGDFEA	KNQHTLYVEV	TNEAPFVLKL	
PTSTATIVVH	VEDVNEAPVF	VPPSKVVVEVQ	EGIPTGEPCV	VYTAEDPDKE	NQKISYRILR	
DPAGWLAMDP	DSGQVTAVGT	LDREDEQFVR	NNIYEVMVLA	MDNGSPPTTG	TGTLLLTLID	
VNDHGPVPEP	RQITICNQSP	VRHVLNITDK	DLSPHSTSPFQ	AQLTDDSDIY	WTAEVNEEGD	
TVVLSKKFL	KQDTYDVHLS	LSDHGNKEQL	TVIRATVCDC	HGHVETCPGP	WKGGFILPVL	
GAVLALLFLL	LVLVLLVRKK	RKIKEPLLLP	EDDTRDNVFY	YGEEGGGEED	QDYDITQLHR	
GLEARPEVVL	RNDVAPTIIP	TPMYRPRPAN	PDEIGNFIIE	NLKAANTDPT	APPYDTLLVF	
DYEGSGSDAA	SLSLLTSSAS	DODODYDYL	EWGSRFKKLA	DMYGGGEDD		

<210> 23
<211> 933
<212> PRT
<213> Homo Sapiens

<400>
MTELKAKGPRAPHVAGGPPSPEVGSPLLCRPAAGPFPGSQTSDEKTDQQLSDVEGAYSRAEATRGAGGSSSSPPEKDSGLLSDVLTLLAPSGPGSQSPPPACEVTSSWCLFGPELPEDPPAAPATQRVLSPMSRSGCKVGDSGGTAAAHKVLPRLGLSPARQLLPASESPHWGAVVKPSQAAAVEEEEDGSESEESAGPLLKGKPRALGGAAGGGAAAVPPGAAAGGVALVPKEDSRFSAPRVALVEQDAPMAPGRSPLATTVMDFIHVPILPLNHALLAARTQLLEDESYDGGAGAASAFAPPRSSPCASSTPVAVGDFPDCAYPPDAEPKDDAYPLYSDFQPPALKIKEEEEAGAEASARSRSYLVAGANPAAFPDFPLGPPPPLPPRATPSRPGEAAVTAAAPASASVSSASSGSTLECILYKAEGAPPOQFAPPCKAPGAGSGCLLPRDGLPSTSASAAAAGAAPALYPALGLNLGPOLGY

QAAVLKEGLPQVPPYLNLYRPDSEASQPSPQYSFESLPQKICLICGDEASGCHYGVLCGSCKVFFKRAMEGQHN
 YLCAGRND CIVDKIRRKNCPACRLRKCQAGMVLGGRKFKKFNKVRVVRALDAVALPQPGVPNESQALSQRFTF
 SPGQDIQLIPPLINLLMSIEPDVIYAGHDNTKPDTSSLLTSLNQLGERQLLSVVKWSKSLPGFRNLHIDDQITL
 IQYSWMSLMVFGLGWRSYKHVGQMLYFAPDLILNEQRMKESSFYSLCLTMWQIPQEFVKLQVSQEEFLCMKVLL
 LLNTIPLEGLRSQTQFEEMRSSYIRELIKAIIGLRQKGVVSSSSQRFYQLTKL LDNLHDLVKQLHLYCLNTFIQSRA
 LSVEFPEMMSEVIAAQLPKILAGMVKPLL FHKK

<210> 24

<211> 805

<212> PRT

<213> Homo Sapiens

<400>

MAFSPWQILSPVQWAKWTWSAVRGGAAGEDEAGGPEGDPEEEDSQAETKSLSFSSDSEGNFETPEAETPIRSPFK
 ESCDPSLGLAGPGAKSQESQEADEQLVAEVVEKCSSKTCSPSENEVPQQAIDSHSVKNFREEPEHDFSKISIVR
 PFSIETKDSTDISAVLGTKAHGCVTAVSGKALPSSPPDALQDEAMTEGSMVTLEASAEADLKAGNSCPELVPS
 RRSKLRKPKPVPLRKAIGGEFSDTNAAVEGTPLPKASYHFSPEELDENTSPLLGDARFQKSPPDIKETPGTLSS
 DTNDSGVELGEESRSSPLKLEFDFTEDTGNIEARKALPRKLGRKLGSTLTPKIQKDGISKSAGLEQPTDPVARDG
 PLSQLTSSKPDPSQWESPSFNPPGSHSVLQNSPPLSSEGSYHFDPDNFDESMDPFKPTTTLTSSDFCSPTGNHVE
 ILESPEKKAKSRLITSGCKVKKHETQSLALDACSREGAVISQISDISNRDGHATDEEKLASTSCGQKSAGAEVKG
 EPEEDLEYFECNSNPVSTINHAFFSSSEAGIEKETCQKMEEDGSTVGLLESSAEKAPVSVSCGGESPLDGICLSE
 SDKTAVLTLIREEIITKEIEANEWWKKYEETRQEVLEMRKIVAEYEKTIAQMIEDEQRTSMTSQKSFQQLTMEKE
 QALADLNSVERSLSDFRRYENLKGVLLEGFKKNEEALKCAQDYLARVQKEEQRYQALKIHAEELDKANEEIAQ
 VRTKAKAESAALHAGLRKEQMKVESLERALQQQNQIEELTKICDELIAKLGKD

<210> 25

<211> 1026

<212> PRT

<213> Homo Sapiens

<400>

MGGSQSLQPAPASDLNLEASEAMSSDSEEAFETPESTTPVKAPPAPPPPPPPEVIPEPEVSTQPPPPEPGCGSETV
 PVPGPRSDSVEGSPFRPPSHFSAVFDEDKPIASSGTYNLDFDNIELVDTFQTLERPRASDAKNQEGKVNRKRS
 TDSVPISKSTLSRSLSLQASDFDGASSSGNPEAVALAPDAYSTGSSSASSTLKRKPRPPSLKKQTTKKPPTET
 PPVKETQEPDEESLVPSPGENLASETKTESAKTEGPSPALLEETPLEPAVGPKAACPLDSES AEGVVPPASGGGR
 VQNSPPVGRKTLPLTTAPEAGEVTPSDGGQEDSPA KGLSVRLEFDYSEDKSSWDNQQENPPPTK KIGKKPVAKM
 PLRRPKMKKTPEKLDNTPASPPRSPAEPNDIPIAKGTYTFDIDKWDPPNFNPFSSSTS KMQESP KLPQQSYNFDPD
 TCDESVPDFKTSKTPSSPSKSPASFEIPASAMEANGVGDGLNKPAKKKTP LKTDTRVKKSPRSPLSDPPS
 QDPTPAATPETPPVISAVVHATDEEKЛАVTNQKWTCTVDLEADKQDYPQPSDLSTFVNETKFSSPTEELDYRNS
 YEIEYMEKIGSSLQDDDAPKKQALYLMFDTSQESPVKSSPVRMSESPTPCSGSSFEETALVNIAKNQHPVPR
 GLAPNQESHLQVPEKSSQKELEAMGLGTPSE AIEITAPEGSFASADALLSLRAHPVSLCGALDYLEPD LAEKNPP
 LFAQKLQEELEFAIMRIEALKLARQIALASRSHQDAKREA AHPTDVSISKTALYSRIGTAEVEK PAGLLFQQPDL
 DSALQIARAEEITKEREVSEWKDKYEE SREV MEMRKIVAEYEKTIAQMIEDEQREKS VSHQTVQQLVLEKEQAL
 ADLNSVEKSLADLFRRYEKMKEVLEGFRKNEEVLRKCAQEYLSRVKKEEQRYQALKVHAEELDRANA EIAQVRG
 KAQQEQAAHQASLRKEQLR DALERTLEQKNKEIEELTKICDELIAKMGKS

<210> 26

<211> 838

<212> PRT

<213> Homo Sapiens

<400>

MSLQVLNDKNVSNEKNTENCDFLFSPPPEVTGRSSVLRVSQKENVPPKNLAKAMKVTFQTPLRDPQTHRILSPSMA
 SKLEAPFTQDDTLGLENSHPWTQKENQQLIKEVDAKTTHGILQKPV EA DTL LGDASPAFGSGSSSES GPGALA
 DLDCSSSSQSPGSSENQMVSPGKVSGSPEQAVEENLSSYSLDRRVTPASETLEDPCRTEQHKAETPHGAAEECK
 AETPHGAAEEC RHGGVCAPAAVATSPPGAI PKEACGGAPLQGLPGEALGCPAGVGT PVPADGTQTLTCAHTSAPE
 STAPTNHLVAGRAMTLSPQEEVAAGQMASSSRSGPVKLEFDVSDGATSKRAPP RRLGERSGLKPPLRKA AVRQQ
 KAPQEVEEEDGRSGAGEDPPMPASRGSYHLDWDKMDP NFIPFGGDTKSGCSEAQPPES PETRLGQPAAEQL HAG
 PATEEPGPCLSQQLHSASAEDTPVVLQAAETPTAESKERALNSASTSLPTSCPGSEPVP THQQGQPALELK EESF
 RDPAEV LGTGAEV DYLEQFGTSSFKESALRKQSLYLKFDPLLRDSPGP RVPVATETSSMHGANETPSGRPREAKL
 VEFDFLGALDIPVPGPPPGVPA PGGPPLSTGP IV DLLQYSQKDLDAVVKATQEE NRELRSRCEELHGKNLELGK I
 MDRFEV VYQAMEEVQKQKELSKAEIQKVLKEKDQLTDLNSMEKSFS DLFKRFEKQKEVIEGYRKNEE SLKKCV

EDYLARITQEGQRYQALKAHAEKLQLANEEIAQVRSKAQAEALALQASLRKEQMRIQSLEKTVEQTKENEELT
RICDDLISKMEKI

<210> 27
<211> 564
<212> PRT
<213> Homo Sapiens

<400>
MASTTTIRSHSSSRGFSANSARLPGVSRSRGFSSVSVRSRGSGGLGGA
CGGAGFGSRSLYGLGGSKRISIGGSCAISGGYGSRAGGSYGFGGAGSGF
GFGGGAGIGFGLGGAGLAGGFGGPGFPVCPPGIQEVTVNQSLLTPLNL
QIDPTIQRVRAEEREQIKTLNNKFASFIDKVRFLEQQNKVLETWTLLQE
QGTKTVRQNLEPLFEQYINNLRQLDSIVGERGRLDSELRGMQDLVEDFK
NKYEDEINKRTAAEENEFVTLKKDVDAAYMNKVELQAKADTLTDEINFLRA
LYDAELSQMOTHISDTSVVLMSMDNNRNLDLDSIIAEVKAQYEEEIAQRSRA
EAESWYQTKEELOVQTAGRHGDDLRTNKQEIAEINRMIQRLRSEIDHVKK
QCANLQAAIADAERGEMALKDAKNKLEGLEDALQAKQDLARLLKEYQE
LMNVKLALDVEIATYRKLEGECRNLNGEVGVQVNISVVQSTVSSGYGGA
SGVGSGLGLGGGSSYSYGSGLGVGGFSSSSGRAIGGLSSVGGGSTIK
YTTTSSSSRKSYKH

<210> 28
<211> 430
<212> PRT
<213> Homo Sapiens

<400>
MSFTTRSTFSTNYRSLGSVQAPSYGARPVSSAASVYAGAGGSGSRISVSRSTSFRGGMGSGLATGIAGGLAGMG
GIONEKETMQSLNDRLASYLDVRSLTENRRLESKIREHLEKKGPQVRDWSHYFKIIEDLRAQIFANTVDNARI
VLQIDNARLAADDFRVKYETELAMRQSVENDIHGLRKVIDDTNITRLQLETEIEALKEELLFMKKNHEEEVKGLQ
AQIASSGLTVEVDAPKSQDLAKIMADIRAQYDELARKNREELDKYWSQQIEESTTVVTQSAEVGAAETTLTELR
RTVQSLEIDLDSMRNLKASLENSLREVEARYALQMEQLNGILLHLESELAQTRAEGQRQAQYEALLNIKVLEA
EIATYRRLLEDGEDFNLGDALDSSNSMQTIQKTTTRIVDGKVVSETNDTKVLRH

<210> 29
<211> 498
<212> PRT
<213> Homo Sapiens

<400>
1 mtvflsfaf1 aailthigcs nqrspensg rrynrighqg caytfilpeh dgncresttd
61 qyntnalqrd aphvepdfss qklqhlehvm enytqwlqkl enyivenmks emaqiqnav
121 qnhtatmlei gtsllsqtae qtrkltdvet qvlnqtsrle iqqlenslst yklekqlqq
181 tneilkihk nsllehhile megkhkeeld tlkeekenlq glvtrqtyii qelekqlnra
241 ttnnsvlqkq qlelmdtvhn lvnltkegv llkggkreee kpfrdcadvy qagfnksgiy
301 tiyinnmpep kkvfcnmvdn gggwtvighr edgslfdfqrg wkeykmqfgn psgeywlgne
361 fifaitsqrq ymlrielmw egnraysqyd rfhighnekqn yrlylkghtg tagkqsslil
421 hgadfstkda dndncmckca lmltggwwfd acgpsnlngm fyttagqnhgk lngikwhyfk
481 gpsyslrstt mmirpldf

<210> 30
<211> 344
<212> PRT
<213> Homo Sapiens

<400>
1 maqkensypw pygrqtapsq 1stlpqrblr kepvtosalv lmsrsnvqpt aapgqkvmen
61 ssgtpdilir hftiddfeig rplgkgkfqn vylarekksh fivalkvlfk sqiekegveh
121 qlrreieiga hlhhpnirl ynyfydrri ylileyaprg elykelqksc tfdeqrtati
181 meeladalmy chgkkvihrd ikpenllgl kgelkiadfg wsvhapslrr ktmcgtldyl
241 ppemiegrmh nekvdlwcig vlcyclvgn ppfesashne tyrrivkvdl kfpasvptga

301 qdlistkllrh npserlplaq vsahpwvran srrvlppsal qsva

<210> 31
<211> 655
<212> PRT
<213> Homo Sapiens

<400>
1 msssnvevfi pvsqgnngf patvsndlka ftegavlsfh nicyrvklks gflpcrkpv
61 keilsningi mkpqlnailg ptgggkssl1 dvlaarkdps glsgdvling aprpanfkcn
121 sgvvqddvv mgltvrenl qfsaalrlat tmtnhekner inrvieelgl dkvadskvgt
181 qfirgvsgge rkrtsigmel itdpsilsls epttgdssst anavllllkr mskqgrtiif
241 sihqprysif klfdsllla sg1mfhgpa qealgyfesa gyhceaynnp adffldiing
301 dstavalnre edfkateiie pskqdcklie klaeiyvnss fyketkaeh qlsggekkkk
361 itvfkeisyts tfchqlrwv skrsfknl1g npqasiaqii vtvvlglvig aiyfglknds
421 tgiquragvl ffltnqcfsvsavelfvv ekklifiheyi sggyrvssyf lgkllsdllp
481 mrmlpsiift civyfmlglk pkadaffvmm ftlmmvaysa ssmalaiaag qsvvsvatll
541 mticfvfmimi fsllvnltt iaswlswlqy fsiprygfta lqhneflgqn fcpglnatgn
601 npcnyatctg eeylvkqgid lspwglwknh valacmivif ltiaylkllf lkys

<210> 32
<211> 412
<212> PRT
<213> Homo Sapiens

<400>
1 mqpsslpla lcllaapasa lvriplhkft sirrtmsevg gsvedliakg pvsqyavp
61 avtegpipev lknymdaqyy geigigtpq cftvvfdtgs snlwpsihi klldiacwi
121 hkynsdsst yvkngtsfdi hygsgs1sgy lsqdtvsvpc qsassasalg gvkverqvfg
181 eatkqpgitf iaakfdgilg mayprisvnn vlpvfdnlmq qklvdqnifs fylsrpdpaq
241 pggelmlggt dskykgs1s ylnvtrkayw qvhldqveva sgtlckege eaivdtgt1
301 mvgpvdevre lqkaigavpl iqgeymipce kvstlpaitl klggkgykls pedytlkvsq
361 agktlclsgf mgmdipppsg plwilgdvfi gryytfdrd nnrvgfæaa rl

<210> 33
<211> 750
<212> PRT
<213> Homo Sapiens

<400>
1 mgksesqmdi tdintpkpkk kqrwtrleis lsvlvl1ti iavrmiyalta tyddgickss
61 dciksaarli qnmdattec rffffkyacgg wlkrnvipet ssrygnfdil rdelevvlkd
121 vlqepktedi vavqkakaly rscinesaid srggeplkl lpidygwpva tenweqkyga
181 swtaekaiqaq lnskykkvl inlfvgtddk nsrnvhviid qprlgipsrd yyectgiyke
241 actayvdfmi svarlirqee rlpidenqla lemnnkvmele keianatakp edrndpmilly
301 nkrmrlaqiqn nfsleingkp fswlnftnei mstvnisitn eedvvvyape yltklkpilt
361 kysardlqn1 mswwrfimdlv ss1srtypes rnafrkalyg ttsetatwrr canyvngnme
421 navgrlyvea afageskhvv edliaqirev fiqtlddltw mdaetkkrae ekalaikeri
481 gypddivsnd nklnneylel nykedeyfen iiqnlkfsqs kqlkkrekv dkdeiwsgaa
541 vvnavfyssgr nqivfpagil qppffsaqqs nslnyggigm vigheithgf ddngrfnkd
601 gd1vdwwtqq sasnkeqsq cmvyqygnfs wdlaggqhlnt gintlgenia dngglgqayr
661 ayqnyikkng eekllpgld1 nhkqlfflnf aqvwcgtyp eyavnsiktd vhspgnfrii
721 gtlqnsaefs eafhcrknsy mnpekkcrvw

<210> 34
<211> 493
<212> PRT
<213> Homo Sapiens

<400>
1 mdkfwwhaaw glclvplsla qidlnitcrf agvfhevng rysisrteaa dlckafnstl
61 ptmaqmekal sigfetcryg fieghvvipr ihpnsicaan ntgyvilytyn tsqydtycfn
121 asappleedct svtdlpnafd gpititivnr dgtryvqkge yrtnpediyp snptdddss
181 gsserssts ggyifytfst vhpiptdedsp wittdstdrip rtnmdsshst tlqptanpnt
241 glvedldrtg plsmmttqgsn sqsfstshieg leedkdhptt stltssnrnd vtggrrdpnh
301 segsthllleg ytshyphktke srtfipvtsa ktgsfgvtav tvgdnsnvn rsldgdqdtf
361 hpsgshtht gsesdghshg sqeggantts gpirtpqipe wliilaslla lalilavcia
421 vnsrrrcgqk kklvinsngn avedrkpsgl ngeasksqem vhlnkesse tpdqfmtade
481 trnlqnvdmk igv

<210> 35
<211> 472
<212> PRT
<213> Homo Sapiens

<400>
1 mttcsrqfts sssmkgs.cgi gggigggssr issvlaggsc rapstygggl svsssrffssg
61 gayglgggyg ggfssssssf gsgfgggygg glgaglgggf gggfaggdgl lvgsekvtmq
121 nlndrlasyl dkvrableean adlevkirdw yqrqrpaek dyspyfktie dlrnkiltat
181 vdnanvllqi dnarlaaddf rtkyeteinl rmsveading lrrvldeltl aradlemqie
241 slkeelaylk knheeemnal rgqvggdvnv emdaapgvdl srilnemrdq yekmaeknrk
301 daeewfftkt eelnrevatn selvqsgkse iselrrtmqn leielqsqsls mkaslensle
361 etkgrycmql aqiqemigs veeqlaqlrce meqqnqeyki lldvktrleq eiatyrrlle
421 gedahlsssq fssgsqssrd vtsssrqirt kvmdvhdgkv vstheqvlrt kn

<210> 36
<211> 604
<212> PRT
<213> Homo Sapiens

<400>
1 mlaralllca vlalshtanp ccshpcqnrg vcmsvgfdqy kcdctrtrgy gencstpefl
61 triklflkpt pntvhyilth fkfwnvvnn ipflrnaims yvltsrshli dsppptynady
121 gyksweafsn lsyytralpp vpddcptplg vkgkkqlpds neiveklllr rkfpdpqgs
181 nmmfaffaqh fthqffktdh krgpaftngl ghgvdlnhiy getlarqrkl rlfkdgkmky
241 qidgemypv tkvdtqaemi yppqvpehrl favgqevfgl vpglmmyati wlrehnrvcd
301 vlkqehpewg deqlfqtsrl iligetikiv iedyvqhlsg yhfklkfdpe llfnkqfqyq
361 nriaaefntl yhwhpllpdt fqihdqkyny qqfiylnnsil lehgitqfve sftrqiaagr
421 aggrnvppav qkvsqasidq srqmkyqsfn eyrkrfmlkp yesfeeltge kemsaeleal
481 ygdidavely pallvekprp daifgetmve vgapsfslkgl mgnvicspay wkpstfggev
541 gfqintasi qslicnnvkg cpftsfsvpd peliktvtn assrsrglhd inptvllker
601 stel

<210> 37
<211> 288
<212> PRT
<213> Homo Sapiens

<400>
1 mvvgggdve dvtprrggcq isgraargcn gipgaaawea alprrrprrh psvnprsraa
61 gsprtrgrrt eerpsgsrlg drgrgralpg grlggrgrgr apervggrgr grgtaapraa
121 paargsrpgr agtmaagsit tlpalpedgg sgafppghfk dpkrlyckng gfflrihpdg
181 rvdgvrekst phiklqlqae ergvvsikgv canrylamke dgrellaskcv tdecffffrl
241 esnnnyntyrs rkytswyval krtgqyklgs ktgpgqkail flpmsaks

<210> 38

<211> 442
<212> PRT
<213> Homo Sapiens

<400>
1 myqlamaan hgpppgayqa ggppgfmhga gaasspvylp tprvpssvlg lsylqggag
61 sasggpsggs pggasgagp gtqqgspgws qagatgaayt pppvsprfsf pggtgslaaa
121 aaaaaareaa ayssggagaag aglagreqyg ragfagsyss pypaymadvg aswaaaaas
181 agpfdsplvh slpgrapaa rhpnldmfdd fsegrecvnc gamstplwrr dgtghylcna
241 cglyhkmngi nrplikpqrr lsasrrvgls cancqtttt lwrRNAegep vcnacgymk
301 lhgvprplam rkegiqtrkr kpknlnkskt paapsgsesl ppasgassns snattssee
361 mrpikepgl sshyghsssv sqtfsvsams ghgpsihpvl salklspqgy aspvsqspqt
421 sskqdswns1 vladshgdii ta

<210> 39
<211> 826
<212> PRT
<213> Homo Sapiens

<400>
1 megaggandk kkiserrke ksrdaarsrr skesevfyel ahqlplphnv sshldkasvm
61 rltisylrvr klldagdldi eddmkaqmnc fylkaldfv mvltddgdmi yisdvnkym
121 gltqfeltgh svfdftpcd heemremlth rnlgvkgke qntqrssflr mkctltsrgr
181 tmniksatwk vlhctghihv ydtntsnpqc gykkppmtcl vliceipiph snieipldk
241 tflsrhslm kfsycderit elmyepeel lgrsiyeyh aldsdhltkt hhdmftkgqv
301 ttgqyrmak rggvvwvetq atviyntkns qpqcivcvny vvgiiqhdi ifslqqtecv
361 lkpvessdmk mtqlftkves edtsslfdkl kkepdalt11 apaagdtiis ldfgsndtet
421 ddqgleevpl yndvmlpspn eklqninlam splptaetpk plrssadpal ngevalklep
481 npeslelsft mpqiqdqtps psdgstrqss pepnspseyc fyvdsmvne fkleveklf
541 aedteaknpl stqtdlldle mlapyipmdd dfqlrsfdql splessasp esaspqstvt
601 vfqqtqiqep tanatttat tdelktvtd rmedikilia spspthihke ttsatsspyr
661 dtqsrtaspn ragkvieqt ekshprspnv lsvalsqrtt vpeeelnpi lalqnaqrkr
721 kmehdgslfq avgigtlqq pddhaattsl swkrvgcks seqngmeqkt iilipsdlac
781 rllgqsmdes glpqltsydc evnapiqgsr nllqgeellr aldqvn

<210> 40
<211> 707
<212> PRT
<213> Homo Sapiens

<400>
1 mslwqplvlv llvlgccfaa prqrqstlvl fpgdlrtnlt drqlaeeyly rygytrvaem
61 rgeskslgpa llllqkqlsl petgeldsat lkamrtpcrg vpdlgrfqtf egdlkwhhhn
121 itywqnyse dlpravidda farafalwsa vtpltfrvy srdadiviqf gvaehgdgyp
181 fdgkdgllah afppgpgiqg dahfdddew slgkvvvpt rfgnadgaac hfpfifegr
241 ysacttdgrs dglpwcssta nytdddrfgf cpserlytqd gnadgkpcqf pfifqgqsys
301 acttdgrsdg yrwcattany drdklfgfcpr tradstvmgg nsagelcvfp ftflgkeyst
361 ctsegrgdgr lwcattsnfd sdkwgfcpr qgyslflvaa hefghalglid hssvpealmy
421 pmyrftegpp lhkddvngir hlygprpepe prpppttpq ptapptvcpt gpptvhps
481 ptagptgpps agptgpptag pstattvpls pvddacnvni fdaiaeignq lylfkdgkyw
541 rfsegrgsrp qgpfliaidkw palprkldsv feerlsklf ffsgrqvwy tgasvlgprr
601 ldklgadv aqvtgalrsg rgkmlfsgsrlwrfdvkaq mvdprsasev drmfpvgpld
661 thdvcqyrek ayfcqdrfyw rvssrselnq vdqvgvtyd ilqcped

<210> 41
<211> 715
<212> PRT
<213> Homo Sapiens

<400>

1 maanmyrvgd yvyfensssn pylirrieel nktangnvea kvvcfyrrrd isstlialad
61 khatlsvcyk agpgadngee geieeemeng emvdlepk hqlrhrelfl sqleslp
121 hirgkcsvtl lneteslksy leredffys lvydpqqktl ladkgeirvg nryqaditdl
181 lkeeedgrd qsrletqvwe ahnpltdkqi dqflvvarsv gtfaraldcs ssvrqpslhm
241 saaaaasrdit lfhamdtlhk niydiskais alvpqggpvl crdemeeewsa seanlfeeal
301 ekygkdftdi qqdflpwksl tsieyyymw kttdryvqqk rlkaaeaeask lkqvyipnyn
361 kpnpnqisvn nvkagvvngt gapgqspgag racescyttq syqwyswgpp nmqcrilcasc
421 wtywkkyggl kmptrllder pgpnrsnmmp hglparssgs pkfamktrqa fylhttkltr
481 iarrlcreil rpwhaaripy lpinsaaika ectarlepas qsplvlkqav rkpleavly
541 lethprppkp dpvksvssvl sslltpakvap vinngsptil gkrseyeqhng vdgnmkkrl1
601 mpsrglanhg qtrhmgpsrn lllngksypt kvrlirggsl ppvkrrrmnw idapgdvfym
661 pkeetrkirk llsssetkra arrpykpiat rqsgalpprp pppapvndep ivied

<210> 42

<211> 180

<212> PRT

<213> Homo Sapiens

<400>

1 cceprgsrar fgcwrlqpef kpkqlegtma ncertfiaik pdgvqrglvg eiikrfeqkg
61 frlvglkfmc asedllkeh vdlkdrpffa glvkmhsgp vvamvwegln vvktgrvmlg
121 etnpadskpg tirgdfciqv grniihgsds vesaekeigl wfhpeelvdy tscaqnwiye

<210> 43

<211> 640

<212> PRT

<213> Homo Sapiens

<400>

1 mserkegrgk gkgkkkergs gkkpesaags qspalppqlk emksqesaag sklvlrcets
61 seysslrfkw fkngnelnrk nkpnikiqk kpgselrin kasladsgey mckviskln
121 dsasanitiv esneiitgmp astegayvss espirisvst egantsssts tttgtshlv
181 kcaekektfc vnggecfmvk dlsnpsrylc kcpgftgar ctenpmkvq nqekaeelyq
241 krvlititgic iallvvgimc vvaycktkq rkkldrlrq slrserrnnmm niangphhp
301 pppenvqlvn qyvsknviss ehivereaat sfstshytst ahhsttvqt pshwsngh
361 esilseshsv ivmssvensr hssptggprg rlngtggpre cnslrhare tpdssyrdsp
421 seryvsamtt parmpspvdh tpsspkspss emspvssmt vsmpsmavsp fmeeeerp11
481 vtpprlrekk fdhhpqfqfss fhhnpahdsn slpasplriv edeeyettqe yepaqepvkk
541 lansrrakrt kpnhianrl evdsntssqs snsesetede rvgedtpflg iqnpblaasle
601 atpafrlads rtncagrfst qeeiqarlss vianqdpiav

<210> 44

<211> 645

<212> PRT

<213> Homo Sapiens

<400>

1 mserkegrgk gkgkkkergs gkkpesaags qspalppqlk emksqesaag sklvlrcets
61 seysslrfkw fkngnelnrk nkpnikiqk kpgselrin kasladsgey mckviskln
121 dsasanitiv esneiitgmp astegayvss espirisvst egantsssts tttgtshlv
181 kcaekektfc vnggecfmvk dlsnpsrylc kcpneftgdr cqnyvmasfy khlgiefmea
241 eelyqkrvt itgiciallv vgimcvvayc ktkkqrkkh drlrqslrse rnmnniang
301 phhpnnppen vqlvnqyvsk nvissehive reaetssts hytstahhst tvttqtpshsw
361 sngthesils eshsvivmss vensrhsspt ggprgrlntg ggprecnsl rharetlds
421 rdsphseryv samtparms pvdftppsp ksppsemssp vssmtvsmps mavspfmeee
481 rplllvtppr lrekkfdhhp qqfssfhhp ahdsnslpas plrivedeey ettqeyepaq
541 epvkklansr rakrtkpngh ianrlevdsn tssqssnses etedervged tpflgiqnpl

601 aasleatpaf rladsrtnpa grfstqeeiq arlssvianq dpiav

<210> 45
<211> 198
<212> PRT
<213> Homo Sapiens

<400>
1 msnrvsngs pslermdarq aehpkpsacr nlfgpvthee ltrdlehcr dmeeasqrkw
61 nfdfqnhkpl egkyewqeve kgslpfyyr pprppkgack vpaquesqdvs gsrpaaplig
121 apansedthl vdpktdpsds qtglaeqcaq irkrpatdds stqnkranrt eenvsdgspn
181 agsveqtpkk pglrrrqqt

<210> 46
<211> 465
<212> PRT
<213> Homo Sapiens

<400>
1 mivfvrfnss hgfpvevdsd tsifqlkevv akrqgvpadq lrvifagkel rndwtvqncd
61 ldqqsivhiv qrpwrkqgem natggddprn aaggcerepq sltrvdlsst vlpgdsvbla
121 vilhtdsrkd sppagspagr siynsfyvyc kgpcqrqvpg klrvqcstcr qatlrltqgp
181 scwddvlipn rmsgecqspn cpgtsaefff kcgahptsdk etpvalhlia tnsrnitcit
241 ctdvrspvlg fqcnrhvic ldcfhlycvr rlndrqfvhd pqlgyslpcv agcpnslike
301 lhhfrilgee qynryqqyga eecvlqmggv lcprpgcag llpepdqrkv tceggnglgc
361 gafafcrecke ayhegecsav feasgttqa yrveraaeq arweaasket ikkttkpcpr
421 chvpekngg cmhmkcpqpq crlewcwnncg cewnrvcmgd hwfdv
<210> 47
<211> 395
<212> PRT
<213> Homo Sapiens

<400>
1 mrallarlll cvlvsdskg snelhqvpky sktcyeqngf fyrgkastdt mgrpclpwns
61 atvllqqtyna hrsdalqlgl gkhnycrnpd nrrrpwcyvq vglklvqec mvhdcadgkk
121 psspeelkf qcggktlrpr fkiiggeftt ienqpwfaai yrrhrggsvt yvcggslisp
181 cwvisathcf idypkkediy vylgrsrlns ntqgemkfev enlilhkdy adtlahhndi
241 allkirskeg rcaqpsrtiq ticlpsmynd pqfgtseit gfgkenstdy lypeqlkmtv
301 vklishrecq qphyygsevt tkmlcaadpq wktscqgds ggplvcsllqg rmtltgivsw
361 grgcalkdkp gvytrvshfl pwirshtkee nglal

<210> 48
<211> 94
<212> PRT
<213> Homo Sapiens

<400>
1 mgseletame tlinvhahs gkegdkykls kkelkellqt elsgfldaqk dv davdkvmk
61 eldengdgev dfqeyvvla altvacnnff wens

<210> 49
<211> 1630
<212> PRT
<213> Homo Sapiens

<400>
1 mlkiplwrc nrhvesvdkr hcsllqavpee iyrysrslee llldanqlre lpkpffrlln

61 lrklglsdne iqrlpppevan fmqlveldvs rndipeipes ikfckaleia dfsgnplsrl
 121 pdgftqlrsl ahlalndvsl qalpgdvgnl anlvtleire nllkslpas1 sflvkleqld
 181 lggndlevlp dtlgalpnlr elwldrnqls alppelgnlr rlvclvdvsen rleelpaelg
 241 glvlldll1 sqnllrrl1d gigqlkqlsi lkvdqnrlce vteaigdcen lseliltenl
 301 lmalprslgk ltkltnlnvd rnhealpp eiggcalsvl slrdnrlavl ppelahttel
 361 hvldvagnrl qslpfalh1 nlkalwlaen qaqpmlrfqt eddardgekv ltcylppqqp
 421 plsledagqq gslsetwsda ppsrvsviqf leapigdeda eeaaaekrg1 qrratphpse
 481 lkvmkrssieg rrseacpcqp dsgsplpaee ekrlsaesgl sedsrpsast vseaepggs
 541 aeaaqggsqeq attaggeeda eedyqeqptvh faedallpgd dreieeggpe apwtlpggrq
 601 rlirkdtpy kkhfkisklp qpeavvallq gmqpdgegpv apggwhngph apwapraque
 661 eeeeeegspq eeeeeeeeen raeeeeaste eedkegavvs apsvkgvsfd qannlliepa
 721 rieeeeeltlt ilrqgtgg1gi siaggkgst1 ykgddegifi srveeegpaa ragvrvgdkl
 781 levngvalqq aehheaveal rgagtavqmr vwrermvepe navtitplrp eddysprerr
 841 ggglrlpllp pespgplqr hvaclarser glgfsiaggk gstyragda gifvsriaeg
 901 gaahragtlq vgdrvlsing vdvtearhdh avslltaasp tiallerea ggplppsplp
 961 hsspptaava ttsittatpg vpglpslaps llaaaalegpy pveeirlpra ggplglsivg
 1021 gsdhsshpf1g vqepgvf1sk v1prglaars glrvdrila vngqdvrdat hqeavsallr
 1081 pc1elsllvr rdpappglre lc1qkapger lgisirrggar ghagnprdt degifiskvs
 1141 ptgaagr1dgr lrvglrllev nqqsllglth geavql1rsv gdtltvlvc1 gfeastdaal
 1201 evspgvianp faagighrns lesissidre lspeegpgkek elpgqtlhwg peateaagrg
 1261 lqplkldyra laavpsagsv qr1psgaagg kmaesp1cps1 gqqppsp1ppsp delpanvkqa
 1321 yrafaavpts hppedapaqp pt1pgpaaspe qlsfrerqky felevrvpqa egppk1rvslv
 1381 gadd1rk1m1qe eearklqqkr aqmlre1aaea gaearlal1dg etlgeeeeqed eqppwasppsp
 1441 tsrqspaspp pl1ggap1vrt akaerrhger lrvqspeppa peral1spak1 raleakral
 1501 wraarmksle qdalraqm1l srsqegr1grt1 gplerlaeap spapt1ps1pt1 ved1gpqtst
 1561 spgr1spdf1 eelrslep1sp1 sp1gpqeedge valv1lgrps pgav1pedva lc1ssrrpv1rp
 1621 grr1lg1pvps

<210> 50
<211> 377
<212> PRT
<213> Homo Sapiens

<400>
 1 mcddeettal vcdngsg1vk agfagddapr avfpsivgrp rhqgvmvgmg qkdsyvgdea
 61 qskrgiltlk ypiehg1itn wddmekiwhh tfynelrvap eehpt1ltea plnpkanrek
 121 mtqimfetfn vpamyvaiqa v1slyasgrt tgivldsgdg vthnvpiyeg yalphaimrl
 181 dlagr1ldy1 lm1kiltergy sfvtt1aerei vrdikek1cy valdfenema taasssslek
 241 sy1pdgqvi t1ignerfr1cp etlfq1psfig mesagihett ynsimkcdid irkdlyannv
 301 lsgg1ttmypg iadrmqkeit alapstmkik iiapperkys vw1ggs1las l1tfqqmwi1s
 361 kqeydeagps ivhrkcf

<210> 51
<211> 1170
<212> PRT
<213> Homo Sapiens

<400>
 1 mg1awglg1v1 flmhvcgtnr ipesggdns1 fd1f1eltgaa rkgsgr1lvk gpdpsspafr
 61 iedan1lipp1 pddkf1qd1vd avraekgf11 las1rqmkkt rgt1laler1 dhsgqvf1svv
 121 sngkag1t1dl sltvqgkqhv vsveeallat gq1wks1t1fv qedraql1y1d cekmenaeld
 181 vpiqsvf1trd lasiar1ria kggvndnf1q1g v1qn1vrf1fg t1ped1l1rnk gcssst1sv1l
 241 t1dn1nnv1ngs spa1rt1ny1g hkt1ndlqaic g1scdel1ssm v1el1rl1rt1 v1tl1qds1irk
 301 vteenkelan elrrpplcyh ngvqyrnnee wtvdsctech cqnsvtickk vsc1p1mpcsn
 361 atvpd1geccp rcwp1s1s1add gwspwsewts cstsc1ng1q1g qr1r1sc1ds1n nrcegssvqt
 421 rtch1qecdk r1fkqdg1gwsh w1spwss1c1vt cg1dg1vit1rir lc1nspsp1qmn gkpcegeare
 481 tkackkd1acp inggwgpwsp wd1c1s1vtcg1g gvq1k1rs1lc1n n1ptpqf1gg1kd cvgdvt1enqi
 541 cnkqd1cp1dg clsnpc1f1agv kct1sy1pdg1sw kcg1ac1ppg1ys gng1qct1dvd ec1kevp1dacf
 601 nhngehr1cen tdp1g1nc1lpc ppr1ft1gs1q1pf ggg1vehatan kq1vckpr1n1pc t1dg1thdc1kn
 661 akcn1ylghys dpmyrceckp gyagng1i1cg edtd1dg1wpn en1vc1vanat yhckkd1nc1p1n

721 lpnsgqedyd kdgiacdd dddndkipdd rdncpfhypn aqydydrddv gdrcdnccyn
781 hnpdqadtn ngegdacaad idggilner dncqyvynvd qrtdmdgvg dqcdncpleh
841 npdqldsdsd rigdtcdnnq didedghqnn ldncpyvpna nqadhdkdgk gdacdhhddn
901 dgipddkdnc rlvpnpdqkd sdgdgrgdac kddfdhdsvp diddicpenv disetdfrrf
961 qmipldpkgt sqndpnwwvr hqgkelvqtv ncdpglavgy defnavdfsg tffinterdd
1021 dyagfvfgq sssrfyvvwm kqvtqsywdt nptraqgysg lsvkvvnstt gpgehlrnal
1081 whtgntpgqv rtlwhdprhi gwkdfayrw rlshrpktgf irvvmmyegkk imadsgpiyd
1141 ktyaggrlgl fvfsqemvff sdlkyecrdp

<210> 52
<211> 207
<212> PRT
<213> Homo Sapiens

<400>
1 mapfeplasg il1llwliap sractcvpph pqtafcnsdl virakfvgt evnqttlyqr
61 yeikmtkmyk gfqalgdaad irfytpame svcgyfhrsh nrseefliag klqdgl1hit
121 tcsfvapwns lslaqrqft ktytvgeec tvfpclsipc klqsgthclw tdql1lgsek
181 gfqsrlacl prepglctwq slrsqia

<210> 53
<211> 419
<212> PRT
<213> Homo Sapiens

<400>
1 mhllgffsva csllaallp gpreapaaaa afesgldlsd aepdageata yaskdleeql
61 rsvssvdelm tvlypeywkm ykcqlrkggw qhnreganln srteetikfa aahynteilk
121 sidnewrktq cmprevcidv gkefgvatnt ffkppcvsvy rcggccnseg lqcmntstsy
181 lsktlfeitv plsqqpkpvt isfanhtscr cmskldvyrq vhsiirrslp atlpqcqaan
241 ktcptnymwn nhicrclae dfmfssdagd dstdgfhdc gpnkeldeet cqcvraglr
301 pascgphkel drnscqcvck nkfpsqcg a nrefdentcq cvckrtcprn qplnpgkac
361 ectespqkcl lkgkkfhhqt cscyrrpctn rqkacepgfs yseevcrcvp sywkrpqms

<210> 54
<211> 466
<212> PRT
<213> Homo Sapiens

<400>
1 mstrsvssss yrrmfggpgt asrpssssrsy vttstrtysl gsalrpstsr slyasspggv
61 yatrssavrl rssvpgvrl1 qdsvdfslad aintefkntr tnekvelqel ndrfanyidk
121 vrfileqqnki llaeleqlkg qgksrlgdlly eeemrelrrq vdqltndkar neverdnlae
181 dimrlrekliq eemlqreeae ntlsfrqdv dnaslarldl erkveslqee iaflkkhhee
241 eiqelqaqiq eqhvqidvdv skpdltaaalr dvrqqyesva aknlqeaeew ykskfadlse
301 aanrnndalr qakqesteyr rqvqsltcev dalkgtnesl ergmreemeen faveaanyqd
361 tigrlqdeiq nmkeemarhl reyqdllnvk maldieiaty rkllegeesr islplpnfss
421 lnlretnlds lplvdthskr tfliktvetr dgqvinetsq hhddle