

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 December 2008 (11.12.2008)

PCT

(10) International Publication Number
WO 2008/149088 A2

(51) International Patent Classification:

G01N 33/574 (2006.01) **G06F 19/00** (2006.01)
G01N 33/68 (2006.01)

(21) International Application Number:

PCT/GB2008/001907

(22) International Filing Date: 4 June 2008 (04.06.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

0710636.2 4 June 2007 (04.06.2007) GB
0710768.3 5 June 2007 (05.06.2007) GB

(71) Applicants (for all designated States except US):

THE NOTTINGHAM TRENT UNIVERSITY [GB/GB]; Clifton Lane, Nottingham NG11 8NS (GB).
DEUTSCHES KREBSFORSCHUNGSZENTRUM [DE/DE]; Theodor-Kutzer-Ufer 1, Gebaeude 27, 68167 Mannheim (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **REES, Robert, Charles** [GB/GB]; 48 Longford Road, Bradway, Sheffield S17 4LQ (GB). **BALL, Graham, Roy** [GB/GB]; 15A Churchmoor Lane, Arnold, Nottingham NG5 8HL (GB). **MATHAROO-BALL, Balwir** [GB/GB]; 8 Sanford Brook, Hilton, Derby DE65 5HQ (GB). **CREASER, Colin** [GB/GB]; 8 Mary's Close, Wymeswold, Leicestershire LE12 6TH (GB). **MILES, Amanda** [GB/GB]; 25 Sovereign Way, Heanorn Lane, Derbyshire DE75 7SB (GB). **LANCASHIRE, Lee** [GB/GB]; 20 Saintsbridge Road, Wythenshawe, Manchester M22 1UW (GB). **UGUREL, Selma** [DE/DE]; Klinik fur Dermatologie,

Venerologie und Allergolo, gie Universitätsklinikum Wurzburg, Josef-Scheider-Str. 2., Gebaude D8, 97080 Wurzburg (DE). **SCHADENDORF, Dirk** [DE/DE]; Klinische Kooperationseinheit Dermato-Onklogie des, DKFZ and der Klinik fur Dermatologie, Venerologie, und Allergologie des Universitätsklinikums Mannhei, m Theodor-Kutzer-Ufer 1, Gebaeude 27, 68167 Mannheim (DE). **LEMETRE, Christophe** [FR/FR]; 12, rue des Acacias, F-03000-Avermes (FR).

(74) Agents: **ELSY, David** et al.; Withers & Rogers LLP, Goldings House, 2 Hays Lane, London SE1 2HW (GB).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: MELANOMA ASSAY AND ANTIGENS

(57) Abstract: The invention relates to a mass spectrometry melanoma fingerprint for identifying the presence of a melanoma and/or for identifying the stage of a melanoma. The invention also relates the use of a fingerprint to identify the presence of a melanoma and/or for identifying the stage of a melanoma. Furthermore, the invention relates to a computer comprising the fingerprint.



WO 2008/149088 A2

P715613PCT

MELANOMA ASSAY AND ANTIGENS

5 TECHNICAL FIELD

The invention relates to a mass spectrometry melanoma fingerprint for identifying the presence of a melanoma and/or for identifying the stage of a melanoma. The invention also relates to use of a fingerprint to identify the presence of a melanoma and/or for identifying
10 the stage of a melanoma. Furthermore, the invention relates to a computer comprising the fingerprint.

BACKGROUND TO THE INVENTION

15 The prognosis of advanced metastatic melanoma (American Joint Committee on Cancer (AJCC) stage IV) remains dismal with a 5-year survival-rate of 6-18 %. In the study which led to the present invention, an integrated matrix-assisted laser desorption/ionization (MALDI) mass spectrometric approach combined with artificial neural networks (ANNs) analysis and modeling has been used for the identification of biomarker ions in serum from
20 stage IV melanoma patients allowing discrimination of metastatic disease from healthy status with high specificities of 92 % for protein ions and 100 % for peptide biomarkers, respectively. The inventors' ANNs model also correctly classified 98 % of a blind validation set of AJCC stage I melanoma samples as non-stage IV samples, emphasizing the power of the newly defined biomarkers to identify patients with late stage metastatic melanoma.

25 Sequence analysis identified peptides derived from metastasis-associated proteins; alpha 1-acid glycoprotein precursor-1/2 (AAG-1/2) and complement C3 component precursor-1 (CCCP-1). Furthermore, quantitation of serum AAG – Alpha 1-acid glycoprotein (AAG) by an immunoassay showed a significant ($p < 0.001$) increase in AAG serum concentration in stage IV patients in comparison with healthy volunteers moreover; the quantity of AAG
30 plotted against MALDI-MS peak intensity classified the groups as two distinct clusters. Ongoing studies of other disease stages will provide evidence whether the inventors' strategy is sufficiently robust to give rise to stage-specific protein/peptide signatures in melanoma.

An artificial neural network (ANN), also known as a 'neural network' (NN), is an interconnected group of artificial neurons that uses a mathematical model or computational model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network. ANNs are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. ANNs are well known to the person skilled in the art.

Melanoma is the least common, but most deadly type of skin cancer accounting for 79 % of skin cancer deaths (American Cancer Society, 2002), the incidence of which is increasing worldwide. Melanoma is primarily characterized by its "thickness" or invasion depth, which is used in staging this cancer in conjunction with where it is located, where it has spread and if it is affecting other organs of the body. Melanoma is clinically divided into four stages according the newest AJCC classification [1] with stage I/II containing primary tumours with increasing tumour thickness, stage III being defined as microscopic or macroscopic loco-regional spread of disease and stage IV where the melanoma cells have spread to distant organs including soft tissue, to lymph nodes, or visceral organs. Surgical resection of tumours remains the primary treatment modality for primary melanoma and is usually curative in low-risk patients (AJCC stage I or IIa) [1,2]. However high-risk patients (stage IIb or III) often develop a recurrence which is eventually fatal [1,2], since currently no treatment is available for advanced metastatic melanoma effectively prolonging survival. Serum biomarkers with the ability to differentiate between high- and low-stage melanoma patients are LDH (lactate dehydrogenase), S100B, and MIA (melanoma-inhibitory activity), but these are not sufficiently sensitive to reliably detect patients who have occult metastatic disease [3-5]. These reports support the concept that no single marker will accurately reflect the complex phenotypic changes associated with development of cancer. Therefore there has been an increasing emphasis on the need to determine multiple protein biomarkers for use in the diagnosis of melanoma. Improved biomarkers are required as prognostic indicators for the selection of candidates for adjuvant therapy, and the reliable identification of patients who are at increased risk of disease relapse.

Proteomics is a powerful screening method for protein expression patterns and identifying signatures that are associated with disease stage. Matrix-assisted laser desorption/ionization coupled to time-of-flight mass spectrometry (MALDI-TOF-MS) analysis has yielded novel

and promising data to aid in the detection and identification of diagnostic and prognostic biomarkers of disease from tissue, serum and plasma samples [6-8]. The robustness of MALDI-TOF-MS allows the investigation of complex samples without prior fractionation or chromatographic separation of proteins and peptides. Surface-enhanced laser desorption and ionization (SELDI), introduced by Hutchens and Yip [9], provides a sample platform using solid phase supports coated with chemical or biochemical agents combined with MALDI ionization, which generally results in a low resolution pattern [10] generated by linear TOF-MS. These technologies have been applied successfully for early cancer detection with reported high sensitivity and specificity using a variety of statistical pattern-recognition tools for ovarian [8], colorectal [11], breast [12,13], and prostate cancer [8,14,15], astrocytoma/glioblastoma [19] and melanoma [9]. A limitation of MALDI or SELDI analysis is that protein identification cannot be obtained by protein molecular weight (m/z) measurement alone, but requires enzymatic digestion of serum or plasma proteins, usually carried out separately following 2-dimensional SDS-PAGE analysis of proteins [17-19].

Some concerns have also been voiced regarding biological, technological, and data mining artifacts that may introduce bias [20]. The most common sources of bias in serum proteomics-based biomarker discovery are related to sample collection, processing and storage, instrumentation, and data analytical methods.

To achieve high predictive power, biomarkers need to be multiparametric (measure of multiple proteins) and mathematically modeled to optimize the composition of biomarker panels. This requires the use of highly sophisticated statistical methods with high computational power such as Bayesian analyses, fuzzy logic, and artificial neural networks (ANNs) [21-24]. ANNs, as used in the present study, have the ability to "learn" predictive patterns contained within complex datasets, by iteratively adjusting and updating the interconnecting weights between the layers of the network using the back propagation (BP) algorithm applied to the multi-layer perceptron (MLP) [25]. Once trained, ANN models have been shown to be able to predict the clinical outcome for blinded cases.

In the inventors' efforts to report thoroughly and transparently, and to minimize criticisms leveled recently at clinical proteomic studies [26-28], the inventors carried out rigorous studies prior to biomarker identification to address the issue of chance and bias at every step – in design, conduct, interpretation and validation. In order to find new serum markers for

stage IV metastatic melanoma, which may later on be validated and used in earlier disease stages, the inventors sought to develop a validated, robust and reproducible methodology based on MALDI-TOF-MS combined with bioinformatics to profile serum protein and tryptic peptides. Additionally, the inventors have used a combination of mass spectrometric techniques for the ANNs-directed sequencing of biomarker tryptic peptides, followed by database searches to identify the predominant, disease-associated peptides and proteins. The integrated proteomic approach described here has generic application to the identification of protein and peptide biomarker ions and sequences associated with other disease states.

10 SUMMARY OF THE INVENTION

Therefore, a first aspect of the invention provides a mass spectrometry melanoma fingerprint for identifying the presence of a melanoma and/or for identifying the stage of a melanoma, the fingerprint comprising:

(a) peaks derived from the following ions:

(i) ions derived from a protein, the ions having (m/z) values of: 12000, 14847, 1649, 15477, 13255, 3031 and 4791 ($\pm 0.4\%$) and optionally one or more the following additional ions, the additional ions having (m/z) values of: 9913, 4835, 4155, 4565 and/or 4673 ($\pm 0.4\%$);

(ii) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 1753, 1161, 1505 and 854 ($\pm 0.02\%$) and optionally one or more of the following additional ions, the additional ions having (m/z) values of: 1444 and/or 1093 ($\pm 0.02\%$);

and/or

(b) peaks derived from the following ions:

(i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 2624, 2715, 1753, 1251, 1285, 2999, 3161, 1312, 3326 and 1371 ($\pm 0.02\%$);

and/or

(c) peaks derived from the following ions:

(i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 1251, 1283, 3443, 3432, 1968, 1299, 2244, 2411, 2468, and 2492 ($\pm 0.02\%$);

and/or

(d) peaks derived from the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 980, 3220, 864, 2966, 2886, 1299, 2309, 3489, 3430 and 933 ($\pm 0.02\%$);

and/or

5 (e) peaks derived from the following ions:

- (i) ions derived from a tryptically digested peptide, the ions having (m/z) values selected from: 1978 and 1825 ($\pm 0.02\%$) and optionally one or more the following additional ions, the additional ions having (m/z) values of: 1731, 1251 and/or 2053 ($\pm 0.02\%$);

10 and/or

(f) peaks derived from the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 861 and 903 ($\pm 0.02\%$);

and/or

15 (g) peaks derived from one or more of the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 877, 903 and 1625 ($\pm 0.02\%$) and optionally one or more the following additional ions, the additional ions having (m/z) values of: 2754, and/or 2064 ($\pm 0.02\%$);

20 and/or

(h) peaks derived from one or more of the following ions:

- (i) ions derived from a protein, the ions having (m/z) values of 1161 and/or 1753 ($\pm 0.02\%$).

25 Preferably the fingerprint comprises two or more, more preferably three, four, five, six, seven, eight, nine, ten, eleven or more of the peaks listed for one of (a)(i), (a)(ii), (b), (c), (d), (e), (f), (g) or (h).

For the fingerprint of (a)(i), preferably the additional ions comprise:

30

- (1) an ion having an (m/z) value of: 9913 ($\pm 0.4\%$); or
- (2) ions having (m/z) values of: 9913 and 4835 ($\pm 0.4\%$); or
- (3) ions having (m/z) values of: 9913, 4835 and 4155 ($\pm 0.4\%$); or
- (4) ions having (m/z) values of: 9913, 4835, 4155 and 4565 ($\pm 0.4\%$); or

(5) ions having (m/z) values of: 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$).

For the fingerprint of (a)(ii), preferably the additional ions comprise:

- 5 (1) an ion having an (m/z) value of: 1444 ($\pm 0.02\%$); or
(2) ions having (m/z) values of: 1444 and 1093 ($\pm 0.02\%$).

For the fingerprint of (e), preferably the additional ions comprise:

- 10 (1) an ion having an (m/z) value of: 1731 ($\pm 0.02\%$); or
(2) ions having (m/z) values of: 1731 and 1251 ($\pm 0.02\%$); or
(3) ions having (m/z) values of: 1731, 1251 and 2053 ($\pm 0.02\%$);

For the fingerprint of (g), preferably the additional ions comprise:

- 15 (1) an ion having an (m/z) value of: 2754 ($\pm 0.02\%$); or
(2) ions having (m/z) values of: 2754 and 2064 ($\pm 0.02\%$);

Most preferably, the mass spectrometry melanoma fingerprint comprises all of the peaks listed in one of (a), (b), (c), (d), (e), (f), (g), or (h).

20 Preferably peaks comprise one or more peaks from (a)(i) and/or (a)(ii). More preferably, the fingerprint comprises all of the peaks of (a)(i) and (a)(ii).

Two or more of the fingerprints may be used in combination.

25 The fingerprint allows determination of whether a subject, from which a test sample has been taken, has melanoma and/or a defined stage of melanoma. For example:

- (a) one or more ions/peaks may be present in a control sample but absent in the test sample;
(b) one or more ions/peaks may be present in a control sample and present in the test sample;
30 (c) one or more ions/peaks may be absent in a control sample but present in the test sample;
(d) one or more ions/peaks may be absent in a control sample and absent in the test sample.

For example see Tables 3 and 4.

‘Absent’ and ‘present’ may be absolute. Furthermore, these terms also refer to an increase or decrease in intensity of one or more peaks/ions necessary for the presence/absence of melanoma and/or melanoma stage to be determined.

Preferably, proteolytic digestion is carried out using trypsin. Other proteases known to the skilled person would also be suitable. Furthermore, a chemical digestion procedure may be used. For example, the chemical digestion may result in the same fragments as a tryptic digestion.

Preferably a tryptic digestion is carried out under the following conditions:

An aliquot (25 μ l) of diluted serum (*e.g.* diluted 1 in 20 with 0.1 % TFA), or of AAG solution, may be initially fractionated using a ZipTip C₁₈ (Milipore, Watford, UK) with 25 cycles of binding. This may be followed by two washes in 0.1 % TFA and elution in 4 μ l of 80 % acetonitrile/0.1 % TFA. This volume may be combined with ammonium bicarbonate (16.6 μ l, 100 mM), water (7.6 μ l), and trypsin Gold Mass Spectrometry Grade (1.3 μ l, 0.5 μ g/ μ l, Promega, Southampton, UK) and incubated at 37 °C overnight. The reaction may be quenched with 1 % TFA (1 μ l) and the sample cleaned-up using a C₁₈ ZipTip following the procedure described above. An aliquot of the eluate (1 μ l) may be spotted onto a MALDI target using the dried droplet method with matrix, α -cyano-4-hydroxycinnamic acid (LaserBio Labs, Cedex, France), prepared as a 10 mg/ml solution in 50 % acetonitrile + 0.1 % TFA. Duplicate samples may be applied to the target plate in a randomized order used for sample preparation and analysed by MALDI-TOF-MS. A range of appropriate blank and control samples are preferably prepared alongside the serum digests. For every 50 patient samples processed, 10 aliquots of 0.1 % TFA (25 μ l) may be taken, as reagent blanks, through the entire ZipTip clean-up prior to digestion. Serum blanks (4 per 100 serum samples processed) may also be diluted 1/10 with 0.1 % TFA (25 μ l), combined with the digestion buffer without the incorporation of trypsin before, and after, ZipTip clean-up and incubated overnight at 37 °C.

Preferably the fingerprint comprises peaks showing an (m/z) of 1161 ($\pm 0.02\%$), 1753 ($\pm 0.02\%$), 4565 ($\pm 0.4\%$), 4673 ($\pm 0.4\%$), and/or 4786 ($\pm 0.4\%$). Preferably the fingerprint does not comprise a peak showing an (m/z) of 4155 ($\pm 0.4\%$).

- 5 A second aspect of the invention relates to a method of identifying a subject having a melanoma, the method comprising identifying whether a test sample obtained from a test subject suspected of having melanoma, when analysed by mass spectrometry, shows a fingerprint as described for the first aspect of the invention.
- 10 Preferably the method further comprises obtaining a fingerprint from a control sample, and comparing the fingerprint from the control sample with the fingerprint obtained from the test sample. Preferably, the comparison comprises identification of an elevation or depression in one or more of the ions of the fingerprint to determine whether the test subject has melanoma. Preferably one or more of the peaks is elevated or depressed as shown in Table
- 15 3. The comparison may comprise identification of the presence/absence of one more of the ions of the fingerprint to determine whether the test subject has melanoma. More preferably the comparison takes into account, *i.e.* comprises identification of an elevation or depression in one or more peaks and/or comprises identification of the presence/absence of one or more peaks. More preferably, the comparison takes into account two, three, four, five, six, seven,
- 20 eight, nine, ten, eleven or twelve peaks. Most preferably, the comparison takes into account all of the peaks of the fingerprint concerned.

Preferably the subject(s) is a human or a fragment or cell obtained therefrom. Preferably the sample(s) comprises one or more of proteins, peptides or fragments thereof and/or a nucleic

25 acid molecule.

In order to determine whether a sample, *e.g.* a test sample or a control sample, shows a fingerprint as described for the first aspect of the invention, the fingerprint of the sample may be compared with a fingerprint of the first aspect of the invention either manually (*e.g.*

30 visually) and/or by using a computer on which data corresponding to one or more of the fingerprints of the first aspect of the invention are stored. The data may be accessed, for example, *via* a removable disk or *via* the internet.

A fingerprint according to the first aspect of the invention may be used for one of the following purposes:

- 5 (a) the fingerprint of (a)(i) and/or (a)(ii) may be used to distinguish between Stage IV melanoma and non-melanoma;
- (b) the fingerprint of (b) may be used to distinguish between Stage III melanoma and non-melanoma;
- (c) the fingerprint of (c) may be used to distinguish between Stage II melanoma and non-melanoma;
- 10 (d) the fingerprint of (d) may be used to distinguish between Stage I melanoma and non-melanoma;
- (e) the fingerprint of (e) may be used to distinguish between Stage I melanoma and Stage II melanoma;
- (f) the fingerprint of (f) may be used to distinguish between Stage II melanoma and
15 Stage III melanoma;
- (g) the fingerprint of (g) may be used to distinguish between Stage II melanoma and Stage III melanoma;
- (h) the fingerprint of (h) may be used to distinguish between Stage IV melanoma and control.

20

Preferably the control sample is obtained from a subject not having melanoma. Alternatively, or in addition, a control sample may be obtained from a subject having melanoma, for example a subject having one of the following stages of melanoma: Stage 0, Stage IA, Stage IB, Stage IIA, Stage IIB, Stage IIC, Stage III and Stage IV. Stage IV melanoma may consist
25 of Stage IV M1a, Stage IV M1b or Stage IV M1c.

A third aspect of the invention relates to a method of identifying a subject having a melanoma comprising:

- 30 (i) determining the level of Alpha 1-Acid Glycoprotein or a fragment thereof in a sample obtained from a test subject suspected of having a melanoma; and
- (j) comparing the level of Alpha 1-Acid Glycoprotein or fragment thereof in the test sample with the level of Alpha 1-Acid Glycoprotein or fragment thereof in a sample from a control subject.

Preferably the fragment of alpha 1-acid glycoprotein has an ion having an m/z of 1161, 1093, 1753 and/or 1444 ($\pm 0.02\%$). Preferably the fragment has the sequence selected from:

- 5 (i) A polypeptide consisting of a sequence selected from:
- (a) WFYIASAFR [SEQ ID No. 1],
 - (b) a derivative sequence of (a) wherein the derivative sequence has an (m/z) of 1161 ($\pm 0.02\%$).
- 10 (ii) A polypeptide comprising a sequence selected from:
- (a) NTLIIYLDK [SEQ ID No. 2],
 - (b) a derivative sequence of (a) wherein the derivative sequence has an (m/z) of 1093 ($\pm 0.02\%$).
- 15 (iii) A polypeptide consisting of a sequence selected from:
- (a) YVGGQEHFAHLLILR [SEQ ID No. 3],
 - (b) a derivative sequence of (a) wherein the derivative sequence has an (m/z) of 1753 ($\pm 0.02\%$).
- 20 (iv) A polypeptide consisting of a sequence selected from:
- (a) TYMLAFDVNDEK [SEQ ID No. 4],
 - (b) a derivative sequence of (a) wherein the derivative sequence has an (m/z) of 1444 ($\pm 0.02\%$).

- 25 Suitable databases include NCBI, when using the MASCOT search engine and UNIPROT, when using the SEQUEST search engine.

Preferably the alpha 1-acid glycoprotein ion is derived from AAG-1/2 or a fragment thereof, *e.g.* i, ii, and iii (above).

30

The sequences of AAG-1 and AAG-2 are publicly available:

AAG-1 (SEQ ID No. 5): NCBI Accession No. P02763 (last modified 1 January 1998), Version P02763.1 GI:112877. Alternative Accession No. Q8TC16.

AAG-2 (SEQ ID No. 6): NCBI Accession No. P19652 (last modified 1 April 1993),
Version P19652.2 GI:231458. Alternative Accession No. Q16571.

- 5 Preferably the melanoma being detected is a Stage IV melanoma.

The control sample may be obtained from a subject not having melanoma. Alternatively, or in addition, a control sample may be obtained from a subject having Stage IV melanoma. One or more control samples may be used. That is, one or more negative and/or one or more
10 positive control samples may be used.

Preferably the fragment of Alpha 1-Acid Glycoprotein, when analysed by mass spectrometry, generates an ion having an (m/z) of 1161 and/or 1753 ($\pm 0.02\%$).

- 15 The method of the third aspect of the invention may further comprise comparing the relative level of Lactate Dehydrogenase (LDH) in the test sample compared to the control sample to identify whether the melanoma of the test sample is a Stage I, Stage II, Stage III, Stage IV, Stage IV M1a, Stage IV M1b and Stage IV M1c.

- 20 The method for assaying LDH may be based on an ELISA colorimetric test on serum/plasma samples. LD reagent (using lactate as substrate) utilizes an enzymatic rate method to measure LD activity in biological fluids. In the reaction, the LD catalyzes the reversible oxidation of L-Lactate to Pyruvate with the concurrent reduction of β -nicotinamide adenine dinucleotide (NAD) to β -nicotinamide adenine dinucleotide hydroxylase (reduced form;
25 NADH). The system monitors the rate of change in absorbance at 340 nm over a fixed time interval. The rate of change in absorbance is directly proportional to the activity of LD in the sample.

- Lactate dehydrogenase measurements are used in the diagnosis and treatment of liver
30 diseases, cardiac diseases, and tumours of the lung or kidney.

Alternatively, and/or in addition, one or more of nephelometry, ELISA and Western Blotting may be used.

Throughout this specification, the terms 'elevation' and 'depression' are defined based on a consistent difference occurring between the ion in the types of samples being compared.

- 5 It is preferred that the test subject shows an elevation of one or both of the peptides having an ion with an (m/z) of 1161 and 1753 ($\pm 0.02\%$).

Preferably the peptide ion having an m/z of 1161 ($\pm 0.02\%$) has the sequence WFYIASAFR [SEQ ID No. 1]. Preferably the peptide ion having an m/z of 1753 ($\pm 0.02\%$) has the
10 sequence YVGGQEHFAHLLILR [SEQ ID No. 3]. Preferably the peptide ion having an m/z of 1093 ($\pm 0.02\%$) has the sequence NTLIYLDK [SEQ ID No. 2].

Preferably the test subject shows an elevated level of one or more of the proteins having an m/z of 4565, 4673 and 4786 ($\pm 0.4\%$). Preferably test subject shows an elevated level of two
15 or more of the proteins having an m/z of 4565, 4673 and 4786 ($\pm 0.4\%$).

Preferably the test subject shows an elevated level of a protein ion having an m/z of 4565 ($\pm 0.4\%$), a protein ion having an m/z of 4673 ($\pm 0.4\%$) and a protein ion having an m/z of 4786 ($\pm 0.4\%$).
20

Preferably the test subject shows a depression of a protein ion having an m/z of 4155 ($\pm 0.4\%$).

Preferably the test subject shows:

- 25 (a) an elevation of the peptide having an ion with an (m/z) of 1160 ($\pm 0.02\%$),
(b) an elevation of the peptide having an ion with an (m/z) of 1753 ($\pm 0.02\%$),
(c) an elevated level of the protein having an m/z of 4565 ($\pm 0.4\%$),
(d) an elevated level of the protein having an m/z of 4673 ($\pm 0.4\%$),
(e) an elevated level of the protein having an m/z of 4786 ($\pm 0.4\%$), and
30 (f) a depression of a protein having an ion having an m/z of 4155 ($\pm 0.4\%$).

That is, preferably a positive diagnosis of a melanoma comprises the identification of the elevations and depressions of the peptides and proteins as listed above.

Preferably the proteolytically digested peptide is a tryptic peptide, although other proteolytically digested peptides may also be used. As an alternative, or in addition, chemically digested peptides may be used.

5

Preferably the melanoma is a metastatic melanoma, more preferably a Stage IV metastatic melanoma.

10

Preferably the method comprises detecting and/or quantifying one or more of the disclosed peptides and/or proteins. Such detection and/or quantification may be carried out, for example, using mass spectrometry and/or an immunological assay. Examples of types of suitable mass spectrometry apparatus and methods are provided above. Immunological assays are known to the skilled person and include, for example, an enzyme linked immunosorbent assay (ELISA).

15

Preferably the immunological assay comprises use of a labelled antibody. Suitable labels are well known to the skilled person.

The methods described herein may optionally involve one or more of the following steps:

20

- Isolation of serum from a blood sample by allowing the blood sample to clot at room temperature for at least 30 minutes but for no longer than 60 minutes. Preferably the clotted sample is centrifuged, preferably at 2500 g for 10 minutes. The serum is then harvested and subsequently frozen, preferably at -20°C or -80°C . Preferably the sample is initially frozen at -20°C and subsequently thawed once and aliquoted to a smaller volume and the smaller aliquot frozen at -80°C . Preferably the sample is not re-thawed until immediately prior to analysis.
- Subjecting the sample to three or fewer, preferably two or fewer, freeze thaw cycles.
- Following receipt of a sample from a subject, e.g. from a clinical centre, storing the sample at -80°C .
- Following storing of the sample at -80°C , thawing, storing on ice and processing within 2 hours of thawing.

30

- Exclusion of spectra having a signal to noise ratio of $<5:1$ for the peak at (m/z) 6632 (± 13) prior to analysis of the mass spectra.
- Carrying out mass spectrometry using an instrument in raster mode.
- Carrying out mass spectrometry using an instrument in auto-quality mode.

5

A further aspect of the invention provides the following polypeptides and mixture of 2 or 3 of these polypeptides:

WFYIASAFR [SEQ ID No. 1], and a derivative sequence of WFYIASAFR wherein the derivative sequence has an (m/z) of 1161 ($\pm 0.02\%$).

10

NTLIYLDK [SEQ ID No. 2], and a derivative sequence of (NTLIYLDK) wherein the derivative sequence has an (m/z) of 1093 ($\pm 0.02\%$).

YVGGQEHAHLLLR [SEQ ID No. 3], and a derivative sequence of
YVGGQEHAHLLLR wherein the derivative sequence has an (m/z) of 1753 ($\pm 0.02\%$).

15

The mixture may further comprise other fragments listed herein, such as TYMLAFDVNDEK [SEQ ID No. 4], and a derivative sequence of TYMLAFDVNDEK wherein the derivative sequence has an (m/z) of 1444 ($\pm 0.02\%$).

20

A fourth aspect of the invention provides a kit for use in a method of detecting melanoma, the kit comprising positive controls for mass spectrometry analysis in which the kit comprises:

(a) a control sample representative of a subject not having melanoma;
and/or

(b) a control sample representative of a subject having melanoma.

Preferably the kit comprises a control sample representative of a subject not having melanoma comprises proteins and/or proteolytically digested peptides which when analysed by mass spectrometry generate two or more ions as described in (a)(i), (a)(ii), (b), (c), (d), (e), (f), (g) or (h) of the first aspect of the invention.

30

More preferably, the control sample generates three, four, five, six, seven, eight, nine, ten, eleven or more of the peaks listed for one of (a)(i), (a)(ii), (b), (c), (d), (e), (f), (g) or (h) of the first aspect of the invention.

- 5 It is further preferred that the control sample generates one or more of the additional ions, or combinations of additional ions, as described for the first aspect of the invention.

It is further preferred that, in the kit:

- 10 (a) one or more of the ions having an (m/z) of 2225, 1753, 1754, 1161 and 1444 is derived from Alpha-Acid Glycoprotein 1 Precursor;
(b) the ion having an (m/z) of 1251 is derived from Kininogen-1 Precursor;
(c) the ion having an (m/z) of 1283 is derived from Apolipoprotein A-I Precursor;
(d) the ion having an (m/z) of 1093 is derived from Complement C3 precursor.

15

The kit may comprise positive controls for mass spectrometry analysis, the positive controls being two or more species selected from:

- (a) a protein having an (m/z) selected from 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$);
20 (b) a proteolytically digested peptide having an (m/z) selected from 1753, 1161, 1505, 854, 1444, 1093 ($\pm 0.02\%$);

Preferably the kit comprises positive controls which comprise a peptide having an ion with an m/z of 1093 ($\pm 0.02\%$) and a peptide having an ion with an m/z of 1161 ($\pm 0.02\%$).

25

More preferably the positive controls comprise:

- (a) a peptide having an ion with an (m/z) of 1161 ($\pm 0.02\%$),
(b) a peptide having an ion with an (m/z) of 1753 ($\pm 0.02\%$),
(c) a protein having an ion with an m/z of 4565 ($\pm 0.4\%$),
30 (d) a protein having an ion with an m/z of 4673 ($\pm 0.4\%$),
(e) a protein having an ion with an m/z of 4786 ($\pm 0.4\%$), and
(f) a protein having an ion having an m/z of 4155 ($\pm 0.4\%$).

Other preferred controls include comprise:

- (a) proteins having an ion with an (m/z) of 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913 and 4835 ($\pm 0.4\%$) and/or
- (b) peptides having an ion with an (m/z) of 1753, 1161, 1505, 854, 1444 and 1093 ($\pm 0.4\%$).

A particularly preferred control comprises a protein having an ion having an m/z of 1444, a protein having an ion having an m/z of 1753 and a protein having an ion having an m/z of 1161.

Increasing the number of positive controls and/or increasing the number of ions represented in each positive control increases the accuracy of the method.

A fifth aspect of the invention relates to a method of identifying a drug for treating melanoma comprising:

- (a) exposing a test sample from a subject having melanoma to a candidate compound;
- (b) subjecting the exposed test sample, from (a), to mass spectrometry analysis to generate a fingerprint comprising one or more peaks;
- (c) comparing the fingerprint obtained from the exposed test sample with a fingerprint obtained from a control sample in which the control sample:
 - (i) is obtained from a subject having melanoma; and
 - (ii) has not been exposed to the candidate compound
- (d) determining whether the candidate compound decreases or elevates the level of one or more of the peaks of the fingerprint of the test sample relative to the fingerprint of the control sample.

A compound which is suitable for treating melanoma would cause the peaks of the fingerprint to match more closely to a fingerprint from a control representing a non-melanoma sample than to a fingerprint from a control representing a melanoma sample.

Preferably, the method further comprises comparing the fingerprint of the test sample with a fingerprint obtained from one or more further control samples in which the one or more further control samples is obtained from:

- (a) a subject not having melanoma; and/or
- (b) a subject having Stage IV melanoma.

Preferably the fingerprint is as described for the first aspect of the invention.

5

It is further preferred that the fingerprint comprises one or more peaks derived from an ion derived from a protein, the ion having an (m/z) of 1161 and/or 1753 (± 0.02).

Other methods may be used in place of mass spectrometry, *e.g.* chromatography.

10

The peaks may be selected from one or more of:

- (a) a protein having an (m/z) selected from 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$);
- (b) a proteolytically digested peptide having an (m/z) selected from 1753, 1161, 1505, 854, 1444, 1093 ($\pm 0.02\%$).

15

The peaks are listed in ranked order. Therefore the peaks are particularly useful if two or more of the peaks are combined in the order listed in a cumulative manner. The cumulative effect of the ions is detailed in Tables 3 and 4.

20

A sixth aspect of the invention relates to a computer system having a processor, a memory and an input for receiving data from a mass spectrometer wherein the memory comprises data corresponding to a fingerprint according to the first aspect of the invention and the computer configured to compare data from the input against the fingerprint data and to identify whether the inputted data is indicative of a melanoma based on the results of the comparison.

25

The invention also provides a disk, for example CD or diskette, on which one or more of the fingerprints of the first aspect of the invention is stored. Furthermore, the invention provides one or more fingerprints accessible *via* the internet.

30

Preferably, the fingerprint is for Stage IV melanoma.

A seventh aspect of the invention relates to a melanoma identification apparatus comprising the computer system of the sixth aspect of the invention and a mass spectrometer connected to the input.

5

Further aspects of the invention are detailed below. These aspects may be combined with one or more of the first, second, third, fourth, fifth, sixth and seventh aspects of the invention.

10 (1) A method of identifying one or more biomarkers of a specified biological phenotype the method comprising:

(a) obtaining:

(i) a sample obtained from each of a plurality of subjects having the specified biological phenotype,

15 (ii) a sample obtained from each of a plurality of control subject not having the specified biological phenotype of (i),

to generate intact samples,

(b) proteolytically digesting a portion of each of the samples from (a) to generate digested samples,

20 (c) analyzing an aliquot of each of the intact and digested samples by a method which separates components of the mixture relative to their molecular weight to generate a profile of peaks representing the relative abundance of the components in the mixture,

25 (d) analyzing the profile obtained from (c) using an artificial neural network to identify peaks of the profile which distinguish between a subject having the specified biological phenotype and a subject not having the specified biological phenotype where ions are added to a model based on their predictive performance for a data set excluded from the modelling process (blind to the model).

30 An 'intact' sample is one which has not been subjected to a digestion step. The analysis step (d) may be replaced by an alternative analysis step to identify peaks of the profile which distinguish between a subject having the specified biological phenotype and a subject not having the specified biological phenotype.

Throughout this specification, the term 'one or more' includes preferably 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 or more. Likewise, the term 'two or more' includes preferably 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 or more.

5

The preferred ANN method of the present invention uses a stepwise approach to determine of a panel of predictive ions coupled with random sample cross validation. This has particular use for the modelling and analysis of highly dimensional complex data *e.g.* mass spectrometry data.

10

Preferably, a set of models using different randomly extracted data set splits for training, stopping training, and testing (normally 50 different sets) are trained for each ion in the data set – this is known as random sample cross validation. Thus it is possible to able to determine a mean performance across 50 models for each ion in the mass spectrometry profile for predicting the blind test data sets. The mean performance may be then ranked for the ion set and the best ion determined. The relative performance of the single ions may be determined at this stage and their distribution can be determined. The ion may then be selected for a second round of training with random sample cross validation using the best ion combined with all the ions in the profile as a pair. Thus it is possible to determine the best pair of ions. This process is repeated no significant improvement is observed by the addition of further ions.

15

20

(2) A method according to statement (1) wherein step (c) comprises mass spectrometry and/or 2-dimensional gel electrophoresis.

25

Both mass spectrometry and 2D GE separate components of a mixture relative to their molecular weight.

(3) A method according to statement (1) or (2) wherein step (a) further comprises

30

(iii) obtaining experimental control samples.

(4) A method according to statement (3) wherein the experimental control samples consist of one or more of: trifluoroacetic acid, and bovine serum albumin.

An experimental control may also include using one or more of the samples from a test subject and/or a control subject, not adding a protease and then subjecting the sample(s) to conditions which are the same as those used for digesting the samples from the test and/or control subject(s). This allows the activity of the protease to be assessed.

- (5) A method according to any preceding statement wherein digestion is carried out using trypsin.
- 10 Other proteases known to the skilled person would also be suitable.

- (6) A method according to any preceding statement further comprising determining the amino acid sequence of one or more of the peaks identified.
- 15 (7) A method according to statement (6) wherein determination of amino acid sequence comprises mass spectrometry and database mining.

Sequencing may be carried out by any suitable method. Sequencing methods are known to the skilled person, *e.g.* a tandem mass spectrometry method or Edman degradation method.

20 Sequencing allows identification of the parent protein that belongs to the ion having the predicted m/z value by ANNs. It also allows verification of the results using another method *e.g.* an immunoassay. Furthermore, it provides the identity of the protein to which antibodies could be developed. Additionally, identification of the sequence allows the DNA and RNA sequence encoding the protein and/or peptide to be determined thus enabling molecular

25 biology analyses to be carried out.

Examples of databases include the NCBI database (available at *e.g.* <http://www.ncbi.nlm.nih.gov>) and the Mascot database (available at *e.g.* <http://www.matrixscience.com>).

30

- (8) A method according to statement (7) wherein the mass spectrometry is selected from one or more of:
- (a) tandem mass spectrometry

- (b) matrix assisted laser/desorption ionization time of flight mass spectrometry with post source decay (MALDI-TOF-PSD)
- (c) atmospheric pressure MALDI/ion trap mass spectrometry (AP-MALDI-QIT)
- (d) liquid chromatography combined with electrospray ion trap tandem mass spectrometry (LC-ESI-QIT MS/MS).

(9) A method according to any preceding statement further comprising randomizing the samples prior to analysis.

An advantage of this randomizing step is that it reduces, and preferably eliminates, bias which may be otherwise generated.

(10) A method according to any preceding statement wherein mass spectrometry is carried out by one or more of:

- (a) matrix assisted laser/desorption ionization time of flight mass spectrometry with post source decay (MALDI-TOF-PSD)
- (b) atmospheric pressure MALDI/ion trap mass spectrometry (AP-MALDI-QIT)
- (c) liquid chromatography combined with electrospray ion trap tandem mass spectrometry (LC-ESI-QIT MS/MS).

(11) A method according to any preceding statement wherein the sample in one or both of steps (a)(i) and (a)(ii) is from a human or animal.

(12) A method according to any preceding statement wherein the sample in one or both steps (a)(i) and (a)(ii) is a serum sample.

(13) A method according to any preceding wherein the samples are diluted in trifluoroacetic acid prior to analysis by mass spectrometry.

(14) A method according to statement (13) wherein the samples are diluted 1 in 10 in 0.1 % trifluoroacetic acid.

(15) A method according to any preceding statement wherein the samples are fractionated prior to digestion.

(16) A method according to any preceding statement wherein the samples are fractionated prior to analysis by mass spectrometry.

5 (17) A method according to any of statements (15) to (16) wherein the fractionation comprises chromatography.

(18) A method according to statement (17) wherein the chromatography comprises reverse phase liquid chromatography.

10

(19) A method according to statement (18) wherein the fractionation comprises use of a ZipTip C₄ or a ZipTip C₁₈.

15

Alternatively or in addition, the fractionation may comprise use of SPE and/or a HPLC column.

20

(20) A method according to any preceding statement wherein prior to MALDI-MS, the samples are subjected to three or fewer freeze-thaw cycles and the thawed samples are stored on ice for a period of time of two or less hours between thawing and mass spectrometry analysis.

This reduces, and preferably eliminates, variation which can result from multiple freeze-thaw cycles and/or a time delay prior to storage.

25

(21) A method according to any preceding statement wherein the artificial neural network sorts only peaks having a predetermined range of masses.

30

(22) A method according to statement (21) wherein the predetermined range of masses for proteins is from 1 to 30 kDa and/or the predetermined range of masses for peptides is from 800 to 3500 Da.

It is preferred that the artificial neural network sorts only peaks having a predetermined range of masses. It is particularly preferred that the predetermined range of masses for

proteins is from 1 to 30 kDa and/or the predetermined range of masses for peptides is from 800 to 3500 Da. These ranges apply particularly to samples from a subject having a melanoma.

- 5 (23) A method according to any preceding statement wherein the biological phenotype is a disease state.

(24) A method according to statement (23) wherein the disease state is a neoplasia.

- 10 (25) A method according to statement (24) wherein the neoplasia is selected from a melanoma, a breast cancer and a prostate cancer.

Most preferably the cancer is a melanoma.

- 15 (26) A method of identifying a subject having a predisposition to develop melanoma or having a melanoma, the method comprising comparing:

(a) the level of one or more biomarkers in a sample obtained from a test subject suspected of having a predisposition to develop melanoma or suspected of having a melanoma,

20 with

(b) the level of the one or more biomarkers in a control or a sample obtained from a control subject,

(c) wherein the biomarker is selected from the following peptides and/or proteins:

25 (i) a protein having an (m/z) selected from 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$);

(ii) a proteolytically digested peptide having an (m/z) selected from 1753, 1161, 1505, 854, 1444, 1093 ($\pm 0.02\%$);

wherein identification of an elevation or depression in the level of one or more of the biomarkers in the test subject compared with the control indicates the presence or
30 absence of a predisposition to developing melanoma or to the presence or absence of melanoma.

(27) A method of identifying a subject susceptible to melanoma or having a melanoma, the method comprising comparing:

(a) the level of one or more biomarkers in a sample obtained from a test subject suspected of being predisposed to developing melanoma or suspected of having a melanoma,

with

(b) the level of the one or more biomarkers in a control or a sample obtained from a control subject,

wherein the biomarker is selected from the following peptides and/or proteins:

(i) a protein having an (m/z) selected from 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$);

(ii) a proteolytically digested peptide having an (m/z) which is elevated or depressed in a proteolytically digested sample from a test subject compared with a proteolytically digested control sample wherein the samples from the test subject and the control are digested with the same protease;

wherein identification of an elevation or depression in the level of one or more of the biomarkers in the test subject compared with the control indicates the presence or absence of susceptibility to melanoma or to the presence or absence of melanoma.

(28) A method according to statement (25) or (26) comprising a control sample is obtained from a subject not having melanoma.

(29) A method according to any of statements (26) to (28) comprising a control sample obtained from a subject having melanoma.

One or more control samples may be used. That is a negative and/or a positive control sample may be used.

(30) A method according to any of statements (26) to (29) wherein the control sample is obtained from a subject having one of the following stages of melanoma: Stage 0, Stage IA, Stage IB, Stage IIA, Stage IIB, Stage IIC, Stage III and Stage IV.

(31) A method according to any of statements (26) to (30) wherein the method comprises identification of the elevation and/or depression of two or more of the biomarkers.

5 Preferably the method comprises identification of the elevation or depression and/or presence or absence of two or more of the biomarkers listed herein.

10 Preferably the method comprises identification of the elevation or depression and/or presence or absence of 2, 3, 4, 5 or 6 biomarkers. Preferably the method comprises the identification of an elevation or depression of 1, 2, 3, 4, 5 or 6 of the biomarkers.

Elevation and depression are defined based on a consistent difference occurring between the ion in the types of samples being compared.

15 (32) A method according to any of statements (26) to (31) wherein the subject shows an elevation of one or both of the peptides having an ion with an (m/z) of 1161 and 1753 (± 0.02 %).

(33) A method according to any of statements (26) to (32) wherein the peptide ion having an m/z of 1161 (± 0.02 %) has the sequence WFYIASAFR [SEQ ID No. 1].

20 (34) A method according to any of statements (26) to (33) wherein the peptide ion having an m/z of 1753 (± 0.02 %) has the sequence YVGGQEHFAHLLILR [SEQ ID No. 3].

25 (35) A method according to any of statements (26) to (34) wherein the peptide ion having an m/z of 1093 (± 0.02 %) has the sequence NTLIIYLDK [SEQ ID No. 2].

(36) A method according to any of statements (26) to (35) wherein the test subject shows an elevated level of one or more of the proteins having an m/z of 4565, 4673 and 4786 (± 0.4 %).

30 (37) A method according to any of statements (26) to (36) wherein the test subject shows an elevated level of two or more of the proteins having an m/z of 4565, 4673 and 4786 (± 0.4 %).

(38) A method according to any of statements (26) to (37) wherein the test subject shows an elevated level of a protein ion having an m/z of 4565 ($\pm 0.4\%$), a protein ion having an m/z of 4673 ($\pm 0.4\%$) and a protein ion having an m/z of 4786 ($\pm 0.4\%$).

5

(39) A method according to any of statements (26) to (38) wherein the test subject shows a depression of a protein ion having an m/z of 4155 ($\pm 0.4\%$).

(40) A method according to any of statements (26) to (39) wherein the test subject shows:

- 10 (a) an elevation of the peptide having an ion with an (m/z) of 1161 ($\pm 0.02\%$),
(b) an elevation of the peptide having an ion with an (m/z) of 1753 ($\pm 0.02\%$),
(c) an elevated level of the protein having an m/z of 4565 ($\pm 0.4\%$),
(d) an elevated level of the protein having an m/z of 4673 ($\pm 0.4\%$),
(e) an elevated level of the protein having an m/z of 4786 ($\pm 0.4\%$), and
15 (f) a depression of a protein having an ion having an m/z of 4155 ($\pm 0.4\%$).

That is, preferably a positive diagnosis of a melanoma comprises the identification of the elevations and depressions of the peptides and proteins as listed above.

20 (41) A method according to any of statements (26) to (40) wherein the proteolytically digested peptide is a tryptic peptide.

(42) A method according to any of statements (26) to (41) wherein the melanoma is a metastatic melanoma.

25

(43) A method according to statement (42) wherein the melanoma is a stage IV metastatic melanoma.

30 (44) A method according to any of statements (26) to (43) wherein the method comprises detecting and/or quantifying one or more of the peptides and/or proteins using mass spectrometry and/or an immunological assay.

Examples of types of suitable mass spectrometry apparatus and methods are provided above. Immunological assays are known to the skilled person and include, for example, an enzyme linked immunosorbent assay (ELISA).

- 5 (45) A method according to any of statements (26) to (44) wherein the immunological assay comprises use of a labelled antibody.

The methods described herein may optionally involve one or more of the following steps:

- 10
- Isolation of serum from a blood sample by allowing the blood sample to clot at room temperature for at least 30 minutes but for no longer than 60 minutes. Preferably the clotted sample is centrifuged, preferably at 2500 g for 10 minutes. The serum is then harvested and subsequently frozen, preferably at -20°C or -80°C . Preferably the
- 15
- sample is initially frozen at -20°C and subsequently thawed once and aliquoted to a smaller volume and the smaller aliquot frozen at -80°C . Preferably the sample is not re-thawed until immediately prior to analysis.
 - Subjecting the sample to three or fewer, preferably two or fewer, freeze thaw cycles.
 - Following receipt of a sample from a subject, e.g. from a clinical centre, storing the sample at -80°C .
- 20
- Following storing of the sample at -80°C , thawing, storing on ice and processing within 2 hours of thawing.
 - Exclusion of spectra having a signal to noise ratio of $<5:1$ for the peak at m/z 6632 ± 13 prior to analysis of the mass spectra.
 - Carrying out mass spectrometry using an instrument in raster mode.
- 25
- Carrying out mass spectrometry using an instrument in auto-quality mode.

(46) A polypeptide consisting of a sequence selected from:

- 30
- (a) WFYIASAFR [SEQ ID No. 1],
 - (b) a derivative sequence of (a) wherein the derivative sequence has an (m/z) of 1161 ($\pm 0.02\%$).

(47) A polypeptide comprising a sequence selected from:

- (a) NTLIIYLDK [SEQ ID No. 2],

(b) a derivative sequence of (a) wherein the derivative sequence has an (m/z) of 1093 ($\pm 0.02\%$).

(48) A polypeptide consisting of a sequence selected from:

- 5 (a) YVGGQEHFAHLLLR [SEQ ID No. 3],
(b) a derivative sequence of (a) wherein the derivative sequence has an (m/z) of 1753 ($\pm 0.02\%$).

10 (49) A mixture of polypeptides comprising 2 or 3 of the polypeptides according any of statements (46) to (48).

(50) A kit for use in a method of detecting a melanoma, the kit comprising positive controls for mass spectrometry analysis, wherein the kit comprises two or more species selected from:

- 15 (a) a protein having an (m/z) selected from 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$);
(b) a proteolytically digested peptide having an (m/z) selected from 1753, 1161, 1505, 854, 1444, 1093 ($\pm 0.02\%$)

20 (51) A kit according to statement (50) wherein the positive controls comprise a peptide having an ion with an m/z of 1093 ($\pm 0.02\%$) and a peptide having an ion with an m/z of 1161 ($\pm 0.02\%$).

(52) A kit according to statement (51) wherein the positive controls comprise:

- 25 (a) a peptide having an ion with an (m/z) of 1161 ($\pm 0.02\%$),
(b) a peptide having an ion with an (m/z) of 1753 ($\pm 0.02\%$),
(c) a protein having an m/z of 4565 ($\pm 0.4\%$),
(d) a protein having an m/z of 4673 ($\pm 0.4\%$),
(e) a protein having an m/z of 4786 ($\pm 0.4\%$), and
30 (f) a protein having an ion having an m/z of 4155 ($\pm 0.4\%$).

(53) A kit according to any of statements (50) to (52) wherein the positive controls comprise:

- (a) proteins having an (m/z) of 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913 and 4835 ($\pm 0.4\%$) and/or
- (b) peptides having an (m/z) of 1753, 1161, 1505, 854, 1444 and 1093 ($\pm 0.4\%$).

5 Increasing the number of positive controls increases the accuracy of the method.

(54) A method of identifying a nucleic acid biomarker comprising

- (a) using the method according to any of statements (1) to (25) to identify a peptide or protein biomarker,

10 (b) identifying the nucleotide sequence encoding the peptide or protein biomarker.

(55) A nucleic acid molecule identified by a method according to statement (54).

That is, the identification of the nucleotide sequence may be determined by using PCR to
15 identify the mRNA species expressed in the sample in which the peptide or protein biomarker was identified.

(56) A vector comprising a nucleic acid molecule according to statement (55).

20 Such vectors include bacteriophages, phagemids, cosmids and plasmids. Preferably the vectors comprise suitable regulatory sequences, such as promoters and termination sequences which enable the nucleic acid to be expressed upon insertion into a suitable host.

(57) A host cell comprising a vector according to statement (56).

25

Preferably the host is *E. coli*. Furthermore, a host cell comprising that vector is provided. The invention also provides a monoclonal antibody capable of specifically binding to a polypeptide according to those described above.

30 (58) A monoclonal antibody capable of specifically binding to a polypeptide according to any of statements (46) to (48).

The polypeptides of the invention may be used to raise antibodies. In order to produce antibodies to tumour-associated antigens procedures may be used to produce polyclonal antiserum (by injecting protein or peptide material into a suitable host) or monoclonal antibodies (raised using hybridoma technology). In addition PHAGE display antibodies may be produced, this offers an alternative procedure to conventional hybridoma methodology. Having raised antibodies which may be of value in detecting tumour antigen in tissues or cells isolated from tissue or blood, their usefulness as therapeutic reagents could be assessed. Antibodies identified for their specific reactivity with tumour antigen may be conjugated either to drugs or to radioisotopes. Upon injection it is anticipated that these antibodies localise at the site of tumour and promote the death of tumour cells through the release of drugs or the conversion of pro-drug to an active metabolite. Alternatively a lethal effect may be delivered by the use of antibodies conjugated to radioisotopes. In the detection of secondary/residual disease, antibody tagged with radioisotope could be used, allowing tumour to be localised and monitored during the course of therapy.

15

The term "antibody" includes intact molecules as well as fragments such as Fa, F(ab')₂ and Fv.

20

(59) The use of an isolated nucleic acid molecule comprising a sequence according to statement (55) or a nucleic acid molecule capable of specifically hybridising to a sequence according to statement (55) to detect or monitor cancer or to detect the susceptibility of a subject to cancer.

25

Such molecules may be used as probes, *e.g.* using PCR. Such probes may be labeled by techniques known in the art, *e.g.* with radioactive or fluorescent labels.

30

The term "specifically hybridising" is intended to mean that the nucleic acid molecule can hybridise to nucleic acid molecules according to the invention under conditions of high stringency. Typical conditions for high stringency include 0.1 x SET, 0.1 % SDS at 68 °C for 20 minutes.

The expression of genes, and detection of their polypeptide products may be used to monitor disease progression during therapy or as a prognostic indicator of the initial disease status of the patient.

5 (60) A method of detecting or monitoring cancer or detecting the susceptibility of a subject predisposed to developing cancer comprising the step of detecting or monitoring an elevated or depressed level of a nucleic acid molecule comprising a sequence according to statement (55) in a sample from a patient.

10 The sample may be a blood, serum or tissue sample.

(61) A method of detecting or monitoring cancer or detecting the susceptibility of a subject to cancer comprising the use of a nucleic acid molecule according to statement (55) or a nucleic acid probe capable of specifically binding to a nucleic acid sequence according to
15 (55) in combination with a reverse transcription polymerase chain reaction (RT-PCR).

(62) A method of detecting or monitoring melanoma or detecting the susceptibility of a subject to melanoma comprising detecting or monitoring elevated levels of a polypeptide according to any of statements (46) to (48).

20

(63) A method according to statement (62) comprising the use of an antibody selective for a polypeptide as defined in any of statements (46) to (48).

(64) Use or method according to any one of statements (61) to (63), wherein the cancer is a
25 melanoma.

(65) Use or method according to statement (64) wherein the cancer is a stage IV melanoma.

Alternatively, or in addition, the uses or methods may be applied in relation to other
30 neoplasias, such as other cancers.

(66) A method of prophylaxis or treatment of cancer comprising administering to a patient a pharmaceutically effective amount of nucleic acid molecule comprising a nucleic acid

sequence according to statement (55) or a nucleic acid molecule capable of specifically hybridising to nucleic acid molecule comprising a nucleic acid sequence according to statement (56) or a pharmaceutically effective fragment thereof.

5 By pharmaceutically effective fragment, the inventors mean a fragment of the molecule which still retains the ability to be a prophylactant or to treat cancer. The molecules are preferably administered in a pharmaceutically amount. Preferably the dose is between 1 $\mu\text{g/kg}$ to 10 mg/kg .

10 (67) A method of prophylaxis or treatment of a melanoma comprising administering to a patient a pharmaceutically effective amount of a polypeptide as defined in any of statements (46) to (48) or a pharmaceutically effective fragment thereof.

(68) A method of prophylaxis or treatment of cancer comprising the step of administering to
15 a patient a pharmaceutically effective amount of an antibody according to statement (58).

(69) A method according to any one of statements to (59) to (68), wherein the sample is a serum sample.

20 (70) A method according to any of statements (59) to (69) wherein the cancer is a melanoma.

Preferably the method is applied to a stage IV melanoma.

(71) A vaccine comprising a nucleic acid molecule having a nucleic acid sequence as defined
25 in statement (55) or a pharmaceutically effective fragment thereof and a pharmaceutically acceptable carrier.

(72) A polypeptide comprising a carrier covalently attached to a polypeptide according to
any of statements (46) to (48) or a pharmaceutically effective fragment thereof.

30

The carrier may be a further protein or a fragment of a further protein, such as tetanus toxoid, to make it immunogenic (using well-known techniques). That is, the further protein

or protein fragment is not from the polypeptide or protein. Such constructs and nucleic acid molecules encoding such constructs are also part of the invention.

Another aspect of the invention provides a polypeptide comprising a carrier covalently attached to a polypeptide according to any of those described above or a pharmaceutically effective fragment thereof.

(73) A nucleic acid molecule encoding a polypeptide according to statement (72) or any of the polypeptides described herein.

10

(74) A vaccine comprising a polypeptide according to any of statements (46) to (48) or any of the polypeptides described herein or a pharmaceutically effective fragment thereof which may be optionally attached to a further immunogen, and a pharmaceutically acceptable carrier.

15

(75) Use of a biomarker obtainable by a method according to any of statements (1) to (25) or (54) to detect or monitor neoplasia or to identify a subject predisposed to developing a neoplasia.

(76) Use according to statement (75) wherein the neoplasia is selected from a melanoma, a prostate cancer and a breast cancer.

20

(77) A method of identifying a subject having a melanoma comprising determining the level of alpha 1-acid glycoprotein or a fragment thereof in a sample obtained from a subject and comparing that with the level of alpha 1-acid glycoprotein from a control sample.

25

(78) A method according to statement (77) wherein the melanoma is a Stage IV melanoma.

(79) A method according to statement (77) or (78) wherein the fragment of alpha 1-acid glycoprotein has an ion having an m/z of 1161 or 1753 ($\pm 0.02\%$).

30

(80) A method according to statement (77) to (79) wherein the fragment has the sequence selected from the polypeptide of statement (46) or of statement (48).

(81) A method according to statement (80) wherein the alpha 1-acid glycoprotein ion is derived from AAG-1/2.

- 5 (82) A method of identifying a drug for treating melanoma comprising:
- (a) treating a sample from a subject having melanoma with a candidate compound,
 - (b) subjecting the treated sample from (a) to mass spectrometry to generate one or more peaks,
 - 10 (c) comparing the peaks obtained from the treated sample with a control sample from a subject having melanoma wherein the control sample was not treated with the candidate drug,
- to identify candidate compounds which decrease the level of melanoma biomarkers.

15 Other methods may be used in place of mass spectrometry, *e.g.* chromatography.

- (83) A method according to statement (82) further comprising comparing the peaks obtained from the treated sample with peaks obtained from one or more further control samples treated with mass spectrometry to generate one or more peaks wherein the further control sample is obtained from one or more of:
- 20 (a) a subject not having melanoma, or
 - (b) a subject having Stage IV melanoma.

As an alternative, any other suitable method may be used in place of mass spectrometry.

- 25 (84) A method according to statement (83) wherein the peaks are selected from one or more of:
- (a) a protein having an (m/z) selected from 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$);
 - (b) a proteolytically digested peptide having an (m/z) selected from 1753, 1161, 1505,
 - 30 854, 1444, 1093 ($\pm 0.02\%$).

(85) A mass spectrometry melanoma fingerprint for identifying the presence of a melanoma comprising two or more of the following peaks:

(a) a protein having an (m/z) selected from 12000, 14847, 1649, 15477, 13255, 3031, 4791, 9913, 4835, 4155, 4565 and 4673 ($\pm 0.4\%$);

(b) a proteolytically digested peptide having an (m/z) selected from 1753, 1161, 1505, 854, 1444, 1093 ($\pm 0.02\%$).

5

(86) A fingerprint according to statement (85) wherein the fingerprint comprises peaks showing an (m/z) of 1161 ($\pm 0.02\%$), 1753 ($\pm 0.02\%$), 4565 ($\pm 0.4\%$), 4673 ($\pm 0.4\%$), 4786 ($\pm 0.4\%$).

10 (87) A fingerprint according to statement (86) wherein the fingerprint does not comprise a peak showing an (m/z) of 4155 ($\pm 0.4\%$).

(88) A computer system having a processor, a memory and an input for receiving data from a mass spectrometer wherein the system with memory comprises data corresponding to a
15 fingerprint according to statement (85), (86) or (87) and the computer configured to compare data from the input against the fingerprint data and to identify whether the inputted data is indicative of a melanoma based on the results of the comparison.

(89) Melanoma identification apparatus comprising the computer system of statement (89)
20 and a mass spectrometer connected to the input.

(90) A method of identifying one or more biomarkers of a specified biological phenotype and/or of exposure to a specified environmental condition, the method comprising:

(a) obtaining:

25 (i) a sample obtained from each of a plurality of subjects having the specified biological phenotype and/or having been exposed to the specified environmental condition,

(ii) a sample obtained from each of a plurality of control subject not having the specified biological phenotype of (i) and/or having been exposed to the specified
30 environmental condition (i),

to generate intact samples,

- (b) analyzing an aliquot of each of the samples by a method which separates components of the mixture relative to their molecular weight to generate a profile of peaks representing the relative abundance of the components in the mixture,
- (c) analyzing the profile obtained from (b) using an artificial neural network to identify peaks of the profile which distinguish between a subject having the specified biological phenotype and/or having been exposed to the specified environmental condition and/or a subject not having the specified biological phenotype or having been exposed to the specified environmental condition.
- 10 The analysis step, using an ANN in any of the above methods, may be replaced by an alternative analysis step to identify peaks of the profile which distinguish between a subject having the specified biological phenotype and a subject not having the specified biological phenotype.
- 15 (91) A method according to statement (90) wherein the subject is a human, an animal, a plant, a virus, a bacteria or a fragment or cell obtained therefrom.
- (92) A method according to statement (90) or (91) wherein the sample comprises proteins, peptides or fragments thereof.
- 20 (93) A method according to any of statements (90) to (93) wherein the sample comprises a nucleic acid molecule.
- (94) A method of diagnosis of the susceptibility of a subject to a neoplasia comprising determining the level of a biomarker identified according to the method of any of statements (1) to (25) or (54).
- 25 (1) to (25) or (54).

Examples of biological phenotypes include neoplasias such as cancers including melanoma, prostate and breast cancers. Examples of environmental conditions include heat shock, cold shock, exposure to one or more forms of radiation, exposure to one or more pharmaceutical compounds or combinations of the above.

30

Another aspect of the invention provides a method of diagnosis of the susceptibility of a subject to a neoplasia comprising determining the level of a biomarker identified according to the method of any of the above methods. Preferably the method comprises a comparison with a negative or positive control.

5

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be explained by way of example only with reference to the following drawings:

10

Figure 1. Experimental flowchart for profiling serum proteins and tryptic peptides.

15

Figure 2. Replicate MALDI-TOF spectra of serum proteins and tryptic peptides. Replicate MALDI-TOF analysis of aliquots of the same sample (a) diluted 1 in 10 with 0.1 % TFA and analyzed by MALDI-MS in linear mode and (b) prepared by ZipTip clean-up prior to and following tryptic digestion and analyzed in reflectron mode. The m/z and normalized intensities of selected peaks within the ranges 1000-16000 Da for serum proteins (a) and 800-3500 Da for tryptic peptides (b) are presented in Table 1a. and 1b.

20

Figure 3. (a) Representative mass spectra of proteins m/z 1000-15000 and (b) magnified view of protein ions m/z 4000-5000 using unfractionated human serum from melanoma stage IV patients and controls analyzed by MALDI-linear TOF MS. Serum was diluted 1 in 10 in 0.1 % TFA and deposited with an equal volume of sinapinic acid using the dried droplet method on a MALDI sample plate. Protein peaks in the region m/z 4500-5000 (Figure 3b) show both up- and down-regulation of ions in cancer serum samples (a – c), compared with serum from healthy controls (d – f).

25

Figure 4. Representative mass spectra of serum tryptic peptides from melanoma stage IV patients and controls. This spectrum illustrates the observed intensity differences of ions at m/z 1161 and 1753, present in melanoma stage IV sera (a – c), but detected only at very low intensities in the control samples (d – f).

30

Figure 5. (a) Predictive capability of ANNs trained to recognize protein profiles based on a 9 ion ANNs model and (b) predictive capability of ANNs to recognize tryptic peptide profiles based on a 6 ion model. The grey bars indicate control samples and the black bars indicate stage IV Melanoma cancer. A predicted value below 1.5 indicates a control sample, whilst a prediction greater than 1.5 indicates a treated sample. (c) Validation of the ANNs model was confirmed using 50 stage I samples as a blind data set in the developed ANN model for control versus stage IV tryptic peptides. The samples were correctly assigned as non-stage IV samples to a median accuracy over the 50 optimized sub-models of 97 % (inter-quartile range 92 % - 99 %). The white bars indicate controls, the grey bars represent stage I and the black bars are stage IV samples.

Figure 6. Sequence analysis of serum tryptic peptide precursor ion m/z 1753.2 (YVGGQEHLILR, derived from α -1-acid glycoprotein) by (a) MALDI-TOF PSD (parent ion not shown), (b) AP-MALDI-QIT MS/MS and (c) LC-ESI-QIT MS/MS. Inset shows an extracted ion chromatogram (m/z 1753.2) displaying a single peak, which shows that this ion was derived from a single peptide.

Figure 7. (a) representative spectra of commercially available AAG was made up to 1mg/ml and digested and spotted on the MALDI-TOF and analysed in reflectron mode. The two most prominent peaks identified are at m/z 1160.1 and 1753.1 verifying the inventors' results from stage IV patients as belonging to AAG parent protein. (b) Sequence analysis of the ion at m/z 1753.1 after an aliquot of commercial AAG was digested and analysed on the LC-ESI-QIT MS/MS. These data confirm the inventors' results with similar b and y ions identified and similar qualitative spectral pattern.

Figure 8. (a) Box and whisker plot depicts the mean $\pm$ SE (Standard Error of the Mean) (g/L) for serum AAG levels in stage IV versus control patients. Turbidimetric analysis of serum AAG levels reveal that it is significantly increased in stage IV patients in comparison with healthy controls. (b) Plot of AAG concentration versus predictive ions at m/z 1753 and 1160 MALDI mass spectral intensity for 15 control and 12 stage IV patients. It is clear that the two groups can be distinguished as two separate clusters.

Figure 9. Ion intensity of m/z 1753 versus sub-stages of Stage 4 metastatic Melanoma. Sub-stage M1c shows a higher level of AAG in comparison with the other sub-stages which may indicate that AAG can be used to assess the progression to metastatic Stage IV disease.

- 5 Figure 10. Ion intensity of m/z 1160 versus sub-stages of Stage 4 metastatic Melanoma. Sub-stage M1c shows a higher level of AAG in comparison with the other sub-stages which may indicate that AAG can be used to assess the progression to metastatic Stage IV disease.

EXPERIMENTAL PROCEDURES

10

Patient Selection and Serum Sample Collection

Serum samples were selected from a frozen collection of sera from patients with histologically confirmed melanoma. Tumour staging was performed by CT/MRI and the patients were classified according to the staging system of the AJCC [1]. All patients gave informed consent and the study was approved by the local Institutional Review Board (IRB).
15 The serum samples were processed following a standardized protocol: blood was drawn from the patients' cubital vein into gel coated serum tubes (Sarstedt, Nuembrecht, Germany) and allowed to clot at room temperature for at least 30 min, but no longer than 60 min. Thereafter, the tubes were centrifuged at 2500 g for 10 min. The serum phase was harvested
20 and subsequently frozen without any additives in 1 ml aliquots at -20 °C thawed once to aliquot the sample into smaller volumes and transferred to -80 °C, and thereafter it was not thawed until immediately prior to analysis. Serum samples from stage IV melanoma patients (30 male and 20 female) had a mean age $\pm$ SD (standard deviation) of 54.4 $\pm$ 15.6 years. Distribution of patients in Stage IV (n = 50) was as follows: M1a - metastases in skin, subcutaneous and/or lymph nodes and normal LDH - (n = 14), M1b - lung metastases only
25 and normal LDH - (n = 10), M1c - metastases in visceral organs (e.g. liver, brain, etc) and/or elevated serum LDH (n = 26). In this study the inventors have not sought to perform analysis on the sub-stages due to the relatively low numbers and have therefore, combined these patients as having advanced disease. Stage I melanoma patients (24 male and 26 female) had
30 a mean age of 58.8 $\pm$ 13.7 years. The healthy volunteers (32 male and 18 female) had a mean age of 45.3 $\pm$ 16 years.

Sample Preparation

Protein analysis

Prior to sample handling and preparation, the samples were randomized, for position on a 96-well dilution plate and a 384 MALDI target plate, using a computer randomization program (Microsoft Excel). Serum samples were prepared in duplicate according to Figure 1. Briefly, serum was diluted 1 in 10 in 0.1 % trifluoroacetic acid (TFA) and an aliquot was spotted together with SA matrix (10 mg/ml) using the dried droplet method, in which equal volumes (1 µL) of sample and matrix were mixed on the target plate allowed to air dry, and analyzed by MALDI-TOF-MS.

Peptide analysis

Tryptic serum and commercially available AAG (1mg/ml) peptides were prepared according to Figure 1. An aliquot (25 µL) of the diluted serum (1 in 10 with 0.1 % TFA), or alpha 1-acid glycoprotein (AAG) solution was initially fractionated using a ZipTip C₁₈ (Milipore, Watford, UK) with 25 cycles of binding. This was followed by two washes in 0.1 % TFA and elution in 4 µl of 80 % acetonitrile/0.1 % TFA. This volume was combined with ammonium bicarbonate (16.6 µl of 100 mM), water (7.6 µL), and trypsin Gold Mass Spectrometry Grade (1.3 µl of 0.5 µg/µl, Promega, Southampton, UK) and incubated at 37 °C overnight. The reaction was quenched with 1 % TFA (1 µl) and the sample cleaned-up using a C₁₈ ZipTip following the procedure described above and an aliquot of the eluate (1 µl) was spotted onto the MALDI target using the dried droplet method with matrix, CHCA (LaserBio Labs, Cedex, France), prepared as a 10 mg/ml solution in 50 % ACN + 0.1 % TFA. Duplicate samples were applied to the target plate in the randomized order used for sample preparation and analysed by MALDI-TOF-MS. A range of appropriate blank and control samples were prepared alongside the serum digests. For every 50 patient samples processed 10 aliquots of 0.1 % TFA (25 µL) were taken as reagent blanks through the entire ZipTip clean-up prior to digestion. Serum blanks (4 per 100 serum samples processed) were also diluted 1/10 with 0.1 % TFA (25 µL), combined with the digestion buffer without the incorporation of trypsin before, and after ZipTip clean-up and incubated overnight at 37 °C. Bovine serum albumin (BSA) controls (1 per 25 serum samples processed) were used to ensure the efficiency of the digestion procedure. A BSA solution was made up to 1 nmol/µl,

diluted 1/10 with 0.1 % TFA and digested as above, followed by ZipTip clean-up. Samples were applied to the MALDI-TOF target plate using the dried droplet method.

Mass Spectrometric Analysis

- 5 MALDI-TOF experiments were performed on Axima CFR⁺ mass spectrometer (Shimadzu, Manchester, UK). Close external calibration was performed using protein calibration mix 2 and peptide calibration mix 4 (proteomix) supplied by Laser Bio Labs (Cedex, France). The protein calibrants were: Cytochrome C, horse heart, m/z 12361.12; Myoglobin, horse, m/z 16181.06 and Trypsinogen, m/z 23981.98. Insulin beta chain, m/z 3494.65 (3 μ l of 5mM)
- 10 was also added to the calibration protein mix. Mass spectral acquisition in the range 1000-25,000 Da was carried out in 'raster mode' using linear TOF-MS. The resultant mass spectra were all examined visually and spectra were excluded from the data set before being processed for bioinformatic analysis if the signal-to-noise ratio was <5:1 for the peak at m/z 6632 $\pm$ 13. The peptide calibration was based on the monoisotopic masses of [M+H]⁺ of
- 15 bradykinin fragment 1-5, angiotensin II, neurotensin, ACTH clip (18-39) and Insulin B-chain oxidised at m/z 573.31, 1046.54, 1672.91, 2465.19, 3494.65, respectively. The samples were analyzed by MALDI-TOF mass spectrometry in reflectron mode using pre-scanning in 'auto quality' mode.
- 20 Identification of peptides of known m/z values predicted as being discriminatory between stage IV melanoma and control populations by ANNs was carried out using MALDI-TOF with post source decay (MALDI-TOF-PSD), atmospheric pressure MALDI/ion trap mass spectrometry (AP-MALDI-QIT) and liquid chromatography combined with electrospray ion trap tandem mass spectrometry (LC-ESI-QIT MS/MS) analysis. Sample preparation for
- 25 MALDI-TOF-PSD and AP-MALDI was identical to that described above and according to Figure 1. MALDI-PSD was carried out as an accumulation of 500 laser shots using the Axima CFR⁺ mass spectrometer. Laser power for PSD measurements was increased to 50 % above the threshold for observation of the fragment ions. PSD calibration was used, as prescribed by the manufacturer. AP-MALDI experiments were performed using a MassTech
- 30 (Columbia, MD, USA) AP-MALDI source equipped with a 10 Hz, 300 mJ, 337nm N₂ laser interfaced with an LCQ ion trap mass spectrometer (ThermoElectron, San Jose, CA, USA). For all experiments, the ion trap automatic gain control was deactivated, the gate time set to 500 ms, and the heated capillary inlet temperature was set to 350°C. The AP-MALDI target

plate was held at 2.7 kV and no auxiliary or sheath gases were used. Typical collision energy used for MS/MS was between 35-38 %. The experimental masses of the precursor ion and fragment ions, were used as inputs for MASCOT sequence query (www.matrixscience.com), using the following parameter settings: 1 maximum missed cleavage, 0.8 Da (for PSD-MALDI and AP-MALDI) tolerance for the singly charged precursor ion, and 0.7 Da (PSD-MALDI) and 0.6 Da (AP-MALDI) for the fragment ion mass. NCIBnr was used as the reference database (human taxonomy). Trypsin was selected as the proteolytic enzyme.

LC-ESI-QIT MS/MS analyses were carried out using on-line reverse-phase nano-scale liquid chromatography in conjunction with ion trap tandem mass spectrometry. Human serum samples from Stage IV melanoma and control patients, and commercially available AAG, were digested as described above, followed by dilution with 100 µl 0.1 % TFA. Samples were injected (100 µl volume) onto a precolumn (LC Packings, C₁₈-PepMap, 100 Å, 3µm particle size, 300 µm ID x 5 mm, Dionex Ltd, UK) via a six-port automated switching valve, using a loading mobile-phase (0.1 % TFA, 30 µL/min), supplied by a loading pump (LC Packings, Dionex Ltd, UK), to effect sample preconcentration and desalting (total of 14.0 min.). The six-port valve was switched to introduce a counter-current solvent flow (180 nL/min, split flow) to the pre-column from an UltiMate gradient pump (LC Packings, Dionex Ltd, UK) to direct the sample onto an reverse-phase capillary LC column (LC Packings, C₁₈-PepMap, 100 Å, 3µm particle size, 75µm ID x 150 mm, Dionex Ltd, UK) connected to the mass spectrometer interface by a fused silica transfer line (20 µm ID, 300 mm). On-line sample separation prior to mass spectrometric detection was carried out using a linear gradient (Solvent A: 0.1 % formic acid in water; Solvent B: 80 % acetonitrile in 0.1 % formic acid in water) from 5 % B (at time of six-port valve switch) to 75 % B (over 60 min., hold for 2 min.) then to 5 % B (over 2 min.) and re-equilibrate for the next analytical run. Tandem mass spectrometric analysis was carried out using a Finnigan LCQ Classic ion trap mass spectrometer (ThermoElectron, San Jose, CA, USA) equipped with a dynamic nano-electrospray ion source, operated in positive ion mode. Analytical performance of the hyphenated LC-QIT MS/MS system was assessed by analysis of a BSA tryptic digest as standard, bracketing replicate serum samples, to provide validation for human serum analysis. Data was acquired for human serum tryptic peptides following LC introduction using either full scan mode (m/z 300-2000, 3 microscans) or targeted tandem mass spectrometric (MS/MS) mode (200 ms activation time, isolation m/z of 3.0). Automatic gain

control (AGC) was applied in all data acquisition modes. Automated database search of fragment ion spectra was carried out using the SEQUEST algorithm (Bioworks software, ThermoFinnigan, San Jose, CA, USA), the parameters were: trypsin was selected as the proteolytic enzyme and m/z tolerances were set to 0.9 Da for the parent ion and 0.2 Da for fragment ions. A total m/z range of 200-3500 was specified, with a minimum total ion chromatogram count of 5×10^4 counts. Peptide assignment from SEQUEST output satisfied a confidence level value > 2.0 and probability score (Sp) > 200 , matching a minimum of 4 product ions above m/z 400 upon manual inspection. Sequence identities were confirmed using the Mascot database and the search parameter settings for the MASCOT sequence query routine were as follows: 1 maximum missed cleavage, 0.8 Da tolerance was used for the singly and doubly charged precursor ion and 0.5 Da for the fragment ion mass. NCIBnr was used as the reference database (human taxonomy). Trypsin was set as the proteolytic enzyme.

To validate the developed MALDI/bioinformatic protocols, the experiments were repeated for serum proteins/peptides by a different operator four months for proteins and one week for peptides after the first analysis of all samples to assess intra-laboratory reproducibility. Fresh serum samples were prepared and spotted onto the target plate as above and analysed by MALDI-TOF-MS. The resulting data was used to validate the ANNs model developed for the first set of samples.

Artificial Neural Networks (ANNs)

Protein Data

25 *Data pre-processing and model development*

Prior to ANN training the raw data from the mass spectrometer were imported as ASCII text files. Linear TOF mass spectral data were binned across the 2-25 kDa mass range with the median intensity value across a 1 Da range determined to represent the intensity of a given mass value. A three layer MLP ANN was used with a feed forward BP algorithm utilising a sigmoidal transfer function a momentum of 0.5 and a learning rate of 0.1. Data were scaled between 0 and 1 using minimums and maximums, where raw values were scaled linearly so that the smallest value in the dataset was the minimum, and the largest the maximum. This scaling method ensures that all potential relationships amongst variables are kept identical.

therefore not introducing bias into the data. Inputs to the network represented binned m/z values between 2-25 kDa together with their corresponding intensity values. Two hidden nodes were used in the hidden layer. The output later consisted of a single node, where control samples were coded as "1", and Stage IV melanoma samples were coded as "2".

5

For ANN analysis, the data consisted of 100 samples each with 23001 corresponding variables specifying the intensity at a given binned m/z value. Models were trained using random sample cross validation. Here, the samples were randomly split into three groups; training, test, and independent validation. From a total of 100 samples (50 per class) 60 samples were randomly assigned to the training set, 20 were randomly assigned to the test set, and the remaining 20 were completely removed and used for an independent validation. This process has been termed random sample cross validation. The ANNs were trained using the training set and network error with regards to predictive performance was monitored with the test set, which was unseen during training. Once this error failed to improve for a pre-determined number of training events, training was terminated, and the model validated on the independent validation set, which was truly blind to the model. The random sample cross validation process was repeated 50 times so that each sample was part of the independent validation data subset a number of times. This enabled confidence intervals to be calculated for the network predictions on blind data, and to provide a large set of models for comparing predictive performances and identified biomarker ions.

10
15
20

Modelling process

The modelling process involved a novel stepwise approach. Initially, each variable from the dataset was used as an individual input in a network, thus creating n (23001) individual models. These n models were then trained, using the random sample cross validation process described above, creating 50 sub-models for each of the n models. These models were ranked in ascending order based upon their mean squared error values for test data. The model with the lowest predictive error identified the most important single ion which was selected for inclusion in the subsequent additive step.

25
30

The remaining inputs were then sequentially added to the previous best input, creating $n-1$ models each containing two inputs. Training was repeated and performance evaluated for these as described above. The model which showed the best performance was then selected

and the process repeated, creating $n-2$ three input models. This process was repeated until no statistically significant improvement in model performance was gained from the addition of further inputs for three complete steps, thus resulting in a final model containing the proteomic pattern which most accurately predicted between the two outcomes. This process defines a best subset of ions (that are ordinal to each other) that explain the variation within the melanoma versus control classes.

Digested Peptide Data

For the tryptic digested peptides, the modelling protocol was identical to that for the protein data outlined above, except that data used were between the range m/z 800-3500.

Validation of the ANN models

To address the reproducibility of the ANNS procedure over multiple experiments, both the proteins and tryptic peptides were re-analyzed on a separate occasion by a different operator.

The spectra from this second experiment were then used for validation of samples as a second cohort of blind data to assess the reproducibility of the proteomic and bioinformatics methodologies. Further, 50 stage I samples were analysed by MALDI-TOF mass spectrometry and validated in the developed model to assess the ability of the model to discriminate between early (stage I) and late (stage IV) stage metastatic melanoma samples and to show that the biomarker ions were specific to late stage melanoma compared to controls.

Quantitation of AAG

Randomly selected $n = 15$ stage IV and $n = 15$ controls were used for quantitation of AAG (g/L) using an automated turbidimetric assay with a Randox kit (Randox Laboratories Ltd, Co Antrim, UK) based on a specific antibody to AAG. The formation of the antibody complex during the reaction results in an increase in turbidity, which is measured as the amount of light absorbed at 340 nm. The method was run on a Kone analyser (Labmedics, UK).

RESULTS

Optimization of Serum Protein and Peptide protocols

Mass spectral quality and reproducibility has been shown to be central to the performance of proteomic approaches based on mass spectrometry. A comprehensive strategy to optimize all aspects of sample preparation, deposition, calibration and instrumentation parameters to produce standardized protocols for the analysis of protein and tryptically digested peptides of serum samples is reported. The inventors' results indicated that 1 in 10 dilution of serum with 0.1 % TFA gave the highest number of peaks and signal-to-noise ratios in the range across the m/z 1000-25000 using linear TOF-MS (data not shown). ZipTip clean-up methods caused protein loss corresponding to 84 % and 75 % respectively for the C₄ and C₁₈ tips at a 1 in 10 serum dilution. For the analysis of tryptic peptides, the use of ultrafiltration devices to remove high molecular weight proteins prior to enzymatic digestion led to inconsistent results, as did acetonitrile precipitation. In reflectron TOF-MS mode, peptide signal intensities were comparable for samples prepared by acetonitrile precipitation and by C₁₈ ZipTip clean-up before digestion. However, a greater number of peptide ions and improved reproducibility were observed in the spectra obtained using ZipTip clean-up. Prior to ANNs analysis MALDI linear TOF and reflectron TOF mass spectra were visually examined and spectra were excluded if the signal-to-noise criteria were not met.

Reproducibility studies were carried out through replicate analyses of aliquots of the same serum sample, which were taken through the full analytical procedure before being spotted randomly on the MALDI target plate. Representative spectra are shown in Figures.-2a. and -2b. The mean and coefficient of variation (CV) for the m/z values and normalized intensity ratios (NIR) were calculated as shown in Table-1a. for serum proteins (n = 10) and Table-1b. for digested serum peptides (n = 6).

Tables 1(a) and (b). Reproducibility data for (a) serum proteins and (b) tryptic peptides. Masses (m/z) and intensities (after normalization) acquired by MALDI-TOF-MS in linear mode for serum proteins and in reflectron mode for the digested peptides are shown as mean values and their respective coefficient of variation (CV).

- 5 The CV for 5 selected protein peaks with normalized intensities in each of the regions between the range m/z 2000-16000 was less than 25 %. The tryptic peptides lead to CV of normalized intensities between 10.9 – 36.5 %. The mean CV for m/z using linear mode analysis of serum proteins was 0.1 % whereas the digested peptides lead to a CV of 0.01 % for all peaks. *NIR = Normalized Intensity Ratio.*

10

(a)

PROTEINS

	m/z	NIR	m/z	NIR	m/z	NIR	m/z	NIR	m/z	NIR
	2751.0	0.26	4570.4	1.30	10400.0	0.2	11321.3	0.74	15872.2	0.10
	2752.1	0.23	4568.6	1.16	10398.0	0.2	11320.1	0.67	15863.2	0.13
	2758.4	0.24	4571.6	1.46	10396.9	0.3	11320.1	0.60	15867.3	0.17
	2757.5	0.16	4572.4	0.94	10397.8	0.3	11319.1	0.66	15867.6	0.12
	2751.8	0.23	4577.8	1.03	10397.5	0.2	11318.0	0.69	15874.5	0.13
	2754.6	0.23	4578.6	1.12	10392.1	0.3	11319.0	0.61	15869.9	0.12
	2748.5	0.29	4567.5	1.54	10405.2	0.3	11327.5	0.68	15872.4	0.12
	2749.8	0.33	4569.6	1.60	10407.2	0.3	11328.0	0.75	15874.9	0.11
	2754.8	0.21	4567.4	1.63	10393.1	0.2	11329.6	0.81	15876.1	0.20
	2756.3	0.24	4570.3	1.19	10395.7	0.2	11394.3	0.56	15863.9	0.10
MEAN	2753.48	0.24	4571.42	1.30	10398.35	0.24	11329.70	0.68	15870.20	0.13
CV %	0.12	18.73	0.09	19.04	0.05	12.90	0.20	11.19	0.03	24.28

Mean CV = 17.23 % (NIR) and 0.098 % (m/z)

(b)

**DIGESTED
PEPTIDES**

	<i>m/z</i>	NIR	<i>m/z</i>	NIR	<i>m/z</i>	NIR	<i>m/z</i>	NIR
	842.4	0.32	1716.8	0.50	2451.6	0.16	2806.8	0.23
	842.4	0.34	1716.8	0.32	2451.3	0.16	2806.6	0.20
	842.4	0.45	1716.8	0.23	2451.7	0.18	2806.5	0.21
	842.3	0.29	1716.7	0.20	2451.9	0.14	2806.5	0.24
	842.4	0.46	1716.7	0.25	2451.5	0.24	2806.5	0.18
	842.3	0.22	1716.7	0.27	2451.3	0.12	2807.3	0.19
MEAN	842.4	0.35	1716.8	0.29	2451.6	0.17	2806.7	0.21
CV %	0.01	27.16	0.01	36.48	0.01	24.88	0.01	10.89

Mean CV = 24.85 % (NIR) and 0.01 % (*m/z*)

Table 2: Ions predicted by ANN analysis discriminating between melanoma stage IV cancer and control patients. As the number of ions added to the ANN model was increased the accuracy of prediction increased and the error decreased. The ANNs predictive capability plateaued once an accuracy of 92 % for proteins and 100 % for peptides respectively for blind samples was achieved. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and N.S. $p > 0.05$.

Control vs Stage IV proteins	<i>m/z</i>	Test Performance %	<i>p</i> -value	Control vs Stage IV tryptic peptides	<i>m/z</i>	Test Performance %	<i>p</i> - value
1	12000	64	-	1	1753	78	-
2	14847	73	***	2	1161	92	***
3	1649	80	***	3	1505	95	*
4	15477	80	N.S.	4	854	97	**
5	13255	82	N.S.	5	1444	99	*
6	3031	84	N.S.	6	1093	100	***
7	4791	87	** (visually detected)				
8	9913	87	N.S. (visual)				
9	4835	92	*** (visual)				

To simulate realistic factors that potentially determine the “biomarker profile” of retrospective clinical serum samples, the effects of storage temperature on serum proteins/peptides were studied using MALDI-TOF-MS. No significant changes in intensity were observed for protein or peptide peaks for samples subjected to three freeze thaw cycles, (data not shown). However, slow degradation of proteins was apparent for samples incubated at 4 °C for 24 hours, resulting in peptide fragment peaks of greater intensity in comparison

with fresh samples; room temperature-aging of serum also showed significant changes in protein and peptide profiles after a period of 96 hours (data not shown). In order to avoid such 'aging' problems, serum samples were stored at -80°C after receipt from the clinical centre, kept on ice following thawing and processed within 2 hours.

5

MALDI-TOF analysis of serum biomarker ions

Metastatic lesions occur in up to 20 % of patients with melanoma. Currently there is no reliable way to detect/predict which patients at time of tumour freedom are at risk of metastatic melanoma. Thus it would prove of great clinical value if predictive biomarkers
10 could be identified which distinguish patients with early stage disease from those who progress to metastatic disease after a defined time-interval. Analysis of 50 serum samples from stage IV melanoma and 50 serum samples from healthy controls using the integrated MALDI-MS analysis reported here showed reproducible and visible variations between control and stage IV sera for both proteins and tryptic peptides, as shown in Figure 3a.,
15 Figure 3b. and Figure 4., respectively. The most prominent spectral discriminatory pattern in stage IV melanoma compared with unaffected individuals was defined by ions in the range m/z 4000-5000 for proteins and between m/z 1100 and 1800 for tryptic peptides. The spectral analysis of both proteins and tryptic peptides resulted in clear and distinct patterns that differentiate disease from controls. Up-regulation of protein ions at m/z ~ 4565, 4673 and
20 4786 (Figure 3b) and tryptic peptides at m/z 1160 and 1753 (Figure 4) and down-regulation of protein ions m/z ~4155 (Figure 3b) was observed in metastatic melanoma patients compared with healthy controls. Due to the difficulty in visually analyzing the complex and highly dimensional spectral data (Figures 3a, 3b and 4) and in order to remove operator subjectivity in data analysis a comprehensive analysis using the full dataset for both the
25 protein and tryptic peptides was undertaken using bioinformatic ANNs analysis.

ANNs Model development and biomarker identification

Using a stepwise approach, ANNs identified biomarker patterns containing nine ions from the protein mass spectral profiles and six ions from the tryptically digested peptide profiles,
30 which correctly discriminated between control and stage IV samples to a median accuracy of 92 % (inter-quartile range 89.4 – 94.8 %, sensitivity of 91 % and specificity of 93 %; (Figure 5a) and 100 % (inter-quartile range 96.7 – 100 %, sensitivity of 100 % and specificity of 100 %; (Figure 5b) respectively for blind data sets. Table 2. shows the performance of the model

at each step of the analysis for both protein and tryptic peptide data. With the continual addition of key ions, an overall improvement in both the predictive error and the median accuracies for samples correctly classified was observed. No further steps were conducted after the inclusion of nine and six ions for the proteins and peptides respectively, because no statistically significant improvement in model performance was achieved with the addition of further ions. Therefore the biomarker ions within these models were considered to be those which most accurately distinguished between stage IV melanoma patient and healthy control sera for this dataset.

10 *ANNs Model Validation*

To address the reproducibility of the analytical procedure over multiple experiments [30], both the proteins and peptides were re-analyzed on a separate occasion by a different operator. The spectra from this second experiment were used to validate the proteomics and bioinformatics methodology by evaluating the ability of the models to assign these new samples accurately to the correct class. For the protein data, the optimised model set correctly classified 85 % (inter-quartile range 83 % - 89 %) of these blind samples correctly, with sensitivity and specificity values of 82 and 88 % respectively, with an area under the curve (AUC) value of 0.9 when evaluated with a Receiver Operating Characteristic (ROC) curve. For peptides, the model correctly classified 94 % (inter-quartile range 93 % - 97 %), with sensitivity and specificity values of 100 and 92 % respectively; the AUC value for this dataset was 0.98. This suggests that due to the mass accuracy being higher for peptides, the peptide data was more reproducible than the protein data on mass spectrometry analysis. However, the acquisition of both protein and peptide data allows the two sets of biomarker ions to be combined to provide a highly specific, sensitive, and a rapid diagnostic tool for discriminating stage IV melanoma from healthy controls with 100 % accuracy for the sample cohort used in this study. In addition, 50 samples from stage I melanoma were analysed by MALDI mass spectrometer and treated as blind in the developed ANN model. The samples were correctly assigned into group 1 as non-stage IV samples to a median accuracy over the 50 optimized sub-models of 97 % (inter-quartile range 92 % - 99 %) with specificity being 95 %. (Figure 5c). This demonstrates that the biomarker ions identified are extremely powerful predictors, and therefore are potential indicators of late stage metastatic melanoma.

Sequences of Predictive Biomarkers associating with late stage melanoma.

The predictive peptide ions were sequenced and identified using PSD and tandem mass spectrometry. The MALDI-TOF-PSD spectrum for the m/z 1753.2 ion is shown in Figure 6a. A Mascot search (NCIBnr database) identified the sequence as derived from AAG-1/2 with a score of 48 ($p < 0.0021$, 12 matched peptides for the 1753.2 ion). This identification was verified by AP-MALDI ion trap tandem mass spectrometry (Figure 6b). The capillary LC-ESI-QIT MS/MS analysis of the singly charged ion at m/z 1753.2 (Figure 6c) confirmed the assignment with higher selectivity and showed that this ion was derived from a single chromatographically-resolved peptide (inset Figure 6c), which was also assigned to AAG-1/2 using a SEQUEST data search. Confidence in this assignment was further strengthened by targeting the doubly charged ion for this candidate peptide (m/z 877.1, with a score of 62, $p < 0.00018$, 13 matched peptides) by LC-ESI-QIT MS/MS, which confirmed the peptide sequence as AAG-1/2 precursor. The same methodologies were applied to the identification of the peptide ion at m/z 1160.8 which yielded the sequence WFYIASAFR with a score of 41 ($p < 0.031$, 10 matched peptides), also derived from AAG-1/2. Conclusive confirmation that ions at m/z 1160.8 and 1753.2 from serum studies belonged to AAG parent protein were obtained from studies of commercially available AAG, which was digested and analysed in reflectron mode MALDI-TOF-MS (Figure 7a). This showed that the two most prominent peaks were at m/z 1160.1 and 1753.1. These ions were further targeted for sequence analysis using the LC-ESI-QIT MS/MS and an example of sequence data at m/z 1753.1 is shown in Figure 7b, which shows that the spectrum gives a good qualitative match to the product ion spectrum from the serum studies (Figure 6c). The peptide ion at m/z 1093 was identified as the tryptic peptide NTLIYLDK with a score of 41 ($p < 0.033$, 10 matched peptides) derived from complement C3 component precursor-1 (CCCPP-1). Tandem LC-ESI-QIT MS/MS analysis of the biomarker ions at m/z 1505, 854 and 1444 did not yield significant sequence matches ($p < 0.05$) using Sequest or Mascot search engines. PSD identification was not undertaken due to the low intensities of these biomarker ions.

Immunoassay Quantitation of serum AAG

AAG was quantified by immunoassay for a selected subset stage IV and control patients' serum. The AAG concentrations (mean $\pm$ SE) measured in cancer patients and in the control group are shown in Figure 8a. The serum AAG concentrations were significantly increased ($p < 0.001$) in the stage IV cancer patients (1.28 ± 0.13 g/L) compared with the control

group (0.74 ± 0.06 g/L). A plot of ion intensity for the m/z 1161 and 1753 ions versus AAG levels (Figure 8b) shows that serum samples fell into two distinct clusters except for one sample which was misclassified on the basis of the immunoassay and the mass spectral data. This supports the inventors' observation that both biomarker ion intensities and AAG concentrations can distinguish stage IV melanoma patients from controls with high predictability.

Specificity of Fingerprints

Table 3: Ions derived from the ANN analysis of Mass Spectrometry data that can classify between Melanoma Stages. Ion mass, classifier performance and the class of melanoma in which the ion is present are shown.

Study	Stepwise ANNs modelling					
	Step	1	2	3	4	5
Stage 1 vs. Stage 2						
	Ion (m/z)	1978	1825	1731	1251	2053
	Classifier performance (% cases correct for unseen data)	94.70 %	97.10 %	96.50 %	97.60 %	98.40 %
	Class in which the ion is present	present in Stage 2	absent in Stage 2	present in Stage 2	absent in Stage 2	present in Stage 2
Stage 2 vs. Stage 3						
	Ions	861	903			
	Classifier performance (% cases correct for unseen data)	73.80 %	99.90 %			
	Class where the ion is present	absent in Stage 3	present in Stage 3			
Stage 3 vs. Stage 4						
	Ion (m/z)	877	903	1625	2754	2064
	Classifier performance (% cases correct for unseen data)	73.30 %	95.60 %	97.30 %	99.20 %	99.90 %
	Class where the ion is present	present in Stage 4	absent in Stage 4	present in Stage 4	absent in Stage 4	present in Stage 4

Table 4 Ions derived from the ANN analysis of Mass Spectrometry data that can classify between healthy control and Melanoma for each stage of Melanoma. Ion mass, classifier performance and the class in which the ion is present are shown.

Study	Stepwise ANNs modelling										
	Step	1	2	3	4	5	6	7	8	9	10
Stage 1 vs. Control	Ion (<i>m/z</i>)	980	3220	864	2966	2886	1299	2309	3489	3430	933
	Classifier performance (% cases correct for unseen data)	65.40 %	66.70 %	66.70 %	69.90 %	68.40 %	71.60 %	74.40 %	80.10 %	84.60 %	83.30 %
	Class where the ion is present	present in Control	absent in Control	present in Control	present in Control	absent in Control	absent in Control	present in Control	absent in Control	present in Control	absent in Control
Stage 2 vs. Control	Ion (<i>m/z</i>)	1251	1283	3443	3432	1968	1299	2244	2411	2468	2492
	Classifier performance (% cases correct for unseen data)	58.30 %	82.50 %	82.40 %	86.40 %	91 %	92 %	95.20 %	95 %	95.30 %	95 %
	Class where the ion is present	absent in Control	present in Control	present in Control	absent in Control	absent in Control	absent in Control	present in Control	absent in Control		

Study	Stepwise ANNs modelling										
	Step	1	2	3	4	5	6	7	8	9	10
Stage 3 vs. Control	Ion (<i>m/z</i>)	2624	2715	1754	1251	1285	2999	3161	1312	3326	1371
	Classifier performance (% cases correct for unseen data)	58.20 %	60.40 %	70 %	80.10 %	83.60 %	87.80 %	89 %	91.40 %	93.30 %	93.60 %
	Class where the ion is present	present in Control	present in Control	present in Control	absent in Control	present in Control	absent in Control	present in Control	absent in Control	absent in Control	present in Control
Stage 4 vs. Control	Ion (<i>m/z</i>)	1753	1161	1505	854	1444	1093				
	Classifier performance (% cases correct for unseen data)	80.7	96.3	97.7	99.6	99.5	99.4				
	Class where the ion is present	present in Control	present in Control	absent in Control	present in Control	absent in Control	absent in Control				

DISCUSSION

- Multiparametric diagnostic assay development holds tremendous promise for fulfilling accurate clinical molecular analyses in modern medicine. A diagnostic platform for

expression differences of serum proteins may prove to be an important variable in the understanding of difficult diagnostic objectives. The identification of the individual differentially expressed proteins that comprise the diagnostic expression profile is essential to facilitating real progress in the development of a robust accurate diagnostic platform, because reliable diagnosis of patients who are at increased risk of metastasizing melanoma disease remains challenging. The inventors have developed a reproducible and standardized integrated sample preparation and mass spectrometry-based proteomic protocol, combined with ANNs modeling, for protein and tryptic peptide biomarker discovery and identification in human serum samples. In order to find new serum markers of stage IV melanoma, which may later be validated and used in earlier disease stages, an integrated MALDI mass spectrometric approach combined with ANNs analysis and modeling was used for the identification of biomarker ions in serum from stage IV melanoma patients allowing discrimination of metastatic disease from healthy status with high specificities of 92 % for protein ions and 100 % for peptide biomarkers, respectively. An independent set of AJCC stage I melanoma serum samples were used to further validate the model. 98 % of these samples were correctly classified as non-stage IV samples, emphasizing the power of the newly defined biomarkers to identify patients with late stage metastatic disease. This contrasts with a reported 20 % false negative rate for S100B using an immunoluminometric assay (LIA-mat) in patients with detectable metastases identified by routine clinical and radiographic studies [4]. Sequence analysis of predictive peptides identified two peptide sequences belonging to AAG-1/2 and a third sequence associated with CCCPP-1.

Recently, there has been considerable controversy concerning the mass spectrometry-based proteomic profiling approach [20, 26, and 28]; including commentaries by several investigators and the analysis of publicly available data posted by Petricoin and co-workers [8]. Diamandis [28, 29] has raised concerns whether the SELDI/MALDI-based approaches are reproducible, whereas Sorace and Zahn [30] and Baggerly et al. [27] have raised study design bias concerns. Critics have also argued that reported discriminatory protein profiles have been based largely on experimental artifacts rather than biological differences [28, 30]. The inventors have designed this study carefully to incorporate criteria set out for clinical biomarker identification by proteomic based technologies outlined by Pepe et al. [31] and elaborated on in a recent review by Ransohoff [26].

The inventors contend that completion of rigorously designed, inclusive of platform reproducibility pre- and analytical variables standardized is the best approach to addressing these concerns, and indeed, none of the concerns being raised are unique to MALDI-TOF-MS profiling and should be important elements of all pre-validation and validation efforts.

5 Before the biomarker discovery phase of the study, the inventors implemented standardized and well-documented sample collection procedures and analytical (including specimen processing, storage, analysis and finally sample randomization throughout the processing and analytical procedures), instrumental and bioinformatic bias have been systematically studied and are discussed in the following sections.

10

Retrospective serum samples accumulated over many years are a valuable resource. However, with the labile nature of many proteins/peptides it is crucial to ensure rigorous care be taken with sample handling. Processing time, clotting times, centrifugation speeds, pre-aliquoting, storage temperature, and the number of freeze-thaw cycles are all critically

15 important as well variation in ZipTips fractionation, MALDI sample crystallization, laser irradiation, and serum sample preparation, as reported here, all influence protein/peptide pattern outcomes. The impact of preanalytical variables on the quality of routine clinical laboratory results is well documented [32-34]. In order to minimize these effects the inventors used standard protocols for sample collection from the same clinical centre into
20 uncoated gel phase tubes, samples were gently mixed and left to clot at room temperature for a defined time. The samples were then centrifuged at a set speed and time and the serum phase aliquoted into clear plastic tubes. The samples were stored at -20°C and, on arrival at the inventors' laboratory, they were thawed and aliquoted further into smaller volumes (50 µl) frozen and stored at -80°C until use. From the inventors' blood aging studies the
25 inventors observed a correlation between the increasing number of hour's incubation and the variation between the observed serum protein and peptide profiles. However, three freeze thaw cycles had a less pronounced effect as previously reported by other groups [8, 35]. Extensive MALDI-TOF MS and data analysis suggested negligible contributions from three freeze thaw cycles, however; in this work freeze thaw was limited to two cycles to
30 accomplish initial freezing and then subsequent aliquoting. the inventors' study and that reported by other groups [6, 34, 36] highlights the importance of precautions that need to be taken to avoid bias from sample processing which was reported recently to be the main peril to biomarker research [26]. The sample bias issue, which caused major controversy in a

previous study [8], was tackled by randomization of samples prior to the initial or analytical handling and this randomization was carried through the sample processing to the deposition on the target plate with appropriate blanks as outlined in Methods and Materials.

5 Systematic sample preparation and optimization using several clean-up methods for removal of high abundant protein were assessed including the use of ultrafiltration devices, precipitation using acetonitrile and C₄ and C₁₈ ZipTip fractionation prior to protein profiling and tryptic peptide investigation. The results indicated that the greatest number of ions and improved reproducibility was observed after a 1 in 10 dilution for protein profiling and after
10 C₁₈ ZipTip clean-up prior to and following, tryptic peptide analysis, which were adopted for all studies. Different sample depositing methods were also tested with the inventors' system and in all cases the dry droplet method gave the most uniform and reproducible peaks for both the protein and peptide profiles. Appropriate blanks were included in a randomized fashion and analysed alongside the serum samples. TFA blanks were included to ensure that
15 there was no contamination of the target plate from a previous acquisition, and serum blanks without the addition of trypsin was included to ensure tryptic enzyme activity. Finally BSA controls were also taken through the tryptic digestion procedure to check efficiency of the digestion and peptide acquisition. These quality assurance (QA) samples were included in all protocols ensuring a robust and reproducible methodology.

20

Instrument acquisition parameters were also tested and optimized separately for protein and tryptic peptide profiling. Surprisingly with the inventors' instrumentation a raster mode for protein profiling resulted in the highest signal-to-noise ratios whereas in reflectron mode the auto-quality mode resulted in the highest intensities. Other instrumental settings such as laser
25 irradiation and automated acquisition modes were optimized to give a robust reproducible method.

Ultimately, classification by MALDI-TOF-MS relies on spectral differences in position and amplitude of mass spectral peaks. Since the reproducibility of quantitation, resolution and
30 mass accuracy of the MALDI-TOF-MS, or any high throughput mass spectrometric technique, has received little attention and the clinical usefulness of the approach has been questioned. The inventors provide evidence in this application of the reproducibility of MS measurements Mean coefficient of variation (CV's) for normalized intensity ratio (NIR)

were 17.2 % for proteins and 24.9 % for tryptic peptides in linear and reflectron MALDI-MS modes respectively. de Noo et al., [35] found similar CVs as reported by the inventors with 6/8 samples giving between 20 % and 29 % using MALDI-TOF-MS for proteins/peptides. The ion intensity variations for the principle biomarker ion identification in this work are
5 larger compared to mass spectrometry reproducibility. These data establish that melanoma stage assignment is not an artifact of the normal intensity variation obtained in MALDI mass spectrometry.

Poor or infrequent calibration can lead to a significant shift in the m/z values for mass
10 spectral peaks. the inventors' protocol uses close external calibration in both linear and reflectron modes, in which data are acquired from a calibration mixture spotted adjacent to each group of four sample spots. Mass measurement reproducibility was therefore maintained well within the instrumental specifications with mean CVs of 0.1 % and 0.01 % for protein and peptides respectively. Prior to bioinformatic analysis acceptance criteria
15 based on visual inspection of mass spectrometric, serum and calibration profiles were assessed and those deemed to have a poor signal were removed from bioinformatic data assessment.

Recent reports [27, 30] have documented the inherent problems of biological data mining of
20 MALDI-TOF signatures and highlighted the need for a systematically reproducible and validated bioinformatic tool for using biomarker expression patterns as predictive indicators. The standard protocol for biomarker identification is to analyze a training set of data and then to assess the predictive power of the models on an independent validation set of cases. In this study the inventors expand upon traditional "leave-one-out" methods [37] by using
25 repeated and extensive random sample cross validation to assess model performance and the validity of the biomarker ions identified from these models. This repeated resampling therefore determines the best subset of predictive biomarker ions that are consistently presented as being discriminatory over a large number of models. Furthermore the application of blinded validation and test data in the modeling approach produces models
30 with good performance on blind data and prevents overfitting. Analysis of these models has identified biomarker ions appropriate for classification of the wider (out of sample) population which are therefore strong representative markers of the presence of late stage melanoma. These have been confirmed by subsequent validation exercises. The probability

of this being a random event is 1.3×10^{-7} , based on the probability of determining two ions from the same fragment within a diagnostic profile containing 2700 variables in two separate analysis steps. Given the relevance of the identified ions to late stage melanoma the probability of identifying this purely by chance is much lower than this but is not easily
5 calculated. One would have to determine the likelihood of a melanoma related protein being identified from all of the protein components of the serum. Data mining by ANNs employed a step-wise analytical approach in the present work which also allowed for the identification of biomarker ion masses with higher predictive performance than in the inventors' previous SELDI analysis [9] which used larger mass ranges to identify regions of the profile which
10 were important in discriminating between groups.

Several reports on detection of potential melanoma serum biomarkers have used SELDI or MALDI-MS, mostly without bioinformatic analysis [6, 38-40]. Two studies reported the importance of protein ions in ranges m/z 10000 – 30000 and m/z 2500-3000. The latter
15 showed significant variations, with some ions increasing in intensity from stage I to stage II melanoma, but absent in controls [38, 39]. Due to the small sample numbers ($n = 17$) used by Ferrari et al., [38] no statistical analysis was performed. Regazzi et al., [39] reported that analysis of the relative abundances of protein ions, according to three clustering criteria, did not allow them to separate areas of interest which distinguished patients from controls.
20 Recently the same group [40] reported the presence of a small number of native peptide ions in stage III melanoma samples versus controls; however, the study did not include data mining/identification. A recent study [6] by the inventors' group using SELDI-MS analysis using a different co-hort of patient samples reported that a region around 11,700 Da was significant in differentiating between stage I and stage IV metastatic melanoma. In this work,
25 the inventors report a rapid and higher resolution MALDI-TOF-MS analysis that has allowed serum protein and peptide signatures from control and stage IV melanoma patients to be identified. Protein profiling data showed an ion at m/z 12000 ± 24 to be the closest in mass to the spectral region identified in the previously reported study [6]. There is no overlap with the m/z values reported in the above studies on proteins, native peptides and the
30 data for tryptic peptides presented here; differences in sample preparation, MS and bioinformatic analysis play a critical role in data evaluation, and therefore it is not surprising that the studies detailed failed to identify common biomarkers. The protein(s) associated with these ions were not identified in any of these studies including the inventors' study, due

to the inherent technical limitations in protein biomarker ion identification derived from SELDI and MALDI, but the tryptic peptide protocol reported here overcomes this limitation by allowing peptide sequences to be identified and hence the parent protein.

- 5 Protein identification from proteolytic digests has been reported for human serum proteome in pancreatic and ovarian cancers [17-19]. Mass spectral quality was improved, in terms of the S/N ratio, for data acquired for tryptic serum peptides in comparison with native serum peptides/proteins [19], consistent with the inventors' observations. Here, ANN analysis identified six tryptic peptide ions capable of distinguishing melanoma stage IV sera from
10 control sera with a median accuracy of 100 %, and these peptide biomarker ions may be used for screening on their own, or in combination with protein biomarker ions for improved confidence.

- Sequence analysis of stage IV melanoma tryptic peptide biomarker ions was achieved using
15 tandem mass spectrometric techniques and the sequences of the tryptic ions at m/z 1753 and 1161 were identified as YVGGQEHFALLLR and WFYIASAFR respectively, derived from AAG-1/2. Interestingly, the same m/z 1753 AAG-1/2 tryptic peptide was identified by Koomen *et al.*, [19] in plasma from pancreatic cancer patients. In the inventors' study the inventors have further demonstrated that the mean serum AAG concentrations in stage IV
20 melanoma cancer were significantly increased compared with the healthy population. Serum AAG levels are shown to be increased in inflammatory and lymphoproliferative disorders and cancer [41-43]. Duche *et al.*, [42] and Bleasby *et al.*, [43] demonstrated that AAG is also up-regulated in plasma samples from breast, ovarian and lung cancer patients analyzed by an immunonephelometric method leading to the assumption that AAG might be produced
25 by cancer cells themselves. MALDI-TOF mass spectrometry has shown the up-regulation of AAG in breast cancer patients by analysis of proteins from nipple aspirate fluid [44]. AAG, a highly heterogeneous glycoprotein, is an acute-phase protein produced mainly in the liver [45], but extrahepatic synthesis has also been reported [46, 47], however, its physiological significance is not fully understood. Confirmation of extrahepatic production of AAG has
30 been reported [47] recently in endothelial cells of blood vessels within or adjacent to bladder tumour tissue. Moreover, AAG was also shown to be present in the urine of patients and particularly in those with invasive bladder cancer. These authors concluded that the endogenous production of AAG may be related to angiogenic activity and marked the switch

from a non-vascularized to a vascularized progression of the bladder tumour. In this regard, it is widely accepted that, during cancer progression, tumours not only require a blood supply for growth, but also use this vehicle for metastatic dissemination [48, 49]. It is therefore conceivable that production of AAG by endothelial cells may be an important factor in promoting neoangiogenesis in human metastatic melanoma progression as well, but the significance of AAG-1/2 production in other stages of melanoma will require investigation. It is also possible that clinical stage of the cancer may affect AAG levels and its variants as demonstrated by Duche *et al* [42] in breast, ovarian and lung cancers. the inventors' sequence analysis led to identification of two peptides derived from AAG-1/2 and the inventors have shown that the intensities of these ions correlate with total AAG increases and disease state. However, it is possible that the identified disease-specific peptides represent degradation fragments derived from these proteins as a result of enzymatic or other activity. Support for this suggestion comes from studies by Villanueva *et al.*, [50] who postulated that fragments of endogenous blood proteins generated *ex vivo* serve as a substrate pool for disease-specific proteinases that arise from the tumour itself or within the tumour microenvironment [51]. The resultant peptide signatures are therefore composed of what could be considered surrogate markers for the detection and classification of certain types of cancer.

The sequence of the tryptic peptide ion at m/z 1093 was identified as NTLIIYLDK derived from CCCPP-1, an abundant serum protein produced within the kidney [52] that may be an important mediator of local inflammatory and immunological injury [53]. The complement system is also important in immuno-surveillance against tumours. The complement system is a complex pathway which through a variety of proteolytic cleavages of C3 and C4 creates fragments which bind to the cell surface. Subsequently the proteolytic cleavage leads to further fragments which remain attached to cell surface and thereafter serve as ligands for receptors on phagocytic and natural killer (NK) cells. Opsonization of target cells with these C3 fragments promotes and enhances antibody-dependent and complement-dependent cytotoxicity, two systems paramount in the elimination of neoplastic cells [54, 55]. Chang *et al.*, [55] identified a marker complement 3 f-fragment (C3f) in nasopharyngeal cancer which is generated by iC3b and released into plasma during C3 inactivation. They therefore, concluded that this may represent an important immune evasion mechanism of tumour cells

through the cleavage of C3b protein from the cell surface causing an inhibition of the complement cascade.

Acute phase proteins have been previously reported in proteomics-based experiments using serum or plasma because they represent proteins present in high abundance. Most of these proteins, including albumin, transthyretin, lipoproteins, c-reactive proteins, among others, are synthesized in the liver and are regarded as non-cancer biomarkers or epiphenomena of tumours resulting from a cascade of inflammatory signals [28, 30, 56]. However, Fung et al. [57] recently re-evaluated the role of transthyretin and inter-alpha trypsin inhibitor heavy chain 4 (ITIH4) and found a high degree of post-translational modifications, including proteolytic truncation, cysteinylolation, and glutathionylation, in sera of patients with ovarian, breast, prostate, and colon cancer. These modified proteins allowed the differentiation of cancer sera from non-cancer sera and the differential diagnosis of tumour type, suggesting that modified 'acute phase' cancer markers, could be considered as diagnostic biomarkers for cancers. Furthermore, the predictive biomarker ions identified in this work are derived from at least two proteins, so disease state prediction using the full set of biomarker ions relies on complex, interrelated changes in the proteome. These complex changes may be tumour/disease specific, unlike changes in a single acute phase protein. Further studies are needed to establish whether there are distinct similarities and/or differences in biomarker ions for different cancers.

In conclusion, the inventors' data shows that an integrated approach to protein and tryptic peptide profiling, with robust, standardized and reproducible pre- and post-analytical protocols using MALDI-MS combined with ANNs, leads to high diagnostic specificity (0 % false negatives for the combined approach) and sensitivity in patients with stage IV melanoma compared with conventional immunoassays. The advantage of this novel methodology is the ability to highlight variations in the peptide mass fingerprint between cancerous and control patients using ANNs and to target m/z values selectively for sequence identification. Quantitative analysis confirms that an increase in AAG correlates with stage IV melanoma compared with control patients. Further studies are in progress to investigate the significance of AAG-1/2, CCCPP-1 precursor ions and serum levels of AAG in stages I to III melanoma. These studies will be crucial to verify whether the detection of these proteins can be used as biomarkers for disease progression and overall survival. The

inventors envision that this methodology has the potential to be transferred to the clinical setting for high throughput detection of targeted protein and peptide biomarker ions and sequences using MALDI-TOF-MS informed by ANNs analysis.

5 REFERENCES:

- [1] Balch, C. M., Buzaid, A. C., Soong, S. J., Atkins, M. B. *et al.*, Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol* 2001, 19, 3635-48.
- 10 [2] Ben-Porat, L., Panageas, K. S., Hanlon, C., Patel, A. *et al.*, Estimates of stage-specific survival are altered by changes in the 2002 American Joint Committee on Cancer staging system for melanoma. *Cancer* 2006, 106, 163-71.
- [3] Utikal, J., Schadendorf, D., Ugurel, S. Serological and immunohistochemical
15 prognostic biomarkers of cutaneous malignancies. *Arch Dermatol Res.* 2007 Jan 13, [Epub ahead of print].
- [4] Djukanovic, D., Hofmann, U., Sucker, A., Rittgen, W. *et al.*, Comparison of S100 protein and MIA protein as serum marker for malignant melanoma. *Anticancer Res*
20 2000, 20, 2203-7.
- [5] Miliotes, G., Lyman, G. H., Cruse, C. W., Puleo, C. *et al.*, Evaluation of new putative tumor markers for melanoma. *Ann Surg Oncol* 1996, 3, 558-63.
- 25 [6] Mian, S., Ugurel, S., Parkinson, E., Schlenzka, I. *et al.*, Serum proteomic fingerprinting discriminates between clinical stages and predicts disease progression in melanoma patients. *J Clin Onco* 2005, 23, 5088-93.
- [7] Hanash, S. Disease proteomics. *Nature* 2003, 422, 226-32.
- 30 [8] Petricoin, E. F., Ardekani, A.M., Hitt, B.A., Levine, P.J. *et al.*, Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002, 16, 572-7.

- [9] Hutchens, T. W., Yip, T. T. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* 1993, 7, 576-580.
- 5 [10] Koomen, J. M., Li, D., Xiao, L. C., Liu, T. C. *et al.*, Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. *J Proteome Res* 2005, 4, 972-81.
- 10 [11] de Noo, M. E., Mertens, B. J., Ozalp, A., Bladergroen, M. R. *et al.*, Detection of colorectal cancer using MALDI-TOF serum protein profiling. *Eur J Cancer* 2006, 42, 1068-76.
- [12] Traub, F., Feist, H., Kreipe, H. H., Pich, A. SELDI-MS-based expression profiling of ductal invasive and lobular invasive human breast carcinomas. *Pathol Res Pract* 2005, 201, 763-70.
- 15 [13] Mian, S., Ball, G., Hornbuckle, J., Holding, F. *et al.*, A prototype methodology combining surface-enhanced laser desorption/ionization protein chip technology and artificial neural network algorithms to predict the chemoresponsiveness of breast cancer cell lines exposed to Paclitaxel and Doxorubicin under in vitro conditions. *Proteomics* 2003, 3, 1725-37.
- 20 [14] Finnskog, D., Jaras, K., Ressine, A., Malm, J. *et al.*, High-speed biomarker identification utilizing porous silicon nanovial arrays and MALDI-TOF mass spectrometry. *Electrophoresis* 2006, 27, 1093-103.
- 25 [15] Adam, B. L., Qu, Y., Davis, J. W., Ward, M. D. *et al.*, Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002, 62, 3609-14.
- 30 [16] Ball, G., Mian, S., Holding, F., Allibone, R. O. *et al.*, An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics* 2002, 18, 395-404.

[17] Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J. *et al.*, Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* 2002, 1, 947-55.

5

[18] Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J. *et al.*, Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics* 2003, 2, 1096-103.

10

[19] Koomen, J. M., Zhao, H., Li, D., Abbruzzese, J., Baggerly, K. *et al.*, Diagnostic protein discovery using proteolytic peptide targeting and identification. *Rapid Commun Mass Spectrom* 2004, 18, 2537-48.

15

[20] Check, E. Proteomics and cancer: running before we can walk? *Nature* 2004, 429, 496-497.

[21] Veltri, R. W., Miller, M. C., An, G. Standardization, analytical validation, and quality control of intermediate endpoint biomarkers. *Urology* 2001, 57, 164-70.

20

[22] Parekattil, S. J., Fisher, H. A., Kogan, B. A. Neural network using combined urine nuclear matrix protein-22, monocyte chemoattractant protein-1 and urinary intercellular adhesion molecule-1 to detect bladder cancer. *J Urol* 2003, 169, 917-20.

25

[23] Ringner, M., Peterson, C. Microarray-based cancer diagnosis with artificial neural networks. *Biotechniques Suppl* 2003, 30-5.

30

[24] Lancashire, L., Schmid, O., Shah, H., Ball, G. Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. *Bioinformatics* 2005, 15, 2191-9.

- [25] Rumelhart, D.E., McClelland, J. L. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, The MIT Press, Cambridge 1986.
- 5 [26] Ransohoff, D. Lessons from controversy: ovarian cancer screening and serum proteomics. *J Natl Cancer Inst* 2005, 97, 315-9.
- [27] Baggerly, K. A., Morris, J. S., Coombes, K. R. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics*
10 2004, 22, 777-85.
- [28] Diamandis, E. P., van der Merwe, D. E. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. *Clin Cancer Res* 2005, 11, 963-5.
- 15 [29] Diamandis, E. P. Proteomic patterns to identify ovarian cancer: 3 years on. *Expert Rev Mol Diagn* 2004, 4, 575-7.
- [30] Sorace, J. M., Zhan, M. A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* 2003, 9, 24.
- 20 [31] Pepe, M.S., Etzioni, R., Feng, Z., Potter, J.D. *et al.*, Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001, 93, 1054-61.
- [32] Dugue, B., Leppanen, E., Grasbeck, R. Preanalytical factors (biological variation) and the measurement of serum soluble intercellular adhesion molecule-1 in humans: influence of the time of day, food intake, and physical and psychological stress. *Clin Chem* 1999 45, 1543-7.
- 25 [33] Ulmert, D., Becker, C., Nilsson, J.A., Piironen, T. *et al.*, Reproducibility and accuracy of measurements of free and total prostate-specific antigen in serum vs plasma after long-term storage at -20 degrees C. *Clin Chem* 2006 52, 235-9.
- 30

- [34] Banks, R.E., Stanley, A.J., Cairns, D.A., Barrett, J.H. *et al.*, Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin Chem* 2005 51, 1637-49.
- 5 [35] de Noo, M. E., Tollenaar, R. A., Ozalp, A., Kuppen, P. J. *et al.*, Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry. *Anal Chem* 2005, 77, 7232-41.
- 10 [36] Semmes, O. J., Feng, Z., Adam, B. L., Banez, L. L., *et al.*, Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* 2005 51, 102-12.
- 15 [37] Michiels, S., Koscielny, S., Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005, 365, 488-492.
- [38] Ferrari, L., Seraglia, R., Rossi, C. R., Bertazzo, A. *et al.*, Protein profiles in sera of patients with malignant cutaneous melanoma. *Rapid Commun Mass Spectrom* 2000, 14, 1149-54.
- 20 [39] Ragazzi, E., Vogliardi, S., Allegri, G., Costa, C. V. L. *et al.*, Cluster analysis of serum proteins in malignant cutaneous melanoma: search for disease markers. *Rapid Commun Mass Spectrom* 2003, 17, 1511-1515.
- 25 [40] Seraglia, R., Vogliardi, S., Allegri, G., Comai, S. *et al.*, Search for melanoma markers in plasma and serum samples. *Eur J Mass Spectrom* 2005, 11, 353-60.
- [41] Kremmer, T., Szollosi, E., Boldizsar, M., Vincze, B. *et al.*, Liquid chromatographic and mass spectrometric analysis of human serum acid alpha-1-glycoprotein. *Biomed Chromatogr* 2004, 18, 323-9.
- 30 [42] Duche, J. C., Urien, S., Simon, N., Malaurie, E. *et al.*, Expression of the genetic variants of human alpha-1-acid glycoprotein in cancer. *Clin Biochem* 2000, 33, 197-202.

- [43] Bleasby, A. J., Knowles, J. C., Cooke, N. J. Microheterogeneity of alpha 1-acid glycoprotein: lack of discrimination between benign and malignant inflammatory disease of the lung. *Clin Chim Acta* 1985, 150, 231-235.
- 5
- [44] Alexander, H., Stegner, A. L., Wagner-Mann, C., Du Bois, G. C. *et al.*, Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. *Clin Cancer Res* 2004, 10, 7500-7510.
- 10
- [45] Sarcione, E. J. Hepatic synthesis and secretory release of plasma alpha-2 (acute phase)-globulin appearing in malignancy. *Cancer Res* 1967, 27, 2025-2033.
- [46] Sorensson, J., Matejka, G. L., Ohlson, M., Haraldsson, B. Human endothelial cells produce orosomucoid, an important component of the capillary barrier. *Am J Physiol*
- 15
- 1999, 276, H530-4.
- [47] Irmak, S., Tilki, D., Heukeshoven, J., Oliveira-Ferrer, L. *et al.*, Stage-dependent increase of orosomucoid and zinc-alpha2-glycoprotein in urinary bladder cancer. *Proteomics* 2005, 5, 4296-4304.
- 20
- [48] Carmeliet, P. Angiogenesis in health and disease. *Nat. Med* 2003, 9, 653-660.
- [49] Jain, R. K. Normalizing tumor vasculature with anti-angiogenic therapy: A new paradigm for combination therapy. *Nat. Med* 2001, 7, 987-989.
- 25
- [50] Villanueva, J., Shaffer, D. R., Philip, J., Chaparro, C. A. *et al.*, Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 2006, 116, 271-284.
- 30
- [51] Liotta, L. A., Petricoin, E. F. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J Clin Invest* 2006, 116, 26-30.

[52] Warner, N. B. *The complement system, Mechanisms of activation and use as a diagnostic tool*, Beckman Coulter Inc. 1998, C-1, pp.10-11.

5 [53] Tang, S., Zhou, W., Sheerin, N. S., Vaughan, R. W. *et al.*, The complement system, mechanisms of activation and use as a diagnostic tool. *J Immunol* 1999, 162, 4336-4341.

[54] Cole, D.S., Morgan, B.P. Beyond lysis: how complement influences cell fate. *Clin Sci (Lond)* 2003, 104, 455-66.

10

[55] Chang, J.T., Chen, L.C., Wei, S.Y., Chen, Y.J. *et al.*, Increase diagnostic efficacy by combined use of fingerprint markers in mass spectrometry-plasma peptidomes from nasopharyngeal cancer patients for example. *Clin Biochem* 2006, 39, 1144-51.

15

[56] Sheth, K., Bankey, P. The liver as an immune organ. *Curr Opin Crit Care* 2001, 7, 99-104.

20

[57] Fung, E. T., Yip, T. T., Lomas, L., Wang, Z. *et al.*, Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. *Int J Cancer* 2005, 115, 783-789.

CLAIMS

1. A mass spectrometry melanoma fingerprint for identifying the presence of a melanoma and/or for identifying the stage of a melanoma, the fingerprint comprising:

(a) peaks derived from the following ions:

- (i) ions derived from a protein, the ions having (m/z) values of: 12000, 14847, 1649, 15477, 13255, 3031 and 4791 ($\pm 0.4\%$) and optionally one or more the following additional ions, the additional ions having (m/z) values of: 9913, 4835, 4155, 4565 and/or 4673 ($\pm 0.4\%$);
- (ii) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 1753, 1161, 1505 and 854 ($\pm 0.02\%$) and optionally one or more of the following additional ions, the additional ions having (m/z) values of: 1444 and/or 1093 ($\pm 0.02\%$);

and/or

(b) peaks derived from the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 2624, 2715, 1753, 1251, 1285, 2999, 3161, 1312, 3326 and 1371 ($\pm 0.02\%$);

and/or

(c) peaks derived from the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 1251, 1283, 3443, 3432, 1968, 1299, 2244, 2411, 2468 and 2492 ($\pm 0.02\%$);

and/or

(d) peaks derived from the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 980, 3220, 864, 2966, 2886, 1299, 2309, 3489, 3430 and 933 ($\pm 0.02\%$);

and/or

(e) peaks derived from the following ions:

- (i) ions derived from a tryptically digested peptide, the ions having (m/z) values selected from: 1978 and 1825 ($\pm 0.02\%$) and optionally one or more the following additional ions, the additional ions having (m/z) values of: 1731, 1251 and/or 2053 ($\pm 0.02\%$);

and/or

(f) peaks derived from the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 861 and 903 ($\pm 0.02\%$);

and/or

- (g) peaks derived from one or more of the following ions:

- (i) ions derived from a proteolytically digested peptide, the ions having (m/z) values of: 877, 903 and 1625 ($\pm 0.02\%$) and optionally one or more the following additional ions, the additional ions having (m/z) values of: 2754, and/or 2064 ($\pm 0.02\%$);

and/or

- (h) peaks derived from one or more of the following ions:

- (i) ions derived from a protein, the ions having (m/z) values of 1161 and/or 1753 ($\pm 0.02\%$).

2. A mass spectrometry melanoma fingerprint according to claim 1 comprising all of the peaks of (a)(i) and (a)(ii).
3. A mass spectrometry melanoma fingerprint according to claim 1 or 2 in which the fingerprint comprises all of the peaks listed in (a), (b), (c), (d), (e), (f), (g), or (h).
4. A method of identifying a subject having a melanoma, the method comprising identifying whether a test sample obtained from a test subject suspected of having melanoma, when analysed by mass spectrometry, shows a fingerprint as described in claim 1, 2 or 3.
5. A method according to claim 4 further comprising obtaining a fingerprint from a control sample, and comparing the fingerprint from the control sample with the fingerprint from the test sample, wherein identification of an elevation or depression in one or more of the ions of the fingerprint to determine whether the test subject has melanoma.
6. A method according to any preceding claim in which:
 - (a) the fingerprint of (a) is used to distinguish between Stage IV melanoma and non-melanoma;

- (b) the fingerprint of (b) is used to distinguish between Stage III melanoma and non-melanoma;
 - (c) the fingerprint of (c) is used to distinguish between Stage II melanoma and non-melanoma;
 - (d) the fingerprint of (d) is used to distinguish between Stage I melanoma and non-melanoma;
 - (e) the fingerprint of (e) is used to distinguish between Stage I melanoma and Stage II melanoma;
 - (f) the fingerprint of (f) is used to distinguish between Stage II melanoma and Stage III melanoma;
 - (g) the fingerprint of (g) is used to distinguish between Stage II melanoma and Stage III melanoma;
 - (h) the fingerprint of (h) is used to distinguish between control and Stage IV melanoma.
7. A method according to any of claims 4 to 6 in which the control sample is obtained from a subject not having melanoma.
8. A method according to any of claims 4 to 7 in which the control sample is obtained from a subject having melanoma.
9. A method according to any of claims 4 to 8 in which the control sample is obtained from a subject having one of the following stages of melanoma: Stage 0, Stage IA, Stage IB, Stage IIA, Stage IIB, Stage IIC, Stage III and Stage IV.
10. A method of identifying a subject having a melanoma comprising:
- (a) determining the level of Alpha 1-Acid Glycoprotein or a fragment thereof in a sample obtained from a test subject suspected of having a melanoma; and
 - (b) comparing the level of Alpha 1-Acid Glycoprotein or fragment thereof in the test sample with the level of Alpha 1-Acid Glycoprotein or fragment thereof in a sample from a control subject.
11. A method according to claim 10 in which the control sample is obtained from a subject not having melanoma.

12. A method according to claim 10 or 11 in which the control sample is obtained from a subject having Stage IV melanoma.
13. A method according to any of claims 10 to 12 in which the fragment of Alpha 1-Acid Glycoprotein has an ion having an (m/z) of 1161 or 1753 ($\pm 0.02\%$).
14. A method according to any one of claims 10 to 13 further comprising comparing the relative level of Lactate Dehydrogenase in the test sample and the control sample to identify whether the melanoma of the test sample is a Stage I, Stage II, Stage III, Stage IV, Stage IV M1a, Stage IV M1b and Stage IV M1c.
15. A kit for use in a method of detecting melanoma, the kit comprising positive controls for mass spectrometry analysis in which the kit comprises:
 - (a) a control sample representative of a subject not having melanoma;and/or
 - (b) a control sample representative of a subject having melanoma.
16. A kit according to claim 15 in which the control sample representative of a subject not having melanoma comprises proteins and/or proteolytically digested peptides which when analysed by mass spectrometry generate two or more ions as described in (a), (b), (c), (d), (e), (f), (g) or (h) of claim 1 or one or more ions described in claim 13.
17. A kit according to claim 16 in which:
 - (a) one or more of the ions having an (m/z) of 2225, 1753, 1754, 1161 and 1444 is derived from Alpha-Acid Glycoprotein 1 Precursor;
 - (b) the ion having an (m/z) of 1251 is derived from Kininogen-1 Precursor;
 - (c) the ion having an (m/z) of 1283 is derived from Apolipoprotein A-I Precursor;
 - (d) ;and/or
 - (e) the ion having an (m/z) of 1093 is derived from Complement C3 precursor.
18. A method of identifying a drug for treating melanoma comprising:

- (a) exposing a test sample from a subject having melanoma to a candidate compound;
 - (b) subjecting the exposed test sample from (a) to mass spectrometry analysis to generate a fingerprint comprising one or more peaks;
 - (c) comparing the fingerprint obtained from the exposed test sample with a fingerprint obtained from a control sample in which the control sample:
 - (i) is obtained from a subject having melanoma; and
 - (ii) has not been exposed to the candidate compound
 - (d) determining whether the candidate compound decreases or elevates the level of one or more of the peaks of the fingerprint of the test sample relative to the fingerprint of the control sample.
19. A method according to claim 18 further comprising comparing the fingerprint of the test sample with a fingerprint obtained from one or more further control samples in which the one or more further control samples is obtained from:
- (a) a subject not having melanoma; and/or
 - (b) a subject having Stage IV melanoma
20. A method according to claim 18 or 19 in which the fingerprint comprises one or more peaks derived from, in order, one or more of the following ions: an ion derived from a protein, the ion having an (m/z) of 1161 and 1753 (± 0.4)
21. A computer system having a processor, a memory and an input for receiving data from a mass spectrometer wherein the memory comprises data corresponding to a fingerprint according to claim 1, 2 or 3 and the computer configured to compare data from the input against data corresponding to a fingerprint and to identify whether the inputted data is indicative of a melanoma based on the results of the comparison.
22. A computer system according to claim 21 in which the fingerprint is for Stage IV melanoma.
23. A computer program product comprising instructions which when run on a computer with a processor, memory and input configure it to the computer of claim 21 or 22.

24. Melanoma identification apparatus comprising the computer system of claim 21 or 22 and/or a computer program product according to claim 23 and a mass spectrometer connected to the input.

Figure 1

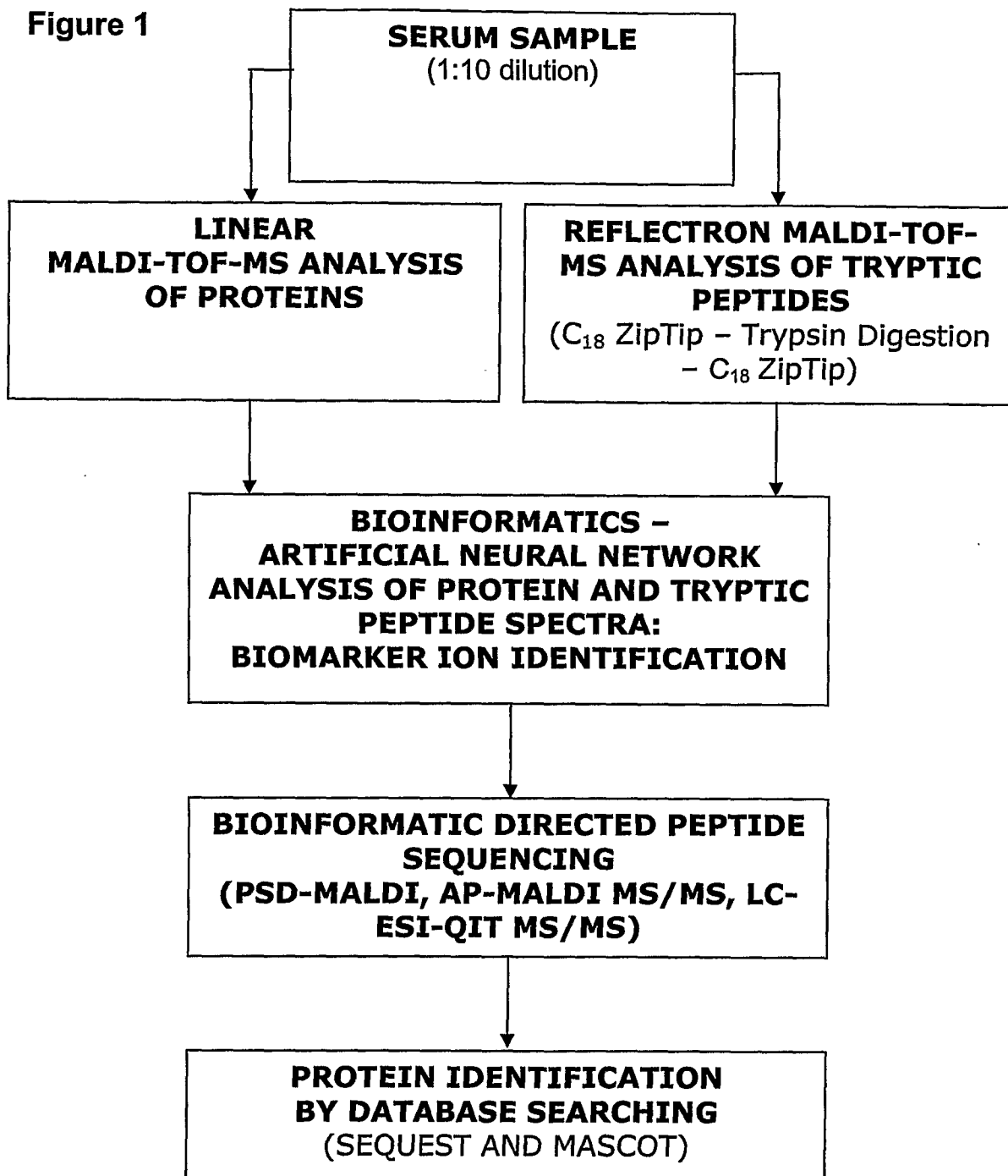


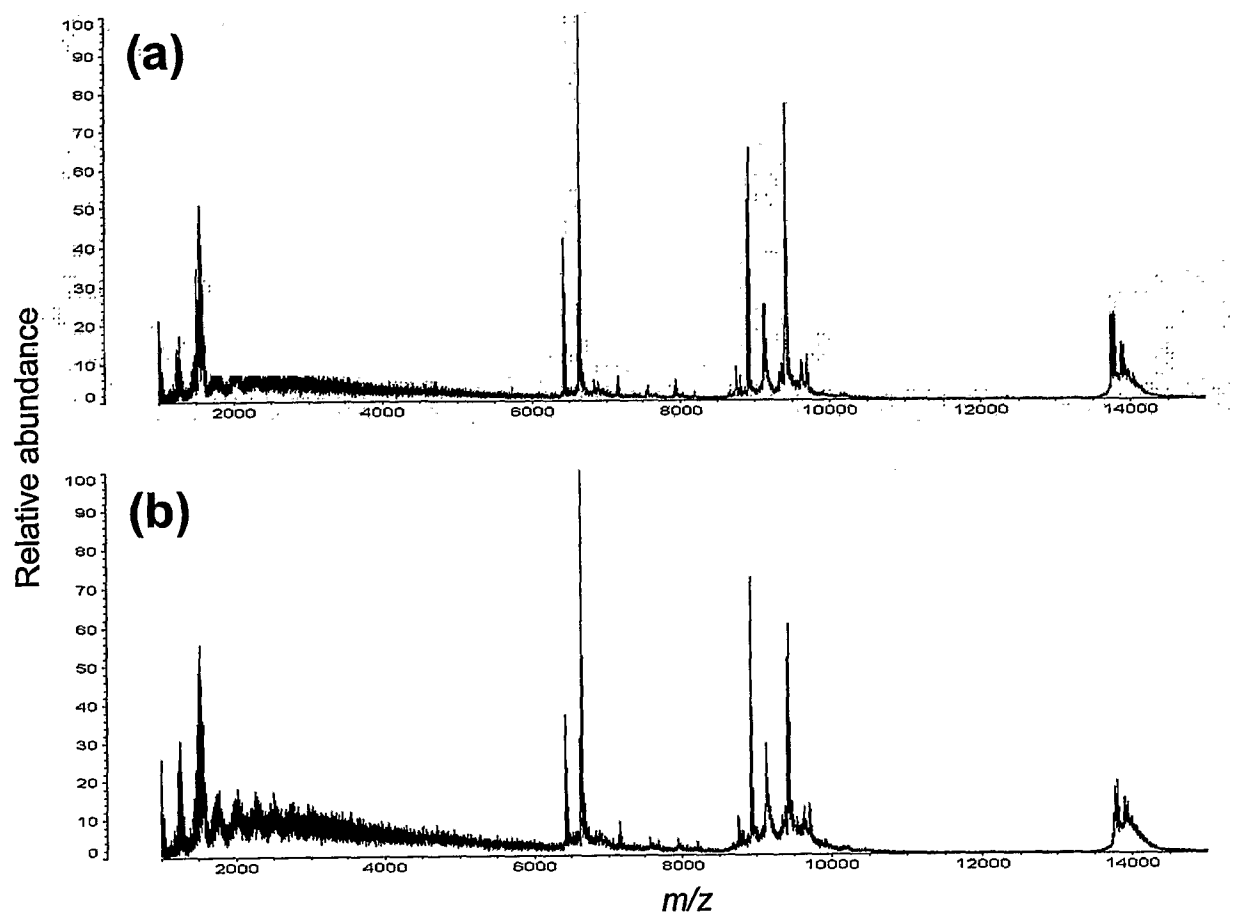
Figure 2.

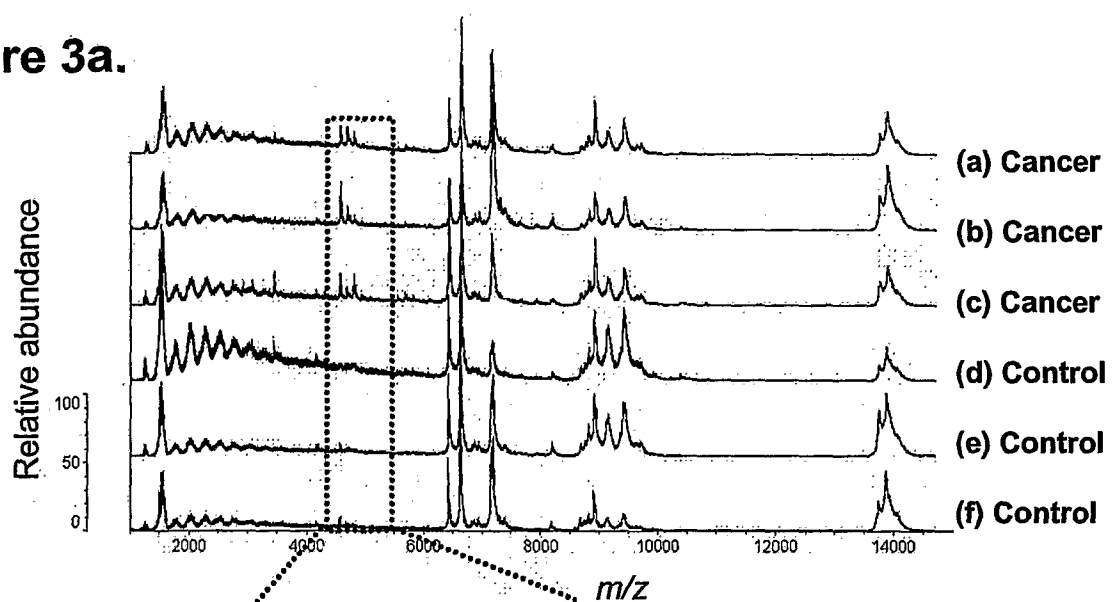
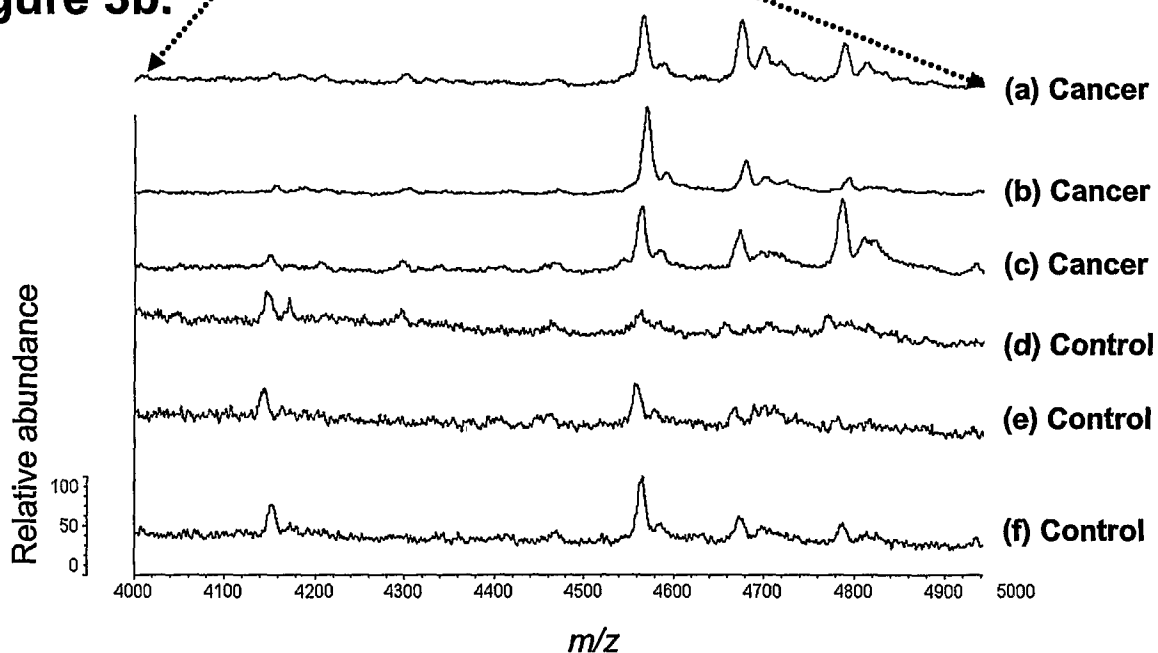
Figure 3a.**Figure 3b.**

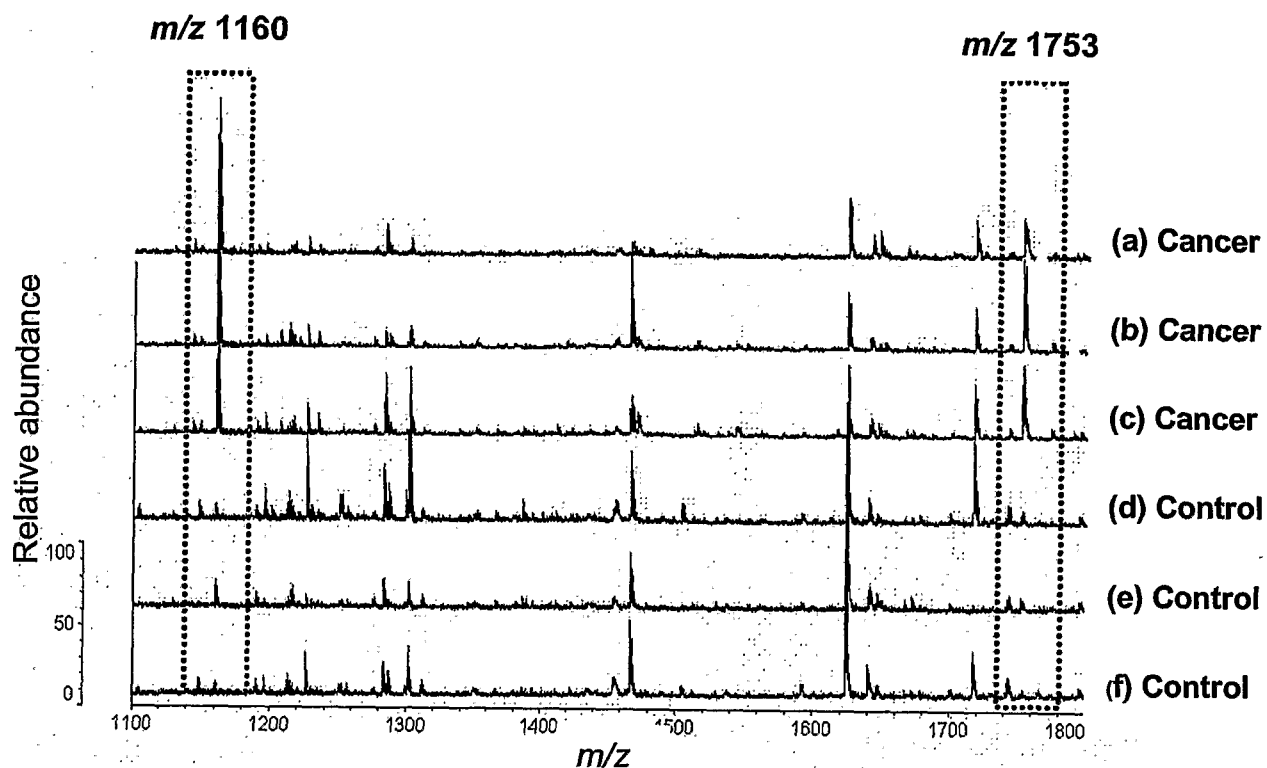
Figure 4.

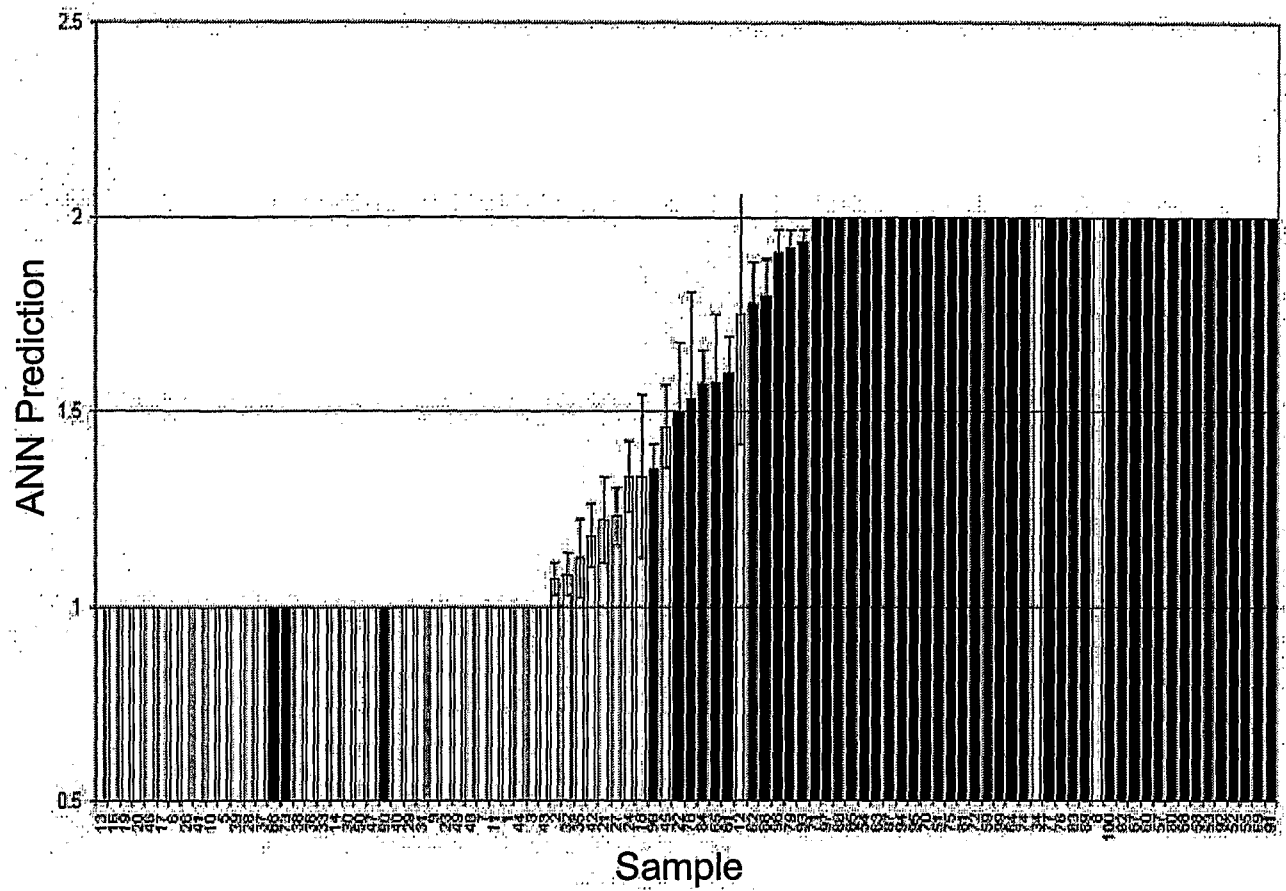
Figure 5a.

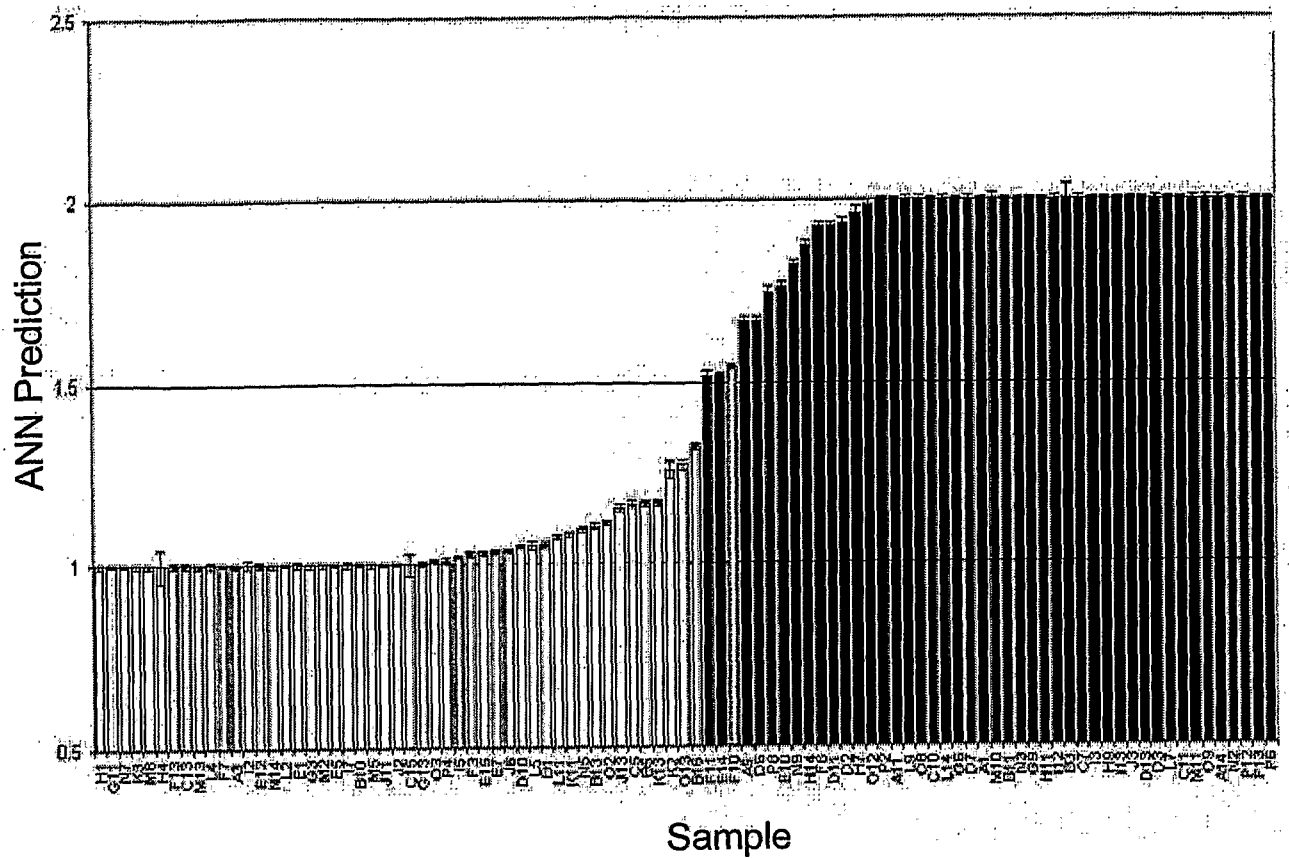
Figure 5b.

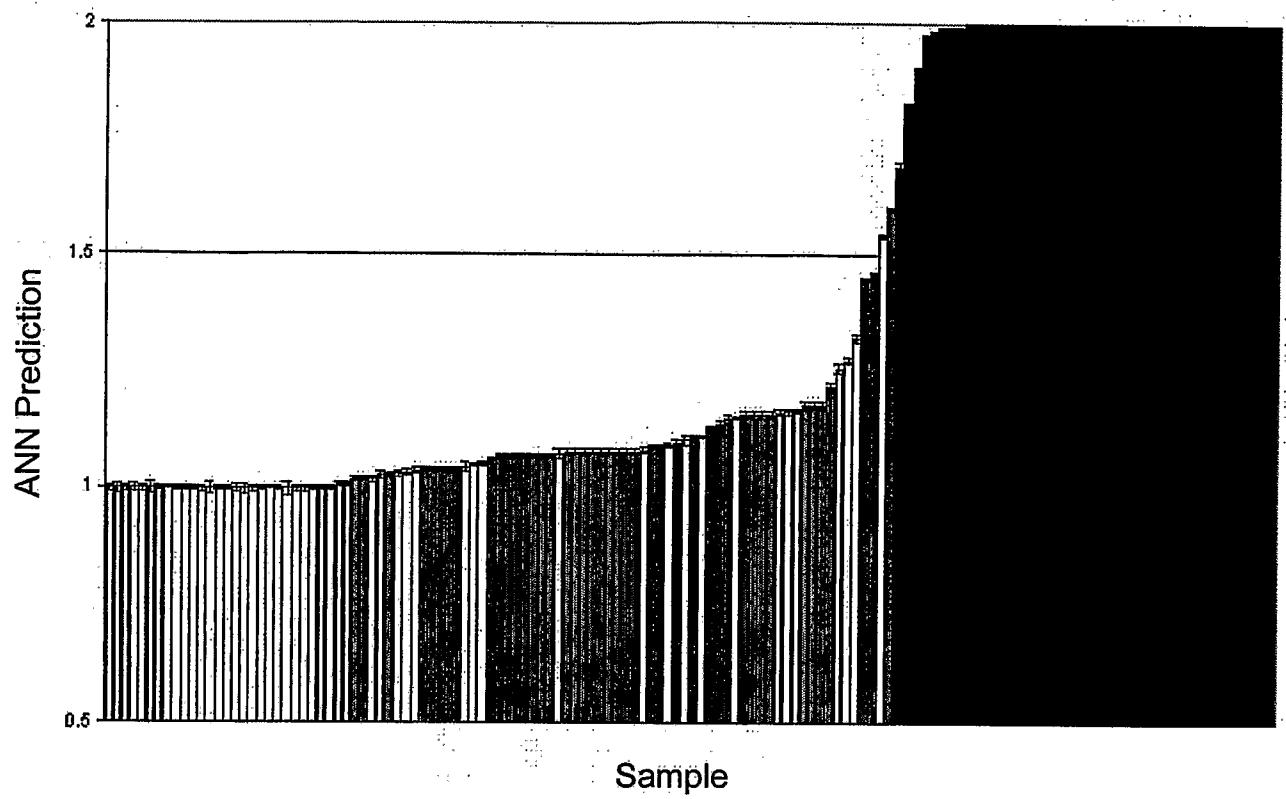
Figure 5c.

Figure 6a.

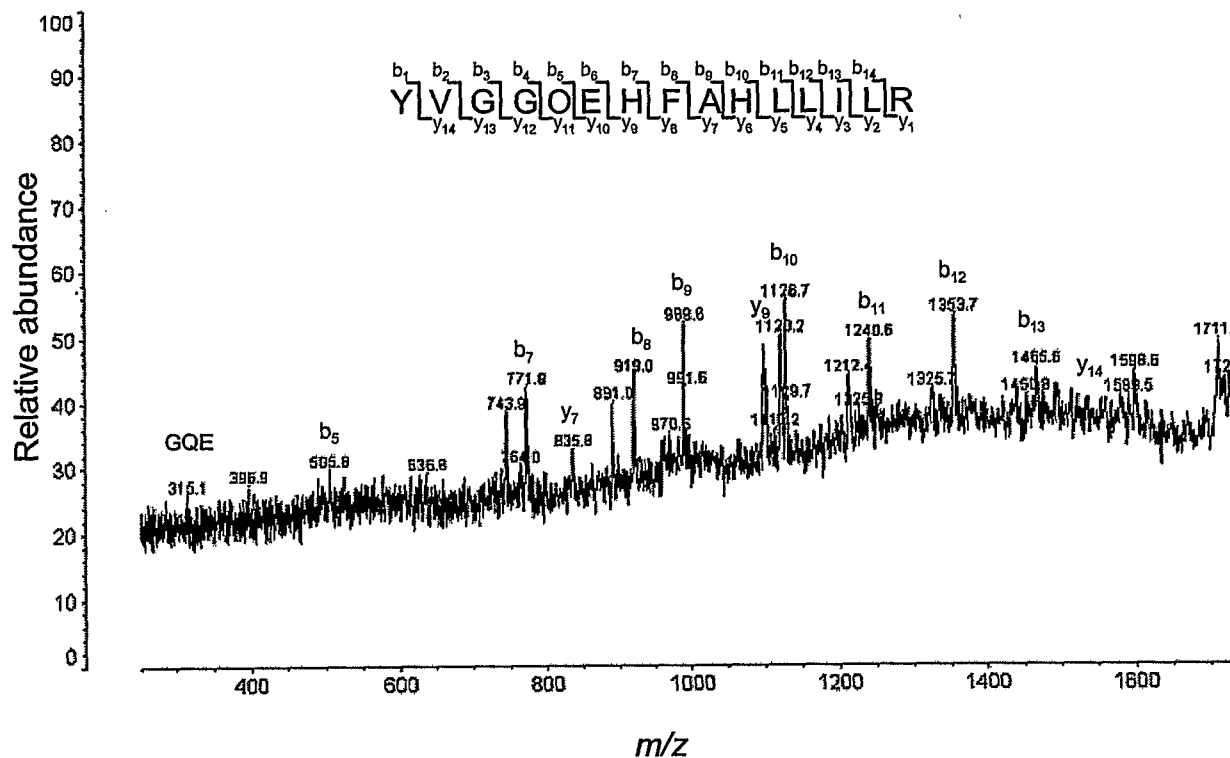


Figure 6b.

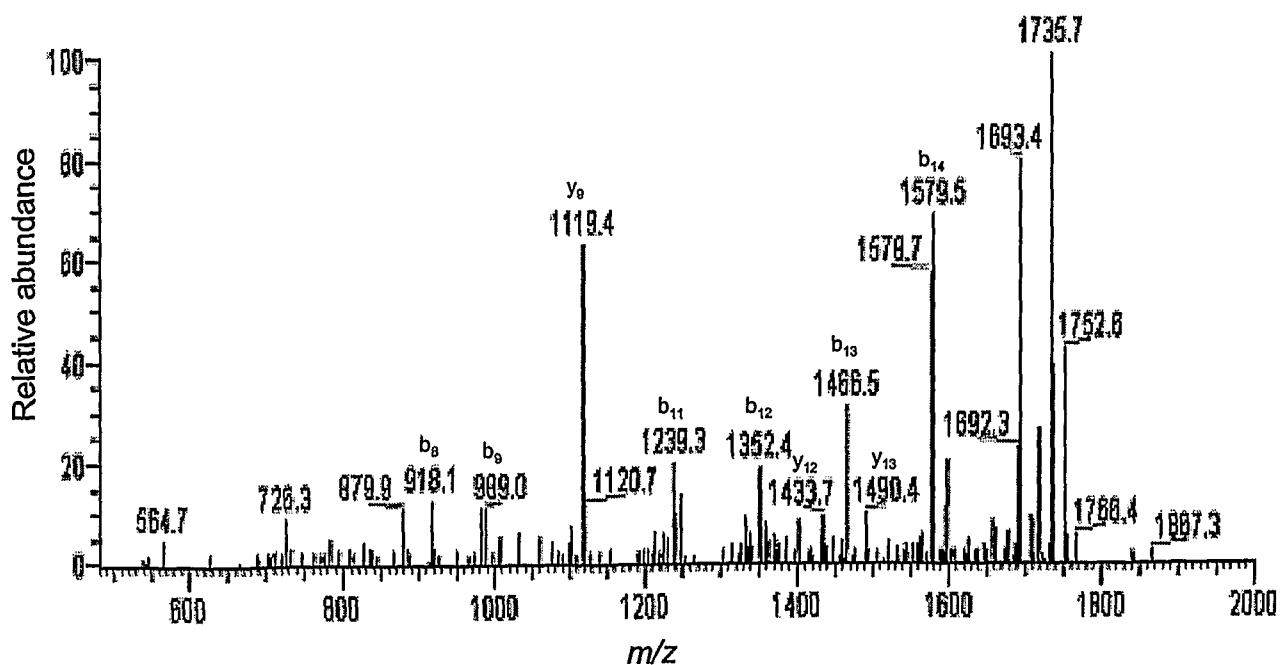


Figure 6c.

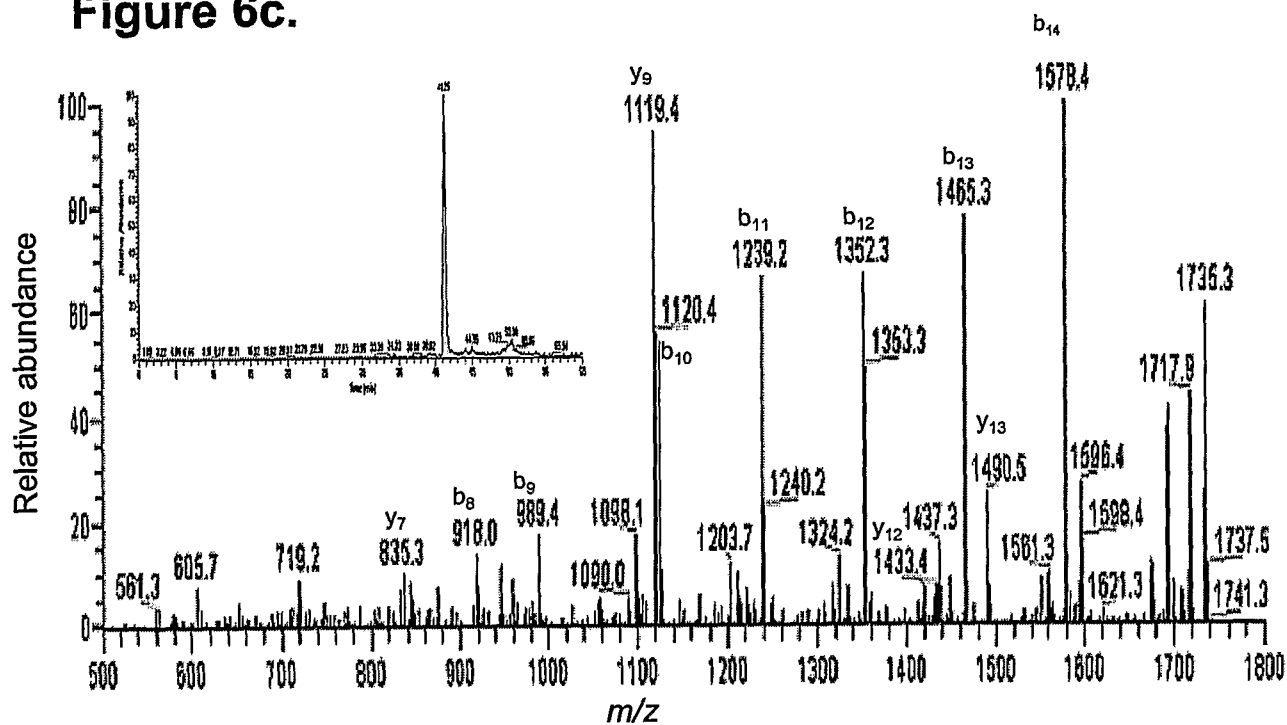


Figure 7a.

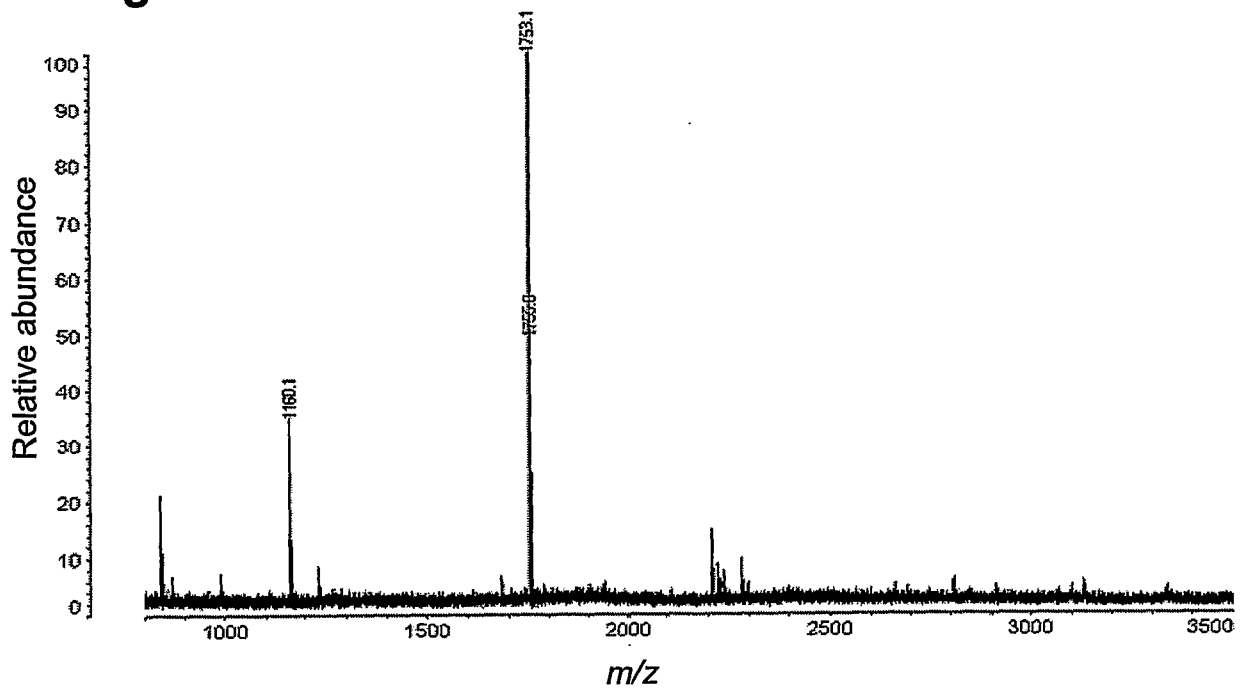


Figure 7b.

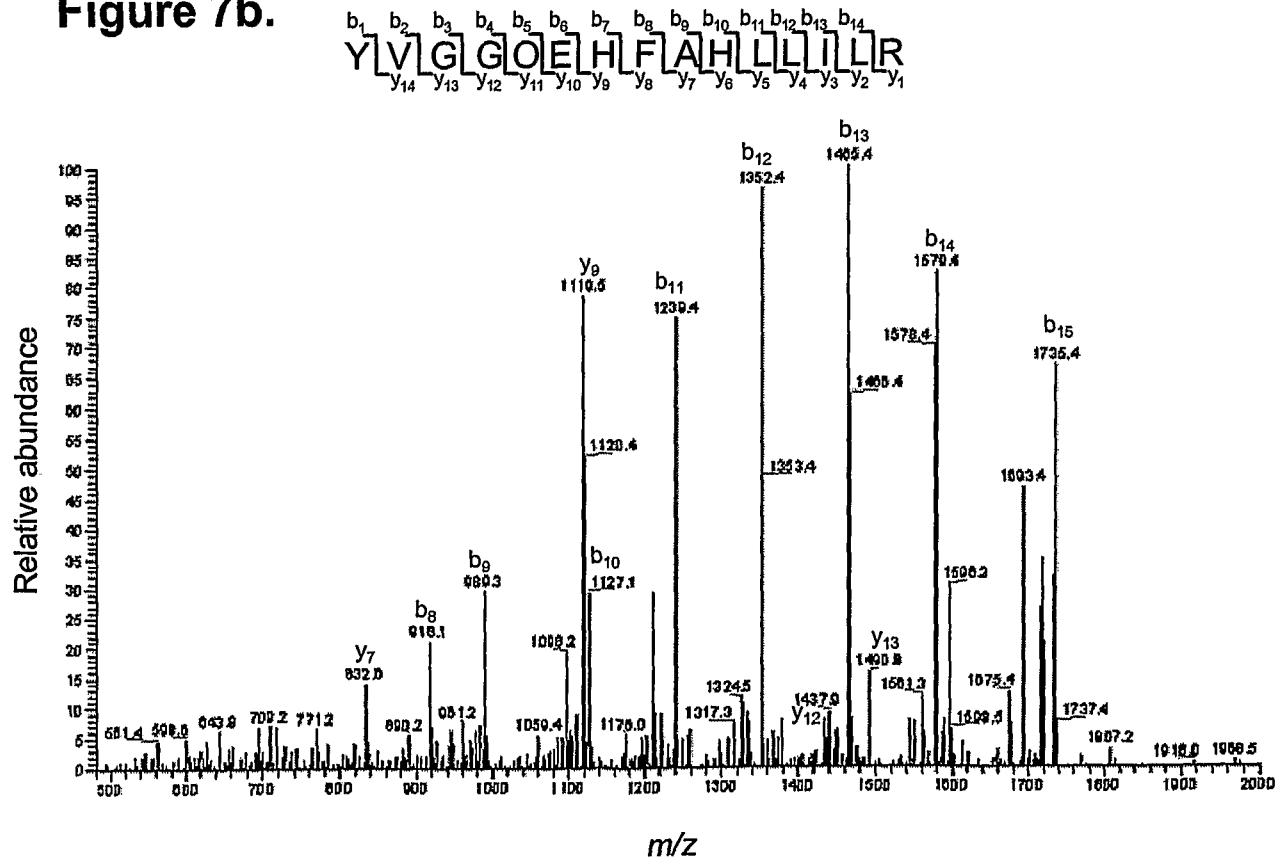


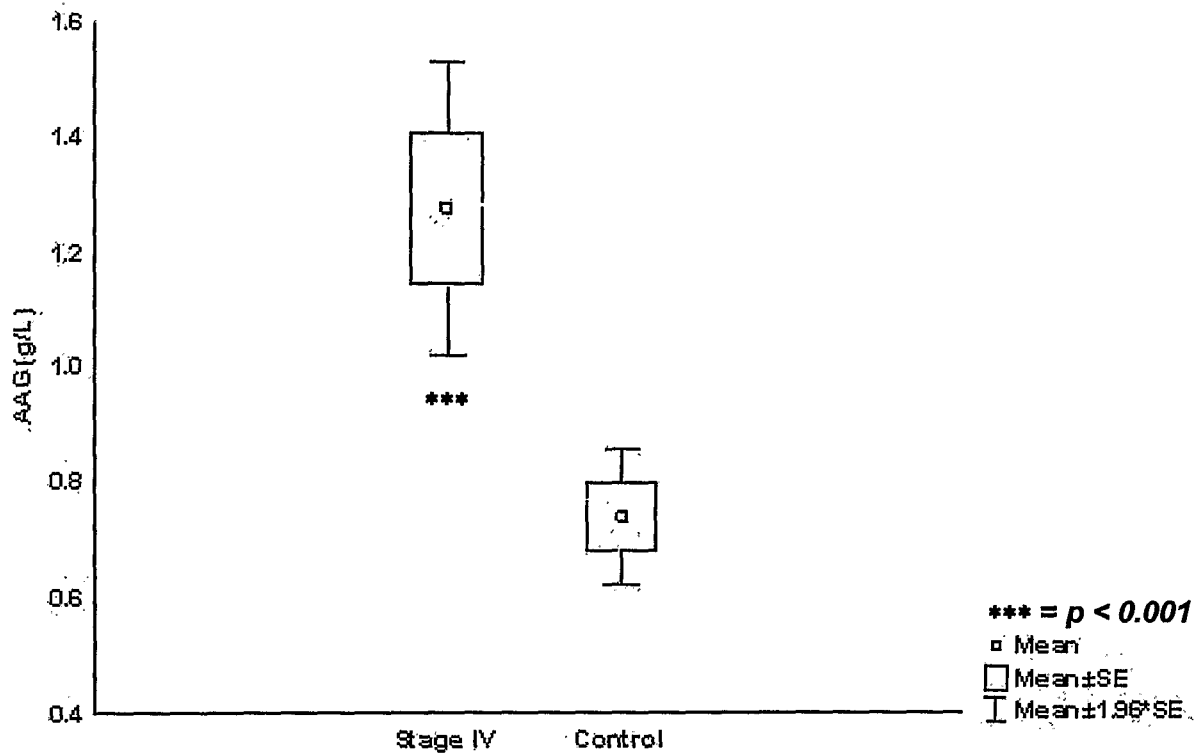
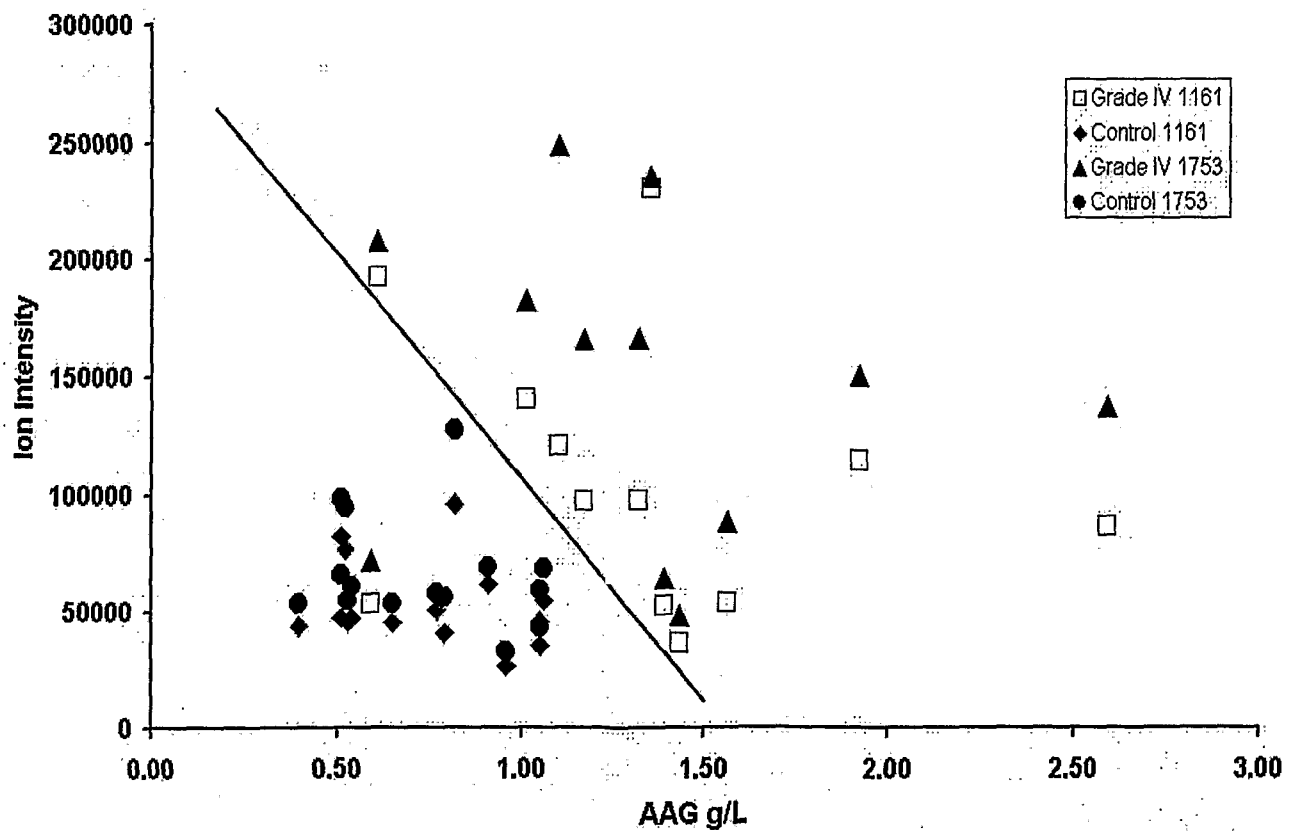
Figure 8a.**Figure 8b.**

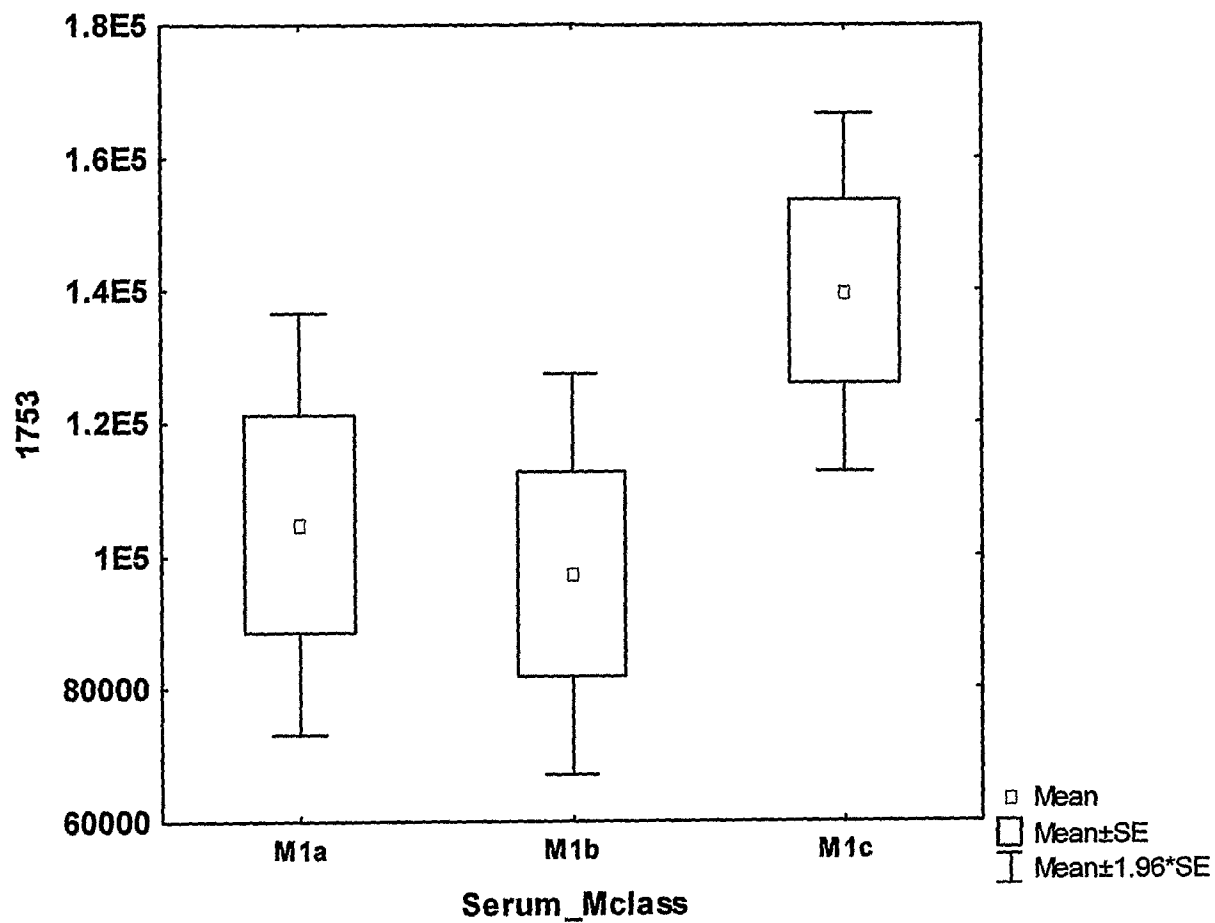
Figure 9.

Figure 10.