



(51) International Patent Classification:

C12N 15/12 (2006.01) G01N 33/50 (2006.01)
C07H 21/00 (2006.01) G01N 33/574 (2006.01)
C07K 14/47 (2006.01) G01N 33/68 (2006.01)
C07K 16/18 (2006.01) C07K 19/00 (2006.01)
C12Q 1/68 (2006.01) C12N 15/62 (2006.01)

(21) International Application Number:

PCT/CA2010/000423

(22) International Filing Date:

29 March 2010 (29.03.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PCTCA2009/000351 27 March 2009 (27.03.2009) CA

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:

US PCTCA2009/000351 (CIP)
Filed on 27 March 2009 (27.03.2009)

(71) Applicant (for all designated States except US): **GENESIS GENOMICS INC.** [CA/CA]; Suite 1000, 290 Munro Street, Thunder Bay, Ontario P7A 7T1 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PARR, Ryan** [CA/CA]; 1282 Hutton Park Drive, Thunder Bay, Ontario P7G 1J4 (CA). **DAKUBO, Gabriel** [CA/CA]; 1016 Brannan Lane, Thunder Bay, Ontario P7J 1H7 (CA). **HARBOTTLE, Andrew** [GB/GB]; 19 Love Lane, Newcastle Upon Tyne, Tyne and Wear NE1 3DW (GB). **REGULY, Brian** [CA/CA]; 937 West 14th Avenue, Apt. #403, Vancouver, British Columbia V5Z 1R3 (CA). **CREED, Jennifer** [CA/CA]; 238 London Drive, Thunder Bay, Ontario P7A 7Z6 (CA). **ROBINSON, Kerry** [CA/CA]; 335 Adelaide Street, Suite 407, Thunder Bay, Ontario P7A 7T8 (CA).

KLEIN, Daniel [CA/CA]; 2041 Sills Street, Thunder Bay, Ontario P7E 5S8 (CA).

(74) Agents: **CHARI, Santosh** et al.; Blake, Cassels & Graydon LLP, Box 25, Commerce Court West, 199 Bay Street, Suite 2800, Toronto, Ontario M5L 1A9 (CA).

(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, CL, 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, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, 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, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

Published:

— with international search report (Art. 21(3))
— with information concerning one or more priority claims considered void (Rule 26bis.2(d))

(54) Title: ABERRANT MITOCHONDRIAL DNA, ASSOCIATED FUSION TRANSCRIPTS AND TRANSLATION PRODUCTS AND HYBRIDIZATION PROBES THEREFOR

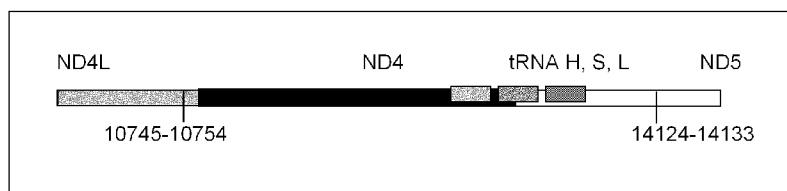


Figure 5a

(57) Abstract: The present invention provides novel mitochondrial fusion transcripts, the parent mutated mtDNA molecules, and the resulting translation products (proteins) for predicting, diagnosing and/or monitoring cancer. Hybridization probes complementary thereto for use in the methods of the invention are also provided.

**ABERRANT MITOCHONDRIAL DNA, ASSOCIATED FUSION TRANSCRIPTS AND
TRANSLATION PRODUCTS AND HYBRIDIZATION PROBES THEREFOR**

CROSS REFERENCE TO PRIOR APPLICATIONS

[0001] This application is a Continuation in Part of PCT application number PCT/CA2009/000351, filed on March 27, 2009, which claims priority from U.S. provisional application number 61/040,616, filed on March 28, 2008. The entire contents of such prior applications are incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The present invention relates to the field of mitochondrial genomics and proteomics. In one aspect, the invention relates to the identification and use of mitochondrial genome fusion transcripts and translation products, as well as probes that hybridize thereto.

BACKGROUND OF THE INVENTION

[0003] Mitochondrial Genome

[0004] The mitochondrial genome is a compact yet critical sequence of nucleic acids. Mitochondrial DNA, or "mtDNA", comprises a small genome of 16,569 nucleic acid base pairs (bp) (Anderson et al., 1981; Andrews et al., 1999) in contrast to the immense nuclear genome of 3.3 billion bp (haploid). Its genetic complement is substantially smaller than that of its nuclear cell mate (0.0005%). However, individual cells carry anywhere from 10^3 to 10^4 mitochondria depending on specific cellular functions (Singh and Modica-Napolitano 2002). Communication or chemical signalling routinely occurs between the nuclear and mitochondrial genomes (Sherratt et al., 1997). Moreover, specific nuclear components are responsible for the maintenance and integrity of mitochondrial sequences (Croteau et al., 1999). All mtDNA genomes in a given individual are identical due to the clonal expansion of mitochondria within the ovum, once fertilization has occurred. However mutagenic events can induce sequence diversity reflected as somatic mutations. These mutations may accumulate in different tissues throughout the body in a condition known as heteroplasmy.

[0005] Mitochondrial Proteome

[0006] About 3,000 nuclear genes are required to construct, operate and maintain mitochondria, with only thirty-seven of these coded by the mitochondrial genome, indicating heavy mitochondrial dependence on nuclear loci. The mitochondrial genome codes for a complement of 24 genes, including 2 rRNAs and 22 tRNAs that ensure correct translation of the remaining 13 genes which are vital to electron transport (see Figure 1). The mitochondrial genome is dependent on seventy nuclear encoded proteins to accomplish the oxidation and reduction reactions necessary for this vital function, in addition to the thirteen polypeptides supplied by the mitochondrial genome. Both nuclear and mitochondrial proteins form complexes

spanning the inner mitochondrial membrane and collectively generate 80-90% of the chemical fuel adenosine triphosphate, or ATP, required for cellular metabolism. In addition to energy production, mitochondria play a central role in other metabolic pathways as well. A critical function of the mitochondria is mediation of cell death, or apoptosis (see Green and Kroemer, 2005). Essentially, there are signal pathways which permeabilize the outer mitochondrial membrane, or in addition, the inner mitochondrial membrane as well. When particular mitochondrial proteins are released into the cytosol, non-reversible cell death is set in motion. This process highlights the multi-functional role that some mitochondrial proteins have. These multi-tasking proteins suggest that there are other mitochondrial proteins as well which may have alternate functions.

[0007] Mitochondrial Fusion Transcriptome/Proteome

[0008] The mitochondrial genome is unusual in that it is a circular, intron-less DNA molecule. The genome is interspersed with repeat motifs which flank specific lengths of sequences. Sequences between these repeats are prone to deletion under circumstances which are not well understood. Given the number of repeats in the mitochondrial genome, there are many possible deletions. The best known example is the 4977 “common deletion.” This deletion has been associated with several purported conditions and diseases and is thought to increase in frequency with aging (Dai et al., 2004; Ro et al., 2003; Barron et al., 2001; Lewis et al., 2000; Muller-Hocker, 1998; Porteous et al., 1998) (Figure 4). The current thinking in the field of mitochondrial genomics is that mitochondrial deletions are merely deleterious by-products of damage to the mitochondrial genome by such agents as reactive oxygen species and UVR. (Krishnan et al 2008, Nature Genetics). Further, though it is recognized that high levels of mtDNA deletions can have severe consequences on the cell’s ability to produce energy in the form of ATP as a result of missing gene sequences necessary for cellular respiration, it is not anticipated that these deleted mitochondrial molecules may be a component of downstream pathways, have an intended functional role, and possibly may be more aptly viewed as alternate natural forms of the recognized genes of the mitochondria.

[0009] The sequence dynamics of mtDNA are important diagnostic tools. Mutations in mtDNA are often preliminary indicators of developing disease. For example, it has been demonstrated that point mutations in the mitochondrial genome are characteristic of tumour foci in the prostate. This trend also extends to normal appearing tissue both adjacent to and distant from tumour tissue (Parr et al. 2006). This suggests that mitochondrial mutations occur early in the malignant transformation pathway.

[0010] For example, the frequency of a 3.4kb mitochondrial deletion has excellent utility in discriminating between benign and malignant prostate tissues (Maki et al. 2008). Furthermore, an investigation of the disease associated deletions and the novel sequences, created through

1 re-closure of the molecule identifies many open reading frames, suggesting the possibility of
2 unique mitochondrial fusion proteins.

3 **[0011]** Mitochondrial fusion transcripts have been reported previously in the literature, first in
4 soybeans (Morgens et al. 1984) and then later in two patients with Kearns-Sayre Syndrome, a
5 rare neuromuscular disorder (Nakase et al 1990). Importantly, these transcripts were not found
6 to have (or investigated regarding) association with any human cancers.

7 **[0012]** Nuclear Fusion Proteome

8 **[0013]** There is important nuclear precedence for fusion proteins and their resulting effects
9 on cancer. Nuclear *MLL* gene partner translocations are well established in correlation with high
10 risk acute leukemia and therapy-related acute myeloid leukemias following treatment with agents
11 that target topoisomerase II (Libura et al., 2005). Currently, around 50 translocations of the
12 human *MLL* gene are known to be associated with these cancers (Meyer et al., 2005). Break
13 points for these mutations, whether partial tandem duplications or translocations, for the majority
14 of these events, occur within nuclear specific repetitive motifs such as *Alu* I. Most of these
15 mutations are reciprocal translocations (84%) and include about 40 different genes (Libura et al.
16 2005).

17 **[0014]** There are known functional chimeric proteins created from some of these
18 rearrangements which affect the course of malignant disease. For example, murine cells which
19 express the protein from *MLL-ENL* accelerate the prevalence of chromosome abnormalities in
20 cells which survive exposure to etoposide (Eguchi et al., 2006). Of particular interest is *MLL-*
21 *SMAP1* and the reciprocal *SMAP1-MLL*. *SMAP1* binds calcium and as such participates in cell
22 signalling and trafficking.

23 **[0015]** Mitochondrial fusion proteins may be assumed to have similar attributes to nuclear
24 fusion proteins, especially since mitochondria and mitochondrial proteins play similar roles in
25 signalling and apoptosis.

26 SUMMARY OF THE INVENTION

27 **[0016]** An object of the present invention to provide aberrant mitochondrial DNA, associated
28 fusion transcripts and translation products and hybridization probes therefor.

29 **[0017]** In accordance with an aspect of the invention, there is provided an isolated
30 mitochondrial fusion transcript associated with cancer.

31 **[0018]** In accordance with another aspect of the invention, there is provided an isolated
32 mtDNA encoding a fusion transcript of the invention.

33 **[0019]** In accordance with another aspect of the invention, there is provided a hybridization
34 probe having a nucleic acid sequence complementary to at least a portion of a mitochondrial
35 fusion transcript or an mtDNA of the invention.

1 **[0020]** In accordance with another aspect of the invention, there is provided a method of
2 detecting a cancer in a mammal, the method comprising assaying a tissue sample from the
3 mammal for the presence of at least one mitochondrial fusion transcript associated with cancer
4 by hybridizing the sample with at least one hybridization probe having a nucleic acid sequence
5 complementary to at least a portion of a mitochondrial fusion transcript according to the
6 invention.

7 **[0021]** In accordance with another aspect of the invention, there is provided a method of
8 detecting a cancer in a mammal, the method comprising assaying a tissue sample from the
9 mammal for the presence of at least one aberrant mtDNA associated with cancer by hybridizing
10 the sample with at least one hybridization probe having a nucleic acid sequence complementary
11 to at least a portion of an mtDNA according to the invention.

12 **[0022]** In accordance with another aspect of the invention, there is provided a kit for
13 conducting an assay for detecting the presence of a cancer in a mammal, said kit comprising at
14 least one hybridization probe complementary to at least a portion of a fusion transcript or an
15 mtDNA of the invention.

16 **[0023]** In accordance with another aspect of the invention, there is provided a mitochondrial
17 fusion protein, the protein having an amino acid sequence resulting from the translation of a
18 mitochondrial fusion transcript of the invention.

19 **[0024]** In accordance with another aspect of the invention, there is provided a method of
20 detecting a cancer in a mammal, the method comprising assaying a tissue sample from the
21 mammal for the presence of at least one mitochondrial fusion protein, the protein having an
22 amino acid sequence resulting from the translation of a mitochondrial fusion transcript according
23 to the invention.

24 **BRIEF DESCRIPTION OF THE DRAWINGS**

25 **[0025]** The embodiments of the invention will now be described by way of example only with
26 reference to the appended drawings wherein:

27 **[0026]** Figure 1 is an illustration showing mitochondrial protein coding genes.

28 **[0027]** Figure 2 shows polyadenalated fusion transcripts in prostate samples invoked by the
29 loss of the 3.4kb deletion.

30 **[0028]** Figure 3 shows polyadenalated fusion transcripts in prostate samples invoked by the
31 loss of the 4977kb common deletion.

32 **[0029]** Figure 4 shows polyadenalated fusion transcripts in breast samples invoked by the
33 loss of the 3.4 kb segment from the mtgenome.

1 **[0030]** Figures 5a and 5b show an example of a mitochondrial DNA region before and after
2 splicing of genes.

3 **[0031]** Figures 6a to 6g illustrate the results for transcripts 2, 3, 8, 9, 10, 11 and 12 of the
4 invention in the identification of colorectal cancer tumours.

5 **[0032]** Figures 7a to 7d illustrate the results for transcripts 6, 8, 10 and 20 of the invention in
6 the identification of lung cancer tumours.

7 **[0033]** Figures 8a to 8j illustrate the results for transcripts 6, 10, 11, 14, 15, 16 and 20 of the
8 invention in the identification of melanomas.

9 **[0034]** Figures 9a to 9h illustrate the results for transcripts 1, 2, 3, 6, 11, 12, 15 and 20 of the
10 invention in the identification of ovarian cancer.

11 **[0035]** Figures 10 to 18 illustrate the results for transcripts 2, 3, 4, 11, 12, 13, 15, 16 and 20
12 of the invention in the identification of testicular cancer.

13 **[0036]** Figure 19 illustrates the SDS PAGE gel of cytosolic and mitochondrial fractions of
14 RWPE1 and WPE1-NA22 cell lines conducted during the fusion protein discovery phase.

15 **[0037]** Figure 20a illustrates the identified protein of fusion transcript P0026 based on the
16 peptides ILYMTDEVNDPSLTIK and STPYECGFDPMSK.

17 **[0038]** Figure 20b illustrates the wild-type CO2 protein identified in mitochondrial NA22 cell
18 line gel slice 5 of Figure 19 after searching the Human (SwissProt) database.

19 **[0039]** Figure 21a illustrates the identified protein of fusion transcript P0062 based on the
20 peptides KGPNVVGPYGLLPFADAMK, YDQLMHLLWK and LITTQQWLIK.

21 **[0040]** Figure 21b illustrates the identified peptides of ND1 identified in gel slice 5 of Figure
22 19 after searching the Human (SwissProt) database.

23 **[0041]** Figure 22 illustrates the identified protein of fusion transcript P0064 based on the
24 peptides KGPNVVGPYGLLPFADAMK and WAIIEEFTK.

25 **[0042]** Figure 23a illustrates the identified protein of fusion transcript P0176 based on the
26 peptides KGPNVVGPYGLLPFADAMK, VFSWLATLHGSNMK and
27 VLMVEEPSMNLEWLYGCPPPYHTFEPPVYMK.

28 **[0043]** Figure 23b illustrates the wild-type CO1 protein identified in mitochondrial NA22 cell
29 line gel slice 4 of Figure 19 after searching the Human (SwissProt) database.

30 **[0044]** Figures 24a to 24d illustrate the results of quantitative measurements of fusion
31 transcripts P0026, P0062, P0064 and P0176, respectively.

1 **DETAILED DESCRIPTION OF THE INVENTION**

2 **[0045]** The present invention provides novel mitochondrial fusion transcripts, the parent
3 mutated mtDNA molecules, and the resulting translation products that are useful for predicting,
4 diagnosing and/or monitoring cancer. The invention further provides hybridization probes for the
5 detection of fusion transcripts and associated mtDNA molecules and the use of such probes.

6 **[0046]** *Definitions*

7 **[0047]** Unless defined otherwise, all technical and scientific terms used herein have the
8 same meaning as commonly understood by one of ordinary skill in the art to which this invention
9 belongs.

10 **[0048]** The terms “comprise”, “comprises”, “comprised” or “comprising” may be used in the
11 present description. As used herein (including the specification and/or the claims), these terms
12 are to be interpreted as specifying the presence of the stated features, integers, steps or
13 components, but not as precluding the presence of one or more other feature, integer, step,
14 component or a group thereof as would be apparent to persons having ordinary skill in the
15 relevant art.

16 **[0049]** As used herein, “aberration” or “mutation” encompasses any modification in the wild
17 type mitochondrial DNA sequence that results in a fusion transcript and includes, without
18 limitation, insertions, translocations, deletions, duplications, recombinations, rearrangements or
19 combinations thereof.

20 **[0050]** As defined herein, “biological sample” refers to a tissue or bodily fluid containing cells
21 from which a molecule of interest can be obtained. For example, the biological sample can be
22 derived from tissue such as prostate, breast, colorectal, lung and skin, or from blood, saliva,
23 cerebral spinal fluid, sputa, urine, mucous, synovial fluid, peritoneal fluid, amniotic fluid and the
24 like. The biological sample may be a surgical specimen or a biopsy specimen. The biological
25 sample can be used either directly as obtained from the source or following a pre-treatment to
26 modify the character of the sample. Thus, the biological sample can be pre-treated prior to use
27 by, for example, preparing plasma or serum from blood, disrupting cells, preparing liquids from
28 solid materials, diluting viscous fluids, filtering liquids, distilling liquids, concentrating liquids,
29 inactivating interfering components, adding reagents, and the like.

30 **[0051]** A “continuous” transcript is a fusion transcript that keeps the reading frame from the
31 beginning to the end of both spliced genes. An “end” transcript is a fusion transcript that results
32 in a premature termination codon before the original termination codon of a second spliced gene.

33 **[0052]** As used herein, “mitochondrial DNA” or “mtDNA” is DNA present in mitochondria.

34 **[0053]** As used herein, the expression “mitochondrial fusion protein” or “fusion protein” refers
35 to a peptide product produced by the transcription and translation of a mutated mitochondrial

DNA, wherein such mutations comprise deletions or other “large-scale” mitochondrial DNA rearrangements. In addition, or alternatively, an in-frame protein may be translated from alternate initiation and termination codons within that sequence.

[0054] As used herein, the expression “mitochondrial fusion transcript” or “fusion transcript” refers to an RNA transcription product produced as a result of the transcription of a mutated mitochondrial DNA sequence wherein such mutations may comprise mitochondrial deletions and other large-scale mitochondrial DNA rearrangements.

[0055] As used herein, the expression “mitochondrial translation product” or “translation product” refers to any amino acid chain derived from a mitochondrial fusion transcript including peptides, polypeptides and proteins. It will be understood that “mitochondrial translation products” comprise “fusion proteins”, as defined above.

[0056] *Computer Analysis and Sequence Targetting*

[0057] As discussed above, mitochondrial fusion transcripts have been reported in soybeans (Morgens et al. 1984) and in humans suffering from a rare neuromuscular disorder (Nakase et al 1990). Fusion transcripts associated with human cancer have not, however, been described.

[0058] Using the knowledge gained from mapping the large-scale deletions of the human mitochondrial genome associated with cancer, the observation of high frequencies of these deletions, and the evidence in another organism and another disease type of transcriptionally active mutated mtDNA molecules, the present inventors hypothesized that such deletions may have importance beyond the DNA molecule and the damage and repair processes as it relates to cancer. To test this hypothesis computer analysis of the mitochondrial genome was conducted, specific for repeat elements, which suggested many potential deletion sites. Following this initial step of identifying unique repeats in the mitochondrial sequence having non-adjacent or non-tandem locations, a filter was then applied to identify those repeats that upon initiating a deletion event in the DNA molecule would then likely reclose or religate to produce a fused DNA sequence having an open reading frame (ORF) and thus capable of being transcribed by the mitochondrial transcription machinery. A subset of 18 of these molecules were then selected for targetting to investigate whether: they existed in the natural biological state of humans; they were polyadenylated and thus expected to proceed to protein synthesis; they had relevance to malignancy. Results from these investigations proved positive for all three queries and are described hereinafter.

[0059] *Genomic Mutations*

[0060] Mitochondrial DNA (mtDNA) dynamics are an important diagnostic tool. Mutations in mtDNA are often preliminary indicators of developing disease and may act as biomarkers indicative of risk factors associated with disease onset. According to the present invention, mutations in the mitochondrial genome result in the generation of fusion transcripts associated

1 with cancer. Thus, the use of mtDNA encoding such transcripts and probes directed thereto for
2 the detection, diagnosis and monitoring of cancer is provided.

3 **[0061]** One of skill in the art will appreciate that the mtDNA molecules for use in the methods
4 of the present invention may be derived through the isolation of naturally-occurring mutants or
5 may be based on the complementary sequence of any of the fusion transcripts described herein.
6 Exemplary mtDNA sequences and fusion transcripts are disclosed in Applicant's co-pending U.S.
7 Application No. 61/040,616 and published PCT application no. PCT/CA2009/000351 (published
8 as WO 2009/117811).

9 **[0062]** Detection of Mutant Genomic Sequences

10 **[0063]** Mutant mtDNA sequences according to the present invention may comprise any
11 modification that results in the generation of a fusion transcript. Non-limiting examples of such
12 modifications include insertions, translocations, deletions, duplications, recombinations,
13 rearrangements or combinations thereof. While the modification or change can vary greatly in
14 size from only a few bases to several kilobases, preferably the modification results in a
15 substantive deletion or other large-scale genomic aberration.

16 **[0064]** Extraction of DNA to detect the presence of such mutations may take place using art-
17 recognized methods, followed by amplification of all or a region of the mitochondrial genome, and
18 may include sequencing of the mitochondrial genome, as described in Current Protocols in
19 Molecular Biology.

20 **[0065]** The step of detecting the mutations can be selected from any technique known in the
21 art. For example, analyzing mtDNA can comprise sequencing the mtDNA, amplifying mtDNA by
22 PCR, Southern, Northern, Western South-Western blot hybridizations, denaturing HPLC,
23 hybridization to microarrays, biochips or gene chips, molecular marker analysis, biosensors,
24 melting temperature profiling or a combination of any of the above.

25 **[0066]** Any suitable means to sequence mitochondrial DNA may be used. Preferably,
26 mtDNA is amplified by PCR prior to sequencing. The method of PCR is well known in the art and
27 may be performed as described in Mullis and Faloona, 1987, Methods Enzymol., 155: 335. PCR
28 products can be sequenced directly or cloned into a vector which is then placed into a bacterial
29 host. Examples of DNA sequencing methods are found in Brumley, R. L. Jr. and Smith, L.M.,
30 1991, Rapid DNA sequencing by horizontal ultrathin gel electrophoresis, Nucleic Acids Res.
31 19:4121-4126 and Luckey, J.A., et al, 1993, High speed DNA sequencing by capillary gel
32 electrophoresis, Methods Enzymol. 218: 154-172. The combined use of PCR and sequencing of
33 mtDNA is described in Hopgood, R., et al, 1992, Strategies for automated sequencing of human
34 mtDNA directly from PCR products, Biotechniques 13:82-92 and Tanaka, M. et al, 1996,
35 Automated sequencing of mtDNA, Methods Enzymol. 264: 407-421.

1 **[0067]** Methods of selecting appropriate sequences for preparing various primers are also
2 known in the art. For example, the primer can be prepared using conventional solid-phase
3 synthesis using commercially available equipment, such as that available from Applied
4 Biosystems USA Inc. (Foster City, California), DuPont, (Wilmington, Del.), or Milligen (Bedford,
5 Mass.).

6 **[0068]** According to an aspect of the invention, to determine candidate genomic sequences,
7 a junction point of a sequence deletion is first identified. Sequence deletions are primarily
8 identified by direct and indirect repetitive elements which flank the sequence to be deleted at the
9 5' and 3' end. The removal of a section of the nucleotides from the genome followed by the
10 ligation of the genome results in the creation of a novel junction point.

11 **[0069]** Upon identification of the junction point, the nucleotides of the genes flanking the
12 junction point are determined in order to identify a spliced gene. Typically the spliced gene
13 comprises the initiation codon from the first gene and the termination codon of the second gene,
14 and may be expressed as a continuous transcript, *i.e.* one that keeps the reading frame from the
15 beginning to the end of both spliced genes. Some known mitochondrial deletions discovered to
16 have an open reading frame (ORF) when the rearranged sequences are rejoined at the splice
17 site are provided in Table 1.

18 **[0070]** Exemplary mtDNA molecules for use in the methods of the present invention are
19 provided below. These mtDNAs are based on modifications of the known mitochondrial genome
20 (SEQ ID NO: 1) and have been assigned a fusion or "FUS" designation, wherein A:B represents
21 the junction point between the last mitochondrial nucleotide of the first spliced gene and the first
22 mitochondrial nucleotide of the second spliced gene. The identification of the spliced genes is
23 provided in parentheses followed by the corresponding sequence identifier. Where provided
24 below, (AltMet) and (OrigMet) refer to alternate and original translation start sites, respectively.

25 **[0071]** FUS 8469:13447 (AltMet) (ATP synthase F0 subunit 8 to NADH dehydrogenase
26 subunit) (SEQ ID No: 2)

27 **[0072]** FUS 10744:14124 (NADH dehydrogenase subunit 4L (ND4L) to NADH
28 dehydrogenase subunit 5 (ND5)) (SEQ ID No: 3)

29 **[0073]** FUS 7974:15496 (Cytochrome c oxidase subunit II (COII) to Cytochrome b (Cytb))
30 (SEQ ID No: 4)

31 **[0074]** FUS 7992:15730 (Cytochrome c oxidase subunit II (COII) to Cytochrome b (Cytb))
32 (SEQ ID No: 5)

33 **[0075]** FUS 8210:15339 (Cytochrome c oxidase subunit II (COII) to Cytochrome b (Cytb))
34 (SEQ ID No: 6)

- 1 **[0076]** FUS 8828:14896 (ATP synthase F0 subunit 6 (ATPase6) to Cytochrome b (Cytb))
 2 (SEQ ID No: 7)
- 3 **[0077]** FUS 10665:14856 (NADH dehydrogenase subunit 4L (ND4L) to Cytochrome b
 4 (Cytb)) (SEQ ID No: 8)
- 5 **[0078]** FUS 6075:13799 (Cytochrome c oxidase subunit I (COI) to NADH de hydrogenase
 6 subunit 5 (ND5)) (SEQ ID No: 9)
- 7 **[0079]** FUS 6325:13989 (Cytochrome c oxidase subunit I (COI) to NADH dehydrogenase
 8 subunit 5 (ND5)) (SEQ ID No: 10)
- 9 **[0080]** FUS 7438:13476 (Cytochrome c oxidase subunit I (COI) to NADH dehydrogenase
 10 subunit 5 (ND5)) (SEQ ID No: 11)
- 11 **[0081]** FUS 7775:13532 (Cytochrome c oxidase subunit II (COII) to NADH dehydrogenase
 12 subunit 5 (ND5)) (SEQ ID No: 12)
- 13 **[0082]** FUS 8213:13991 (Cytochrome c oxidase subunit II (COII) to NADH dehydrogenase
 14 subunit 5 (ND5)) (SEQ ID No: 13)
- 15 **[0083]** FUS 9191:12909 (ATP synthase F0 subunit 6 (ATPase6) to NADH dehydrogenase
 16 subunit 5 (ND5)) (SEQ ID No: 14)
- 17 **[0084]** FUS 9574:12972 (Cytochrome c oxidase subunit III (COIII) to NADH dehydrogenase
 18 subunit 5 (ND5)) (SEQ ID No: 15)
- 19 **[0085]** FUS 10367:12829 (NADH dehydrogenase subunit 3 (ND3) to NADH dehydrogenase
 20 subunit 5 (ND5)) (SEQ ID No: 16)
- 21 **[0086]** FUS 11232:13980 (NADH dehydrogenase subunit 4 (ND4) to NADH dehydrogenase
 22 subunit 5 (ND5)) (SEQ ID No: 17)
- 23 **[0087]** FUS 8469:13447 (OrigMet) (ATP synthase F0 subunit 8 to NADH dehydrogenase
 24 subunit) (SEQ ID No: 18)
- 25 **[0088]** FUS 9144:13816 ((ATP synthase F0 subunit 6 (ATPase6) to NADH dehydrogenase
 26 subunit 5 (ND5)) (SEQ ID No: 54)
- 27 **[0089]** The present invention also provides the use of variants or fragments of these
 28 sequences for predicting, diagnosing and/or monitoring cancer.
- 29 **[0090]** "Variant", as used herein, refers to a nucleic acid differing from an mtDNA sequence
 30 of the present invention, but retaining essential properties thereof. Generally, variants are overall
 31 closely similar, and, in many regions, identical to a select mtDNA sequence. Specifically, the
 32 variants of the present invention comprise at least one of the nucleotides of the junction point of
 33 the spliced genes, and may further comprise one or more nucleotides adjacent thereto. In one
 34 embodiment of the invention, the variant sequence is at least 80%, 85%, 90%, 95%, 96%, 97%,

1 98% or 99% identical to any one of the mtDNA sequences of the invention, or the
2 complementary strand thereto.

3 **[0091]** In the present invention, "fragment" refers to a short nucleic acid sequence which is a
4 portion of that contained in the disclosed genomic sequences, or the complementary strand
5 thereto. This portion includes at least one of the nucleotides comprising the junction point of the
6 spliced genes, and may further comprise one or more nucleotides adjacent thereto. The
7 fragments of the invention are preferably at least about 15 nt, and more preferably at least about
8 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt, at
9 least about 50 nt, at least about 75 nt, or at least about 150 nt in length. A fragment "at least 20
10 nt in length," for example, is intended to include 20 or more contiguous bases of any one of the
11 mtDNA sequences listed above. In this context "about" includes the particularly recited value, a
12 value larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both
13 termini. These fragments have uses that include, but are not limited to, as diagnostic probes and
14 primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000
15 nucleotides) are also contemplated.

16 **[0092]** Thus, in specific embodiments of the invention, the mtDNA sequences are selected
17 from the group consisting of:

18 **[0093]** SEQ ID NO: 2 (FUS 8469:13447; AltMet)
19 **[0094]** SEQ ID NO: 3 (FUS 10744:14124)
20 **[0095]** SEQ ID NO: 4 (FUS 7974:15496)
21 **[0096]** SEQ ID NO: 5 (FUS 7992:15730)
22 **[0097]** SEQ ID NO: 6 (FUS 8210:15339)
23 **[0098]** SEQ ID NO: 7 (FUS 8828:14896)
24 **[0099]** SEQ ID NO: 8 (FUS 10665:14856)
25 **[00100]** SEQ ID NO: 9 (FUS 6075:13799)
26 **[00101]** SEQ ID NO: 10 (FUS 6325:13989)
27 **[00102]** SEQ ID NO: 11 (FUS 7438:13476)
28 **[00103]** SEQ ID NO: 12 (FUS 7775:13532)
29 **[00104]** SEQ ID NO: 13 (FUS 8213:13991)
30 **[00105]** SEQ ID NO: 14 (FUS 9191:12909)
31 **[00106]** SEQ ID NO: 15 (FUS 9574:12972)
32 **[00107]** SEQ ID NO: 16 (FUS 10367:12829)

1 **[00108]** SEQ ID NO: 17 (FUS 11232:13980)

2 **[00109]** SEQ ID NO: 18 (FUS 8469:13447; OrigMet)

3 **[00110]** SEQ ID NO: 54 (FUS 9144:13816),

4 **[00111]** - and fragments or variants thereof.

5 **[00112]** Probes

6 **[00113]** Another aspect of the invention is to provide a hybridization probe capable of
7 recognizing an aberrant mtDNA sequence of the invention. As used herein, the term "probe"
8 refers to an oligonucleotide which forms a duplex structure with a sequence in the target nucleic
9 acid, due to complementarity of at least one sequence in the probe with a sequence in the target
10 region. The probe may be labeled, according to methods known in the art.

11 **[00114]** Once aberrant mtDNA associated with particular disease is identified, hybridization of
12 mtDNA to, for example, an array of oligonucleotides can be used to identify particular mutations,
13 however, any known method of hybridization may be used.

14 **[00115]** As with the primers of the present invention, probes may be generated directly
15 against exemplary mtDNA fusion molecules of the invention, or to a fragment or variant thereof.
16 For instance, the sequences set forth in SEQ ID NOs: 2-18 and 54 and those disclosed in Table
17 1 can be used to design primers or probes that will detect a nucleic acid sequence comprising a
18 fusion sequence of interest. As would be understood by those of skill in the art, primers or probes
19 which hybridize to these nucleic acid molecules may do so under highly stringent hybridization
20 conditions or lower stringency conditions, such conditions known to those skilled in the art and
21 found, for example, in Current Protocols in Molecular Biology (John Wiley & Sons, New York
22 (1989)), 6.3.1-6.3.6.

23 **[00116]** In specific embodiments of the invention, the probes of the invention contain a
24 sequence complementary to at least a portion of the aberrant mtDNA comprising the junction
25 point of the spliced genes. This portion includes at least one of the nucleotides involved in the
26 junction point A:B, and may further comprise one or more nucleotides adjacent thereto. In this
27 regard, the present invention encompasses any suitable targeting mechanism that will select an
28 mtDNA molecule using the nucleotides involved and/or adjacent to the junction point A:B.

29 **[00117]** Various types of probes known in the art are contemplated by the present invention.
30 For example, the probe may be a hybridization probe, the binding of which to a target nucleotide
31 sequence can be detected using a general DNA binding dye such as ethidium bromide, SYBR®
32 Green, SYBR® Gold and the like. Alternatively, the probe can incorporate one or more
33 detectable labels. Detectable labels are molecules or moieties a property or characteristic of
34 which can be detected directly or indirectly and are chosen such that the ability of the probe to
35 hybridize with its target sequence is not affected. Methods of labelling nucleic acid sequences

1 are well-known in the art (see, for example, Ausubel *et al.*, (1997 & updates) *Current Protocols in*
2 *Molecular Biology*, Wiley & Sons, New York).

3 **[00118]** Labels suitable for use with the probes of the present invention include those that can
4 be directly detected, such as radioisotopes, fluorophores, chemiluminophores, enzymes, colloidal
5 particles, fluorescent microparticles, and the like. One skilled in the art will understand that
6 directly detectable labels may require additional components, such as substrates, triggering
7 reagents, light, and the like to enable detection of the label. The present invention also
8 contemplates the use of labels that are detected indirectly.

9 **[00119]** The probes of the invention are preferably at least about 15 nt, and more preferably
10 at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least
11 about 40 nt, at least about 50 nt, at least about 75 nt, or at least about 150 nt in length. A probe
12 of "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases that
13 are complementary to an mtDNA sequence of the invention. Of course, larger probes (e.g., 50,
14 150, 500, 600, 2000 nucleotides) may be preferable.

15 **[00120]** The probes of the invention will also hybridize to nucleic acid molecules in biological
16 samples, thereby enabling the methods of the invention. Accordingly, in one aspect of the
17 invention, there is provided a hybridization probe for use in the detection of cancer, wherein the
18 probe is complementary to at least a portion of an aberrant mtDNA molecule. In another aspect
19 the present invention provides probes and a use of (or a method of using) such probes for the
20 detection of colorectal cancer, lung cancer, breast cancer, ovarian cancer, testicular, cancer,
21 prostate cancer and/or melanoma skin cancer.

22 **[00121]** Assays

23 **[00122]** Measuring the level of aberrant mtDNA in a biological sample can determine the
24 presence of one or more cancers in a subject. The present invention, therefore, encompasses
25 methods for predicting, diagnosing or monitoring cancer, comprising obtaining one or more
26 biological samples, extracting mtDNA from the samples, and assaying the samples for aberrant
27 mtDNA by: quantifying the amount of one or more aberrant mtDNA sequences in the sample and
28 comparing the quantity detected with a reference value. As would be understood by those of skill
29 in the art, the reference value is based on whether the method seeks to predict, diagnose or
30 monitor cancer. Accordingly, the reference value may relate to mtDNA data collected from one or
31 more known non-cancerous biological samples, from one or more known cancerous biological
32 samples, and/or from one or more biological samples taken over time.

33 **[00123]** In one aspect, the invention provides a method of detecting cancer in a mammal, the
34 method comprising assaying a tissue sample from the mammal for the presence of an aberrant
35 mitochondrial DNA described above. The present invention also provides for methods
36 comprising assaying a tissue sample from the mammal by hybridizing the sample with at least

one hybridization probe. The probe may be generated against a mutant mitochondrial DNA sequence of the invention as described herein.

[00124] In another aspect, the invention provides a method as above, wherein the assay comprises:

a) conducting a hybridization reaction using at least one of the probes to allow the at least one probe to hybridize to a complementary aberrant mitochondrial DNA sequence;

b) quantifying the amount of the at least one aberrant mitochondrial DNA sequence in the sample by quantifying the amount of the mitochondrial DNA hybridized to the at least one probe; and,

c) comparing the amount of the mitochondrial DNA in the sample to at least one known reference value.

[00125] Also included in the present invention are methods for predicting, diagnosing or monitoring cancer comprising diagnostic imaging assays as described below. The diagnostic assays of the invention can be readily adapted for high-throughput. High-throughput assays provide the advantage of processing many samples simultaneously and significantly decrease the time required to screen a large number of samples. The present invention, therefore, contemplates the use of the nucleotides of the present invention in high-throughput screening or assays to detect and/or quantitate target nucleotide sequences in a plurality of test samples.

[00126] *Fusion Transcripts*

[00127] The present invention further provides the identification of fusion transcripts and associated hybridization probes useful in methods for predicting, diagnosing and/or monitoring cancer. One of skill in the art will appreciate that such molecules may be derived through the isolation of naturally-occurring transcripts or, alternatively, by the recombinant expression of mtDNAs isolated according to the methods of the invention. As discussed, such mtDNAs typically comprise a spliced gene having the initiation codon from the first gene and the termination codon of the second gene. Accordingly, fusion transcripts derived therefrom comprise a junction point associated with the spliced genes.

[00128] Detection of Fusion Transcripts

[00129] Naturally occurring fusion transcripts can be extracted from a biological sample and identified according to any suitable method known in the art, or may be conducted according to the methods described in the examples. In one embodiment of the invention, stable polyadenylated fusion transcripts are identified using Oligo(dT) primers that target transcripts with poly-A tails, followed by RT-PCR using primer pairs designed against the target transcript.

[00130] The following exemplary fusion transcripts were detected using such methods and found useful in predicting, diagnosing and/or monitoring cancer as indicated in the examples.

1 Likewise, fusion transcripts derived from the ORF sequences identified in Table 1 may be useful
2 in predicting, diagnosing and/or monitoring cancer.

3 **[00131]** SEQ ID NO: 19 (Transcript 1; 8469:13447; AltMet)

4 **[00132]** SEQ ID NO: 20 (Transcript 2; 10744:14124)

5 **[00133]** SEQ ID NO: 21 (Transcript 3; 7974:15496)

6 **[00134]** SEQ ID NO: 22 (Transcript 4; 7992:15730)

7 **[00135]** SEQ ID NO: 23 (Transcript 5; 8210:15339)

8 **[00136]** SEQ ID NO: 24 (Transcript 6; 8828:14896)

9 **[00137]** SEQ ID NO: 25 (Transcript 7; 10665:14856)

10 **[00138]** SEQ ID NO: 26 (Transcript 8; 6075:13799)

11 **[00139]** SEQ ID NO: 27 (Transcript 9; 6325:13989)

12 **[00140]** SEQ ID NO: 28 (Transcript 10; 7438:13476)

13 **[00141]** SEQ ID NO: 29 (Transcript 11; 7775:13532)

14 **[00142]** SEQ ID NO: 30 (Transcript 12; 8213:13991)

15 **[00143]** SEQ ID NO: 31 (Transcript 14; 9191:12909)

16 **[00144]** SEQ ID NO: 32 (Transcript 15; 9574:12972)

17 **[00145]** SEQ ID NO: 33 (Transcript 16; 10367:12829)

18 **[00146]** SEQ ID NO: 34 (Transcript 17; 11232:13980)

19 **[00147]** SEQ ID NO: 35 (Transcript 20; 8469:13447; OrigMet)

20 **[00148]** SEQ ID NO: 53 (Transcript 13; 9144:13816)

21 **[00149]** Fusion transcripts can also be produced by recombinant techniques known in the art.
22 Typically this involves transformation (including transfection, transduction, or infection) of a
23 suitable host cell with an expression vector comprising an mtDNA sequence of interest.

24 **[00150]** Variants or fragments of the fusion transcripts identified herein are also provided.
25 Such sequences may adhere to the size limitations and percent identities described above with
26 respect to genomic variants and fragments, or as determined suitable by a skilled technician.

27 **[00151]** Probes

28 **[00152]** Once a fusion transcript has been characterized, primers or probes can be developed
29 to target the transcript in a biological sample. Such primers and probes may be prepared using
30 any known method (as described above) or as set out in the examples provided below. A probe
31 may, for example, be generated for the fusion transcript, and detection technologies, such as

1 QuantiGene 2.0™ by Panomics™, used to detect the presence of the transcript in a sample.
2 Primers and probes may be generated directly against exemplary fusion transcripts of the
3 invention, or to a fragment or variant thereof. For instance, the sequences set forth in SEQ ID
4 NOs: 19-35 and 53, as well as those disclosed in Table 1, can be used to design probes that will
5 detect a nucleic acid sequence comprising a fusion sequence of interest.

6 **[00153]** As would be understood by those skilled in the art, probes designed to hybridize to
7 the fusion transcripts of the invention contain a sequence complementary to at least a portion of
8 the transcript expressing the junction point of the spliced genes. This portion includes at least
9 one of the nucleotides complementary to the expressed junction point, and may further comprise
10 one or more complementary nucleotides adjacent thereto. In this regard, the present invention
11 encompasses any suitable targeting mechanism that will select a fusion transcript that uses the
12 nucleotides involved and adjacent to the junction point of the spliced genes.

13 **[00154]** Various types of probes and methods of labelling known in the art are contemplated
14 for the preparation of transcript probes. Such types and methods have been described above
15 with respect to the detection of genomic sequences. The transcript probes of the invention are
16 preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at
17 least about 30 nt, and even more preferably, at least about 40 nt, at least about 50 nt, at least
18 about 75 nt, or at least about 150 nt in length. A probe of "at least 20 nt in length," for example, is
19 intended to include 20 or more contiguous bases that are complementary to an mtDNA sequence
20 of the invention. Of course, larger probes (e.g., 50, 150, 500, 600, 2000 nucleotides) may be
21 preferable.

22 **[00155]** In one aspect, the invention provides a hybridization probe for use in the detection of
23 cancer, wherein the probe is complementary to at least a portion of a mitochondrial fusion
24 transcript provided above.

25 **[00156]** In another aspect, the present invention provides probes and a use of (or a method of
26 using) such probes for the detection of colorectal cancer, lung cancer, breast cancer, ovarian
27 cancer, testicular cancer, prostate cancer or melanoma skin cancer.

28 **[00157]** Assays

29 **[00158]** Measuring the level of mitochondrial fusion transcripts in a biological sample can
30 determine the presence of one or more cancers in a subject. The present invention, therefore,
31 provides methods for predicting, diagnosing or monitoring cancer, comprising obtaining one or
32 more biological samples, extracting mitochondrial RNA from the samples, and assaying the
33 samples for fusion transcripts by: quantifying the amount of one or more fusion transcripts in the
34 sample and comparing the quantity detected with a reference value. As would be understood by
35 those of skill in the art, the reference value is based on whether the method seeks to predict,
36 diagnose or monitor cancer. Accordingly, the reference value may relate to transcript data

collected from one or more known non-cancerous biological samples, from one or more known cancerous biological samples, and/or from one or more biological samples taken over time.

[00159] In one aspect, the invention provides a method of detecting a cancer in a mammal, the method comprising assaying a tissue sample from said mammal for the presence of at least one fusion transcript of the invention by hybridizing said sample with at least one hybridization probe having a nucleic acid sequence complementary to at least a portion of the mitochondrial fusion transcript.

[00160] In another aspect, the invention provides a method as above, wherein the assay comprises:

[00161] a) conducting a hybridization reaction using at least one of the above-noted probes to allow the at least one probe to hybridize to a complementary mitochondrial fusion transcript;

[00162] b) quantifying the amount of the at least one mitochondrial fusion transcript in the sample by quantifying the amount of the transcript hybridized to the at least one probe; and,

[00163] c) comparing the amount of the mitochondrial fusion transcript in the sample to at least one known reference value.

[00164] As discussed above, the diagnostic assays of the invention may also comprise diagnostic imaging methods as described herein and can be readily adapted for high-throughput. The present invention, therefore, contemplates the use of the fusion transcripts and associated probes of the present invention in high-throughput screening or assays to detect and/or quantitate target nucleotide sequences in a plurality of test samples.

[00165] Translation Products

[00166] To date, mitochondrial fusion proteins have not been detected or isolated. However, the levels of mitochondrial fusion transcripts observed from the examples provided below and the indications that they are polyadenylated provide further evidence supporting the existence of such mitochondrial fusion proteins. Accordingly, the present invention provides the identification of fusion proteins for predicting, diagnosing, and/or monitoring of cancer.

[00167] Fusion proteins contemplated for use in the disclosed methods may be derived through the isolation of naturally-occurring polypeptides or through gene expression. Such polypeptides can be prepared by methods known in the art, such as purification from cell extracts or the use of recombinant techniques.

[00168] Putative protein sequences corresponding to transcripts 1-17 and 20 are provided below along with their respective sequence identifier. These, as well as the putative protein sequences corresponding to the deletion sequences disclosed in Table 1, are herein contemplated for use in the methods of the present invention.

- 1 **[00169]** SEQ ID NO: 36 (Transcripts 1)
- 2 **[00170]** SEQ ID NO: 37 (Transcript 2)
- 3 **[00171]** SEQ ID NO: 38 (Transcript 3)
- 4 **[00172]** SEQ ID NO: 39 (Transcript 4)
- 5 **[00173]** SEQ ID NO: 40 (Transcript 5)
- 6 **[00174]** SEQ ID NO: 41 (Transcript 6)
- 7 **[00175]** SEQ ID NO: 42 (Transcript 7)
- 8 **[00176]** SEQ ID NO: 43 (Transcript 8)
- 9 **[00177]** SEQ ID NO: 44 (Transcript 9)
- 10 **[00178]** SEQ ID NO: 45 (Transcript 10)
- 11 **[00179]** SEQ ID NO: 46 (Transcript 11)
- 12 **[00180]** SEQ ID NO: 47 (Transcript 12)
- 13 **[00181]** SEQ ID NO: 48 (Transcript 14)
- 14 **[00182]** SEQ ID NO: 49 (Transcript 15)
- 15 **[00183]** SEQ ID NO: 50 (Transcript 16)
- 16 **[00184]** SEQ ID NO: 51 (Transcript 17)
- 17 **[00185]** SEQ ID NO: 52 (Transcripts 20)
- 18 **[00186]** SEQ ID NO: 55 (Transcript 13)

19 **[00187]** Detection of Fusion Proteins

20 **[00188]** Fusion proteins of the invention can be recovered and purified from a biological
 21 sample by well-known methods including ammonium sulfate or ethanol precipitation, acid
 22 extraction, anion or cation exchange chromatography, phosphocellulose chromatography,
 23 hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite
 24 chromatography, hydrophobic charge interaction chromatography and lectin chromatography.
 25 Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

26 **[00189]** Assaying fusion protein levels in a biological sample can occur using a variety of
 27 techniques. For example, protein expression in tissues can be studied with classical
 28 immunohistological methods (Jalkanen et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et
 29 al., J. Cell. Biol. 105:3087-3096 (1987)). Other methods useful for detecting protein expression
 30 include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the
 31 radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include
 32 enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (<125> I, <121> I),

carbon (<14> C), sulfur (<35> S), tritium (<3> H), indium (<112> In), and technetium (<99m> Tc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

[00190] The polypeptides of the invention can also be produced by recombinant techniques known in the art. Typically this involves transformation (including transfection, transduction, or infection) of a suitable host cell with an expression vector comprising a polynucleotide encoding the protein or polypeptide of interest.

[00191] Antibodies

[00192] Protein specific antibodies for use in the assays of the present invention can be raised against the wild-type or expressed mitochondrial fusion proteins of the invention or an antigenic polypeptide fragment thereof, which may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier.

[00193] As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments, or antigen-binding fragments, thereof (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to, or having "specificity to", a mitochondrial fusion protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl et al., J. Nucl. Med. 24:316-325 (1983)). Thus, these fragments are preferred.

[00194] The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing the mitochondrial fusion protein or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In one method, a preparation of mitochondrial fusion protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

[00195] In a related method, the antibodies of the present invention are monoclonal antibodies. Such monoclonal antibodies can be prepared using hybridoma technology (Kohler et al., Nature 256:495 (1975); Kohler et al., Eur. J. Immunol. 6:511 (1976); Kohler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., (1981) pp. 563-681). In general, such procedures involve immunizing an animal (preferably a mouse) with a mitochondrial fusion protein antigen or with a mitochondrial fusion protein-expressing cell.

[00196] The present invention comprises immunological assays using antibodies or antigen-binding fragments having specificity to the fusion proteins described herein. Such immunological assays may be facilitated by kits containing the antibodies or antigen-binding fragments along with any other necessary reagents, test strips, materials, instructions etc.

1 **[00197]** Assays

2 **[00198]** Measuring the level of a translation product such as a fusion protein in a biological
3 sample can determine the presence of one or more cancers in a subject. The present invention,
4 therefore, provides methods for predicting, diagnosing or monitoring cancer, comprising
5 obtaining one or more biological samples, extracting mitochondrial fusion proteins from the
6 samples, and assaying the samples for such molecules by: quantifying the amount of one or
7 more molecules in the sample and comparing the quantity detected with a reference value. As
8 would be understood by those of skill in the art, the reference value is based on whether the
9 method seeks to predict, diagnose or monitor cancer. Accordingly, the reference value may
10 relate to protein data collected from one or more known non-cancerous biological samples, from
11 one or more known cancerous biological samples, and/or from one or more biological samples
12 taken over time.

13 **[00199]** Techniques for quantifying proteins in a sample are well known in the art and include,
14 for instance, classical immunohistological methods (Jalkanen et al., J. Cell. Biol. 101:976-985
15 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987)). Additional methods useful for
16 detecting protein expression include immunoassays such as the radioimmunoassay (RIA) and
17 the enzyme linked immunosorbent assay (ELISA).

18 **[00200]** In one aspect, the invention provides a method of detecting a cancer in a mammal,
19 the method comprising assaying a tissue sample from the mammal for the presence of at least
20 one mitochondrial fusion protein. In another aspect, the present invention provides for the
21 detection of mitochondrial fusion proteins in the diagnosis of colorectal cancer, lung cancer,
22 breast cancer, ovarian cancer, testicular cancer, prostate cancer and/or melanoma skin cancer.

23 **[00201]** *Diagnostic Imaging*

24 **[00202]** Diagnostic Devices

25 **[00203]** The invention includes diagnostic devices such as biochips, gene chips or
26 microarrays used to diagnose specific diseases or identify specific mutations. All sequenced
27 mitochondrial genomes are assessed to create a consensus structure of the base pair
28 arrangement and are assigned a prohibiting index for proportion of base pair deletions and
29 mutations associated with a particular disease or disorder. The diagnostic arrangement is then
30 used to create biochips, gene chips, or microarrays.

31 **[00204]** Once sequences associated with particular diseases, disease states or disorders are
32 identified, hybridization of a mitochondrial nucleotide sample to an array of oligonucleotides can
33 be used to identify particular mutations. Any known method of hybridization may be used.
34 Preferably, an array is used, which has oligonucleotide probes matching the wild type or mutated
35 region, and a control probe. Commercially available arrays such as microarrays or gene chips
36 are suitable. These arrays contain thousands of matched and control pairs of probes on a slide

1 or microchip, and are capable of sequencing the entire genome very quickly. Review articles
2 describing the use of microarrays in genome and DNA sequence analysis are available on-line.

3 **[00205]** Microarray

4 **[00206]** Polynucleotide arrays provide a high throughput technique that can assay a large
5 number of polynucleotides in a sample comprising one or more target nucleic acid sequences.
6 The arrays of the invention are useful for gene expression analysis, diagnosis of disease and
7 prognosis of disease (e.g., monitoring a patient's response to therapy, and the like).

8 **[00207]** Any combination of the polynucleotide sequences of mtDNA indicative of disease, or
9 disease progression are used for the construction of a microarray.

10 **[00208]** The target nucleic acid samples to be analyzed using a microarray are derived from
11 any human tissue or fluid which contains adequate amounts of mtDNA, as previously described.
12 The target nucleic acid samples are contacted with polynucleotide members under hybridization
13 conditions sufficient to produce a hybridization pattern of complementary nucleic acid
14 members/target complexes.

15 **[00209]** Construction of a Microarray

16 **[00210]** The microarray comprises a plurality of unique polynucleotides attached to one
17 surface of a solid support, wherein each of the polynucleotides is attached to the surface of the
18 solid support in a non-identical preselected region. Each associated sample on the array
19 comprises a polynucleotide composition, of known identity, usually of known sequence, as
20 described in greater detail below. Any conceivable substrate may be employed in the invention.

21 **[00211]** The array is constructed using any known means. The nucleic acid members may be
22 produced using established techniques such as polymerase chain reaction (PCR) and reverse
23 transcription (RT). These methods are similar to those currently known in the art (see e.g. PCR
24 Strategies, Michael A. Innis (Editor), et al. (1995) and PCR: Introduction to Biotechniques Series,
25 C. R. Newton, A. Graham (1997)). Amplified polynucleotides are purified by methods well known
26 in the art (e.g., column purification). A polynucleotide is considered pure when it has been
27 isolated so as to be substantially free of primers and incomplete products produced during the
28 synthesis of the desired polynucleotide. Preferably, a purified polynucleotide will also be
29 substantially free of contaminants which may hinder or otherwise mask the binding activity of the
30 molecule.

31 **[00212]** In the arrays of the invention, the polynucleotide compositions are stably associated
32 with the surface of a solid support, wherein the support may be a flexible or rigid solid support.

33 **[00213]** Any solid support to which a nucleic acid member may be attached may be used in
34 the invention. Examples of suitable solid support materials include, but are not limited to, silicates

1 such as glass and silica gel, cellulose and nitrocellulose papers, nylon, polystyrene,
2 polymethacrylate, latex, rubber, and fluorocarbon resins such as TEFLONTM.

3 **[00214]** The solid support material may be used in a wide variety of shapes including, but not
4 limited to slides and beads. Slides provide several functional advantages and thus are a
5 preferred form of solid support. Due to their flat surface, probe and hybridization reagents are
6 minimized using glass slides. Slides also enable the targeted application of reagents, are easy to
7 keep at a constant temperature, are easy to wash and facilitate the direct visualization of RNA
8 and/or DNA immobilized on the solid support. Removal of RNA and/or DNA immobilized on the
9 solid support is also facilitated using slides.

10 **[00215]** The particular material selected as the solid support is not essential to the invention,
11 as long as it provides the described function. Normally, those who make or use the invention will
12 select the best commercially available material based upon the economics of cost and
13 availability, the expected application requirements of the final product, and the demands of the
14 overall manufacturing process.

15 **[00216]** Numerous methods are used for attachment of the nucleic acid members of the
16 invention to the substrate (a process referred as spotting). For example, polynucleotides are
17 attached using the techniques of, for example U.S. Pat. No. 5,807,522, which is incorporated
18 herein by reference for teaching methods of polymer attachment. Alternatively, spotting is carried
19 out using contact printing technology.

20 **[00217]** The amount of polynucleotide present in each composition will be sufficient to provide
21 for adequate hybridization and detection of target polynucleotide sequences during the assay in
22 which the array is employed. Generally, the amount of each nucleic acid member stably
23 associated with the solid support of the array is at least about 0.1 ng, preferably at least about
24 0.5 ng and more preferably at least about 1 ng, where the amount may be as high as 1000 ng or
25 higher, but will usually not exceed about 20 ng.

26 **[00218]** Control polynucleotides may be spotted on the array and used as target expression
27 control polynucleotides and mismatch control nucleotides to monitor non-specific binding or
28 cross-hybridization to a polynucleotide in the sample other than the target to which the probe is
29 directed. Mismatch probes thus indicate whether a hybridization is specific or not. For example, if
30 the target is present the perfectly matched probes should be consistently brighter than the
31 mismatched probes. In addition, if all central mismatches are present, the mismatch probes are
32 used to detect a mutation.

33 **[00219]** Target Preparation

34 **[00220]** The targets for the microarrays, may be derived from one or more biological samples.
35 It may be desirable to amplify the target nucleic acid sample prior to hybridization. One of skill in
36 the art will appreciate that whatever amplification method is used, if a quantitative result is

1 desired, care must be taken to use a method that maintains or controls for the relative
2 frequencies of the amplified polynucleotides. Methods of "quantitative" amplification are well
3 known to those of skill in the art. For example, quantitative PCR involves simultaneously co-
4 amplifying a known quantity of a control sequence using the same primers. This provides an
5 internal standard that may be used to calibrate the PCR reaction. The high density array may
6 then include probes specific to the internal standard for quantification of the amplified
7 polynucleotide. Detailed protocols for quantitative PCR are provided in PCR Protocols, A Guide
8 to Methods and Applications, Innis et al., Academic Press, Inc. N.Y., (1990). Other suitable
9 amplification methods include, but are not limited to polymerase chain reaction (PCR) (Innis, et
10 al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego,
11 (1990)), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren,
12 et al., Science, 241: 1077 (1988) and Barringer, et al., Gene, 89: 117 (1990), transcription
13 amplification (Kwoh, et al., Proc. Natl. Acad. Sci. USA, 86: 1173 (1989)), and self-sustained
14 sequence replication (Guatelli, et al., Proc. Nat. Acad. Sci. USA, 87: 1874 (1990)).

15 **[00221]** The invention provides for labeled target or labeled probe as described above. For
16 the microarrays, any analytically detectable marker that is attached to or incorporated into a
17 molecule may be used in the invention. An analytically detectable marker refers to any molecule,
18 moiety or atom which is analytically detected and quantified. Detectable labels suitable for use in
19 the present invention include any composition detectable by spectroscopic, photochemical,
20 biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present
21 invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g.,
22 DynabeadsTM), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent
23 protein, and the like), radiolabels (e.g., 3H, 125I, 35S, 14C, or 32P), enzymes (e.g., horseradish
24 peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric
25 labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex,
26 etc.) beads. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752;
27 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

28 **[00222]** Means of detecting such labels are well known to those of skill in the art. Thus, for
29 example, radiolabels may be detected using photographic film or scintillation counters,
30 fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic
31 labels are typically detected by providing the enzyme with a substrate and detecting the reaction
32 product produced by the action of the enzyme on the substrate, and colorimetric labels are
33 detected by simply visualizing the colored label.

34 **[00223]** The labels may be incorporated by any of a number of means well known to those of
35 skill in the art. However, in a preferred embodiment, the label is simultaneously incorporated
36 during the amplification step in the preparation of the sample polynucleotides. Thus, for example,
37 polymerase chain reaction (PCR) with labeled primers or labeled nucleotides will provide a

1 labeled amplification product. In a preferred embodiment, transcription amplification, as
2 described above, using a labeled nucleotide (e.g. fluorescein-labeled UTP and/or CTP)
3 incorporates a label into the transcribed polynucleotides. Alternatively, a label may be added
4 directly to the original polynucleotide sample (e.g., mRNA, polyA mRNA, cDNA, etc.) or to the
5 amplification product after the amplification is completed. Means of attaching labels to
6 polynucleotides are well known to those of skill in the art and include, for example nick translation
7 or end-labeling (e.g. with a labeled RNA) by kinasing of the polynucleotide and subsequent
8 attachment (ligation) of a polynucleotide linker joining the sample polynucleotide to a label (e.g.,
9 a fluorophore).

10 **[00224]** In a preferred embodiment, the target will include one or more control molecules
11 which hybridize to control probes on the microarray to normalize signals generated from the
12 microarray. Labeled normalization targets are polynucleotide sequences that are perfectly
13 complementary to control oligonucleotides that are spotted onto the microarray as described
14 above. The signals obtained from the normalization controls after hybridization provide a control
15 for variations in hybridization conditions, label intensity, "reading" efficiency and other factors that
16 may cause the signal of a perfect hybridization to vary between arrays.

17 **[00225]** Image Acquisition and Data Analysis

18 **[00226]** Following hybridization and any washing step(s) and/or subsequent treatments of a
19 conventional nature, the resultant hybridization pattern is detected. In detecting or visualizing the
20 hybridization pattern, the intensity or signal value of the label will be not only be detected but
21 quantified, by which is meant that the signal from each spot of the hybridization will be measured
22 and compared to a unit value corresponding to the signal emitted by a known number of end
23 labeled target polynucleotides to obtain a count or absolute value of the copy number of each
24 end-labeled target that is hybridized to a particular spot on the array in the hybridization pattern.

25 **[00227]** Methods for analyzing the data collected from hybridization to arrays are well known
26 in the art. For example, where detection of hybridization involves a fluorescent label, data
27 analysis can include the steps of determining fluorescent intensity as a function of substrate
28 position from the data collected, removing outliers, i.e., data deviating from a predetermined
29 statistical distribution, and calculating the relative binding affinity of the test polynucleotides from
30 the remaining data. The resulting data is displayed as an image with the intensity in each region
31 varying according to the binding affinity between associated oligonucleotides and/or
32 polynucleotides and the test polynucleotides.

33 **[00228]** Diagnostic Tests

34 **[00229]** Following detection or visualization, the hybridization pattern is used to determine
35 quantitative information about the genetic profile of the labeled target polynucleotide sample that
36 was contacted with the array to generate the hybridization pattern, as well as the state or

condition of the tissue, fluid, organs, cell, etc. from which the sample was derived. In this regard, the invention further provides for diagnostic tests for detecting cancer. The invention also provides for monitoring a patient's condition. According to the method of the invention, the presence of cancer is detected by obtaining a biological sample from a patient. A test sample comprising nucleic acid is prepared from the biological sample. The nucleic acid extracted from the sample is hybridized to an array comprising a solid substrate and a plurality of nucleic acid members, wherein each member is indicative of the presence of disease or a predisposition to cancer. According to this diagnostic test, hybridization of the sample comprising nucleic acid to one or more nucleic acid members on the array is indicative of cancer or a predisposition to a cancer.

[00230] *Diagnostic Monitoring*

[00231] The methods of the present invention may further comprise the step of recommending a monitoring regime or course of therapy based on the outcome of one or more assays. This allows clinicians to practice personalized medicine; e.g. cancer therapy, by monitoring the progression of the patient's cancer (such as by recognizing when an initial or subsequent mutation occurs) or treatment (such as by recognizing when a mutation is stabilized).

[00232] With knowledge of the boundaries of the sequence variation in hand, the information can be used to diagnose a pre-cancerous condition or existing cancer condition. Further, by quantitating the amount of aberrant mtDNA in successive samples over time, the progression of a cancer condition can be monitored. For example, data provided by assaying the patient's tissues at one point in time to detect a first set of mutations from wild-type could be compared against data provided from a subsequent assay, to determine if changes in the aberration have occurred.

[00233] Where a mutation is found in an individual who has not yet developed symptoms of cancer, the mutation may be indicative of a genetic susceptibility to develop a cancer condition. A determination of susceptibility to disease or diagnosis of its presence can further be evaluated on a qualitative basis based on information concerning the prevalence, if any, of the cancer condition in the patient's family history and the presence of other risk factors, such as exposure to environmental factors and whether the patient's cells also carry a mutation of another sort.

[00234] *Biological Sample*

[00235] The present invention provides for diagnostic tests which involve obtaining or collecting one or more biological samples. In the context of the present invention, "biological sample" refers to a tissue or bodily fluid containing cells from which mtDNA, mtRNA and translation products or fusion proteins can be obtained. For example, the biological sample can be derived from tissue including, but not limited to, skin, lung, breast, prostate, nervous, muscle, heart, stomach, colon, rectal tissue and the like; or from blood, saliva, cerebral spinal fluid, sputa,

1 urine, mucous, synovial fluid, peritoneal fluid, amniotic fluid and the like. The biological sample
2 may be obtained from a cancerous or non-cancerous tissue and may be, but is not limited to, a
3 surgical specimen or a biopsy specimen.

4 **[00236]** The biological sample can be used either directly as obtained from the source or
5 following a pre-treatment to modify the character of the sample. Thus, the biological sample can
6 be pre-treated prior to use by, for example, preparing plasma or serum from blood, disrupting
7 cells, preparing liquids from solid materials, diluting viscous fluids, filtering liquids, distilling
8 liquids, concentrating liquids, inactivating interfering components, adding reagents, and the like.

9 **[00237]** One skilled in the art will understand that more than one sample type may be
10 assayed at a single time (i.e. for the detection of more than one cancer). Furthermore, where a
11 course of collections are required, for example, for the monitoring of cancer over time, a given
12 sample may be diagnosed alone or together with other samples taken throughout a test period.
13 In this regard, biological samples may be taken once only, or at regular intervals such as
14 biweekly, monthly, semi-annually or annually.

15 **[00238] Kits**

16 **[00239]** The present invention provides diagnostic/screening kits for detecting cancer in a
17 clinical environment. Such kits may include one or more sampling means, in combination with
18 one or more probes according to the present invention. Alternatively, or in addition thereto, the kit
19 may comprise means for detecting a translation product of the invention.

20 **[00240]** The kits can optionally include reagents required to conduct a diagnostic assay, such
21 as buffers, salts, detection reagents, and the like. Other components, such as buffers and
22 solutions for the isolation and/or treatment of a biological sample, may also be included in the kit.
23 One or more of the components of the kit may be lyophilised and the kit may further comprise
24 reagents suitable for the reconstitution of the lyophilised components.

25 **[00241]** Where appropriate, the kit may also contain reaction vessels, mixing vessels and
26 other components that facilitate the preparation of the test sample. The kit may also optionally
27 include instructions for use, which may be provided in paper form or in computer-readable form,
28 such as a disc, CD, DVD or the like.

29 **[00242]** In one embodiment of the invention there is provided a kit for diagnosing cancer
30 comprising sampling means and a hybridization probe of the invention.

31 **[00243]** In another embodiment, the kits of the present invention may comprise an
32 immunological assay. In such case, the kits may comprise antibodies or antigen-binding
33 fragments having specificity towards the fusion proteins described herein. It will be understood
34 that various other reagents, test strips etc. required for such immunological assay will be
35 contained in the kits as will the required instructions to users.

1 EXAMPLES

2 **[00244]** Various aspects of the invention will be described by illustration using the following
3 examples. The examples provided herein serve only to illustrate certain specific embodiments of
4 the invention and are not intended to limit the scope of the invention in any way.

5 **[00245]** Example 1: Detection of Mitochondrial Fusion Transcripts

6 **[00246]** The mitochondrial 4977 “common deletion” and a 3.4kb deletion previously identified
7 by the present Applicant in PCT application no. PCT/CA2007/001711 (published under number
8 WO 2009/039601, the entire contents of which are incorporated by reference) result in unique
9 open reading frames having active transcripts as identified by oligo-dT selection in prostate
10 tissue (Figures 2 and 3). Examination of breast tissue samples also reveals the presence of a
11 stable polyadenylated fusion transcript resulting from the 3.4kb deletion (Figure 4).

12 **[00247]** *Reverse transcriptase-PCR protocol for deletion transcript detection*

13 **[00248]** *RNA isolation cDNA synthesis*

14 **[00249]** Total RNA was isolated from snap frozen prostate and breast tissue samples (both
15 malignant and normal samples adjacent to tumours) using the AurumTM Total RNA Fatty and
16 Fibrous Tissue kit (Bio-Rad, Hercules, CA) following the manufacturer's instructions. Since in
17 this experiment, genomic DNA contamination was to be avoided, a DNase I treatment step was
18 included, using methods as commonly known in the art. RNA quantity and quality were
19 determined with an ND-1000 spectrophotometer (NanoDrop® technologies). From a starting
20 material of about 100g, total RNA concentrations varied from 100 – 1000ng/ul with a 260/280
21 ratio between 1.89 – 2.10. RNA concentrations were adjusted to 100ng/ul and 2ul of each
22 template were used for first strand DNA synthesis with SuperScriptTM First-Strand Synthesis
23 System for RT-PCR (Invitrogen) following the manufacturer's instructions. In order to identify
24 stable polyadenylated fusion transcripts, Oligo(dT) primers that target transcripts with poly-A tails
25 were used.

26 **[00250]** *PCR*

27 **[00251]** Real time PCR was performed using 5ul of each cDNA template with the iQTM SYBR[®]
28 Green Supermix (Bio-Rad, Hercules, CA) on DNA Engine Opticon[®] 2 Continuous Fluorescence
29 Detection System (Bio-Rad, Hercules, CA). The primer pairs targeting the 4977bp deletion are;
30 8416F 5'- CCTTACACTATTCCTCATCAC- 3', 13637R 5'- TGACCTGTTAGGGTGAGAAG - 3',
31 and those for the 3.4 kb deletion are; ND4LF 5'- TCGCTCACACCTCATATCCTC -3', ND5R 5'-
32 TGTGATTAGGAGTAGGGTTAGG -3'. The reaction cocktail included: 2X SYBR[®] Green
33 Supermix (100mM KCL, 40mM Tris-HCl, pH 8.4, 0.4mM of each dNTP [dATP, dCTP, dGTP, and
34 dTTP], iTaqTM DNA polymerase, 50 units/ml, 6mM MgCl₂, SYBR[®] Green 1, 20nM fluorescein, and
35 stabilizers), 250nM each of primers, and ddH₂O. PCR cycling parameters were as follows; (1)
36 95°C for 2 min, (2) 95°C for 30 sec, (3) 55°C (for the 4977bp deletion) and 63°C (for the 3.4 kb

deletion) for 30 sec , (4) 72°C for 45 sec, (5) plate read, followed by 39 cycles of steps 3 to 5, and final incubation at 4°C. Apart from cycling threshold and melting curve analysis, samples were run on agarose gels for specific visualization of amplification products (see Figures 2 to 4).

[00252] Figure 2 is an agarose gel showing polyadenalated fusion transcripts in prostate samples invoked by the loss of 3.4kb from the mitochondrial genome. Legend for Figure 2: B-blank, Lanes 1-6 transcripts detected in cDNA; lanes 7-12 no reverse transcriptase (RT) controls for samples in lanes 1-6.

[00253] Figure 3 shows polyadenalated fusion transcripts in prostate samples invoked by the loss of the 4977kb common deletion. Legend for Figure 3: B-blank, Lanes 1-6 transcripts detected in cDNA; lanes 7-12 no RT controls for samples in lanes 1-6.

[00254] Figure 4 shows polyadenalated fusion transcripts in breast samples invoked by the loss of 3.4kb from the mtgenome. Legend for Figure 4: Lanes 2-8 transcripts from breast cDNAs; lane 9 negative (water) control; lanes 10 and 11, negative, no RT, controls for samples in lanes 2 and 3.

[00255] These results demonstrate the existence of stable mitochondrial fusion transcripts.

[00256] Example 2: Identification and Targetting of Fusion Products

[00257] Various hybridization probes were designed to detect, and further demonstrate the presence of novel transcripts resulting from mutated mitochondrial genomes, such as the 3.4kb deletion. For this purpose, a single-plex branched DNA platform for quantitative gene expression analysis (QuantiGene 2.0™, Panomics™) was utilized. The specific deletions and sequences listed in this example are based on their relative positions with the entire mtDNA genome, which is recited in SEQ ID NO: 1. The nucleic acid sequences of the four transcripts to which the probes were designed in this example are identified herein as follows: Transcript 1 (SEQ ID NO: 19), Transcript 2 (SEQ ID NO: 20), Transcript 3 (SEQ ID NO: 21) and Transcript 4 (SEQ ID NO: 22).

[00258] An example of a continuous transcript from the 3.4kb mitochondrial genome deletion occurs with the genes ND4L (NADH dehydrogenase subunit 4L) and ND5 (NADH dehydrogenase subunit 5). A probe having a complementary sequence to SEQ ID NO: 20, was used to detect transcript 2. The repetitive elements occur at positions 10745-10754 in ND4L and 14124-14133 in ND5.

[00259] The 3.4kb deletion results in the removal of the 3' end of ND4L, the full ND4 gene, tRNA histidine, tRNA serine2, tRNA leucine2, and the majority of the 5' end of ND5 (see Figure 5a), resulting in a gene splice of ND4L and ND5 with a junction point of 10744(ND4L):14124(ND5) (Figure 5b).

[00260] By starting at the original initiation codon of the first gene, ND4L, the amino acid sequence was translated until a termination codon occurs. In this example the termination codon is the original termination codon of ND5. Therefore, despite splicing two genes together, the reading frame is kept intact resulting in a hypothetical or predicted transcript that is 100 amino acids (or 300 bp) in length. This fusion protein transcript product is identified herein as SEQ ID NO: 37. The nucleotide sequence (SEQ ID NO: 3) encoding such protein corresponds to the mitochondrial genome positions of 10470-10744:14124-14148. SEQ ID NO: 3 is the complementary DNA sequence to the RNA transcript (SEQ ID NO: 20) detected in the manner described above.

[00261] Similarly, transcript 1 is a fusion transcript between ATPase 8 and ND5 associated with positions 8469:13447 (SEQ ID NO: 19). Transcripts 3 and 4 (SEQ ID NO: 21 and SEQ ID NO: 22, respectively) are fusion transcripts between COII and Cytb associated with nucleotide positions 7974:15496 and 7992:15730 respectively. Table 3 provides a summary of the relationships between the various sequences used in this example. Table 3 includes the detected fusion transcript, the DNA sequence complementary to the fusion transcript detected and hypothetical translation products for each transcript.

[00262] Example 3: Application to Prostate Cancer

[00263] Using the four fusion transcripts, i.e. transcripts 1 to 4, discussed above, two prostate tissue samples from one patient were analyzed to assess the quantitative difference of the novel predicted fusion transcripts. The results of the experiment are provided in Table 2 below, wherein "Homog 1" refers to the homogenate of frozen prostate tumour tissue from a patient and "Homog 2" refers to the homogenate of frozen normal prostate tissue adjacent to the tumour of the patient. These samples were processed according to the manufacturer's protocol (*QuantiGene® Sample Processing Kit for Fresh or Frozen Animal Tissues; and QuantiGene® 2.0 Reagent System User Manual*) starting with 25.8 mg of Homog 1 and 28.9 mg of Homog 2 (the assay setup is shown in Tables 5a and 5b).

[00264] Clearly demonstrated is an increased presence of mitochondrial fusion transcripts in prostate cancer tissue compared to normal adjacent prostate tissue. The fusion transcript is present in the normal tissue, although at much lower levels. The relative luminescence units (RLU) generated by hybridization of a probe to a target transcript are directly proportional to the abundance of each transcript. Table 2 also indicates the coefficients of variation, CV, expressed as a percentage, of the readings taken for the samples. The CV comprises the Standard deviation divided by the average of the values. The significance of such stably transcribed mitochondrial gene products in cancer tissue has implications in disease evolution and progression.

1 **[00265]** Example 4: Application to Breast Cancer

2 **[00266]** Using the same protocol from Example 3 but focusing only on Transcript 2, the novel
 3 fusion transcript associated with the 3.4kb mtgenome deletion, analyses were conducted on two
 4 samples of breast tumour tissue and two samples of tumour-free tissues adjacent to those
 5 tumours, as well as three samples of prostate tumour tissue, one sample comprising adjacent
 6 tumour-free tissue. Results for this example are provided in Table 4. The prostate tumour tissue
 7 sample having a corresponding normal tissue section demonstrated a similar pattern to the
 8 prostate sample analyzed in Example 3 in that the tumour tissue had approximately 2 times the
 9 amount of the fusion transcript than did the normal adjacent tissue. The breast tumour samples
 10 demonstrated a marked increase in the fusion transcript levels when compared to the adjacent
 11 non-tumour tissues. A 1:100 dilution of the homogenate was used for this analysis as it
 12 performed most reproducibly in the experiment cited in Example 3.

13 **[00267]** Thus, the above discussed results illustrate the application of the transcripts of the
 14 invention in the detection of tumours of both prostate and breast tissue.

15 **[00268]** Example 5: Application to Colorectal Cancer

16 **[00269]** This study sought to determine the effectiveness of several transcripts of the
 17 invention in detecting colorectal cancer. A total of 19 samples were prepared comprising nine
 18 control (benign) tissue samples (samples 1 to 9) and ten tumour (malignant) tissue samples
 19 (samples 10 to 19). The samples were homogenized according to the manufacturer's
 20 recommendations (Quantigene® Sample Processing Kit for Fresh or Frozen Animal Tissues; and
 21 Quantigene 2.0 Reagent System User Manual). Seven target transcripts and one housekeeper
 22 transcript were prepared in the manner as outlined above in previous examples. The
 23 characteristics of the transcripts are summarized as follows:

24 **[00270]** Table 7: Characteristics of Breast Cancer Transcripts

| Transcript ID | Junction Site | Gene Junction |
|--|---------------|---------------|
| 2 | 10744:14124 | ND4L:ND5 |
| 3 | 7974:15496 | COII:Cytb |
| 10 | 7438:13476 | COI:ND5 |
| 11 | 7775:13532 | COII:ND5 |
| 12 | 8213:13991 | COII:ND5 |
| Peptidylpropyl isomerase B (PPIB) ("housekeeper") | N/A | N/A |

25

26 **[00271]** It is noted that transcripts 2 and 3 are the same as those discussed above with
 27 respect to Examples 3 and 4.

1 **[00272]** Homogenates were prepared using approximately 25mg of tissue from OCT blocks
 2 and diluted 1:1 for transcripts 2 and 4, and 1:8 for transcripts 10 and 11. The quantity of the
 3 transcripts was measured in Relative Luminescence Units RLU on a Glomax™ Multi Detection
 4 System (Promega). All samples were assayed in triplicate for each transcript. Background
 5 measurements (no template) were done in triplicate as well. The analysis accounted for
 6 background by subtracting the lower limit from the RLU values for the samples. Input RNA was
 7 accounted for by using the formula $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$ where a is the target fusion transcript
 8 and h is the housekeeper transcript.

9 **[00273]** The analysis of the data comprised the following steps:

10 **[00274]** a) Establish CV's (coefficients of variation) for triplicate assays; acceptable if \leq
 11 15%.

12 **[00275]** b) Establish average RLU value for triplicate assays of target fusion transcript(a)
 13 and housekeeper transcript (h).

14 **[00276]** c) Establish lower limit from triplicate value of background RLU (l).

15 **[00277]** d) Subtract lower limit (l) from (a).

16 **[00278]** e) Calculate $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$.

17 **[00279]** Summary of Results:

18 **[00280]** The results of the above analysis are illustrated in Figures 6a to 6g, which comprise
 19 plots of the $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$ against sample number. Also illustrated are the respective
 20 ROC (Receiver Operating Characteristic) curves determined from the results for each transcript.

21 **[00281]** Transcript 2: There exists a statistically significant difference between the
 22 means ($p < 0.10$) of the normal and malignant groups ($p > 0.09$), using a cutoff value of 3.6129 as
 23 demonstrated by the ROC curve results in a sensitivity of 60% and specificity of 89% and the
 24 area under the curve is 0.73 indicating fair test accuracy. The threshold value chosen may be
 25 adjusted to increase either the specificity or sensitivity of the test for a particular application.

26 **[00282]** Transcript 3: There exists a statistically significant difference between the
 27 means ($p < 0.05$) of the normal and malignant groups ($p = 0.03$), using a cutoff value of 4.0813 as
 28 demonstrated by the ROC curve results in a sensitivity of 60% and specificity of 78% and the
 29 area under the curve is 0.79 indicating fair to good test accuracy. The threshold value chosen
 30 may be adjusted to increase either the specificity or sensitivity of the test for a particular
 31 application.

32 **[00283]** Transcript 8: There exists a statistically significant difference between the
 33 means ($p < 0.1$) of the normal and malignant groups ($p = 0.06$). Using a cutoff value of -6.0975 as
 34 demonstrated by the ROC curve results in a sensitivity of 60% and specificity of 89% and the

1 area under the curve is 0.76 indicating fair test accuracy. The threshold value chosen may be
2 adjusted to increase either the specificity or sensitivity of the test for a particular application.

3 **[00284]** Transcript 9: There exists a statistically significant difference between the
4 means ($p < 0.1$) of the normal and malignant groups ($p = 0.06$). Using a cutoff value of -7.5555 as
5 demonstrated by the ROC curve results in a sensitivity of 60% and specificity of 89% and the
6 area under the curve is 0.76 indicating fair to good test accuracy. The threshold value chosen
7 may be adjusted to increase either the specificity or sensitivity of the test for a particular
8 application.

9 **[00285]** Transcript 10: There is a statistically significant difference between the means
10 ($p \leq 0.01$) of the normal and malignant groups ($p = 0.01$). Using a cutoff value of -3.8272 as
11 demonstrated by the ROC curve results in a sensitivity of 90% and specificity of 67% and the
12 area under the curve is 0.84, indicating good test accuracy. The threshold value chosen may be
13 adjusted to increase either the specificity or sensitivity of the test for a particular application.

14 **[00286]** Transcript 11: There exists a statistically significant difference between the
15 means ($p < 0.1$) of the normal and malignant groups ($p = 0.06$), using a cutoff value of 3.1753 as
16 demonstrated by the ROC curve results in a sensitivity of 70% and specificity of 78% and the
17 area under the curve is 0.76 indicating fair to good test accuracy. The threshold value chosen
18 may be adjusted to increase either the specificity or sensitivity of the test for a particular
19 application.

20 **[00287]** Transcript 12: There exists a statistically significant difference between the
21 means ($p < 0.1$) of the normal and malignant groups ($p = 0.06$), using a cut-off value of 3.2626 as
22 demonstrated by the ROC curve results in a sensitivity of 70% and specificity of 78% and the
23 area under the curve is 0.76 indicating fair to good test accuracy. The threshold value chosen
24 may be adjusted to increase either the specificity or sensitivity of the test for a particular
25 application.

26 **[00288]** Conclusions:

27 **[00289]** The above results illustrate the utility of transcripts 2, 3, 8, 9, 10, 11, and 12 in the
28 detection of colorectal cancer and in distinguishing malignant from normal colorectal tissue. As
29 indicated above, transcripts 2 and 3 were also found to have utility in the detection of prostate
30 cancer. Transcript 2 was also found to have utility in the detection of breast cancer. Transcript
31 11 was also found to have utility in the detection of melanoma skin cancer. Transcript 10 was
32 also found to have utility in the detection of lung cancer and melanoma. Transcript 8 was also
33 found to have utility in the detection of lung cancer. Any of the 7 transcripts listed may be used
34 individually or in combination as a tool for the detection of characterization of colorectal cancer in
35 a clinical setting.

36 **[00290]** Example 6: Application to Lung Cancer

[00291] This study sought to determine the effectiveness of several transcripts of the invention in the detection of lung cancer. As in Example 5, nine control (benign) tissue samples (samples 1 to 9) and ten tumour (malignant) tissue samples (samples 10 to 19) were homogenized according to the manufacturer's recommendations (Quantigene® Sample Processing Kit for Fresh or Frozen Animal Tissues; and Quantigene 2.0 Reagent System User Manual). Homogenates were diluted 1:8 and the quantity of 4 target transcripts and 1 housekeeper transcript was measured in Relative Luminescence Units RLU on a Glomax™ Multi Detection System (Promega). All samples were assayed in triplicate for each transcript. Background measurements (no template) were done in triplicate as well.

[00292] The following transcripts were prepared for this example:

[00293] Table 8: Characteristics of Lung Cancer Transcripts

| Transcript ID | Junction Site | Gene Junction |
|--|---------------|---------------|
| 6 | 8828:14896 | ATPase6:Cytb |
| 8 | 6075:13799 | COI:ND5 |
| 10 | 7438:13476 | COI:ND5 |
| 20 | 8469:13447 | ATPase8:ND5 |
| Peptidylpropyl isomerase B (PPIB) ("housekeeper") | N/A | N/A |

[00294] The tissue samples used in this example had the following characteristics:

[00295] Table 9: Characteristics of Lung Cancer Samples

| Sample | Malignant | Comments (source of tissue) |
|--------|-----------|---|
| 1 | NO | interstitial lung disease |
| 2 | NO | emphysema |
| 3 | NO | aneurysm |
| 4 | NO | bronchopneumonia, COPD |
| 5 | NO | malignant neoplasm in liver, origin unknown, calcified granulomas in lung |
| 6 | NO | 12 hours post mortem, mild emphysema |
| 7 | NO | 12 hours post mortem, large B cell lymphoma, pulmonary edema, pneumonia |
| 8 | NO | pneumonia, edema, alveolar damage |
| 9 | NO | congestion and edema |
| 10 | YES | adenocarcinoma, non-small cell |
| 11 | YES | small cell |
| 12 | YES | squamous cell carcinoma, NSC, emphysema |
| 13 | YES | adenocarcinoma, lung cancer, nsc, metastatic |
| 14 | YES | squamous cell carcinoma, non-small cell |
| 15 | YES | mixed squamous and adenocarcinoma |
| 16 | YES | non-small cell carcinoma, squamous |
| 17 | YES | small cell carcinoma |
| 18 | YES | adenocarcinoma, lung cancer, nsc |
| 19 | YES | adenocarcinoma, lung cancer, nsc, metastatic |

1 **[00296]** The analysis of data was performed according to the method described in Example 5.
2 The results are illustrated in figures 7a, 7b, 7c and 7d.

3 **[00297]** Summary of Results:

4 **[00298]** Transcript 6: There exists a statistically significant difference between the
5 means ($p < 0.1$) of the normal (benign) and malignant groups ($p = 0.06$), using a cutoff value of -
6 6.5691 as demonstrated by the ROC curve results in a sensitivity of 80% and specificity of 71%
7 and the area under the curve is 0.77, indicating fair test accuracy. The threshold value chosen
8 may be adjusted to increase either the specificity or sensitivity of the test for a particular
9 application.

10 **[00299]** Transcript 8: The difference between the means of the normal and malignant
11 groups is statistically significant, $p < 0.05$ ($p = 0.02$). Using a cutoff value of -9.6166 as
12 demonstrated by the ROC curve results in a sensitivity of 90% and specificity of 86% and the
13 area under the curve is 0.86 indicating good test accuracy. The threshold value chosen may be
14 adjusted to increase either the specificity or sensitivity of the test for a particular application.

15 **[00300]** Transcript 10: The difference between the means of the normal and malignant
16 groups is statistically significant, $p \leq 0.01$ ($p = 0.01$). Using a cutoff value of -10.6717 as
17 demonstrated by the ROC curve results in a sensitivity of 90% and specificity of 86% and the
18 area under the curve is 0.89 indicating good test accuracy. The threshold value chosen may be
19 adjusted to increase either the specificity or sensitivity of the test for a particular application.

20 **[00301]** Transcript 20: The difference between the means of the normal and malignant
21 groups is statistically significant, $p \leq 0.1$ ($p = 0.1$). Using a cutoff value of 2.5071 as demonstrated
22 by the ROC curve results in a sensitivity of 70% and specificity of 71% and the area under the
23 curve is 0.74 indicating fair test accuracy. The threshold value chosen may be adjusted to
24 increase either the specificity or sensitivity of the test for a particular application.

25 **[00302]** Conclusions:

26 **[00303]** The results from example 6 illustrate the utility of transcripts 6, 8, 10, and 20 of the
27 invention in the detection of lung cancer tumours and the distinction between malignant and
28 normal lung tissues. Any of these three transcripts may be used for the detection or
29 characterization of lung cancer in a clinical setting.

30 **[00304]** Example 7: Application to Melanoma

31 **[00305]** This study sought to determine the effectiveness of several transcripts of the
32 invention in the detection of melanomas. In this study a total of 14 samples were used,
33 comprising five control (benign) tissue samples and nine malignant tissue samples. All samples
34 were formalin fixed, paraffin embedded (FFPE). The FFPE tissue samples were sectioned into
35 tubes and homogenized according to the manufacturer's recommendations (Quantigene® 2.0

1 Sample Processing Kit for FFPE Samples; and Quantigene 2.0 Reagent System User Manual)
 2 such that each sample approximated 20 microns prior to homogenization. Homogenates were
 3 diluted 1:4 and the quantity of 7 target transcripts and 1 housekeeper transcript was measured in
 4 Relative Luminescence Units RLU on a Glomax™ Multi Detection System (Promega). All
 5 samples were assayed in triplicate for each transcript. Background measurements (no template)
 6 were done in triplicate as well.

7 **[00306]** The 14 tissue samples used in this example had the following characteristics:

8 **[00307]** Table 10: Characteristics of Melanoma Cancer Samples

| Sample | Malignant | Comments (source of tissue) |
|--------|-----------|--|
| 1 | NO | breast reduction tissue (skin) |
| 2 | NO | breast reduction tissue (skin) |
| 3 | NO | breast reduction tissue (skin) |
| 4 | NO | breast reduction tissue (skin) |
| 5 | NO | breast reduction tissue (skin) |
| 6 | YES | lentigo maligna, (melanoma in situ) invasive melanoma not present |
| 7 | YES | invasive malignant melanoma |
| 8 | YES | nodular melanoma, pT3b, associated features of lentigo maligna |
| 9 | YES | residual superficial spreading invasive malignant melanoma, Clark's level II |
| 10 | YES | superficial spreading malignant melanoma, Clark's Level II |
| 11 | YES | nodular malignant melanoma, Clark's level IV |
| 12 | YES | superficial spreading malignant melanoma in situ, no evidence of invasion |
| 13 | YES | superficial spreading malignant melanoma, Clark's level II, focally present vertical phase |
| 14 | YES | superficial spreading malignant melanoma in situ, Clark's level I |

9

10 **[00308]** The following transcripts were prepared for this example:

11 **[00309]** Table 11: Characteristics of Melanoma Cancer Transcripts

| Transcript ID | Junction Site | Gene Junction |
|--|---------------|---------------|
| 6 | 8828:4896 | ATPase6:Cytb |
| 10 | 7438:13476 | COI:ND5 |
| 11 | 7775:13532 | COII:ND5 |
| 14 | 9191:12909 | ATPase6:ND5 |
| 15 | 9574:12972 | COIII:ND5 |
| 16 | 10367:12829 | ND3:ND5 |
| 20 | 8469:13447 | ATPase8:ND5 |
| Peptidylpropyl isomerase B (PPIB) ("housekeeper") | N/A | N/A |

12

13 **[00310]** As indicated, transcripts 10 and 11 were also used in Example 5. The analysis of
 14 data was performed according to the method described in Example 5. The results are illustrated
 15 in figures 8a -8g.

1 [00311] Summary of Results:

2 [00312] Transcript 6: There exists a statistically significant difference between the
3 means ($p \leq 0.01$) of the normal and malignant groups ($p = 0.01$). Further, using a cutoff value of -
4 5.9531 as demonstrated by the ROC curve results in a sensitivity of 89% and specificity of 80%
5 and the area under the curve is 0.96, indicating very good test accuracy. The threshold value
6 chosen may be adjusted to increase either the specificity or sensitivity of the test for a particular
7 application.

8 [00313] Transcript 10: There exists a statistically significant difference between the
9 means ($p \leq 0.05$) of the normal and malignant groups ($p = 0.05$), using a cutoff value of -4.7572as
10 demonstrated by the ROC curve results in a sensitivity of 89% and specificity of 40% and the
11 area under the curve is 0.82, indicating good test accuracy. The threshold value chosen may be
12 adjusted to increase either the specificity or sensitivity of the test for a particular application.

13 [00314] Transcript 11: There exists a statistically significant difference between the
14 means ($p < 0.05$) of the normal and malignant groups ($p = 0.02$). Further, using a cutoff value of
15 1.6762 as demonstrated by the ROC curve results in a sensitivity of 78% and specificity of 100%
16 and the area under the curve is 0.89, indicating good test accuracy. The threshold value chosen
17 may be adjusted to increase either the specificity or sensitivity of the test for a particular
18 application.

19 [00315] Transcript 14: There exists a statistically significant difference between the
20 means ($p \leq 0.05$) of the normal and malignant groups ($p = 0.05$). Further, using a cutoff value of -
21 4.9118 as demonstrated by the ROC curve results in a sensitivity of 89% and specificity of 60%
22 and the area under the curve is 0.82, indicating good test accuracy. The threshold value chosen
23 may be adjusted to increase either the specificity or sensitivity of the test for a particular
24 application.

25 [00316] Transcript 15: There exists a statistically significant difference between the
26 means ($p < 0.1$) of the normal and malignant groups ($p = 0.07$), using a cutoff value of -7.3107as
27 demonstrated by the ROC curve results in a sensitivity of 100% and specificity of 67% and the
28 area under the curve is 0.80, indicating good test accuracy. The threshold value chosen may be
29 adjusted to increase either the specificity or sensitivity of the test for a particular application.

30 [00317] Transcript 16: There exists a statistically significant difference between the
31 means ($p < 0.05$) of the normal and malignant groups ($p = 0.03$). Further, using a cutoff value of -
32 10.5963as demonstrated by the ROC curve results in a sensitivity of 89% and specificity of 80%
33 and the area under the curve is 0.878, indicating good test accuracy. The threshold value
34 chosen may be adjusted to increase either the specificity or sensitivity of the test for a particular
35 application.

[00318] Transcript 20: There exists a statistically significant difference between the means ($p < 0.05$) of the normal and malignant groups ($p = 0.04$). Further, using a cutoff value of - 8.3543 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity of 80% and the area under the curve is 0.89, indicating good test accuracy. The threshold value chosen may be adjusted to increase either the specificity or sensitivity of the test for a particular application.

[00319] Conclusions:

[00320] The results from example 7 illustrate the utility of transcripts 6, 10, 11, 14, 15, 16 and 20 of the invention in the detection of malignant melanomas. As indicated above, transcripts 10 and 11 were also found have utility in detecting colorectal cancer while transcript 6 has utility in the detection of lung cancer. A transcript summary by disease is provided at Table 6.

[00321] Example 8: Application to Ovarian Cancer

[00322] This study sought to determine the effectiveness of several transcripts of the invention in detecting ovarian cancer. A total of 20 samples were prepared comprising ten control (benign) tissue samples (samples 1 to 10) and ten tumour (malignant) tissue samples (samples 11 to 20). The samples were homogenized according to the manufacturer's recommendations (Quantigene® Sample Processing Kit for Fresh or Frozen Animal Tissues; and Quantigene 2.0 Reagent System User Manual). Eight target transcripts and one housekeeper transcript were prepared in the manner as outlined above in previous examples.

[00323] The 20 tissue samples used in this example had the following characteristics:

[00324] Table 12: Characteristics of Ovarian Cancer Samples

| Sample | Diagnosis | Comments |
|--------|-----------|--|
| 1 | Normal | follicular cyst |
| 2 | Normal | fibroma |
| 3 | Normal | No pathological change in ovaries |
| 4 | Normal | follicular cysts |
| 5 | Normal | cellular fibroma |
| 6 | Normal | benign follicular and simple cysts |
| 7 | Normal | leiomyomata, corpora albicantia |
| 8 | Normal | copora albicantia and an epithelial inclusions cysts |
| 9 | Normal | corpora albicantia |
| 10 | Normal | corpora albicantia, surface inclusion cysts, follicular cysts |
| 11 | Malignant | high grade poorly differentiated papillary serous carcinoma involving omentum |
| 12 | Malignant | endometrioid adenocarcinoma, well to moderately differentiated with focal serous differentiation |
| 13 | Malignant | papillary serous carcinoma |
| 14 | Malignant | mixed epithelial carcinoma predominantly papillary serous carcinoma |
| 15 | Malignant | High grade: serous carcinoma, papillary and solid growth patterns |
| 16 | Malignant | High Grade (3/3) Papillary serous carcinoma |
| 17 | Malignant | papillary serous carcinoma, high nuclear grade |

| | | |
|----|-----------|---|
| 18 | Malignant | Papillary serous cystadenocarcinomas Grade:III |
| 19 | Malignant | poorly differentiated papillary serous carcinoma |
| 20 | Malignant | Well-differentiated adnecarcinoma, Endometrioid type, Grade 1 |

1

2 **[00325]** The characteristics of the transcripts are summarized as follows:3 **[00326]** Table 13: Characteristics of Ovarian Cancer Transcripts

| Transcript ID | Junction Site | Gene Junction |
|---|---------------|---------------|
| 1 | 8469:13447 | ATPase8:ND5 |
| 2 | 10744:14124 | ND4L:ND5 |
| 3 | 7974:15496 | COII:Cytb |
| 6 | 8828:14896 | ATPase6:Cytb |
| 11 | 7775:13532 | COII:ND5 |
| 12 | 8213:13991 | COII:ND5 |
| 15 | 9574:12972 | COIII:ND5 |
| 20 | 8469:13447 | ATPase8:ND5 |
| Ribosomal Protein Large PO (LRP) Housekeeper | N/A | N/A |

4

5 **[00327]** It is noted that transcripts 1, 2, 3, 6, 11, 12, 15 and 20 are the same as those
6 discussed above with respect to Examples 3-7.

7 **[00328]** Homogenates were prepared using approximately 25mg of frozen tissue and diluted
8 1:4. The quantity of the transcripts was measured in Relative Luminescence Units RLU on a
9 Glomax™ Multi Detection System (Promega). All samples were assayed in triplicate for each
10 transcript. Background measurements (no template) were done in triplicate as well. The
11 analysis accounted for background by subtracting the lower limit from the RLU values for the
12 samples. Input RNA was accounted for by using the formula $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$ where a is
13 the target fusion transcript and h is the housekeeper transcript.

14 **[00329]** The analysis of the data comprised the following steps:

- 15 a) Establish CV's (coefficients of variation) for triplicate assays; acceptable if $\leq 15\%$.
16 b) Establish average RLU value for triplicate assays of target fusion transcript(a) and
17 housekeeper transcript (h).
18 c) Establish lower limit from triplicate value of background RLU (l).
19 d) Subtract lower limit (l) from (a).
20 e) Calculate $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$.

1 [00330] Summary of Results:

2 [00331] The results of the above analysis are illustrated in Figures 9a to 9h, which comprise
3 plots of the $\log_2 a$ RLU – $\log_2 h$ RLU against sample number. Also illustrated are the respective
4 ROC (Receiver Operating Characteristic) curves determined from the results for each transcript.

5 [00332] Transcript 1: There exists a statistically significant difference between the
6 means ($p < 0.05$) of the normal and malignant groups ($p = 0.002$). Using a cutoff value of -11.1503
7 as demonstrated by the ROC curve results in a sensitivity of 90% and specificity of 80% and the
8 area under the curve is 0.91 indicating very good test accuracy. The threshold value chosen
9 may be adjusted to increase either the specificity or sensitivity of the test for a particular
10 application.

11 [00333] Transcript 2: There exists a statistically significant difference between the
12 means ($p < 0.01$) of the normal and malignant groups ($p = 0.001$). Using a cutoff value of 0.6962 as
13 demonstrated by the ROC curve results in a sensitivity of 90% and specificity of 100% and the
14 area under the curve is 0.96 indicating very good test accuracy. The threshold value chosen
15 may be adjusted to increase either the specificity or sensitivity of the test for a particular
16 application.

17 [00334] Transcript 3: There exists a statistically significant difference between the
18 means ($p < 0.01$) of the normal and malignant groups ($p = 0.000$). Using a cutoff value of 0.6754
19 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity of 100% and
20 the area under the curve is 1.00 indicating excellent test accuracy. The threshold value chosen
21 may be adjusted to increase either the specificity or sensitivity of the test for a particular
22 application.

23 [00335] Transcript 6: There exists a statistically significant difference between the
24 means ($p < 0.01$) of the normal and malignant groups ($p = 0.007$). Using a cutoff value of -9.6479
25 as demonstrated by the ROC curve results in a sensitivity of 90% and specificity of 70% and the
26 area under the curve is 0.86 indicating good test accuracy. The threshold value chosen may be
27 adjusted to increase either the specificity or sensitivity of the test for a particular application.

28 [00336] Transcript 11: There is a statistically significant difference between the means
29 ($p < 0.01$) of the normal and malignant groups ($p = 0.000$). Using a cutoff value of -1.3794
30 demonstrated by the ROC curve results in a sensitivity of 100% and specificity of 90% and the
31 area under the curve is 0.99, indicating excellent test accuracy. The threshold value chosen may
32 be adjusted to increase either the specificity or sensitivity of the test for a particular application.

33 [00337] Transcript 12: There exists a statistically significant difference between the
34 means ($p < 0.01$) of the normal and malignant groups ($p = 0.001$). Using a cutoff value of -1.2379
35 as demonstrated by the ROC curve results in a sensitivity of 90% and specificity of 100% and the

area under the curve is 0.96 indicating excellent test accuracy. The threshold value chosen may be adjusted to increase either the specificity or sensitivity of the test for a particular application.

[00338] Transcript 15: There exists a statistically significant difference between the means ($p < 0.05$) of the normal and malignant groups ($p = 0.023$). Using a cut-off value of -8.6926 as demonstrated by the ROC curve results in a sensitivity of 70% and specificity of 80% and the area under the curve is 0.80 indicating good test accuracy. The threshold value chosen may be adjusted to increase either the specificity or sensitivity of the test for a particular application.

[00339] Transcript 20: There exists a statistically significant difference between the means ($p < 0.01$) of the normal and malignant groups ($p = 0.000$). Using a cut-off value of 0.6521 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity of 100% and the area under the curve is 0.76 indicating fair to good test accuracy. The threshold value chosen may be adjusted to increase either the specificity or sensitivity of the test for a particular application.

[00340] Conclusions:

[00341] The above results illustrate the utility of transcripts 1, 2, 3, 6, 11, 12, 15, and 20 in the detection of ovarian cancer and in distinguishing malignant from normal ovarian tissue.

Transcripts 1, 2 and 3 were also found to have utility in the detection of prostate cancer.

Transcript 6 was also found to have utility in the detection of melanoma and lung cancer.

Transcript 11 was also found to have utility in the detection of melanoma skin cancer, colorectal

cancer and testicular cancer. Transcript 12 was also found to have utility in the detection of

colorectal cancer and testicular cancer. Transcript 15 was also found to have utility in the

detection of melanoma and testicular cancer. Transcript 20 was also found to have utility in the

detection of colorectal cancer, melanoma, and testicular cancer. Any of the 8 transcripts listed

may be used individually or in combination as a tool for the detection or characterization of

ovarian cancer in a clinical setting.

[00342] Example 9: Application to Testicular Cancer

[00343] This study sought to determine the effectiveness of several transcripts of the invention in detecting testicular cancer. A total of 17 samples were prepared comprising eight control (benign) tissue samples (samples 1 to 8) and 9 tumour (malignant) tissue samples (samples 9 to 17), 5 of the malignant samples were non-seminomas (samples 9-13) and 4 were seminomas (samples 14-17). The samples were homogenized according to the manufacturer's recommendations (Quantigene® Sample Processing Kit for Fresh or Frozen Animal Tissues; and Quantigene 2.0 Reagent System User Manual). 10 target transcripts and one housekeeper transcript were prepared in the manner as outlined above in previous examples.

[00344] The 17 tissue samples used in this example had the following characteristics:

[00345] Table 14: Characteristics of Testicular Cancer Samples

| Sample | General Diagnosis | Stratified Malignant Diagnosis |
|--------|-------------------|--------------------------------|
| 1 | Benign | Benign |
| 2 | Benign | Benign |
| 3 | Benign | Benign |
| 4 | Benign | Benign |
| 5 | Benign | Benign |
| 6 | Benign | Benign |
| 7 | Benign | Benign |
| 8 | Benign | Benign |
| 9 | Malignant | Non-Seminoma |
| 10 | Malignant | Non-Seminoma |
| 11 | Malignant | Non-Seminoma |
| 12 | Malignant | Non-Seminoma |
| 13 | Malignant | Non-Seminoma |
| 14 | Malignant | Seminoma |
| 15 | Malignant | Seminoma |
| 16 | Malignant | Seminoma |
| 17 | Malignant | Seminoma |

1

2 **[00346]** The characteristics of the transcripts are summarized as follows:

3 **[00347]** Table 15: Characteristics of Testicular Cancer Transcripts

| Transcript ID | Junction Site | Gene Junction |
|-----------------------------------|---------------|---------------|
| 2 | 10744:14124 | ND4L:ND5 |
| 3 | 7974:15496 | COII:Cytb |
| 4 | 7992:15730 | COII:Cytb |
| 11 | 7775:13532 | COII:ND5 |
| 12 | 8213:13991 | COII:ND5 |
| 13 | 9144:13816 | ATPase6:ND5 |
| 15 | 9574:12972 | COIII:ND5 |
| 16 | 10367:12829 | ND3:ND5 |
| 20 | 8469:13447 | ATPase8:ND5 |
| Peptidylpropyl isomerase B (PPIB) | N/A | N/A |

4

5 **[00348]** It is noted that transcripts 2, 3, 4, 11, 12, 15, 16 and 20 are the same as those
6 discussed above with respect to Examples 3-8.

7 **[00349]** Homogenates were prepared using approximately 25mg of frozen tissue and diluted
8 1:4. The quantity of the transcripts was measured in Relative Luminescence Units RLU on a

1 Glomax™ Multi Detection System (Promega). All samples were assayed in triplicate for each
 2 transcript. Background measurements (no template) were done in triplicate as well. The
 3 analysis accounted for background by subtracting the lower limit from the RLU values for the
 4 samples. Input RNA was accounted for by using the formula $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$ where a is
 5 the target fusion transcript and h is the housekeeper transcript.

6 **[00350]** The analysis of the data comprised the following steps:

7 a) Establish CV's (coefficients of variation) for triplicate assays; acceptable if $\leq 15\%$.

8 b) Establish average RLU value for triplicate assays of target fusion transcript(a) and
 9 housekeeper transcript (h).

10 c) Establish lower limit from triplicate value of background RLU (l).

11 d) Subtract lower limit (l) from (a).

12 e) Calculate $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$.

13 **[00351]** Summary of Results:

14 **[00352]** The results of the above analysis are illustrated in Figures 10 to 18, which comprise
 15 plots of the $\log_2 a \text{ RLU} - \log_2 h \text{ RLU}$ against sample number. Also illustrated are the respective
 16 ROC (Receiver Operating Characteristic) curves determined from the results for each transcript.

17 **[00353]** While some transcripts distinguish between benign and malignant testicular tissue,
 18 others demonstrate distinction between the tumour subtypes of seminoma and non-seminoma
 19 and/or benign testicular tissue. It is therefore anticipated that combining transcripts from each
 20 class will facilitate not only detection of testicular cancer but also classification into subtype of
 21 seminoma or non-seminomas.

22 **[00354]** Transcript 2: There exists a statistically significant difference between the
 23 means ($p < 0.05$) of the normal group and the malignant seminomas ($p = 0.02$). Using a cutoff value
 24 of 1.5621 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity of
 25 100% and the area under the curve is 1.00 indicating excellent test accuracy. There also exists a
 26 statistically significant difference between the means ($p < 0.05$) of the malignant seminomas and
 27 the malignant non-seminomas ($p = 0.024$). Using a cutoff value of 2.1006 as demonstrated by the
 28 ROC curve results in a sensitivity of 100% and specificity of 80% and the area under the curve is
 29 0.90 indicating excellent test accuracy. The threshold value chosen may be adjusted to increase
 30 either the specificity or sensitivity of the test for a particular application.

31 **[00355]** Transcript 3: There exists a statistically significant difference between the
 32 means ($p < 0.05$) of the normal group and the malignant seminomas ($p = 0.018$). Using a cutoff
 33 value of 0.969 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity
 34 of 87.5% and the area under the curve is 0.969 indicating excellent accuracy. There also exists a
 35 statistically significant difference between the means ($p < 0.05$) of the malignant seminomas and
 36 the malignant non-seminomas ($p = 0.017$). Using a cutoff value of 1.8181 as demonstrated by the

1 ROC curve results in a sensitivity of 100% and specificity of 80% and the area under the curve is
2 0.9 indicating excellent test accuracy. The threshold value chosen may be adjusted to increase
3 either the specificity or sensitivity of the test for a particular application.

4 **[00356]** Transcript 4: There exists a statistically significant difference between the
5 means ($p<0.05$) of the normal and malignant groups ($p=0.034$). Using a cutoff value of -9.7628
6 as demonstrated by the ROC curve results in a sensitivity of 67% and specificity of 100% and the
7 area under the curve is 0.833 indicating good test accuracy. The threshold value chosen may be
8 adjusted to increase either the specificity or sensitivity of the test for a particular application.

9 **[00357]** Transcript 11: There exists a statistically significant difference between the
10 means ($p<0.05$) of the normal group and the malignant seminomas ($p=0.016$). Using a cutoff
11 value of 0.732 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity
12 of 100% and the area under the curve is 1.00 indicating excellent test accuracy. There also
13 exists a statistically significant difference between the means ($p<0.05$) of the malignant
14 seminomas and the malignant non-seminomas ($p=0.016$). Using a cutoff value of 0.9884 as
15 demonstrated by the ROC curve results in a sensitivity of 100% and specificity of 80% and the
16 area under the curve is 0.90 indicating excellent test accuracy. The threshold value chosen may
17 be adjusted to increase either the specificity or sensitivity of the test for a particular application.

18 **[00358]** Transcript 12: There exists a statistically significant difference between the
19 means ($p<0.1$) of the normal group and the malignant seminomas ($p=0.056$). Using a cutoff value
20 of 1.5361 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity of
21 87.5% and the area under the curve is 0.969 indicating excellent test accuracy. There also exists
22 a statistically significant difference between the means ($p<0.05$) of the malignant seminomas and
23 the malignant non-seminomas ($p=0.044$). Using a cutoff value of 1.6039 as demonstrated by the
24 ROC curve results in a sensitivity of 100% and specificity of 80% and the area under the curve is
25 0.9 indicating excellent test accuracy. The threshold value chosen may be adjusted to increase
26 either the specificity or sensitivity of the test for a particular application.

27 **[00359]** Transcript 13: There exists a statistically significant difference between the
28 means ($p<0.05$) of the normal group and the malignant group ($p=0.019$). Using a cutoff value of -
29 9.8751 as demonstrated by the ROC curve results in a sensitivity of 87.5% and specificity of 78%
30 and the area under the curve is 0.875 indicating very good test accuracy. There also exists a
31 statistically significant difference between the means ($p<0.01$) of the malignant non-seminomas
32 and the benign group ($p=0.000$). Using a cutoff value of -13.9519 as demonstrated by the ROC
33 curve results in a sensitivity of 100% and specificity of 87.5% and the area under the curve is
34 0.975 indicating excellent test accuracy. There also exists a statistically significant difference
35 between the means ($p<0.01$) of the malignant seminomas and the malignant non-seminomas
36 ($p=0.001$). Using a cutoff value of -15.8501 as demonstrated by the ROC curve results in a
37 sensitivity of 100% and specificity of 100% and the area under the curve is 1.00 indicating

1 excellent test accuracy. The threshold value chosen may be adjusted to increase either the
2 specificity or sensitivity of the test for a particular application.

3 **[00360]** Transcript 15: There exists a statistically significant difference between the
4 means ($p < 0.1$) of the normal and malignant groups ($p = 0.065$). Using a cut-off value of -5.4916
5 as demonstrated by the ROC curve results in a sensitivity of 75% and specificity of 89% and the
6 area under the curve is 0.833 indicating good test accuracy. The threshold value chosen may be
7 adjusted to increase either the specificity or sensitivity of the test for a particular application.

8 **[00361]** Transcript 16: There exists a statistically significant difference between the
9 means ($p < 0.05$) of the normal and malignant groups including both seminomas and non-
10 seminomas ($p = 0.037$). Using a cut-off value of -6.448 as demonstrated by the ROC curve
11 results in a sensitivity of 89% and specificity of 75% and the area under the curve is 0.806
12 indicating good test accuracy. There also exists a statistically significant difference between the
13 means ($p < 0.05$) of the normal and malignant seminomas ($p = 0.037$). Using a cut-off value of -
14 7.4575 as demonstrated by the ROC curve results in a sensitivity of 100% and specificity of
15 87.5% and the area under the curve is 0.938 indicating excellent test accuracy. The threshold
16 value chosen may be adjusted to increase either the specificity or sensitivity of the test for a
17 particular application.

18 **[00362]** Transcript 20: There exists a statistically significant difference between the
19 means ($p < 0.01$) of the normal group and the malignant seminomas ($p = 0.006$). Using a cutoff
20 value of 1.8364 as demonstrated by the ROC curve results in a sensitivity of 100% and
21 specificity of 100% and the area under the curve is 1.00 indicating excellent test accuracy. There
22 also exists a statistically significant difference between the means ($p < 0.01$) of the malignant
23 seminomas and the malignant non-seminomas ($p = 0.004$). Using a cutoff value of 1.6065 as
24 demonstrated by the ROC curve results in a sensitivity of 100% and specificity of 100% and the
25 area under the curve is 1.00 indicating excellent test accuracy. The threshold value chosen may
26 be adjusted to increase either the specificity or sensitivity of the test for a particular application.

27 **[00363]** Conclusions:

28 **[00364]** The above results illustrate the utility of transcripts 2, 3, 4, 11, 12, 13, 15, 16, and 20
29 in the detection of testicular cancer, and testicular cancer subtypes, and in distinguishing
30 malignant from normal testicular tissue. Transcript 2 was also found to have utility in the
31 detection of prostate, breast, colorectal and ovarian cancer. Transcript 3 was also found to have
32 utility in the detection of prostate, breast, melanoma, colorectal, and ovarian cancers. Transcript
33 4 was also found to have utility in the detection of prostate and colorectal cancers. Transcript 11
34 was also found to have utility in the detection of colorectal, melanoma, and ovarian cancers.
35 Transcript 12 was also found to have utility in the detection of colorectal and ovarian cancers.
36 Transcript 15 was also found to have utility in the detection of melanoma and ovarian cancers.
37 Transcript 16 was also found to have utility in the detection of melanoma skin cancer. Transcript

20 was also found to have utility in the detection of colorectal cancer, melanoma, and ovarian cancer. Any of the 9 transcripts listed may be used individually or in combination as a tool for the detection or characterization of ovarian cancer in a clinical setting.

[00365] In one aspect, the invention provides a kit for conducting an assay for determining the presence of cancer in a tissue sample. The kit includes the required reagents for conducting the assay as described above. In particular, the kit includes one or more containers containing one or more hybridization probes corresponding to transcripts 1 to 17, and 20 described above. As will be understood, the reagents for conducting the assay may include any necessary buffers, salts, detection reagents etc. Further, the kit may include any necessary sample collection devices, containers etc. for obtaining the needed tissue samples, reagents or materials to prepare the tissue samples for example by homogenization or nucleic acid extraction, and for conducting the subject assay or assays. The kit may also include control tissues or samples to establish or validate acceptable values for diseased or non-diseased tissues.

[00366] Example 10: Detection of Fusion Protein

[00367] Cell lines

[00368] The presence of fusion proteins was investigated in two human prostate cell lines. Firstly the normal prostate cell line RWPE-1 (ATCC Cat# CRL-11609), these cells are non tumorigenic in nude mice and were established by infection with human papilloma virus 18 of histologically normal adult human prostate cells. Secondly a tumorigenic cell line WPE1-NA22 were examined (ATCC Cat# CRL-2849). These cells were derived from the RWPE-1 cells following exposure to N-methyl-N-nitrosourea. These cells are tumorigenic in nude mice unlike it's parent cell line RWPE-1.

[00369] Both cell lines were grown in Keratinocyte Serum Free Medium (Invitrogen Cat#17005-042), medium is supplemented with bovine pituitary extract and human recombinant epidermal growth factor. Cells were grown to 90% confluence then trypsinised using TrypLE Select (Invitrogen Cat#12563029). Cells were then counted using an automated counting system (Invitrogen Countess Cat#C10227), aliquots were then snap frozen and stored at -80°C.

[00370] Protein Extraction

[00371] Cell fractions were extracted from both RWPE1 and WPE1-NA22 cell lines using the Qproteome Mitochondria Isolation Kit (Qiagen Cat#37612). Both mitochondrial and cytoplasmic fractions were extracted from 1×10^7 cells. Protein concentration was then calculated using a fluorescent protein assay (Quant-IT Protein, Invitrogen Cat#Q33211) measured on a Qubit fluorometer (Invitrogen Cat#Q32857).

[00372] SDS-Page Gel electrophoresis

1 **[00373]** SDS-Page electrophoresis was carried out on mitochondrial and cytosolic fractions
 2 prepared using the Qproteome mitochondrial isolation kit. 20 μ g of protein was run in each lane
 3 on a 4-12% precast (Invitrogen Nupage Cat#NP0321) bis-tris gel reducing gel, using a MES
 4 running buffer (Invitrogen Cat# NP00020). The gel was stained overnight with colloidal blue gel
 5 stain (Invitrogen Cat# LC6025). The results are illustrated in Figure 19. The approximate size
 6 (kD) range of proteins predicted to be contained in each of the 8 gel slices illustrated in Figure 19
 7 are as follows:

| | |
|---|--------|
| 1 | 60-80 |
| 2 | 50-60 |
| 3 | 40-50 |
| 4 | 30-40 |
| 5 | 20-30 |
| 6 | 15-20 |
| 7 | 10-15 |
| 8 | 3.5-10 |

8 **[00374]** LCMS

9 **[00375]** Eight gel slices were cut out from each lane of a colloidal blue stained 1D SDS-PAGE
 10 (Figure 19) and in gel digested with trypsin following standard procedures.

11 **[00376]** The digestion products were eluted from the gel and evaporated. An aliquot was
 12 injected onto an LCMS system (Dionex / LC Packings Ultimate3000 coupled online to a Thermo
 13 LTQ XL orbitrap) and separated on a 25 cm (75 μ m ID) PepMap (Dionex) column at a flow rate
 14 of 300 ml/min with formic acid as a ion pairing agent and a linear gradient starting at 5% MeCN
 15 going to 40% MeCN over 110 min. MS spectra were collected in the orbitrap at a resolution of
 16 60000 (400 Da) and MSMS spectra in the linear ion trap at low resolution.

17 **[00377]** Data were processed using Thermo Proteome Discoverer to generate .mgf (mascot
 18 generic format) peak list files, which were submitted in house to X!Tandem, searching a custom
 19 database comprised of the human proteome (ensembl) and predicted fusion proteins based on
 20 the fusion transcripts described previously. To calculate a false discovery rate (FDR), the
 21 searched database also included the reverse sequence of all proteins.

22 **[00378]** Protein Complexes Analysis

23 **[00379]** Upon completion of the X!Tandem custom database search all identified proteins and
 24 fusion transcripts were returned. The proteins were scored by their $\log(e)^+$ values and classified
 25 as significant when the $\log(e)^+$ was less than negative one, with preference given to proteins with
 26 a $\log(e)^+$ less than negative three. Fusion proteins were identified by the presence of at least one
 27 peptide from each of the contributing genes of the fusion transcript present in the same gel slice.
 28 Protein sequence coverage from the identified peptides by the LC/MS-MS are displayed in red.
 29 The sequence of the protein which may be difficult to observe a peptide due to experimental

conditions are indicated in green. Finally, protein sequence that is displayed in black represents a neutral possibility of identifying a peptide.

[00380] Examples of Identified Fusion Proteins

[00381] Many mitochondrial fusion proteins were identified using this methodology. Four of such fusion proteins are described below as representative examples.

[00382] Example Fusion Protein 1

[00383] Figures 20a illustrates the amino acid sequence of the fusion protein corresponding to the fusion transcript identified as P0026, which was identified ($\log(e)^+ = -13.2$) in slice 7 of the mitochondrial NA22 cell line (Figure 19) from the presence of the Cytochrome c oxidase subunit 2 (CO2) N-terminus peptide ILYMTDEVNDPSLTIK and the NADH-ubiquinone oxidoreductase chain 3 (ND3) C-terminus peptide STPYECGFDPMS (Figure 20a).

[00384] The most C-terminus tryptic peptide of wild-type CO2, IFEMGPVFTL, was searched against all mitochondrial NA22 cell line gel slices .xml data. This peptide was only observed in gel slice 5 (Figure 19). This was further confirmed by identifying CO2 wild-type ($\log(e)^+ = -42.9$) (Figure 20b) in mitochondrial NA22 cell line gel slice 5 only after searching the Human (SwissProt) database (no fusion transcripts) with all gel slices.

[00385] Cytochrome c oxidase subunit 2 peptide ILYMTDEVNDPSLTIK was observed in gel slices 5 and 7. This indicates that wild type CO2, with a molecular weight of ~25kDa is present in the 20-30kDa gel slice 5, and a fragment of CO2 N-terminus exists in gel slice 7. The tryptic peptide STPYECGFDPMS from ND3 is only identified in gel slice 7 (10-15kDa), which identifies the wild-type gene (13kDa) and the C-terminus of P0026.

[00386] The sequences for the fusion transcript P0026, the mutant DNA from which it is derived and the resulting protein are provided herein, respectively, as SEQ ID NO: 56, SEQ ID NO: 57 and SEQ ID NO: 58.

[00387] Example Fusion Protein 2

[00388] Figure 21a illustrates the amino acid sequence of the fusion protein corresponding to fusion transcript P0062, which was identified ($\log(e)^+ = -41.2$) in slice 5 (20-30kDa), shown in Figure 19, of the mitochondrial NA22 cell line from the presence of the NADH dehydrogenase subunit 1 (ND1) N-terminus peptides KGPNVVGPYGLLQPFADAMK and YDQLMHLLWK and the ATP synthase subunit 6 C-terminus peptide LITTQQWLIK. All three peptides were identified in the mitochondrial NA22 cell line gel slice 5 (Figure 19) but due to the most C-terminus peptide of ND1 (YDQLMHLLWK) being present only in gel slice 5, the presence of both wild-type (Figure 21b) and the fusion protein corresponding to fusion transcript P0062 is possible.

1 **[00389]** The sequences for the fusion transcript P0062, the mutant DNA from which it is
 2 derived and the resulting protein are provided herein, respectively, as SEQ ID NO: 59, SEQ ID
 3 NO: 60 and SEQ ID NO: 61.

4 **[00390]** Example Fusion Protein 3

5 **[00391]** Figure 22 illustrates the amino acid sequence of the fusion protein corresponding to
 6 fusion transcript P0064, which was identified in slice 4 ($\log(e)^+ = -22$), shown in Figure 19, of the
 7 mitochondrial NA22 cell line with the peptide KGPNVVGPYGLLPFADAMK from the N-terminus
 8 of ND1 and the NADH dehydrogenase subunit 2 (ND2) C-terminus peptide WAIEEFTK. The
 9 ND1 C-terminus peptide YDQLMHLLWK was not observed in gel slice 4, and based on the
 10 expected sizes of P0064 and ND2 it is suggested that gel slice 4 contains P0064 and ND2.

11 **[00392]** The sequences for the fusion transcript P0064, the mutant DNA from which it is
 12 derived and the resulting protein are provided herein, respectively, as SEQ ID NO: 62, SEQ ID
 13 NO: 63 and SEQ ID NO: 64.

14 **[00393]** Example Fusion Protein 4

15 **[00394]** Figure 23a illustrates the amino acid sequence of the fusion protein corresponding to
 16 fusion transcript P0176, which was identified in slice 4 ($\log(e)^+ = -33.8$), shown in Figure 19, of
 17 the mitochondrial NA22 cell line with the peptide KGPNVVGPYGLLPFADAMK from the N-
 18 terminus of ND1 and the Cytochrome c oxidase subunit 1 (CO1) C-terminus peptides
 19 VFSWLATLHGSNMK and VLMVEEPSMNLEWLYGCPPPYHTFEPPVYMK. Both of the CO1
 20 peptides were only observed together in gel slice 4 (30-40kDa) of the mitochondrial NA22 cell
 21 line despite an expected size of 55 kDa. This was further confirmed by identifying CO1 wild-type
 22 ($\log(e)^+ = -14.6$) (Figure 23b) in mitochondrial NA22 cell line gel slice 4 only after searching the
 23 Human (SwissProt) database (no fusion transcripts) with all gel slices.

24 **[00395]** The only ND1 peptide observed in gel slice 4 was KGPNVVGPYGLLPFADAMK.
 25 Since the ND1 C-terminus peptide YDQLMHLLWK was not present, wild-type ND1 is not present
 26 in the slice, which supports the presence of P0176.

27 **[00396]** The sequences for the fusion transcript P0176, the mutant DNA from which it is
 28 derived and the resulting protein are provided herein, respectively, as SEQ ID NO: 65, SEQ ID
 29 NO: 66 and SEQ ID NO: 67.

30 **[00397]** Corresponding Fusion Transcripts

31 **[00398]** Quantitative measurements of the fusion transcripts associated with each of these
 32 four fusion proteins were conducted in a series of cell lines of which two were those used in the
 33 LC-MS/MS experiment, specifically RWPE-1 and WPE1-NA22, which is a malignant cell line with
 34 low invasive potential. The results of these measurements is illustrated in Figures 24a-24d,
 35 corresponding to the four proteins discussed above. In Figures 24a-d, cell line RWPE-1 is

1 indicated as NO and cell line WPE1-NA22 is indicated as LI. The additional cell lines included in
2 this experiment represent a continued progression of malignancy with moderate invasive
3 potential (MI), high invasive potential (HI), and very high invasive potential (VH).

4 **[00399]** The cells were lysed and assayed using custom probes specific to each of the fusion
5 transcripts on the branching DNA platform as described herein or previously in PCT application
6 no. PCT/CA2009/000351 (published under number WO 2009/117811), the entire contents of
7 which are incorporated herein by reference. Results indicated high levels of expression (with
8 RLU values ranging from 10^6 - 10^8). A general trend was observed in the quantity of each fusion
9 transcript in that the initial transformation from normal cells to malignant cells (NO-LI) was
10 punctuated by a marked change in quantity of the transcript, followed by either a continued
11 increase or continued decrease in the quantity as malignant progression proceeds from LI to
12 VH).

13
14 **[00400]** Although the invention has been described with reference to certain specific
15 embodiments, various modifications thereof will be apparent to those skilled in the art without
16 departing from the purpose and scope of the invention as outlined in the claims appended
17 hereto. Any examples provided herein are included solely for the purpose of illustrating the
18 invention and are not intended to limit the invention in any way. Any drawings provided herein
19 are solely for the purpose of illustrating various aspects of the invention and are not intended to
20 be drawn to scale or to limit the invention in any way. The disclosures of all prior art recited
21 herein are incorporated herein by reference in their entirety.

- 1 **[00401]** Bibliography
- 2 **[00402]** The following references, amongst others, were cited in the foregoing description.
- 3 The entire contents of these references are incorporated herein by way of reference thereto.

| <u>Author</u> | <u>Journal</u> | <u>Title</u> | <u>Volume</u> | <u>Date</u> |
|-------------------------|---------------------------|--|--------------------|-------------|
| Anderson et al | Nature | Sequence and Organization of the Human Mitochondrial Genome | 290(5806):457-65 | 1981 |
| Andrews et al | Nat Genet | Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. | 23(2):147 | 1999 |
| Modica-Napolitano et al | Expert Rev Mol Med | Mitochondria as targets for detection and treatment of cancer | 4:1-19 | 2002 |
| Sherratt et al | Clin Sci (Lond) | Mitochondrial DNA defects: a widening clinical spectrum of disorders. | 92(3):225-35 | 1997 |
| Croteau et al | Mutat Res | Mitochondrial DNA repair pathways. | 434(3):137-48 | 1999 |
| Green and Kroemer | J Clin Invest | Pharmacological manipulation of cell death: clinical applications in sight? | 115(10): 2610–2617 | 2005 |
| Dai et al | Acta Otolaryngol | Correlation of cochlear blood supply with mitochondrial DNA common deletion in presbycusis. | 24(2):130-6 | 2004 |
| Ro et al | Muscle Nerve | Deleted 4977-bp mitochondrial DNA mutation is associated with sporadic amyotrophic lateral sclerosis: a hospital-based case-control study. | 28(6):737-43 | 2003 |
| Barron et al | Invest Ophthalmol Vis Sci | Mitochondrial abnormalities in ageing macular photoreceptors. | 42(12):3016-22 | 2001 |
| Lewis et al | J Pathol | Detection of damage to the mitochondrial genome in the oncocyctic cells of Warthin's tumour. | 191(3):274-81 | 2000 |
| Muller-Hocker et al | Mod Pathol | The common 4977 base pair deletion of mitochondrial DNA preferentially accumulates in the cardiac conduction system of patients with Kearns-Sayre syndrome. | 11(3):295-301. | 1998 |
| Porteous et al | Eur J Biochem | Bioenergetic consequences of accumulating the common 4977-bp mitochondrial DNA deletion. | 257(1):192-201 | 1998 |
| Parr et al | J Mol Diagn | Somatic mitochondrial DNA mutations in prostate cancer and normal appearing adjacent glands in comparison to age-matched prostate samples without malignant histology. | 8(3):312-9. | 2006 |
| Maki et al | Am J Clin Pathol | Mitochondrial genome deletion aids in the identification of false- and true-negative prostate needle core biopsy specimens. | 129(1):57-66 | 2008 |
| Nakase et al | Am J Hum Genet | Transcription and translation of deleted mitochondrial genomes in Kearns-Sayre syndrome: implications for pathogenesis. | 46(3):418-27. | 1990 |
| Libura et al | Blood | Therapy-related acute myeloid leukemia-like MLL rearrangements are induced by etoposide in primary human CD34+ cells and remain stable after clonal expansion. | 105(5):2124-31 | 2005 |

| | | | | |
|---------------|--------------------------|---|-----------------|------|
| Meyer et al | Proc Natl Acad Sci U S A | Diagnostic tool for the identification of MLL rearrangements including unknown partner genes. | 102(2):449-54 | 2005 |
| Eguchi et al | Genes Chromosomes Cancer | MLL chimeric protein activation renders cells vulnerable to chromosomal damage: an explanation for the very short latency of infant leukemia. | 45(8):754-60 | 2006 |
| Hayashi et al | Proc Natl Acad Sci U S A | Introduction of disease-related mitochondrial DNA deletions into HeLa cells lacking mitochondrial DNA results in mitochondrial dysfunction | 88: 10614-10618 | 1991 |

Table 1: Known mitochondrial deletions having an ORF

| Deletion Junction (mt:nt) | Deletion Size (bp) | Repeat Location (mt:nt) | Number of Repeats | References |
|---------------------------|--------------------|-------------------------|-------------------|---|
| COX I - ND5 | | | | |
| 6075:13799 | -7723 | 6076-6084/13799-13007 | D, 9/9 | •Mita, S., Rizzuto, R., Moraes, C.T., Shanske, S., Arnaudo, E., Fabrizio, G.M., Koga, Y., DiMauro, S., Schon, E.A. (1990) "Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA" <i>Nucleic Acids Research</i> 18(3):561-567 |
| 0200:14100 | -7004 | 6235-6238/14099-14102 | D, 4/4 | •Bink, R.R., Timburn, D.R., Thompson, G.N., Dahl, H.H. (1995) "A lipoburnase II cleavage site is associated with a novel mitochondrial DNA deletion" <i>Human Genetics</i> 95 (1): 75-81 |
| 6325:13989 | -7663 | 6326-6341/13889-14004 | D, 16/17 | •Larsson, N.O., Holme, E., Kristianson, B., Oldfors, A., Tulinius, M. (1990) "Progressive increase of the mutated mitochondrial DNA fraction in Kearns-Sayre syndrome" <i>Pediatric Research</i> 28 (2): 131-136 |
| 6325:13989 | -7663 | 6326-6341/13889-14004 | D, 16/17 | •Larsson, N.O., Holme, E. (1992) "Multiple short direct repeats associated with single mtDNA deletions" <i>Biochimica et Biophysica Acta</i> 1133 (4): 311-314 |
| 6331:13994 | -7663 | 6331-6341/13994-14004 | D, 11/11 | •Mita, S., Rizzuto, R., Moraes, C.T., Shanske, S., Arnaudo, E., Fabrizio, G.M., Koga, Y., DiMauro, S., Schon, E.A. (1990) "Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA" <i>Nucleic Acids Research</i> 18(3):561-567 |
| COX II - ND5 | | | | |
| 7829:14135 | -6305 | 7829-7829/14129-14134 | D, 6/6 | •Bet, L., Moggio, M., Comi, G.P., Mariani, C., Prella, A., Checcarelli, N., Bordoni, A., Bresolin, N., Scarpi, E., Scarlato, G. (1994) "Multiple sclerosis and mitochondrial myopathy: an unusual combination of diseases" <i>Journal of Neurology</i> 241 (8): 511-516 |
| 0210:10991 | -5777 | 8214-8220/13991-13997 | D, 7/7 | •Iinokio, Y., Suzuki, S., Konata, K., Ohtomo, M., Onoda, M., Matsumoto, M., Hirai, S., Sato, Y., Akai, H., Abe, K., Tomyita, T. (1995) "A new mitochondrial DNA deletion associated with diabetic amyotrophy, diabetic myopathy and diabetic fatty liver" <i>Muscle and Nerve</i> 18 (9): 1142-1149 |
| ATPase - ND5 | | | | |
| 0001:10510 | -4001 | 8625-8631/13506-13512 | D, 7/7 | •Zhang, C., Baumer, A., Mackay, I.R., Linnane, A.W., Nagley, P. (1995) "Unusual pattern of mitochondrial DNA deletions in skeletal muscle of an adult human with chronic fatigue syndrome" <i>Human Molecular Genetics</i> 4 (4): 751-754 |
| 9144:13816 | 4671 | 9137-9144/13000-13815 | D, 8/8 | •Ota, Y., Tanaka, M., Sato, W., Ohno, K., Yamamoto, T., Maehara, M., Negoro, T., Watanabe, K., Aoyama, S., Ozawa, T. (1991) "Detection of platelet mitochondrial DNA deletions in Kearns-Sayre syndrome" <i>Investigative Ophthalmology and Visual Science</i> 32 (10): 2007-2015 |
| 9191:12909 | -3717 | 9189-9191/12906-12908 | D, 3/3 | •Tanaka, M., Sato, W., Ohno, K., Yamamoto, T., Ozawa, T. (1989) "Direct sequencing of mitochondrial DNA in myopathic patients" <i>Biochemical and Biophysical Research Communications</i> 164 (1): 156-163 |
| COX III - ND5 | | | | |

| | | | | |
|--------------------------|-------|-------------------------|---------|--|
| 10190:13753 | -3502 | 10191-10198/13753-13760 | D, 0/0 | <p>•Rotig, A., Bourdaron, T., Christien, D., Rustin, P., Munnich, A. (1995) "Spectrum of mitochondrial DNA rearrangements in the Pearson marrow pancreas syndrome" Human Molecular Genetics 4 (8): 1327-1330</p> <p>•Rotig, A., Cormier, V., Koll, F., Mize, C. E., Saudubray, J. M., Veerman, A., Pearson, H. A., Munnich, A. (1991) "Site specific deletions of the mitochondrial genome in Pearson marrow-pancreas syndrome" Genomics 10 (2): 502-504</p> <p>•Wolke, R., Thompson, G.N., Thorburn, D.R., Dahl, H.H., Marzuki, S., Byrne, E., Blak, R.B. (1994) "A novel mtDNA deletion in an infant with Pearson syndrome" Journal of Inherited Metabolic Disease 17 (5): 521-528</p> |
| 10367:12629 ND4 - ND5 | -2461 | 10365-10367/12626-12628 | D, 3/3 | <p>•Cormier-Daire, V., Bonnerfont, J.P., Rustin, P., Maurage, C., Odier, H., Schmitz, J., Ricour, C., Saudubray, J.M., Munnich, A., Rotig, A. (1991) "Mitochondrial DNA rearrangements with onset as chronic diarrhea with villous atrophy" Journal of Pediatrics 121 (1): 63-70</p> |
| 10714:14124 ND4 - ND5 | 3379 | 10715-10754/14124-14133 | D, 9/10 | <p>•Rotig, A., Cormier, V., Koll, F., Mize, C. E., Saudubray, J. M., Veerman, A., Pearson, H. A., Munnich, A. (1991) "Site-specific deletions of the mitochondrial genome in Pearson marrow-pancreas syndrome" Genomics 11 (2): 317-314</p> <p>•Young, A., Cormier, V., Blanchet, S., Bonnerfont, J.P., Lefebvre, T., Monnier, N., Schmitz, J., Rustin, P., Fischer, A., Saudubray, J.M. (1993) "Pearson's marrow-pancreas syndrome: A multi-system mitochondrial disorder in infancy" Journal of Clinical Investigation 85 (3): 1501-1508</p> <p>•Cormier, V., Rotig, A., Coroni, R., Malic, M., Saudubray, J.M., Munnich, A., Rotig, A., Forzi, G.L. (1990) "Widespread multilocus deletions of the mitochondrial genome in Pearson marrow pancreas syndrome" Journal of Pediatrics 117 (1): 589-602</p> <p>•Ywata, T., Matsumoto, T., Iwamoto, Y., Matsuda, A., Kuzuya, T., Saito, T. (1993) "Japanese case of diastolic mellitus and deafness with mutations in mitochondrial tRNA^{Leu}(UUR) gene (letter)" Journal 341 (8855): 1281-1282</p> |
| 11232:13080 | -2747 | 11234-11242/13081-13080 | D, 0/0 | |

Table 2: Prostate Cancer Detection with Novel Mitochondrial Fusion Transcripts

| | RNA | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 |
|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Transcript | Transcr ipt 1 | Transcr ipt 1 | Transcr ipt 1 | Transcr ipt 2 | Transcr ipt 2 | Transcr ipt 2 | Transcr ipt 3 | Transcr ipt 3 | Transcr ipt 3 | Transcr ipt 4 | Transcr ipt 4 | Transcr ipt 4 |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| No dilution | A | 2957 | 353 | 233 | 144838 | 75374 | 348424 | 333189 | 213844 | 509 | 565 | 207 |
| Replicate A | B | 3174 | 475 | 298 | 202793 | 100062 | 320877 | 278137 | 210265 | 401 | 676 | 250 |
| 1:10 dilution | C | 1041 | 262 | 114 | 106195 | 98403 | 238467 | 248677 | 123497 | 181 | 486 | 168 |
| Replicate C | D | 1040 | 272 | 176 | 120308 | 116930 | 239231 | 262520 | 129778 | 153 | 467 | 149 |
| 1:100 dilution | E | 318 | 170 | 110 | 25155 | 64823 | 100345 | 164606 | 85287 | 72 | 265 | 119 |
| Replicate E | F | 287 | 150 | 109 | 23500 | 50524 | 100856 | 178527 | 84731 | 83 | 251 | 120 |
| 1:1000 dilution | G | 100 | 76 | 123 | 3002 | 12960 | 29203 | 102309 | 137 | 31 | 143 | 66 |
| Replicate G | H | 94 | 83 | 91 | 1263 | 5796 | 29092 | 97257 | 96 | 45 | 110 | 94 |
| %CV A | | 5.0 | 20.9 | 17.3 | 23.6 | 19.9 | 5.8 | 12.7 | 1.2 | 16.9 | 12.7 | 13.3 |
| %CV C | | 0.1 | 2.5 | 30.1 | 8.8 | 12.2 | 0.2 | 3.8 | 3.5 | 12.0 | 2.8 | 8.3 |
| %CV E | | 7.1 | 9.0 | 0.6 | 4.8 | 17.5 | 0.4 | 5.7 | 0.5 | 9.8 | 3.8 | 0.6 |
| %CV G | | 4.7 | 6.0 | 20.8 | 57.7 | 54.0 | 0.3 | 3.6 | 25.0 | 27.0 | 18.2 | 24.9 |

* unit results in table are RLU (relative luminescence units); Data read on Glorunner™.

%CV = Coefficient of variation (as %).

Legend: Homog = homogenate.

Homog 1: Prostate tumour tissue sample from patient;

Homog 2: Histologically normal tissue adjacent to tumour from patient.

RNA: Control: Total RNA from prostate tissue (Ambion p/n 7988).

Shading: Background measurement.

Table 3: Deletion/Transcript/Hypothetical translation product relationships

| Deletion | RNA transcript | DNA sequence with deletion complementary to RNA transcript | Transcript No. | Hypothetical Fusion Protein |
|--|-----------------------|---|-----------------------|------------------------------------|
| ATP synthase F0 subunit 8 to NADH dehydrogenase subunit mitochondrial positions 8366-14148 (with reference to SEQ ID NO: 1). Translated sequence begins at position 8389 | SEQ ID NO: 19 | SEQ ID NO: 2 | 1 | SEQ ID NO: 36 |
| NADH dehydrogenase subunit 4L (ND4L) to NADH dehydrogenase subunit 5 (ND5); Mitochondrial positions 10470-14148 (with reference to SEQ ID NO: 1) | SEQ ID NO: 20 | SEQ ID NO: 3 | 2 | SEQ ID NO: 37 |
| Cytochrome c oxidase subunit II (COII) to Cytochrome b (Cytb); Mitochondrial positions 7586-15887 (with reference to SEQ ID NO: 1) | SEQ ID NO: 21 | SEQ ID NO: 4 | 3 | SEQ ID NO: 38 |
| Cytochrome c oxidase subunit II (COII) to Cytochrome b (Cytb); Mitochondrial positions 7586-15887 (with reference to SEQ ID NO: 1) | SEQ ID NO: 22 | SEQ ID NO: 5 | 4 | SEQ ID NO: 39 |

Table 4: Breast and Prostate Cancer Detection

| | Breast Tumour 1 | Normal adjacent Breast Tumour 1 | Breast Tumour 2 | Normal Adjacent to Breast Tumour 2 | Prostate Tumour 3 | Prostate Tumour 4 | Prostate Tumour 5 | Normal Adjacent to Prostate Tumour 5 |
|--------------------------|-----------------|---------------------------------|-----------------|------------------------------------|-------------------|-------------------|-------------------|--------------------------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1:100 dilution | 68920 | 2971 | 49108 | 1245 | 46723 | 56679 | 99836 | 35504 |
| 1:100 dilution replicate | 92409 | 3017 | 60637 | 1512 | 53940 | 56155 | 100582 | 44221 |
| | 420 | 3 | 31 | 6 | 26 | 25 | 44 | 23 |
| | 518 | 3 | 4 | 5 | 5 | 3 | 4 | 2 |
| | | | | | | | | |
| %CV | 20.6 | 1.1 | 14.9 | 13.7 | 10.1 | 0.7 | 0.5 | 15.5 |
| | | | | | | | | |

- unit results in table are RLU (relative luminescence units)

- background G1, H1

- empty well G2-G8, H2- H8

Table 5a: Assay Conditions

| Template for the assay | | | | | | | | | | | |
|------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| RNA | Homogen 1 | Homogen 2 | RNA | Homogen 1 | Homogen 2 | RNA | Homogen 1 | Homogen 2 | RNA | Homogen 1 | Homogen 2 |
| Transcript 1 | Transcript 1 | Transcript 1 | Transcript 2 | Transcript 2 | Transcript 2 | Transcript 3 | Transcript 3 | Transcript 3 | Transcript 4 | Transcript 4 | Transcript 4 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| A | RNA | Homog 1 | Homog 2 | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 |
| B | Dil 1 | Dil 1 | Dil 1 | Dil 1 | Dil 1 | Dil 1 | Dil 1 | Dil 1 | Dil 1 | Dil 1 | Dil 1 |
| C | RNA | Homog 1 | Homog 2 | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 |
| D | Dil 2 | Dil 2 | Dil 2 | Dil 2 | Dil 2 | Dil 2 | Dil 2 | Dil 2 | Dil 2 | Dil 2 | Dil 2 |
| E | RNA | Homog 1 | Homog 2 | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 | RNA | Homog 1 | Homog 2 |
| F | Dil 3 | Dil 3 | Dil 3 | Dil 3 | Dil 3 | Dil 3 | Dil 3 | Dil 3 | Dil 3 | Dil 3 | Dil 3 |
| G | RNA | Homog 1 | Transcript 1 | Homog 1 | Transcript 1 | RNA | Homog 1 | Transcript 1 | RNA | Homog 1 | Transcript 1 |
| H | Dil 4 | Dil 4 | Dil 4 | Dil 4 | Background | Dil 4 | Dil 4 | Background | Dil 4 | Dil 4 | Background |

Homogenate1- Used 26 mg of tissue to homogenize in 700ul H soln with Proteinase K (PK). Used Qiagen TissueRuptor. Used 40ul homogenate supernatant, 20, 10 and 5 ul for dilution
Homogenate1= Tumour tissue from the tumorous Prostate

Homogenate2- Used 29 mg of tissue to homogenize in 700ul H soln with PK. Used Qiagen TissueRuptor. Used 40ul homogenate supernatant, 20, 10 and 5 ul for dilution
Homogenate2= Normal tissue from the tumorous Prostate

RNA dilution was made as below . RNA was from Prostate Normal from Ambion.
Assay was done in duplicates.

Table 5b: RNA dilution

| RNA Dilution | ng/ul |
|--------------|-------|
| Dil 1 | 3000 |
| Dil 2 | 1000 |
| Dil 3 | 333 |
| Dil 4 | 111 |

Table 6: Transcript Summary by Disease

| Probe | Prostate Cancer | Breast Cancer | Colorectal Cancer | Melanoma Skin Cancer | Lung Cancer | Ovarian Cancer | Testicular Cancer |
|-------|--------------------|------------------|----------------------|----------------------------|----------------|-------------------|----------------------|
| 1 | • | | | | | • | |
| 2 | • | • | • | | | • | • |
| 3 | • | | • | • | | • | • |
| 4 | • | | • | | | | • |
| 5 | | | • | • | | | |
| 6 | | | | • | • | • | |
| 7 | | | • | • | | | |
| 8 | | | | | • | | |
| 9 | | | | • | | | |
| 10 | | | • | | • | | |
| 11 | | | • | • | | • | • |
| 12 | | | • | | | • | • |
| 13 | | | | | | | • |
| 14 | | | | • | | | |
| 15 | | | | • | | • | • |
| 16 | | | | • | | | • |
| 17 | | | | • | | | |
| 20 | | | • | • | | • | • |

We Claim:

1. An isolated mitochondrial fusion protein, the protein having an amino acid sequence resulting from the translation of a mitochondrial fusion transcript corresponding to a mutation in mitochondrial DNA.
2. The mitochondrial fusion protein of claim 1, wherein the fusion protein has an amino acid sequence as set forth in any one of SEQ ID NOs: 36-52, 55, 58, 61, 64 or 67.
3. A method of detecting a cancer in a mammal, the method comprising assaying a tissue sample from the mammal for the presence of at least one mitochondrial fusion protein, the protein having an amino acid sequence resulting from the translation of a mitochondrial fusion transcript corresponding to a mutation in mitochondrial DNA.
4. The method of claim 3, wherein the cancer is selected from the group consisting of prostate cancer, testicular cancer, ovarian cancer, breast cancer, colorectal cancer, lung cancer and melanoma skin cancer.
5. The method of claim 3, wherein the fusion protein has an amino acid sequence as set forth in any one of SEQ ID NOs: 36-52, 55, 58, 61, 64 or 67.
6. An antibody or an antigen-binding fragment thereof having a specificity to the protein of claim 1 or 2.
7. The antibody of claim 6, wherein said antibody is polyclonal or monoclonal.
8. The method of claim 3, wherein said assay comprises an immunological assay.
9. The method of claim 8, wherein said assay is conducted with the antibody or antigen-binding fragment as claimed in claim 7.
10. A kit for conducting an assay for detecting the presence of a cancer in a mammal, said kit comprising the antibody or antigen-binding fragment of claim 6.
11. An isolated mitochondrial fusion transcript having a nucleic acid sequence as set forth in SEQ ID NO: 56, 59, 62 or 65.
12. An isolated mtDNA encoding the fusion transcript of claim 11.

13. The isolated mtDNA of claim 12 having a sequence as set forth in any one of SEQ ID NO: 57, 60, 63 or 66.
14. A hybridization probe having a nucleic acid sequence complementary to at least a portion of the mitochondrial fusion transcript according to claim 9 or the mtDNA of claim 12 or 13.
15. A method of detecting a cancer in a mammal, the method comprising assaying a tissue sample from the mammal for the presence of at least one mitochondrial fusion transcript associated with cancer by hybridizing the sample with at least one hybridization probe having a nucleic acid sequence complementary to at least a portion of the mitochondrial fusion transcript according to claim 11.
16. The method of claim 15, wherein the cancer is selected from the group consisting of prostate cancer, testicular cancer, ovarian cancer, breast cancer, colorectal cancer, lung cancer and melanoma skin cancer.
17. A kit for conducting an assay for detecting the presence of a cancer in a mammal, said kit comprising at least one hybridization probe complementary to at least a portion of the fusion transcript of claim 11.

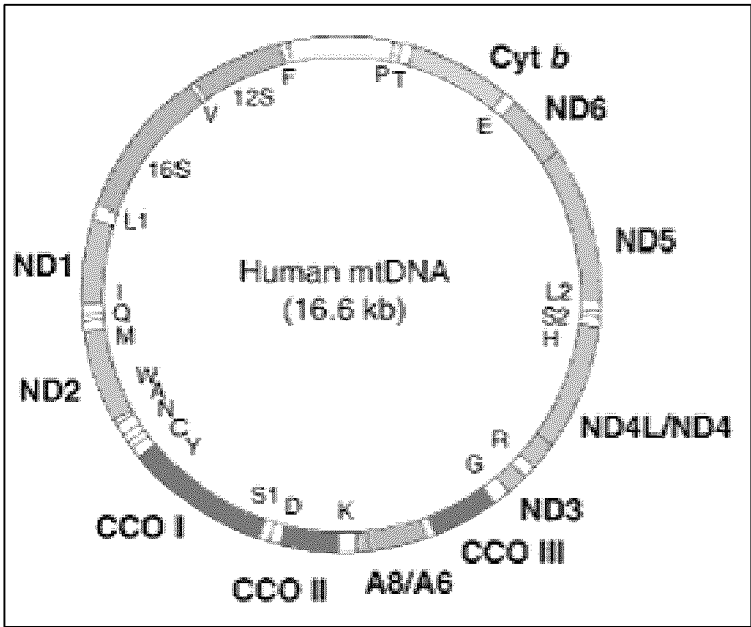


Figure 1

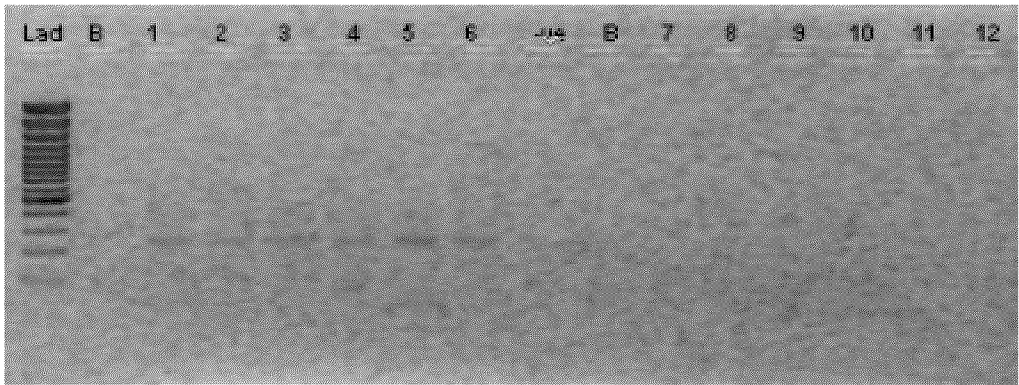


Figure 2

2/61

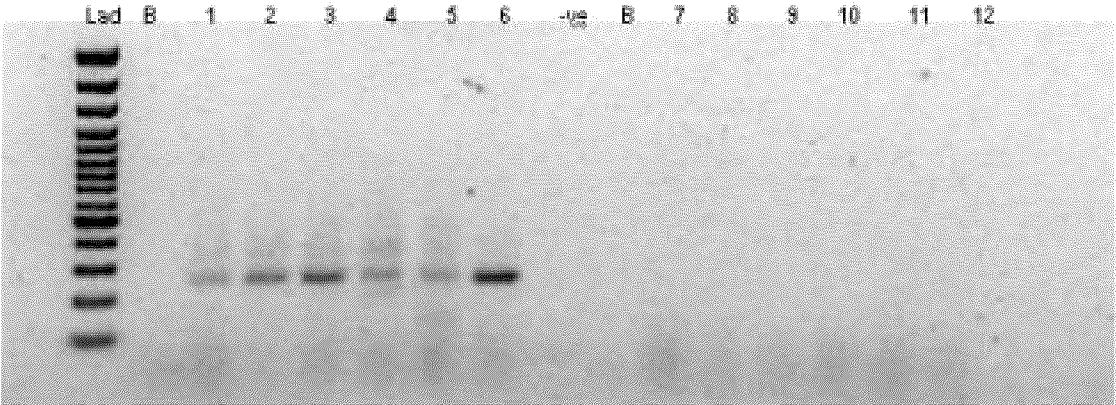


Figure 3

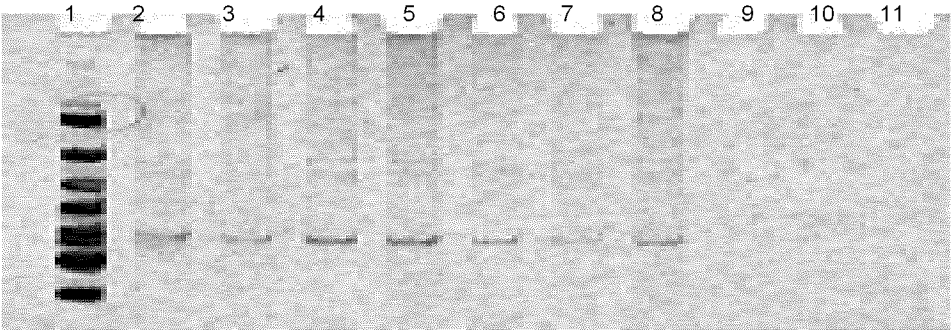


Figure 4

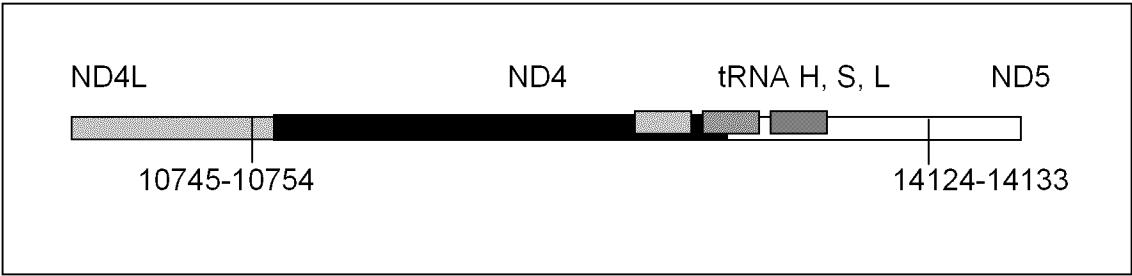


Figure 5a

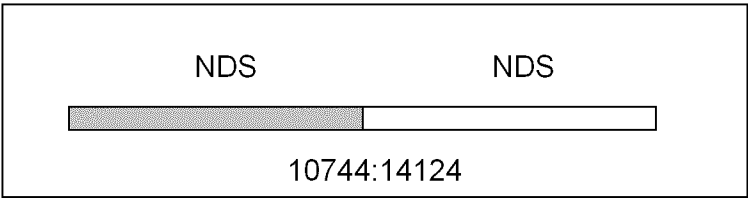
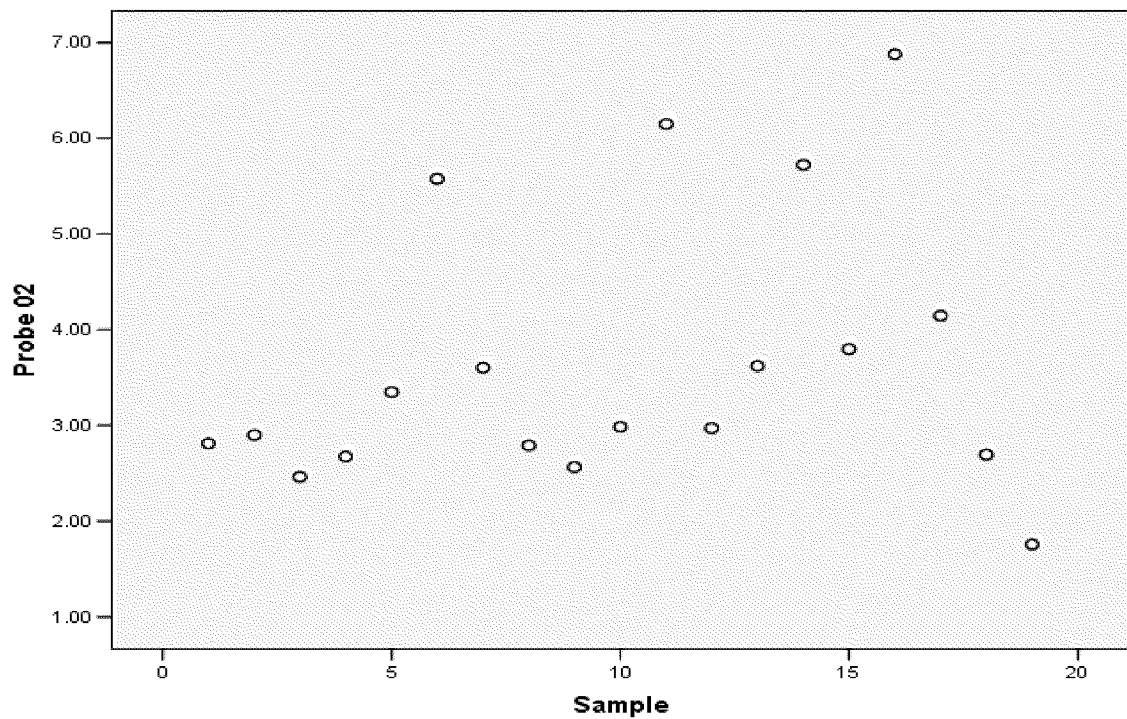


Figure 5b

4/61



ROC Curve

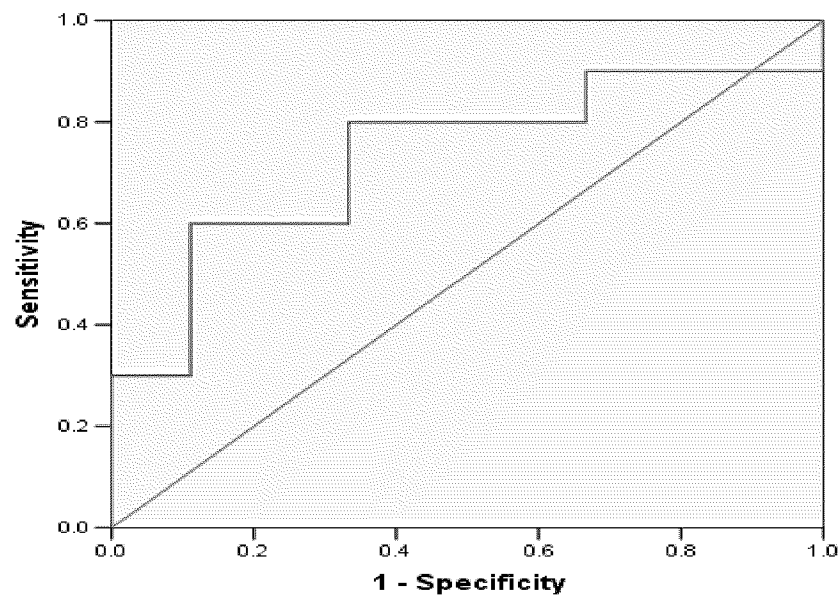
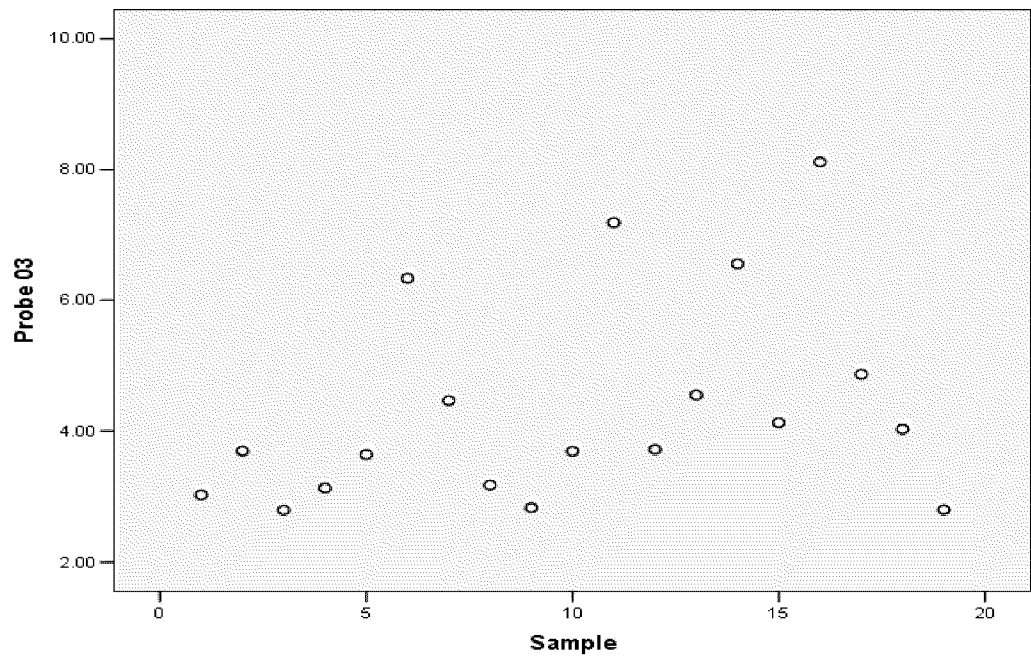


Figure 6a

5/61



ROC Curve

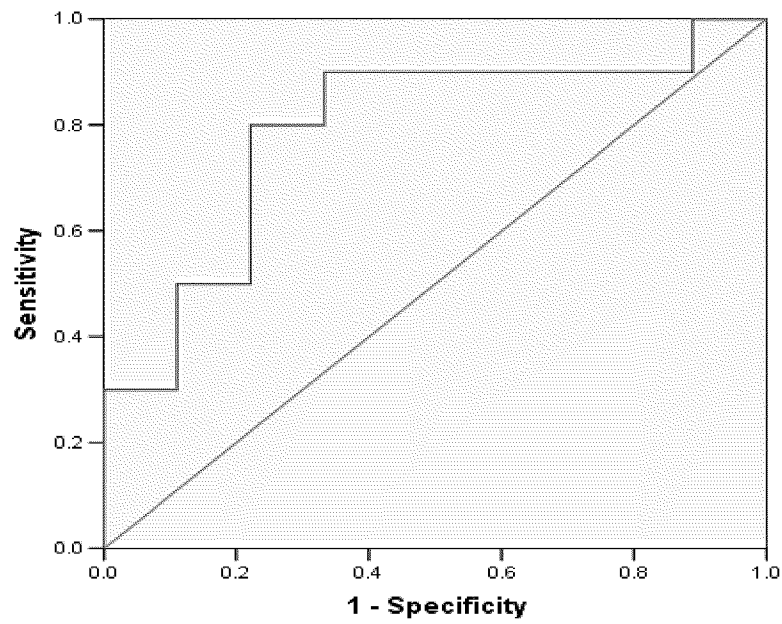
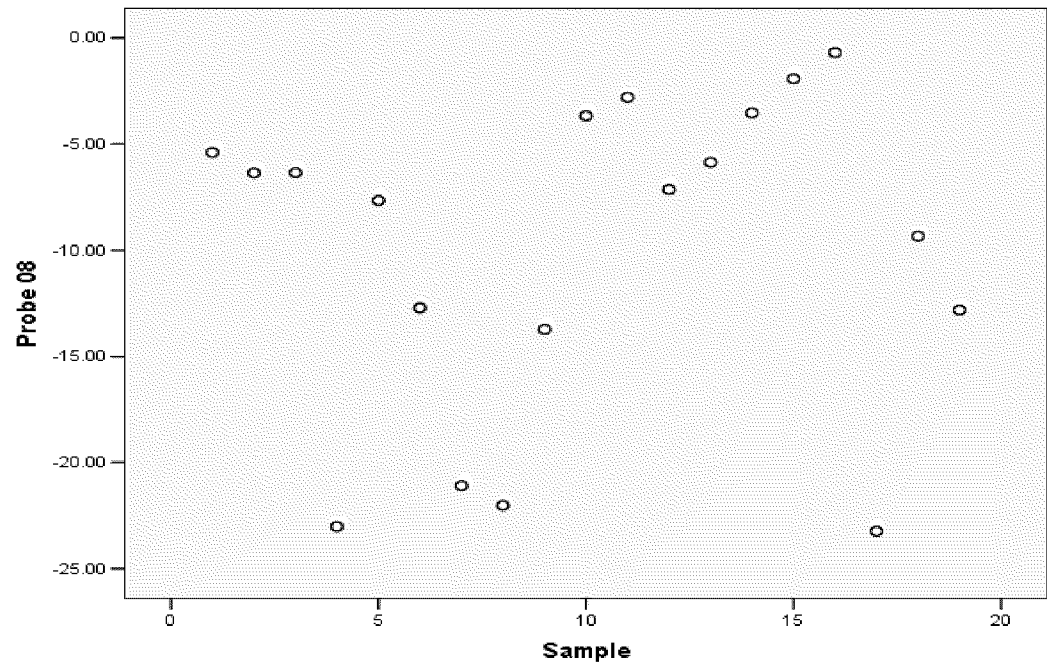


Figure 6b

6/61



ROC Curve

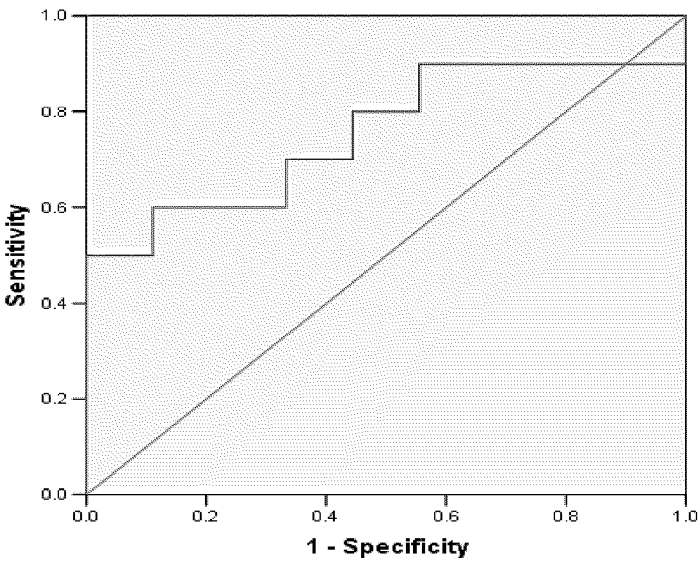
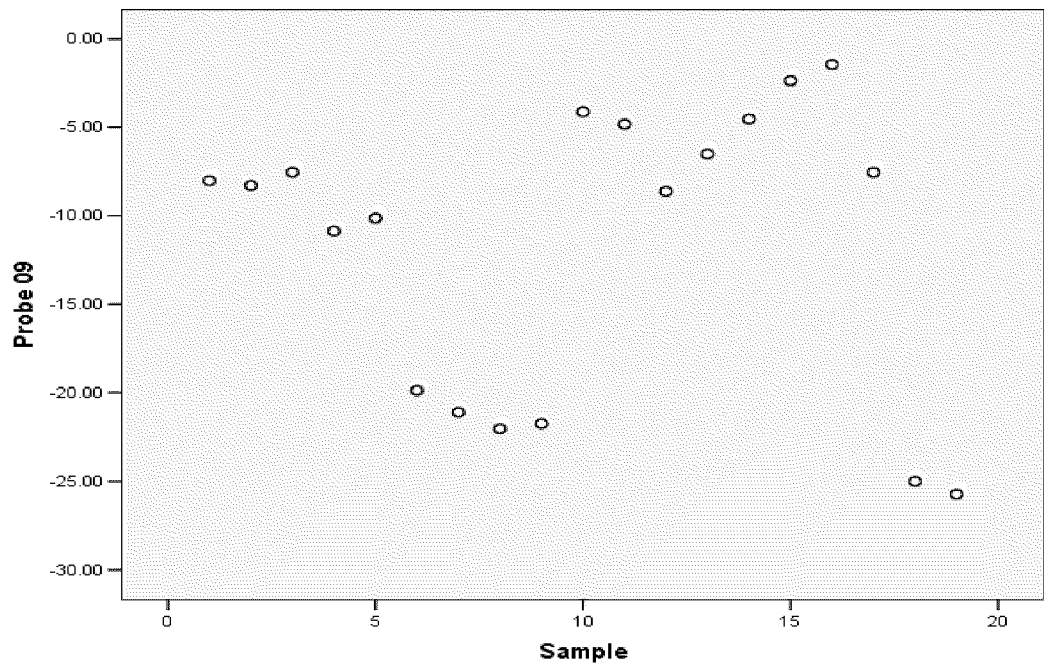


Figure 6c

7/61



ROC Curve

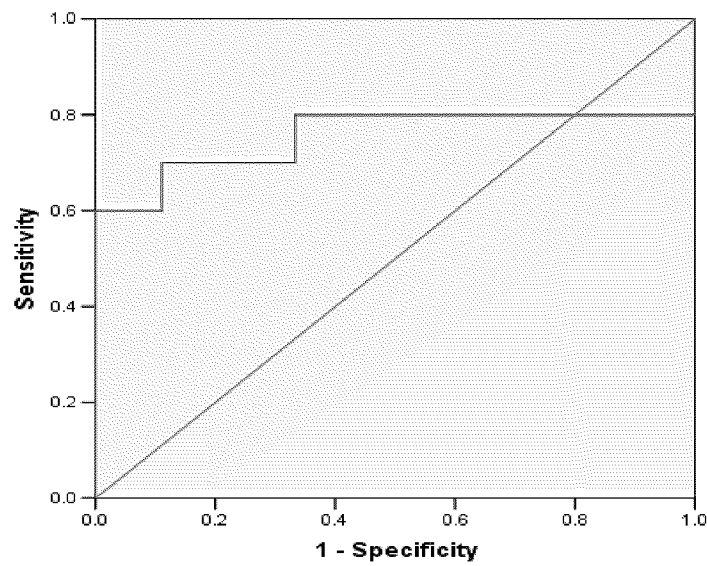
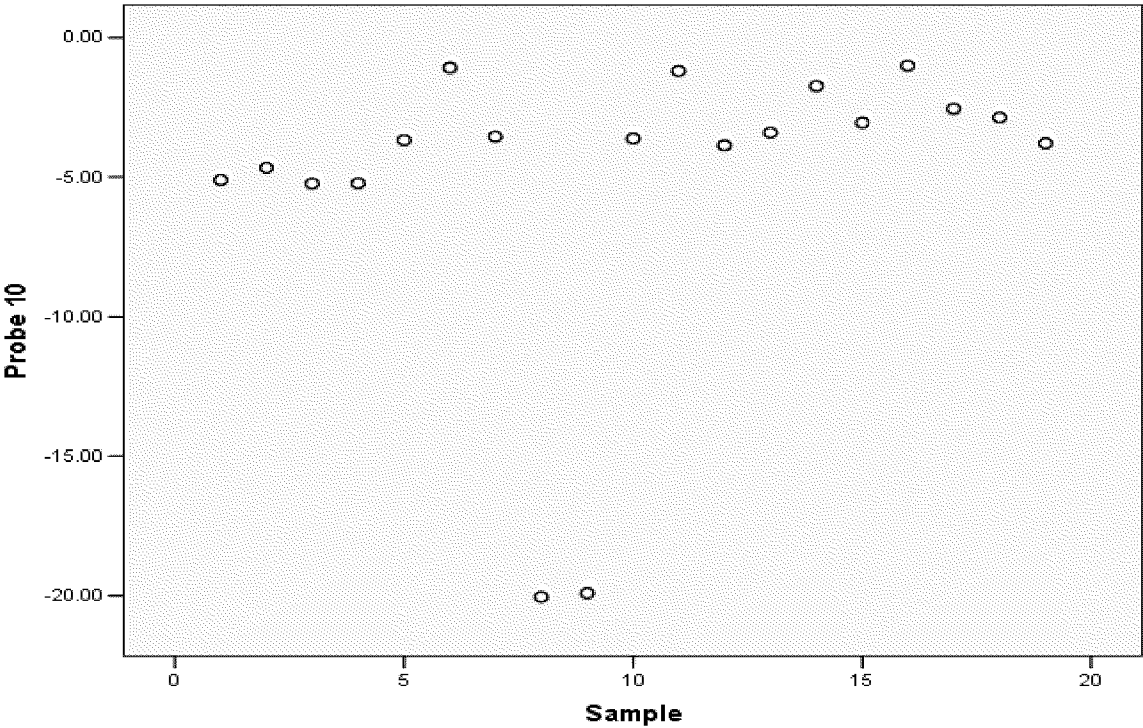


Figure 6d

8/61



ROC Curve

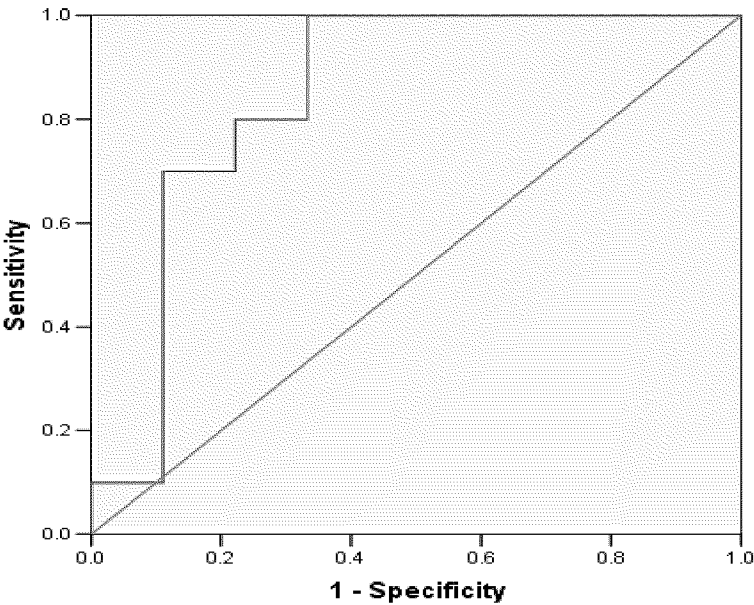
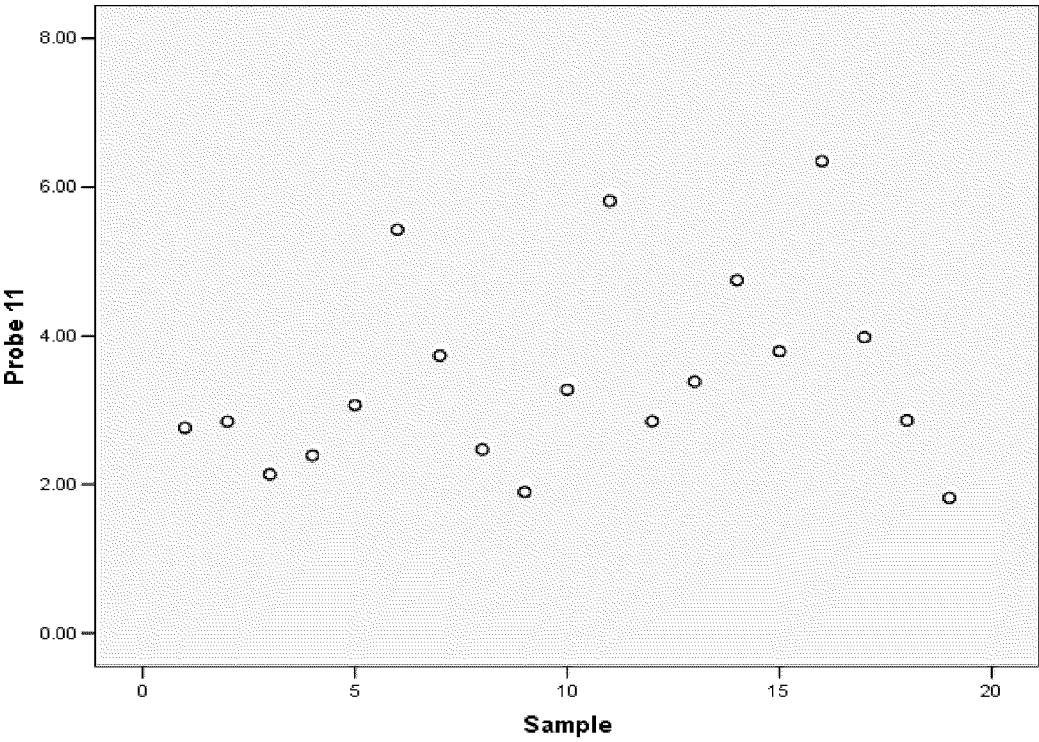


Figure 6e

9/61



ROC Curve

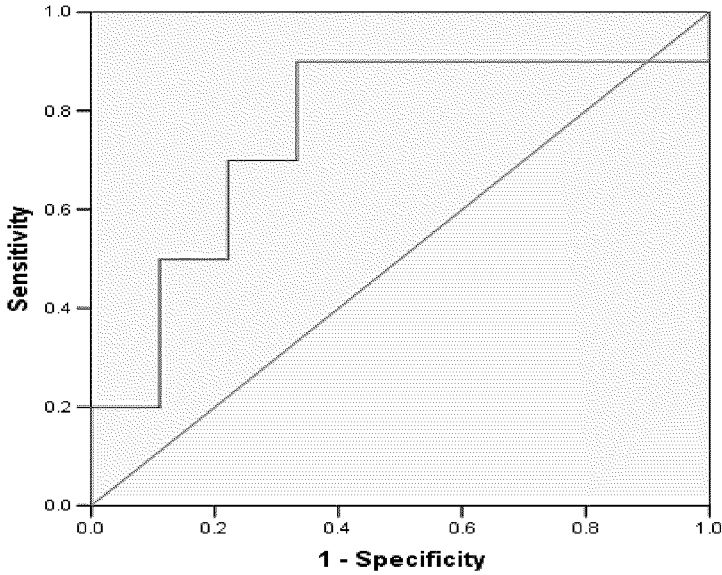
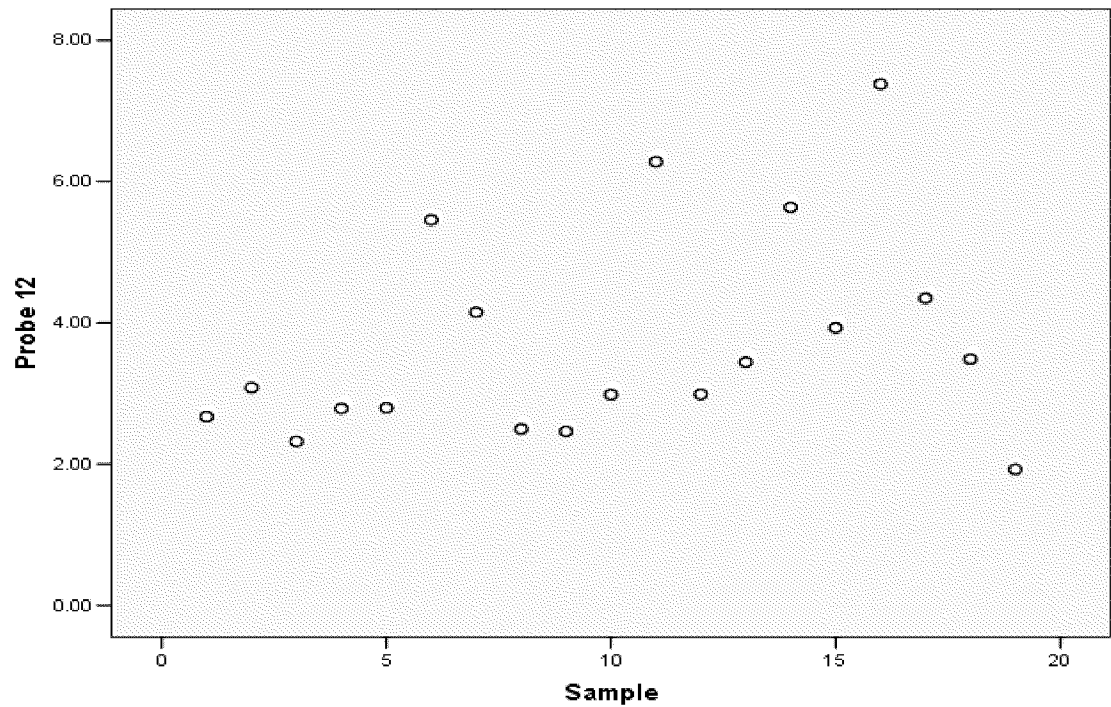


Figure 6f

10/61



ROC Curve

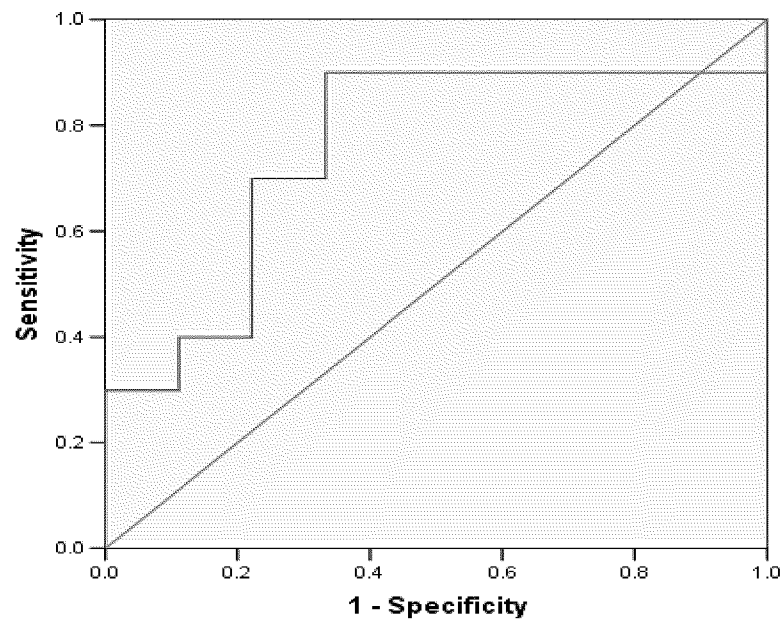


Figure 6g

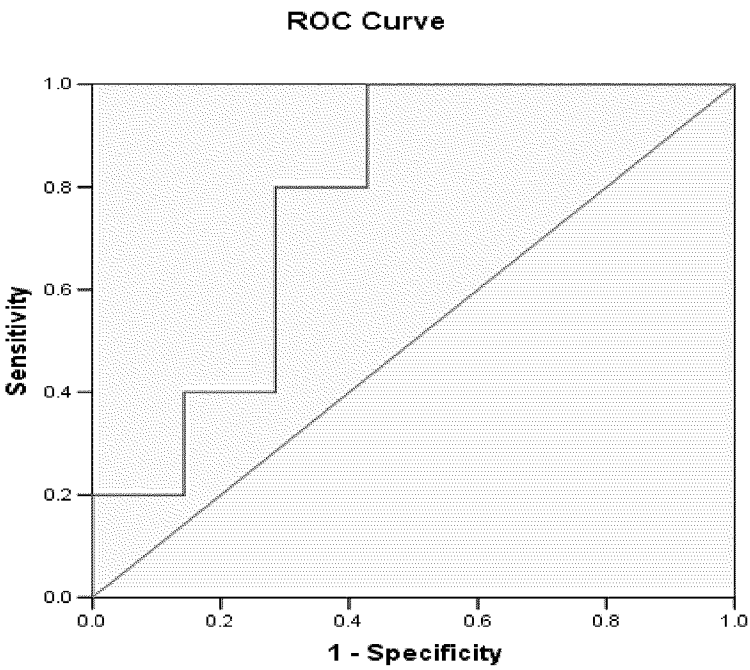
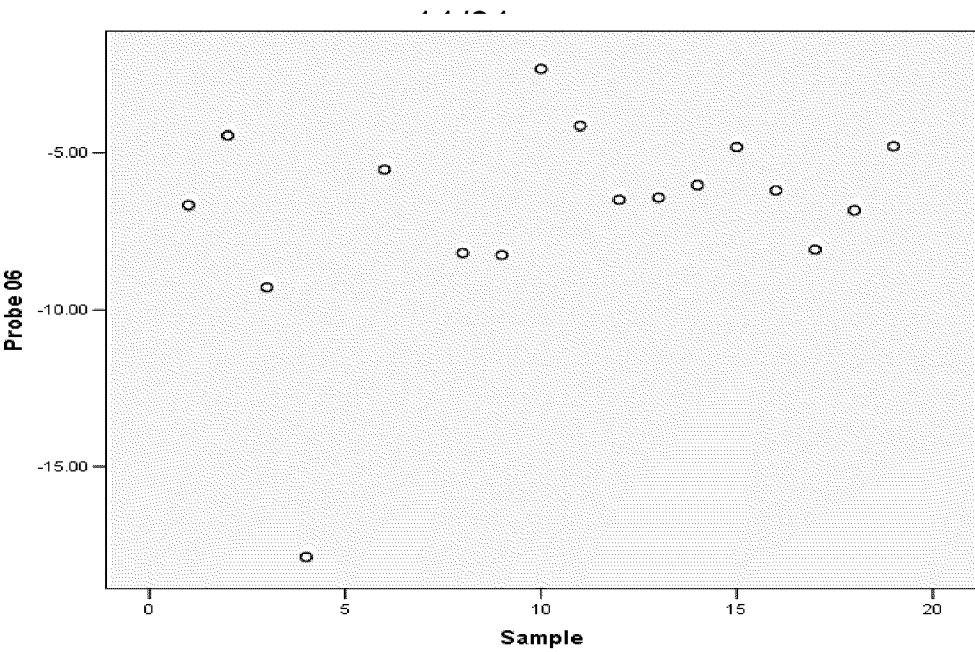
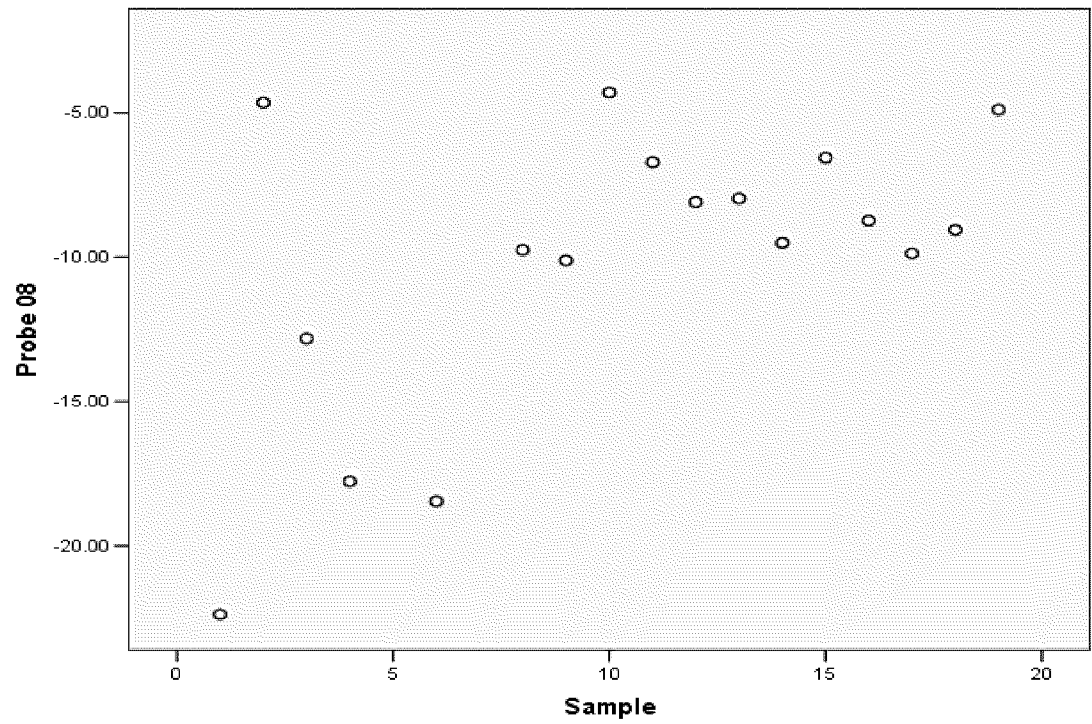


Figure 7a

12/61



ROC Curve

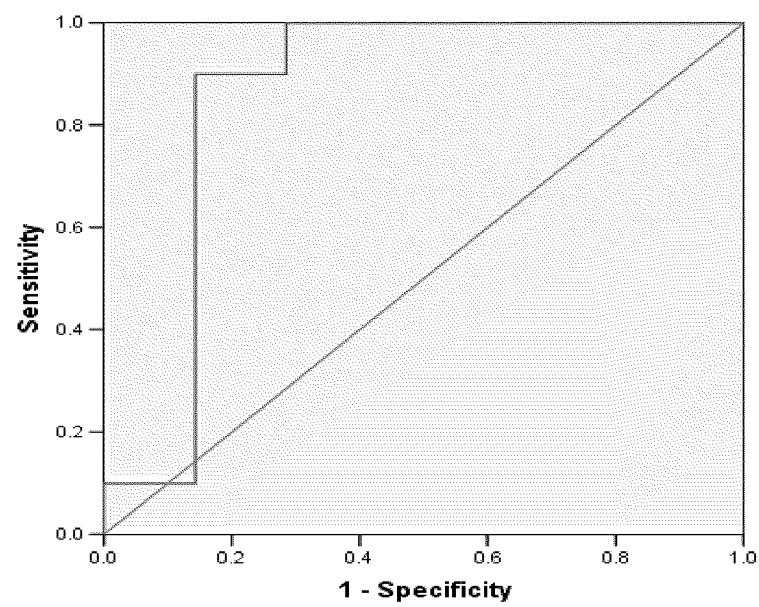
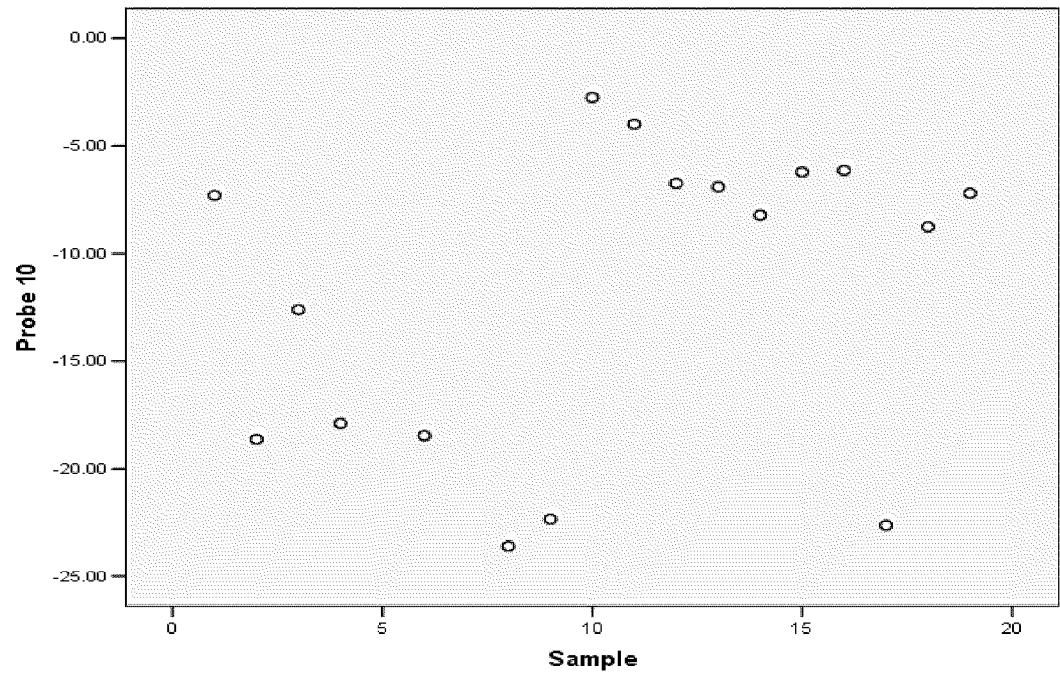


Figure 7b

13/61



ROC Curve

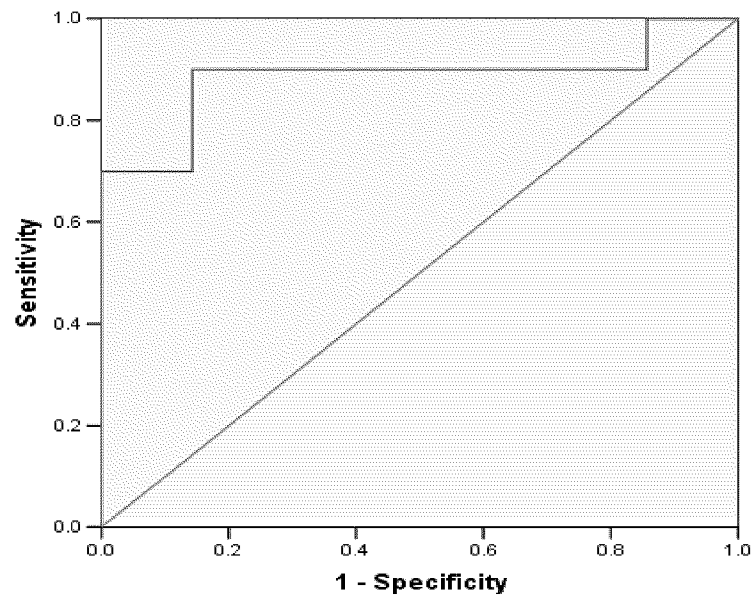
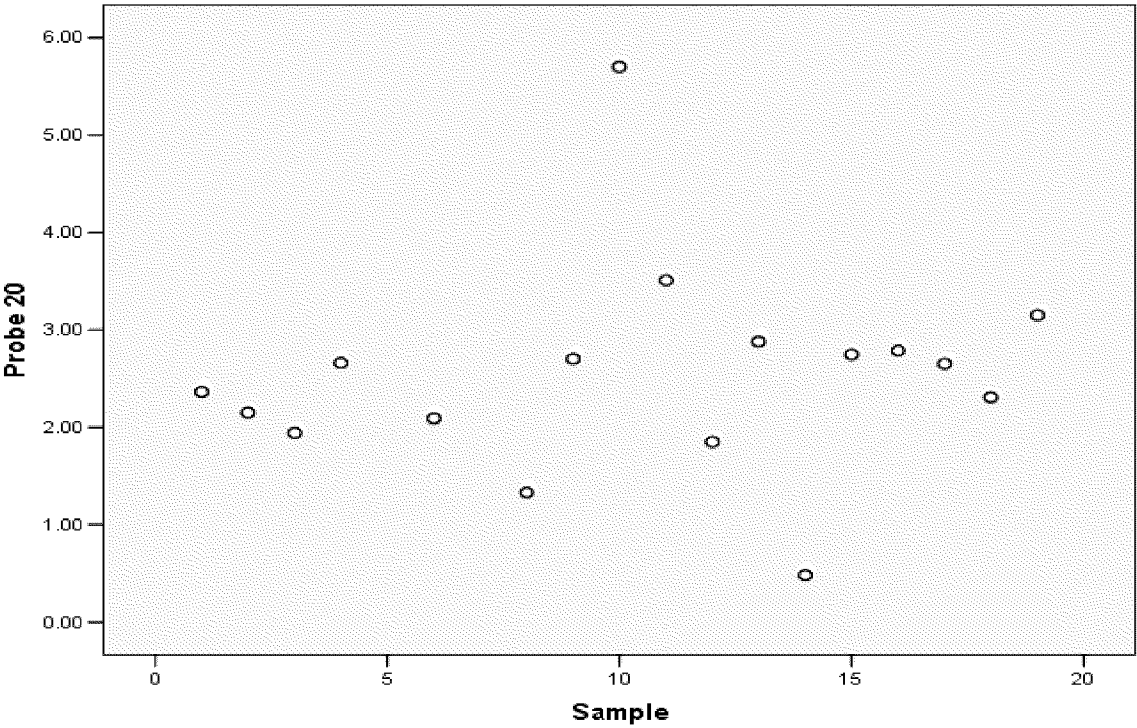


Figure 7c

14/61



ROC Curve

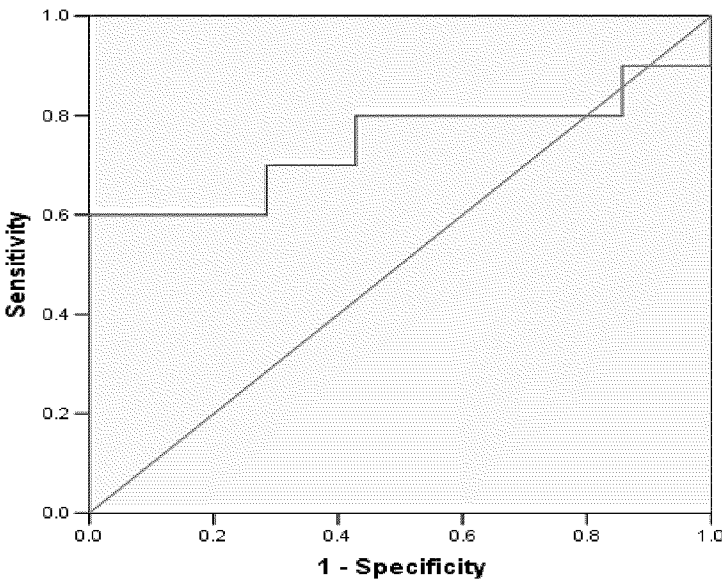
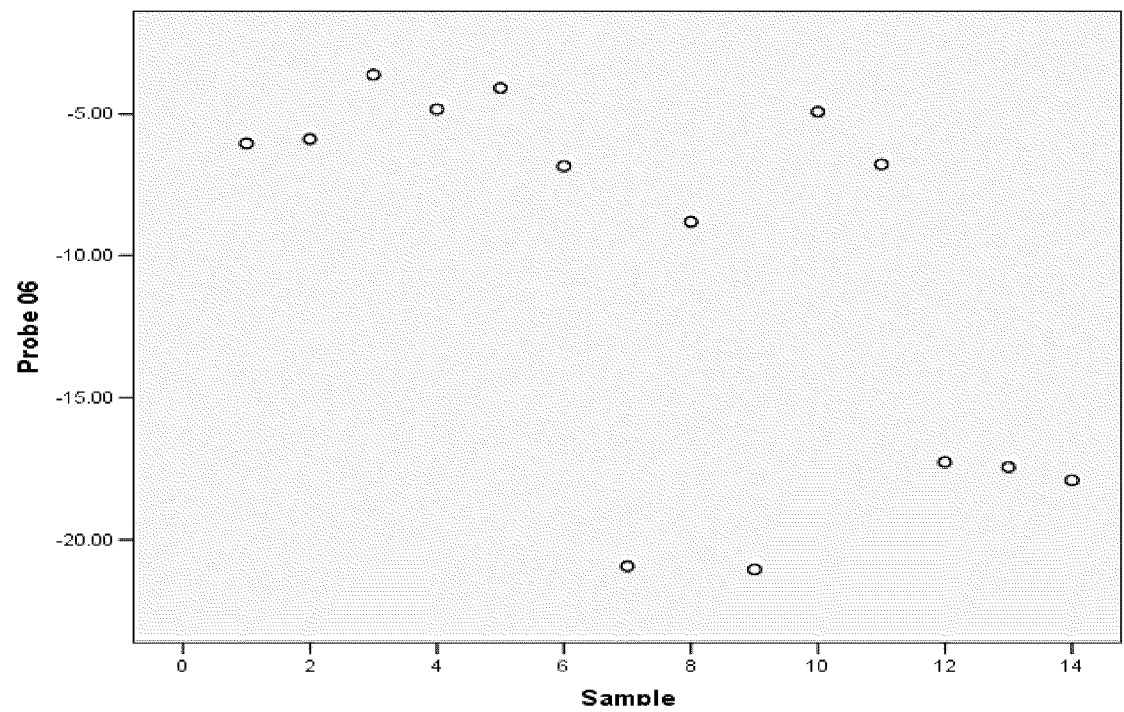


Figure 7d

15/61



ROC Curve

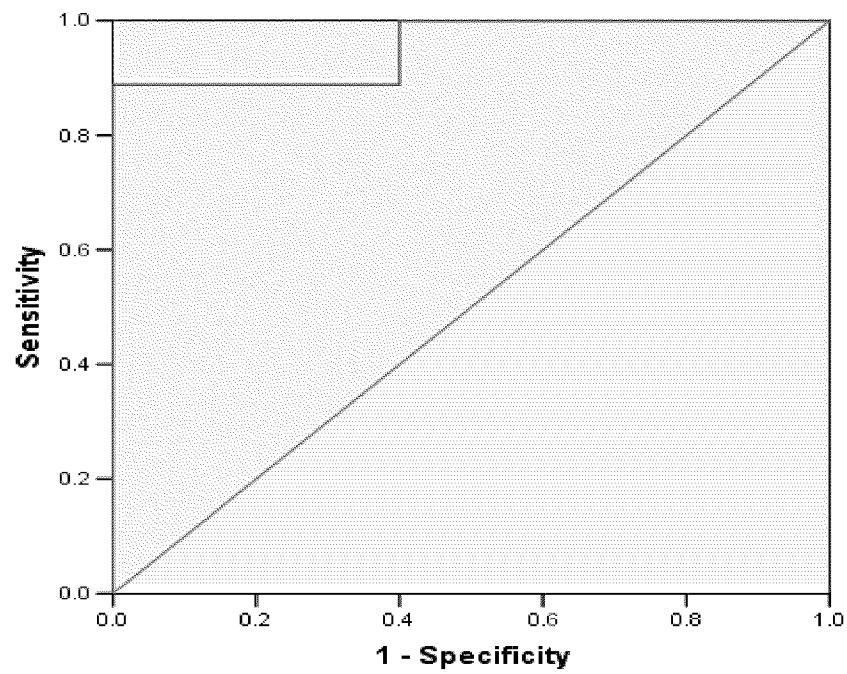
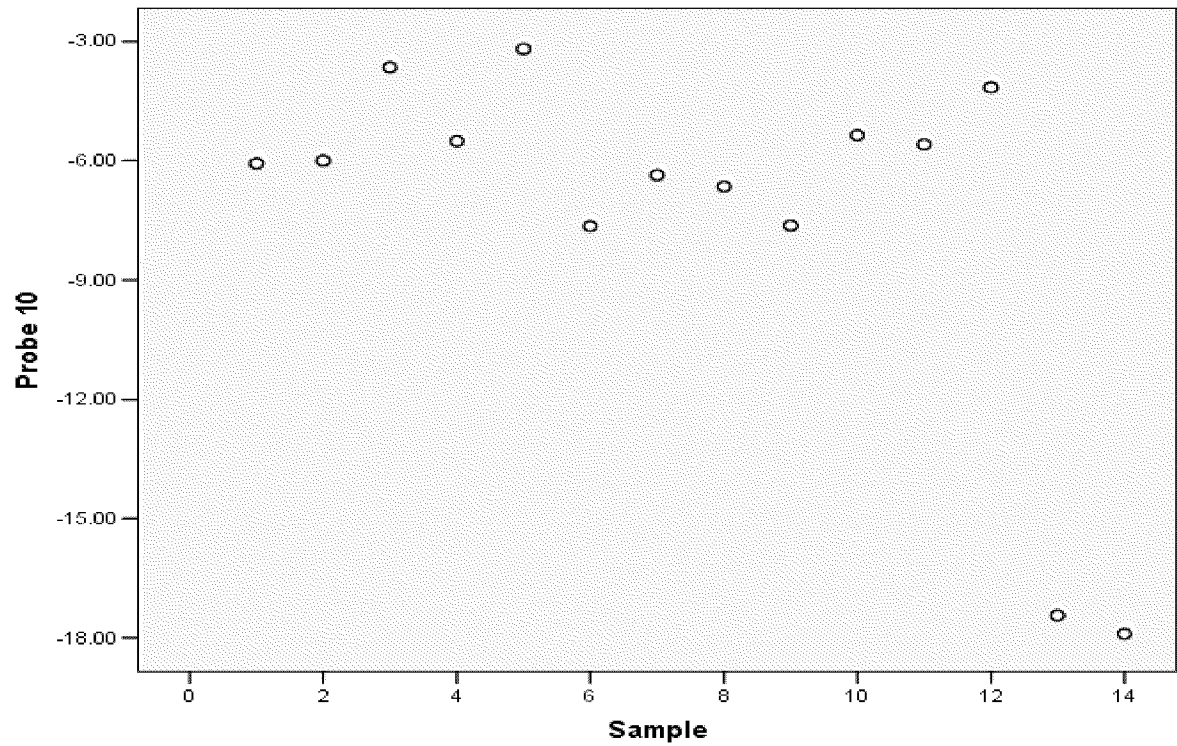


Figure 8a

16/61



ROC Curve

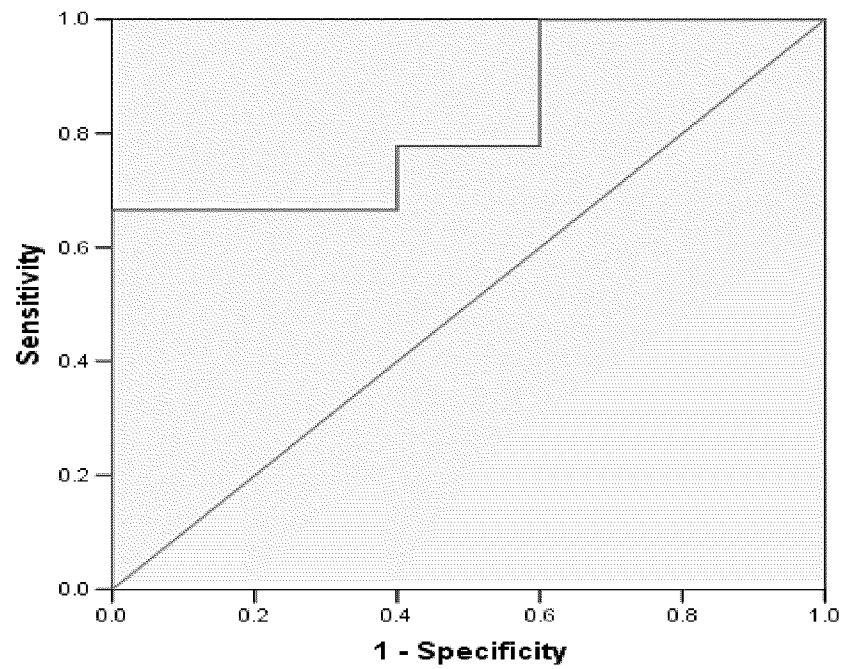
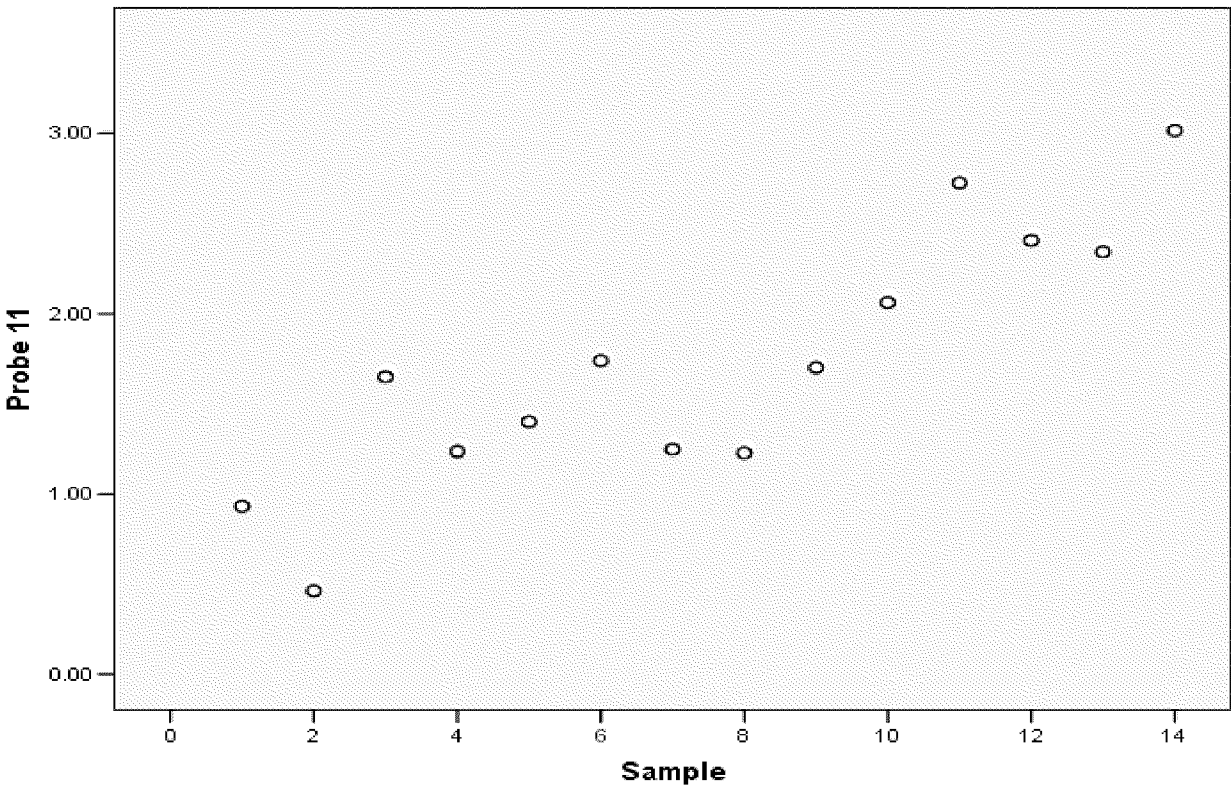


Figure 8b

17/61



ROC Curve

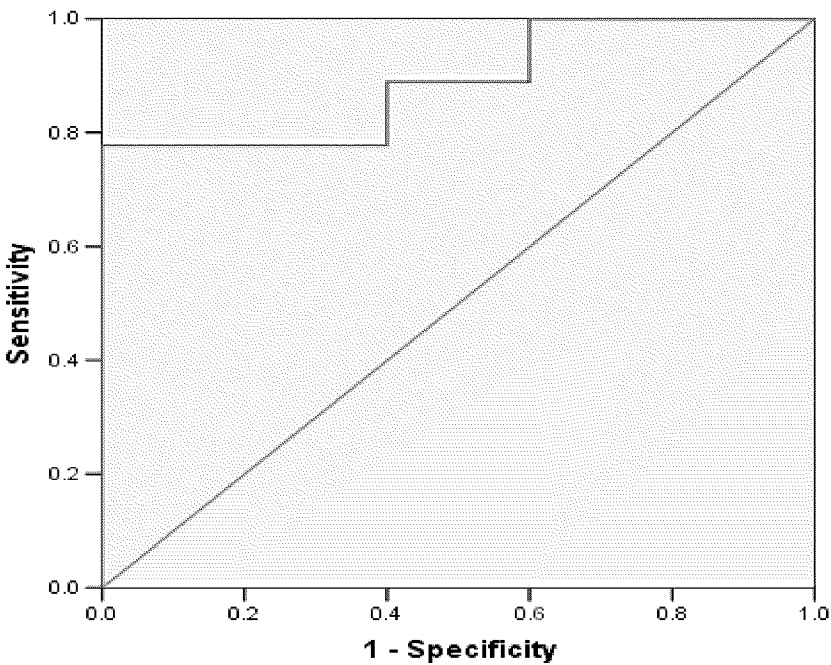
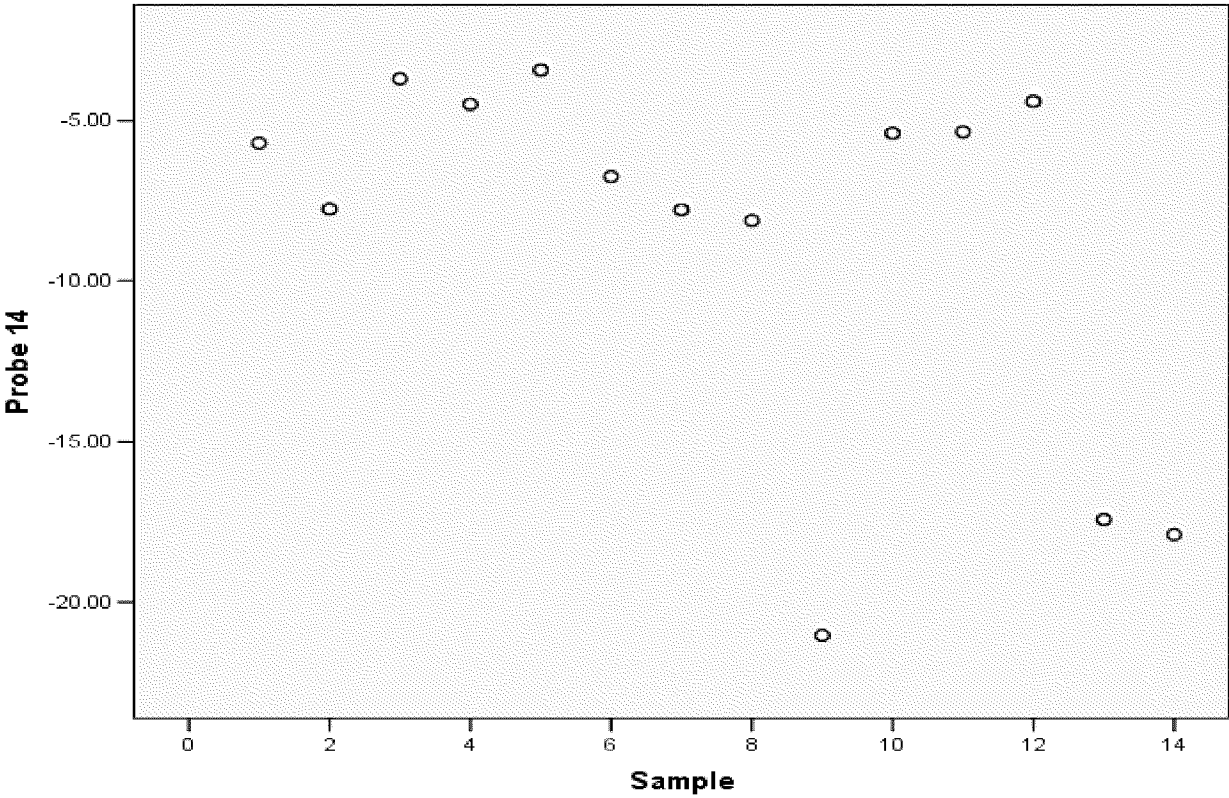


Figure 8c

18/61



ROC Curve

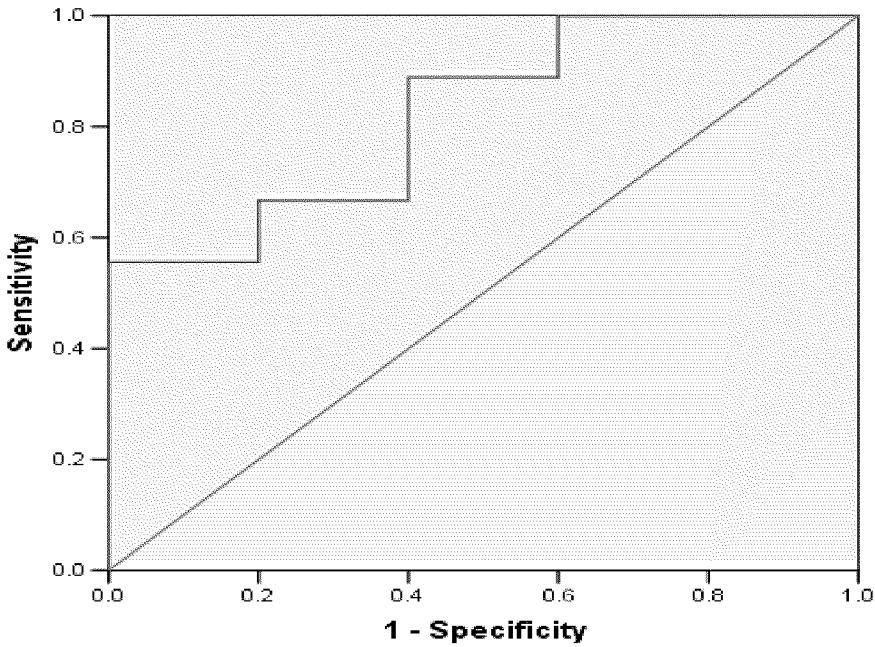
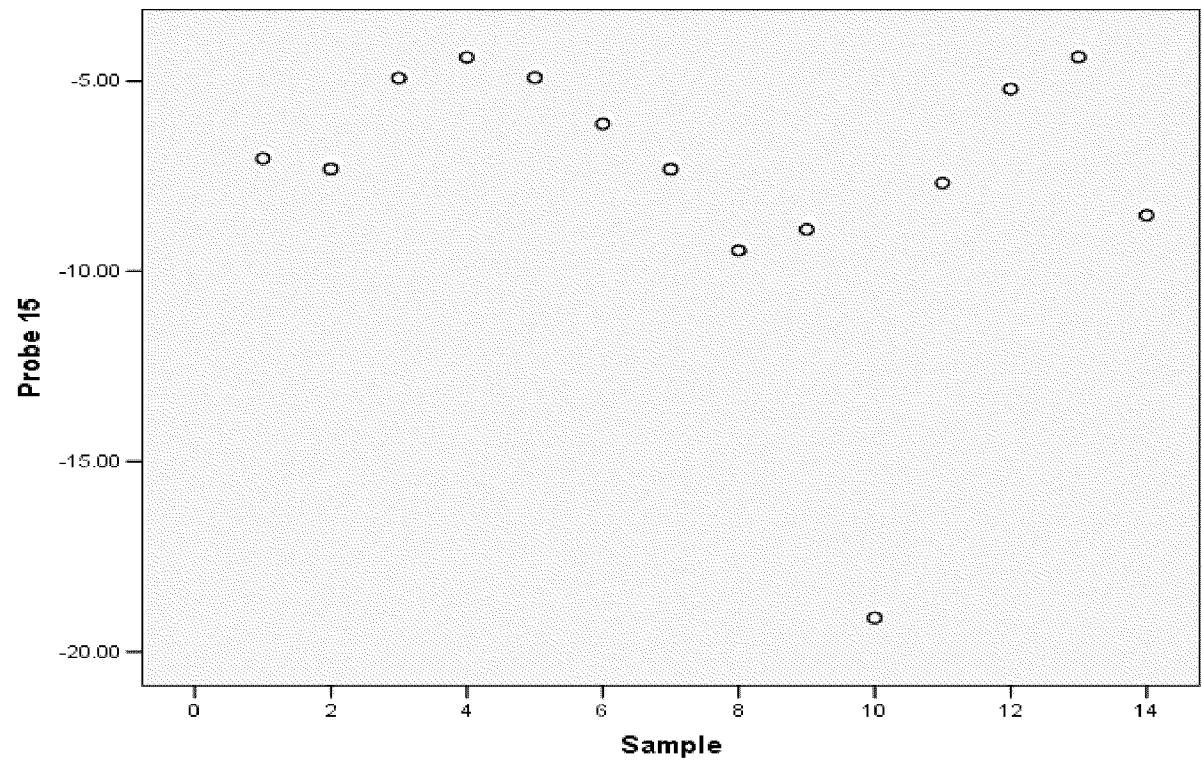


Figure 8d

19/61



ROC Curve

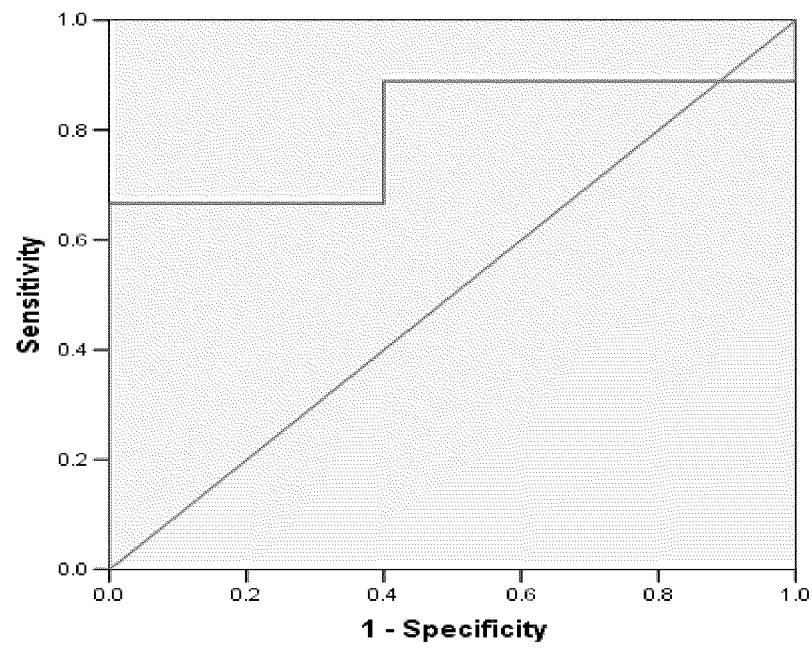
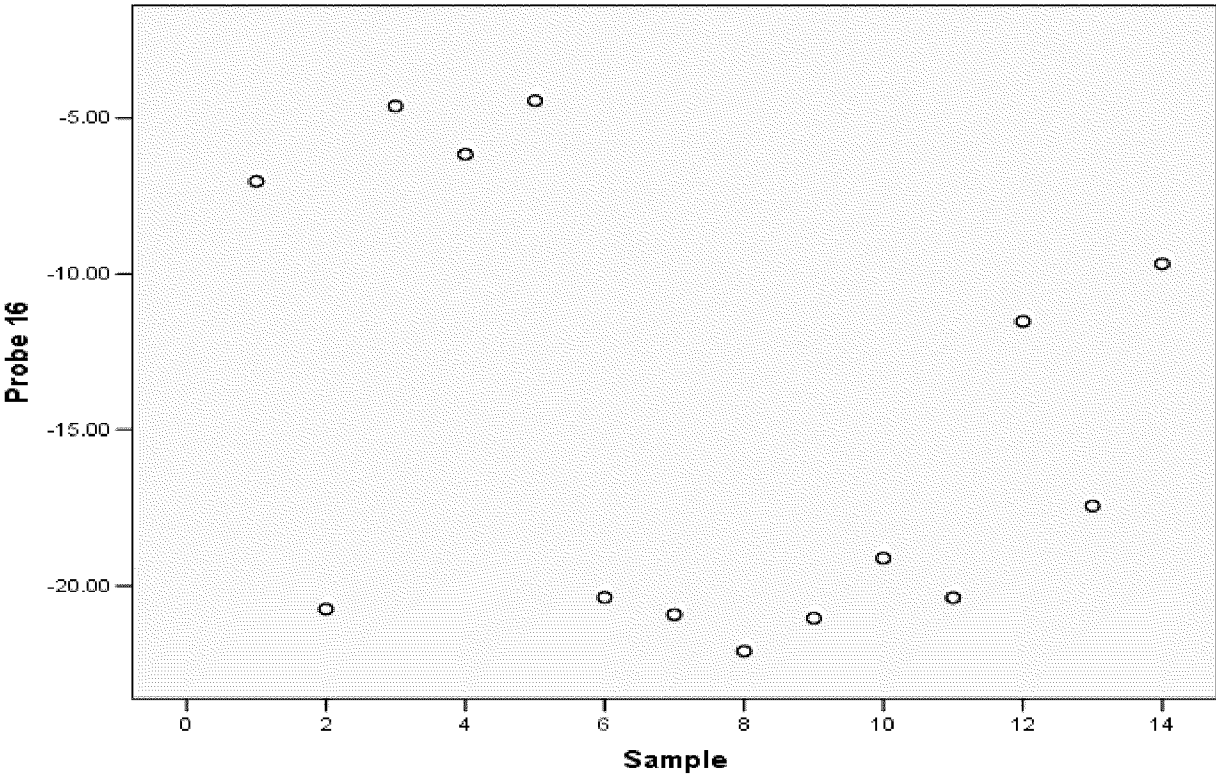


Figure 8e

20/61



ROC Curve

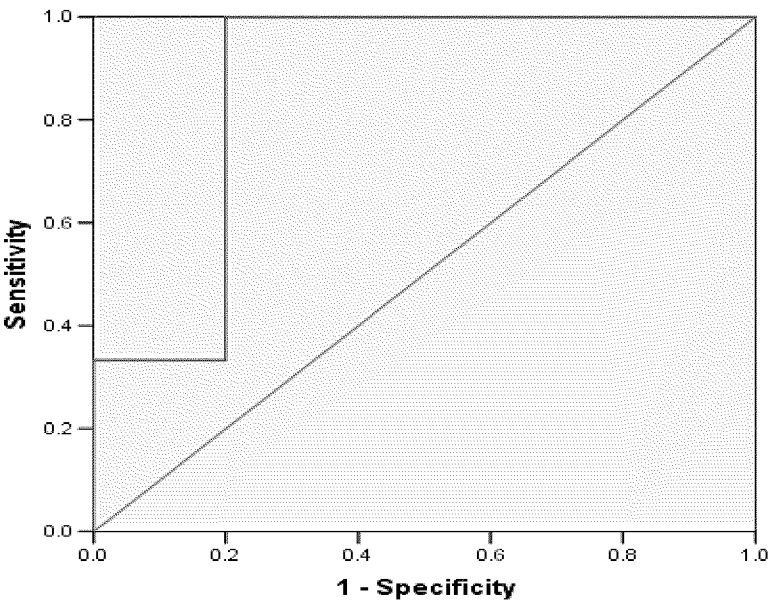
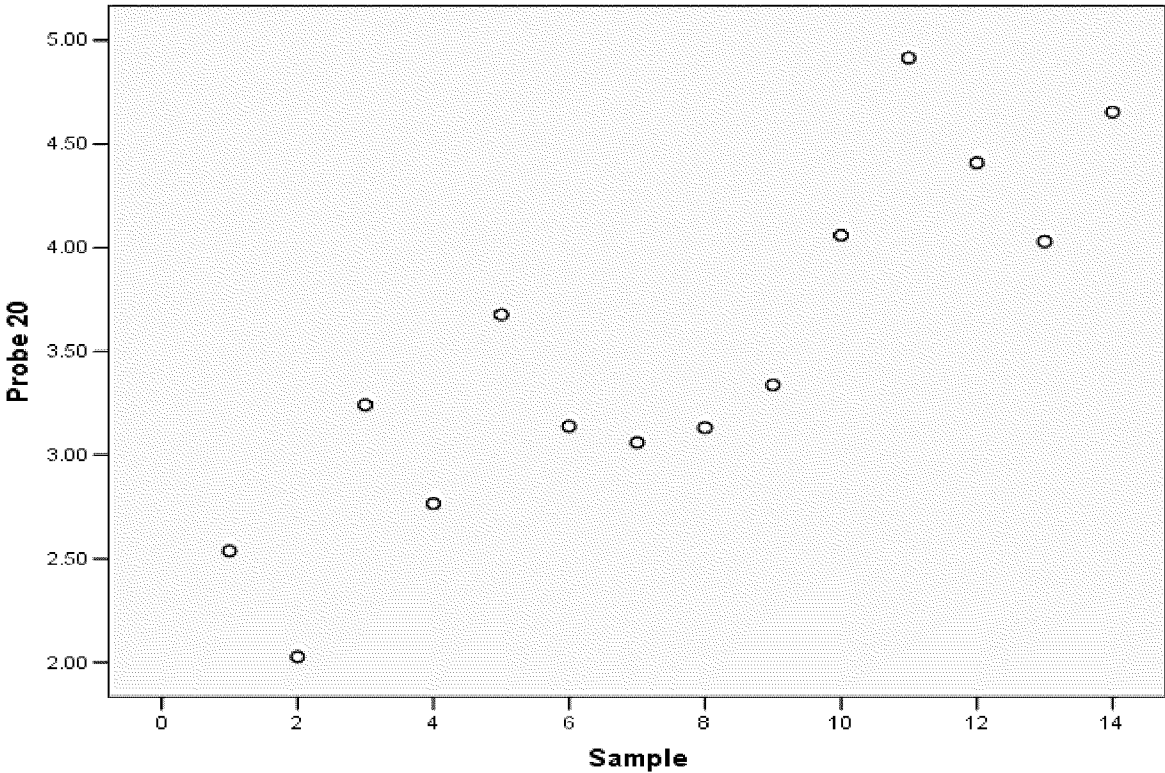


Figure 8f

21/61



ROC Curve

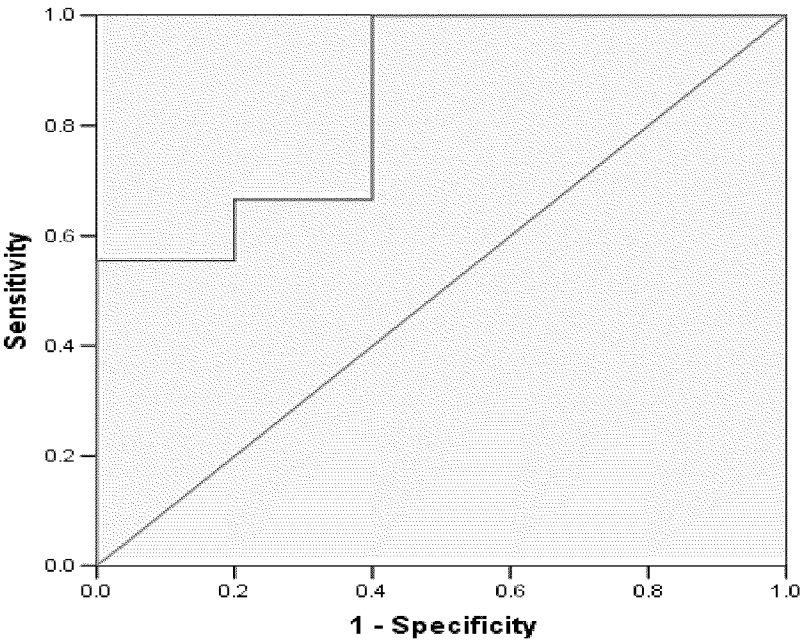


Figure 8g

22/61

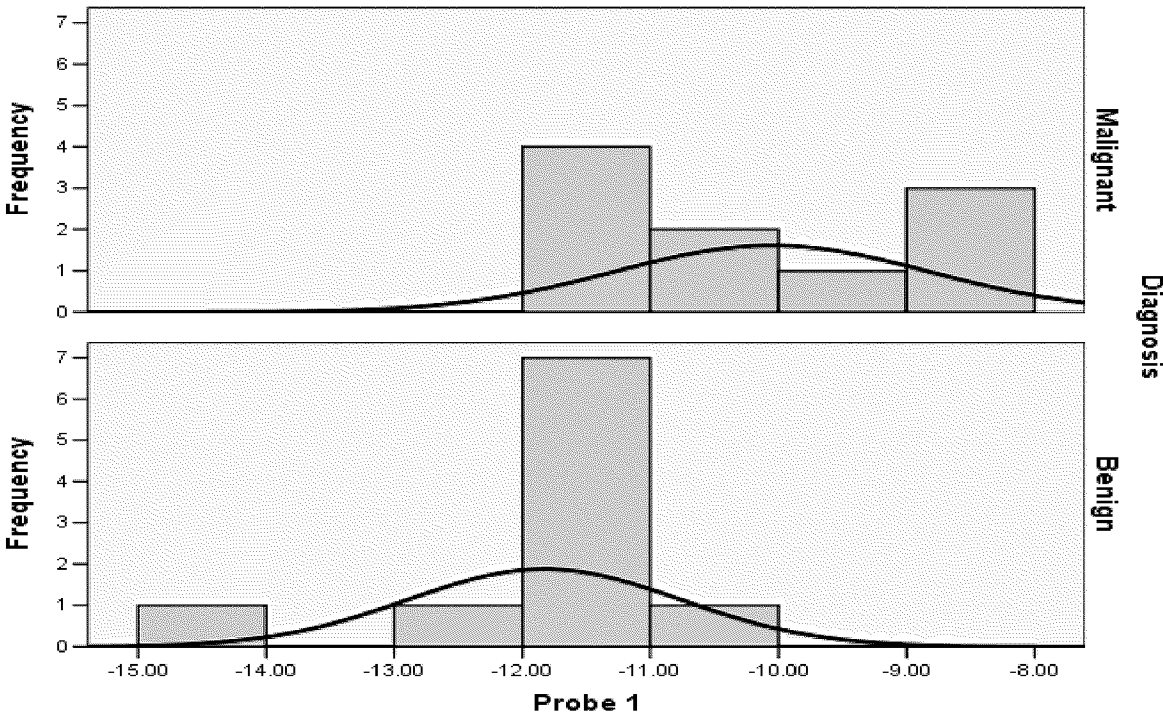
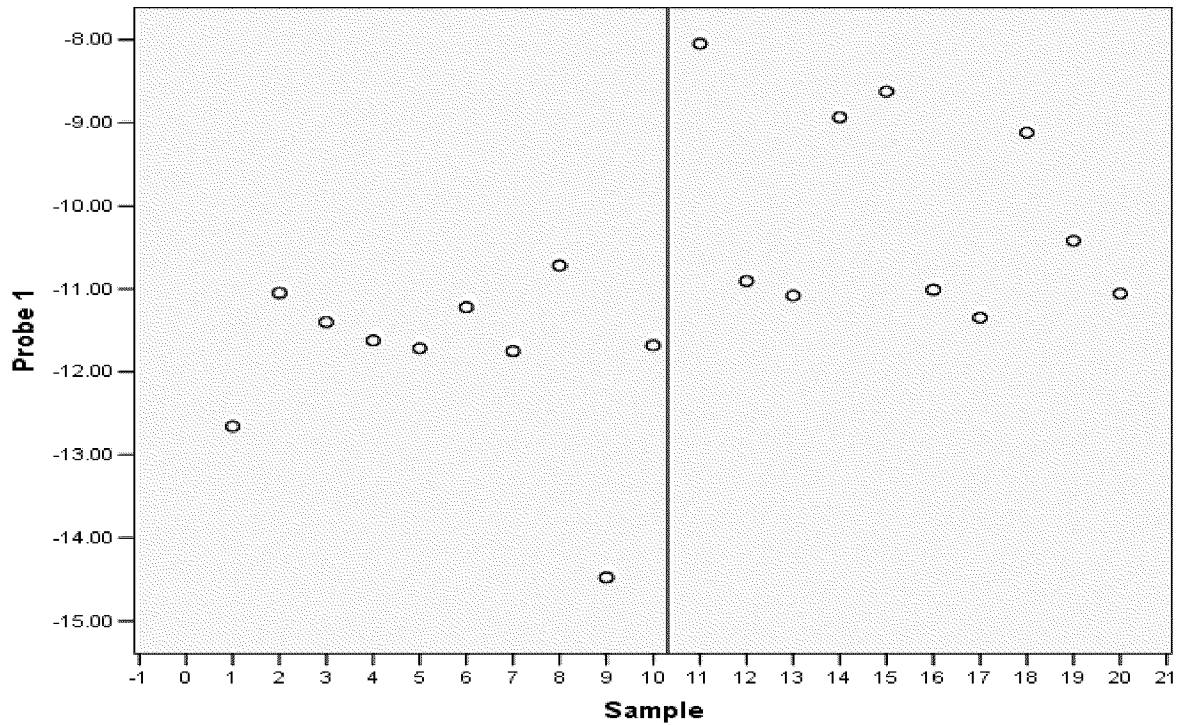


Figure 9a

23/61



ROC Curve

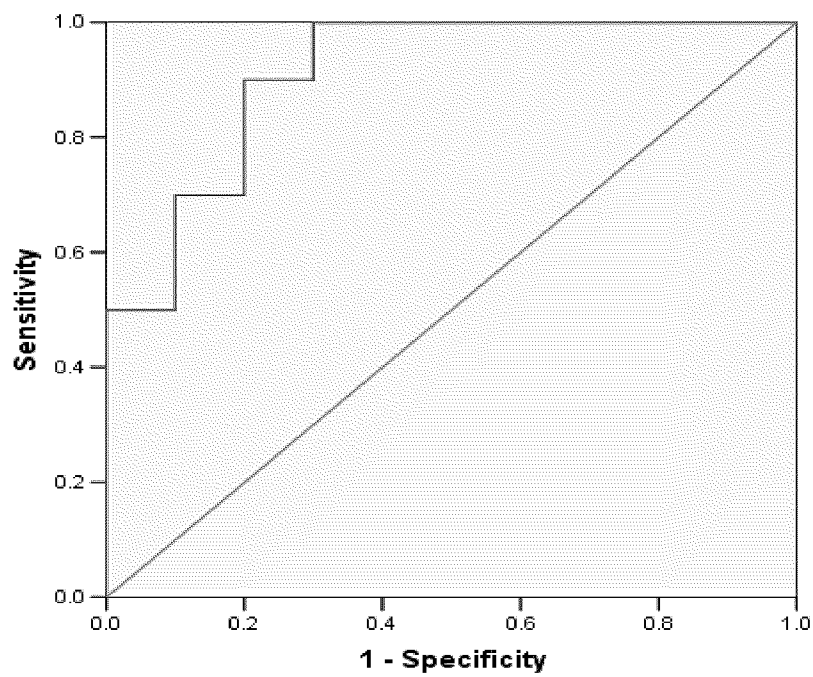


Figure 9a (cont'd)

24/61

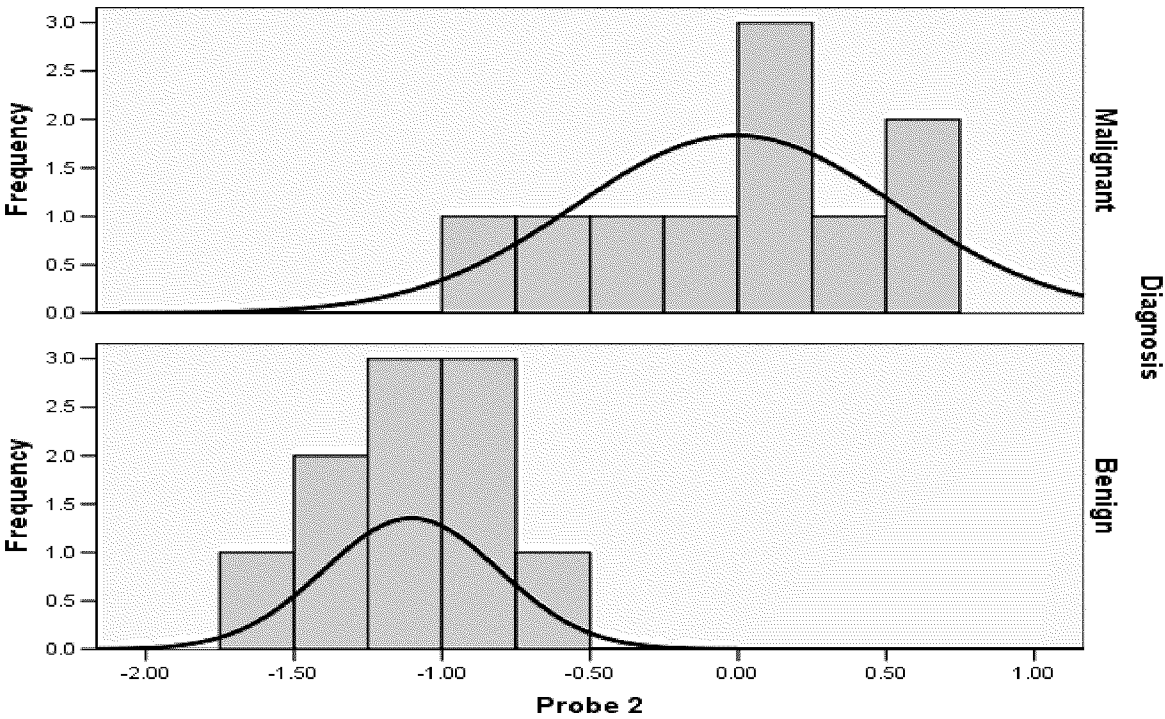
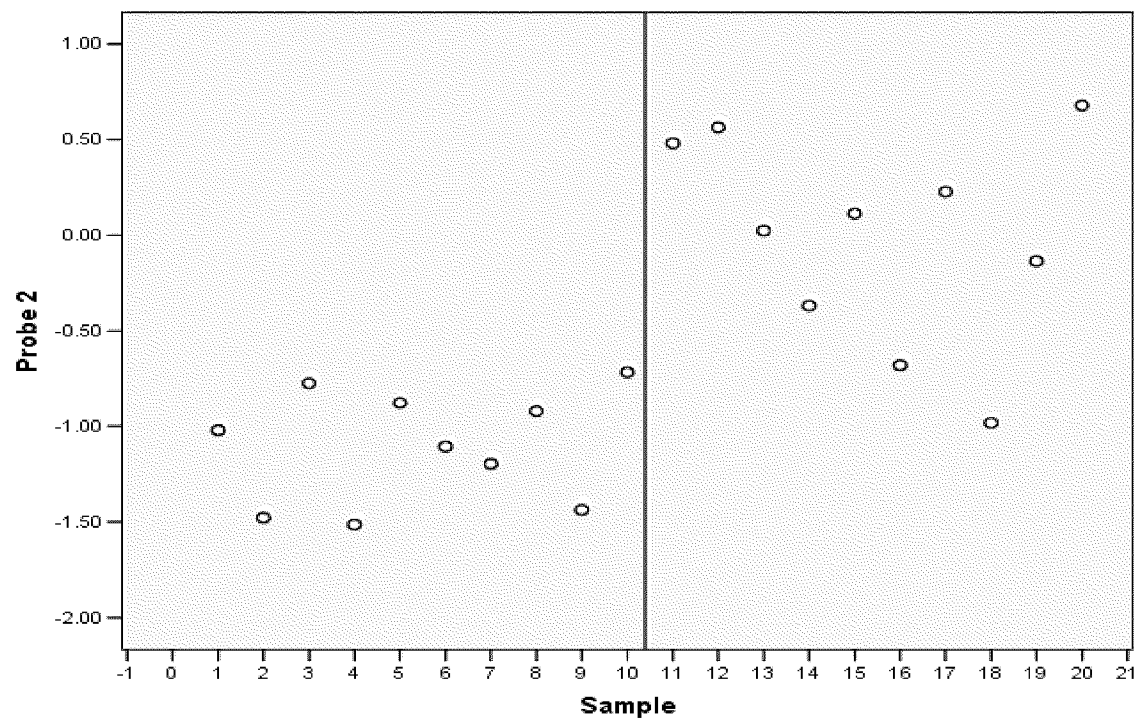


Figure 9b

25/61



ROC Curve

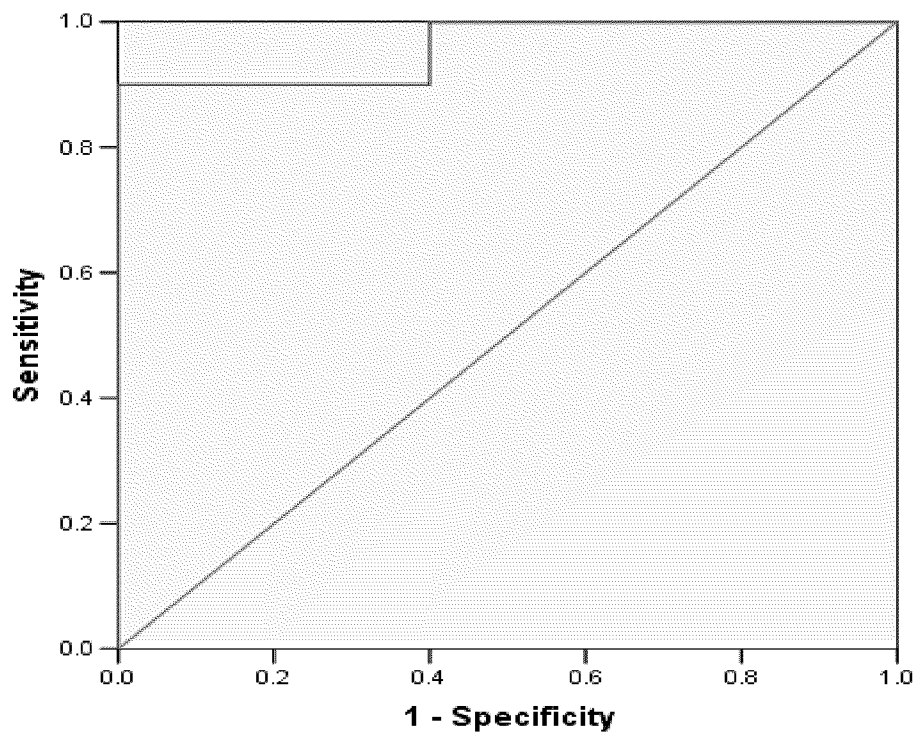


Figure 9b (cont'd)

26/61

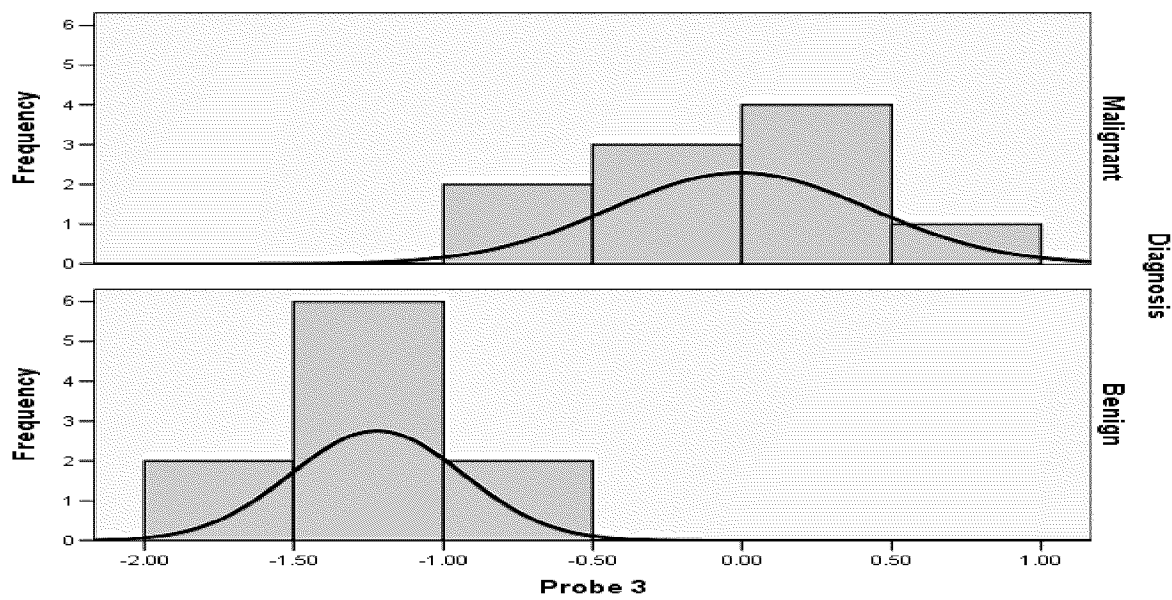
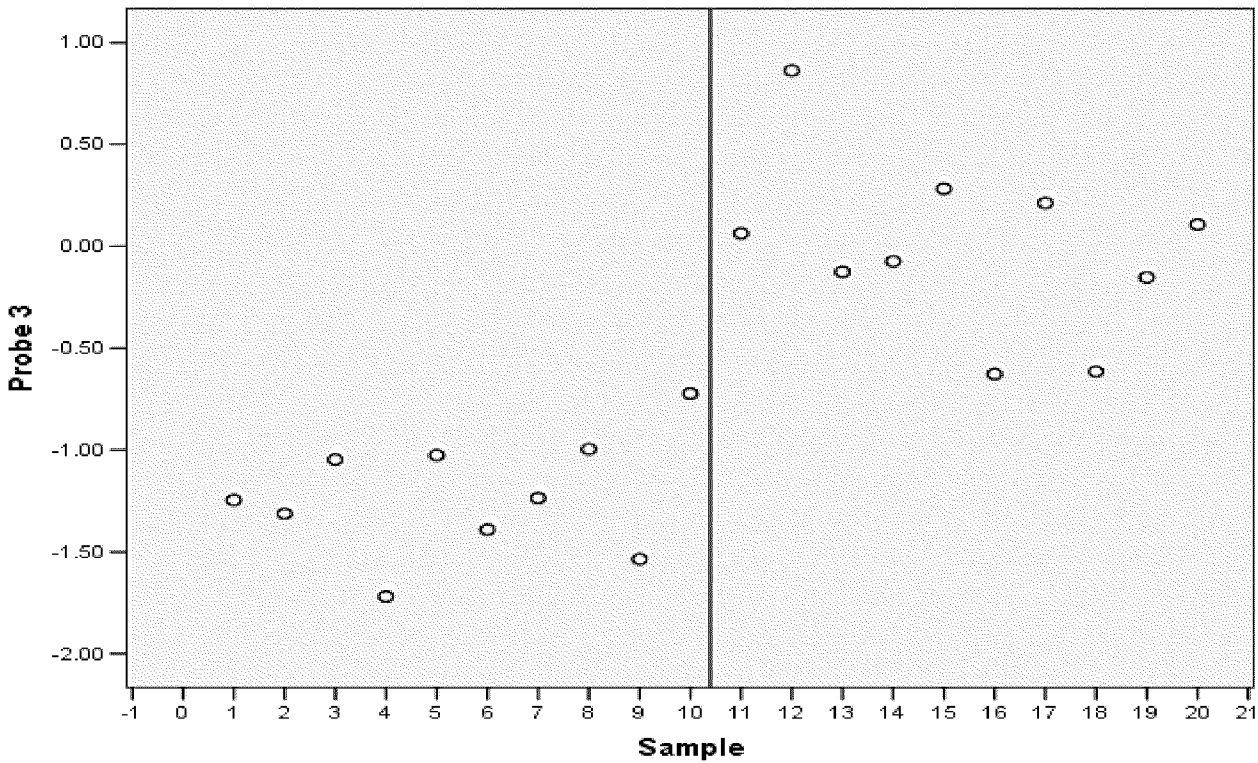


Figure 9c

27/61



ROC Curve

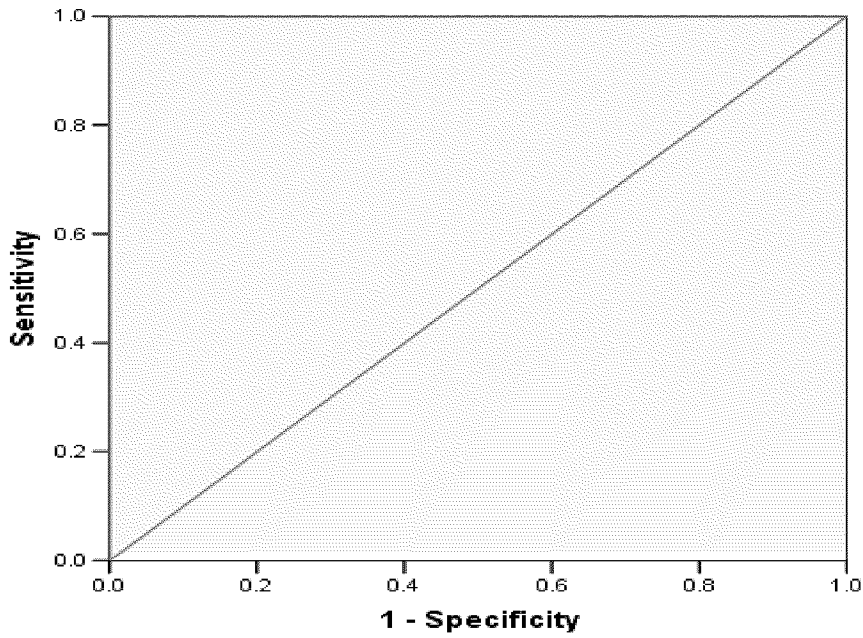


Figure 9c (cont'd)

28/61

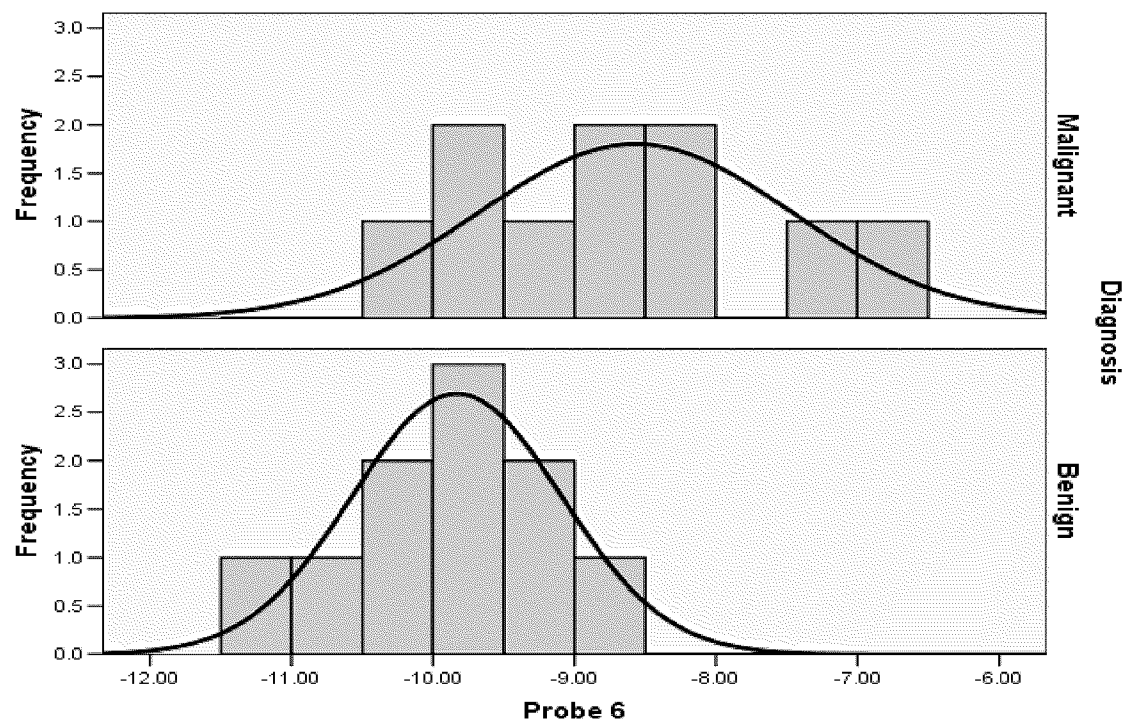


Figure 9d

29/61

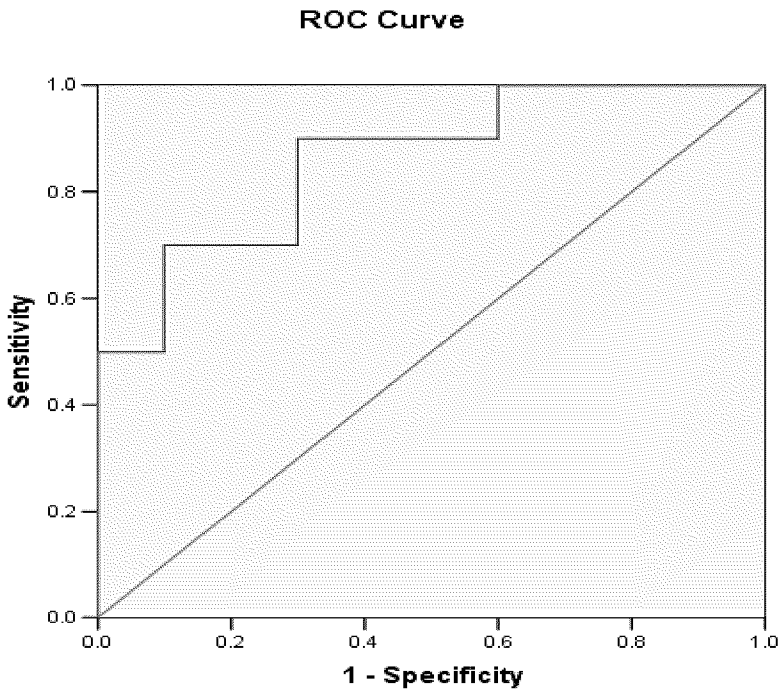
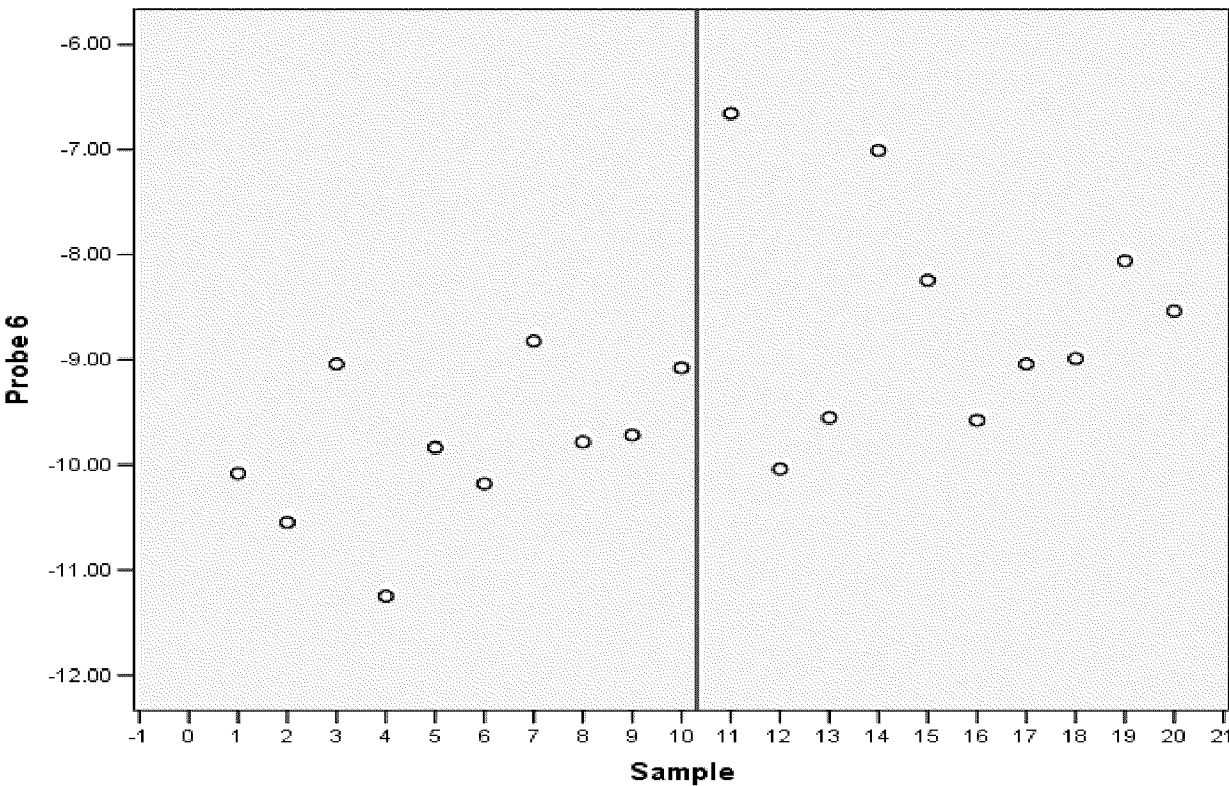


Figure 9d (cont'd)

30/61

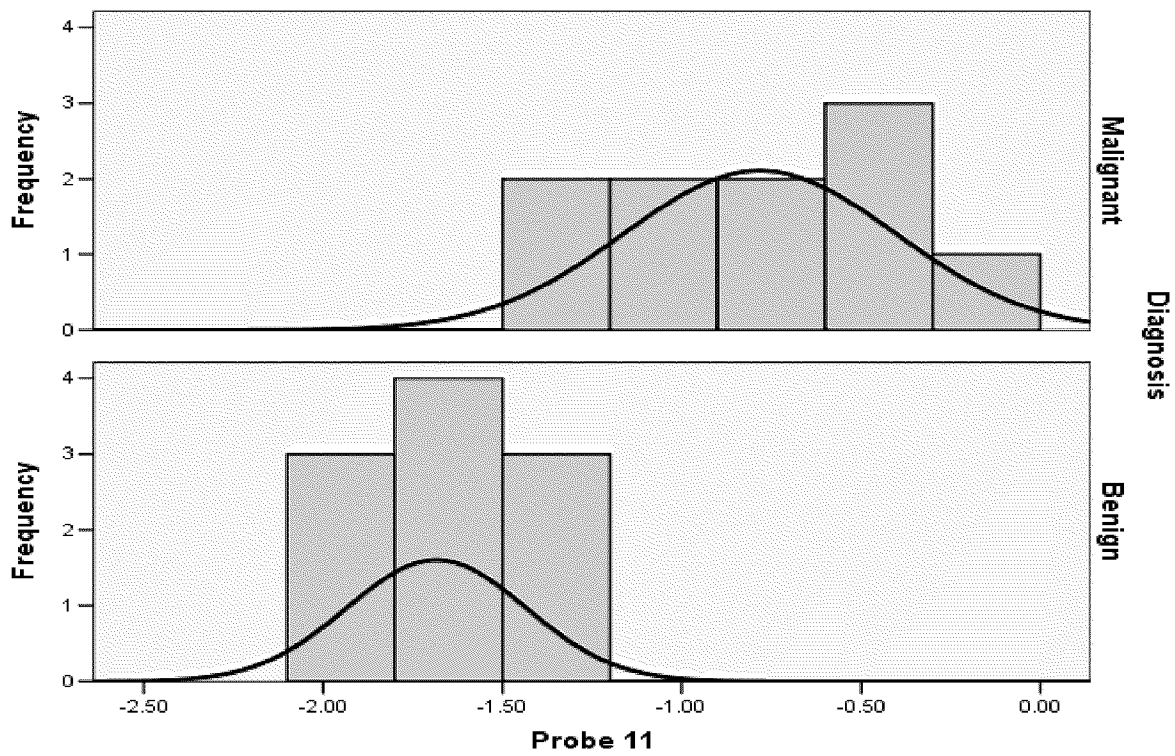
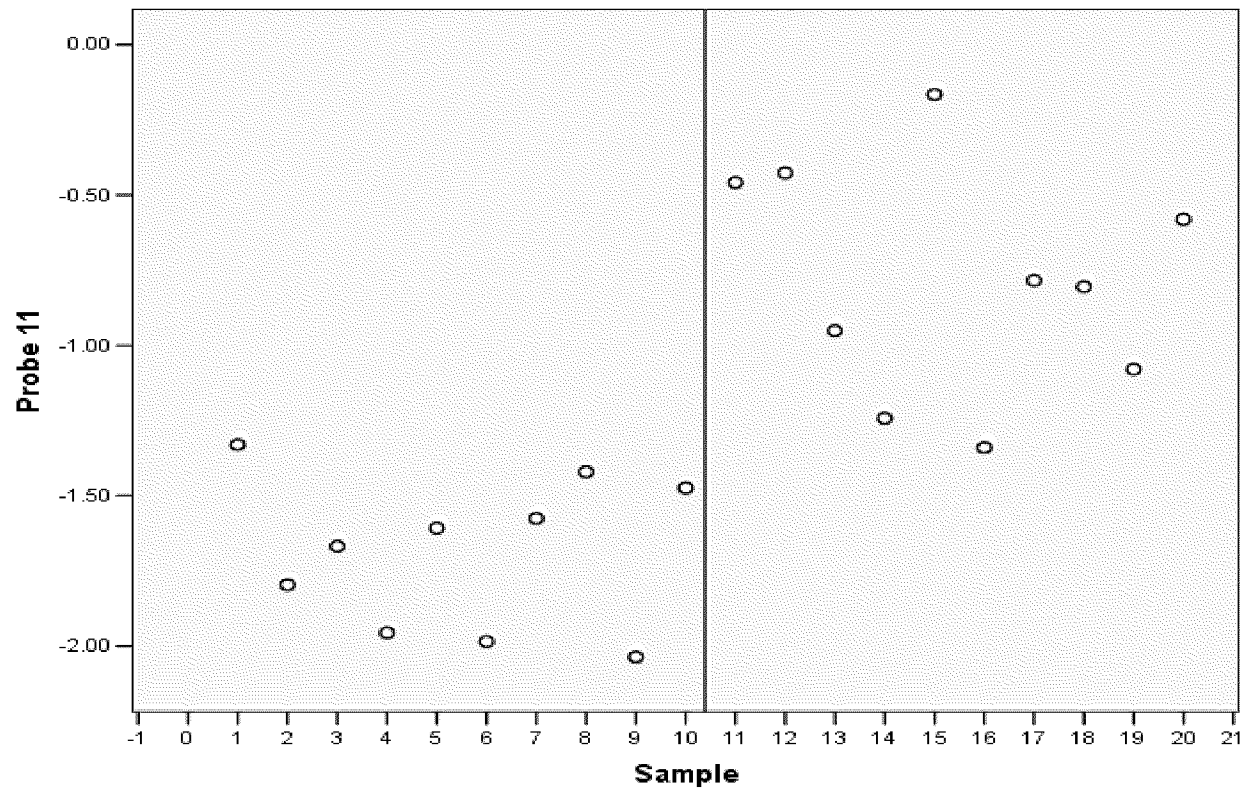


Figure 9e

31/61



ROC Curve

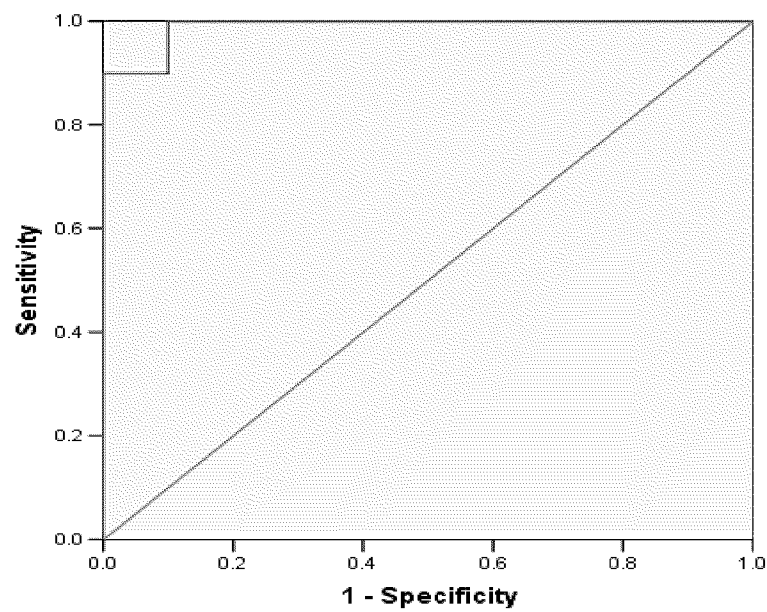


Figure 9e (cont'd)

32/61

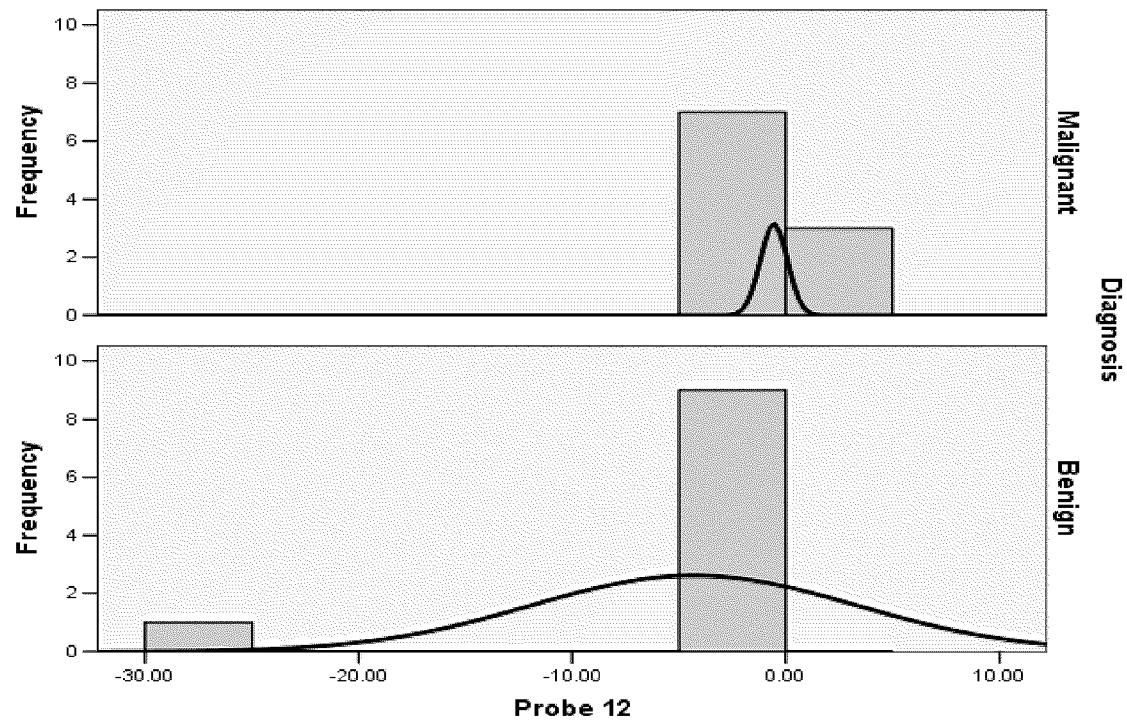
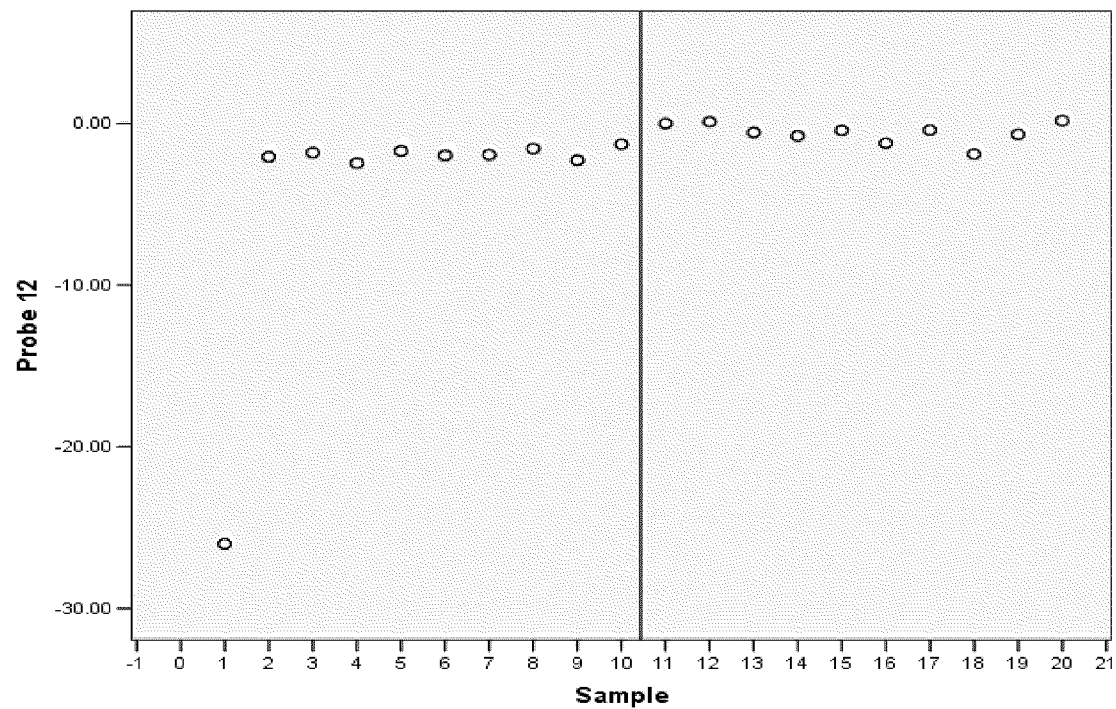


Figure 9f

33/61



ROC Curve

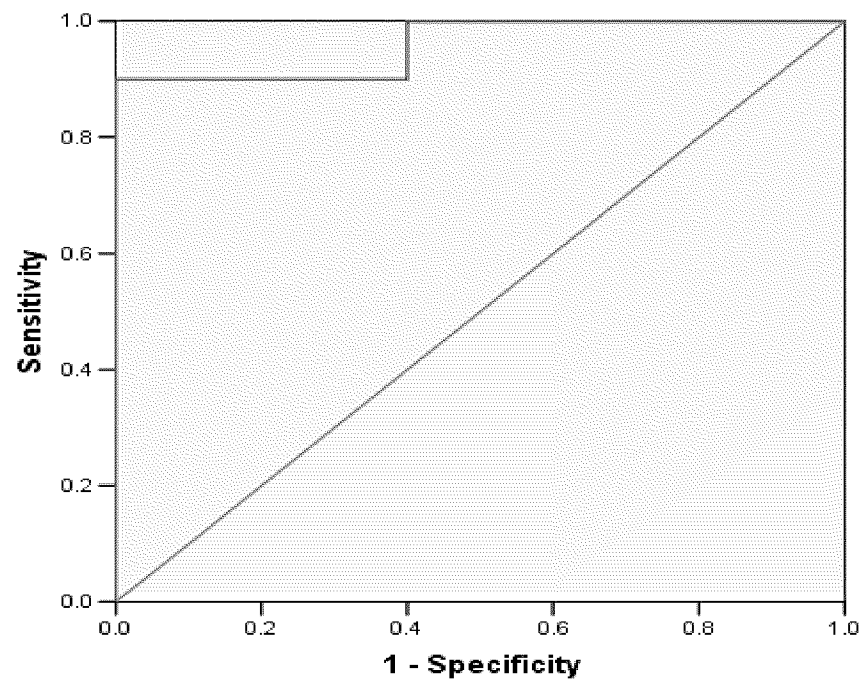


Figure 9f (cont'd)

34/61

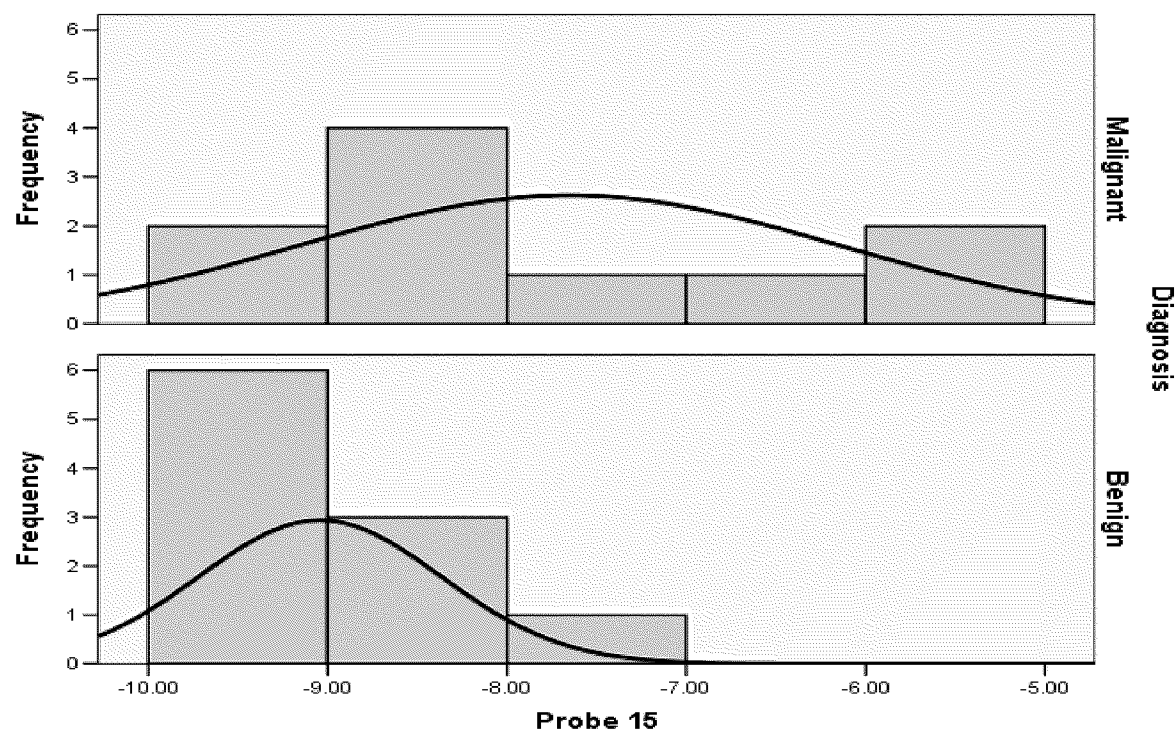
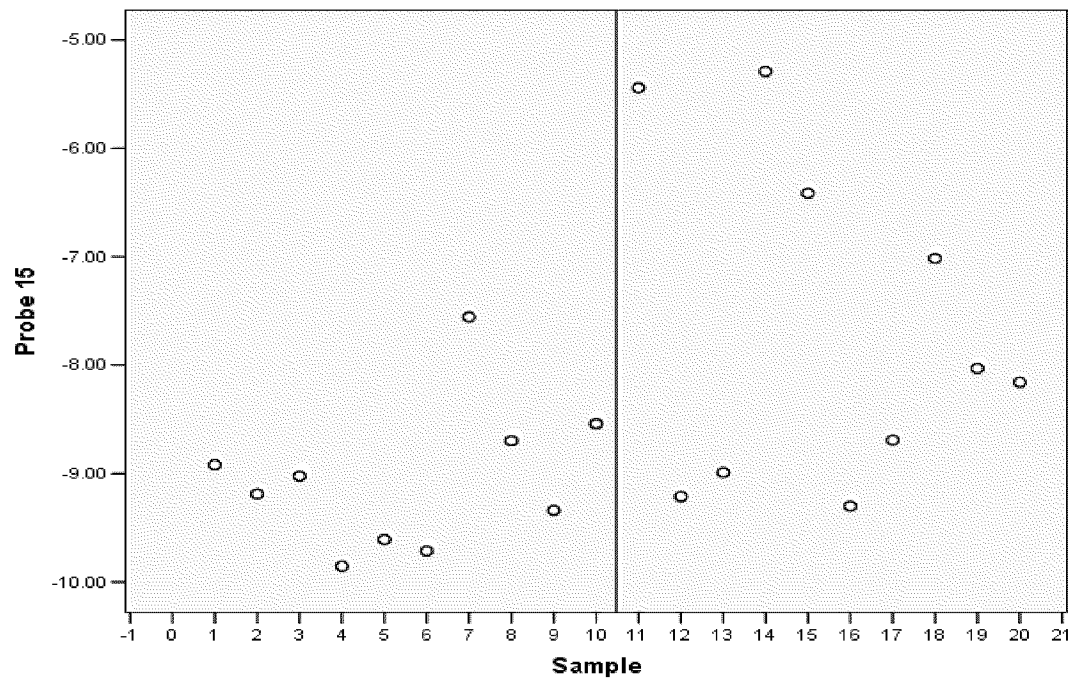


Figure 9g

35/61



ROC Curve

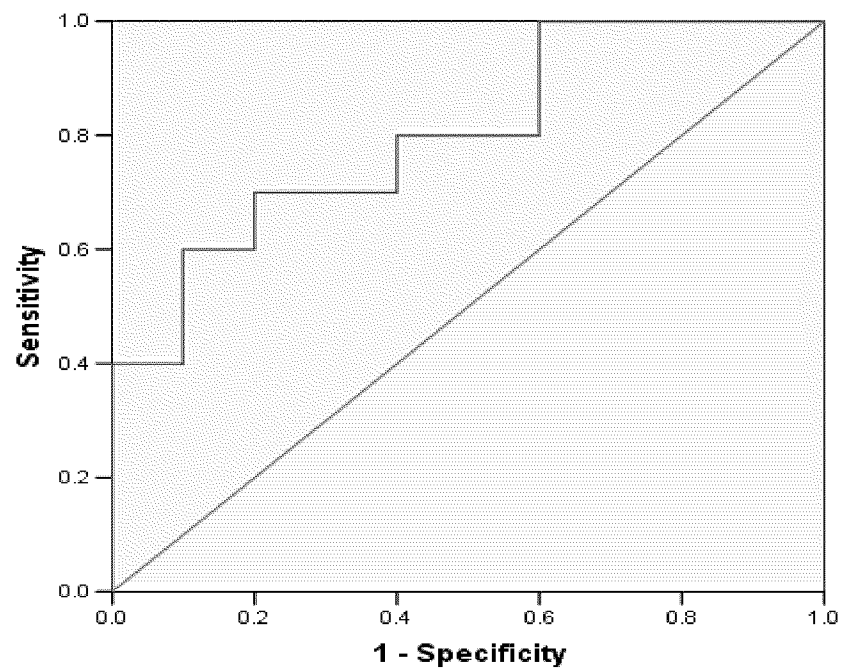


Figure 9g (cont'd)

36/61

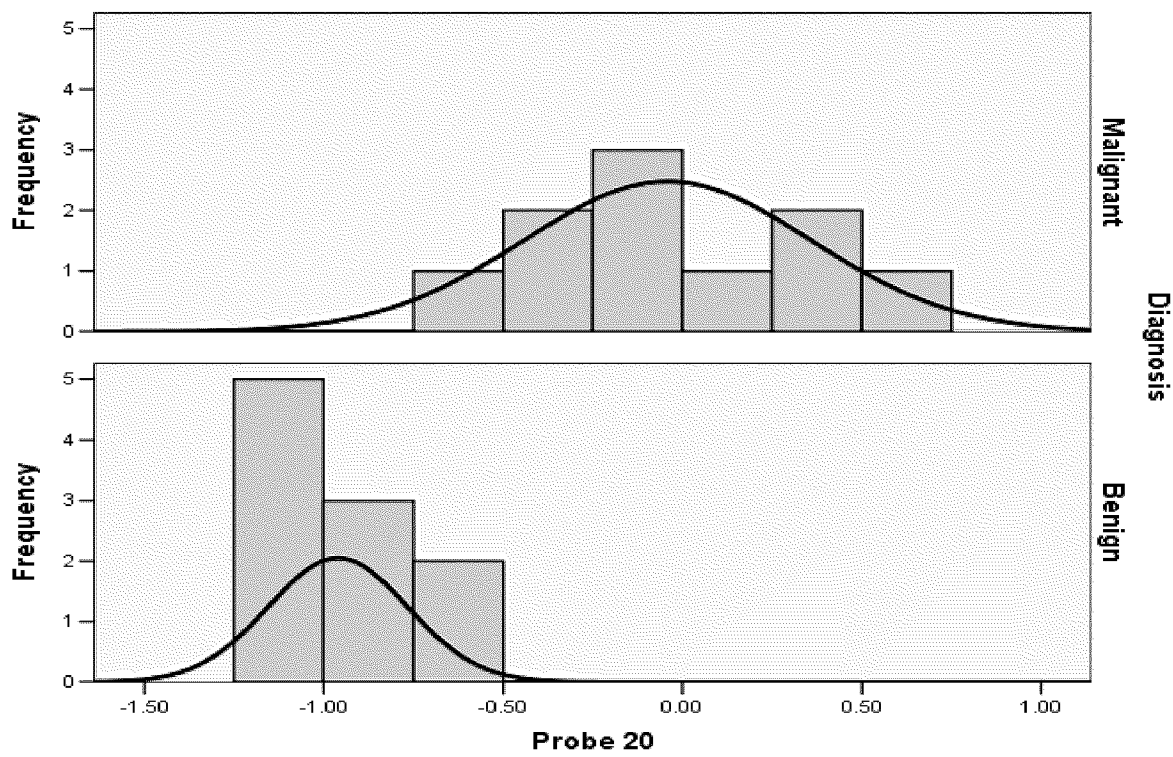
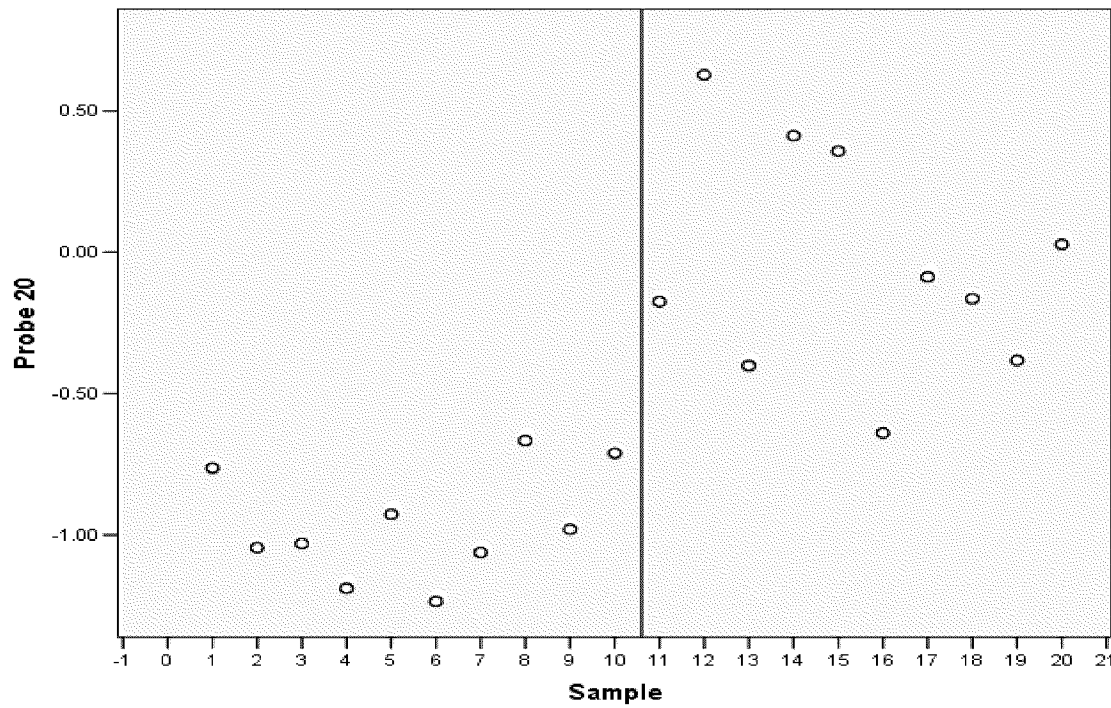


Figure 9h

37/61



ROC Curve

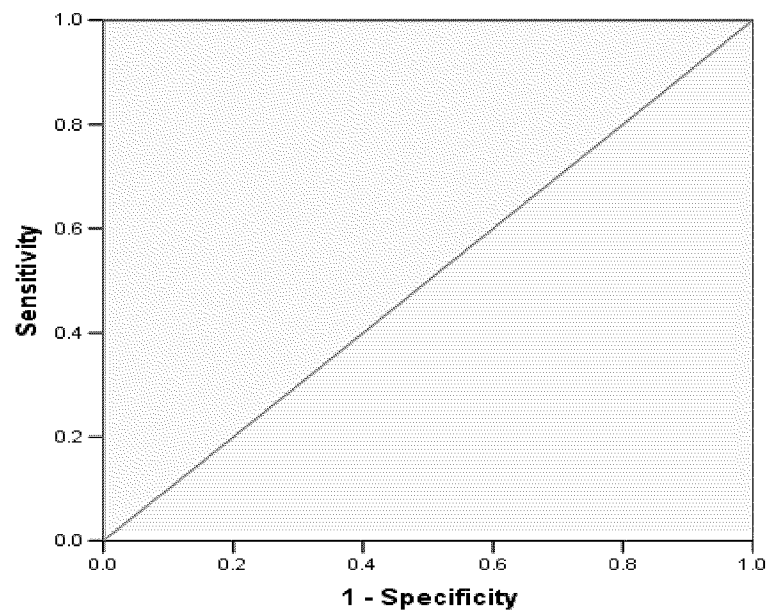


Figure 9h (cont'd)

38/61
Transcript 2

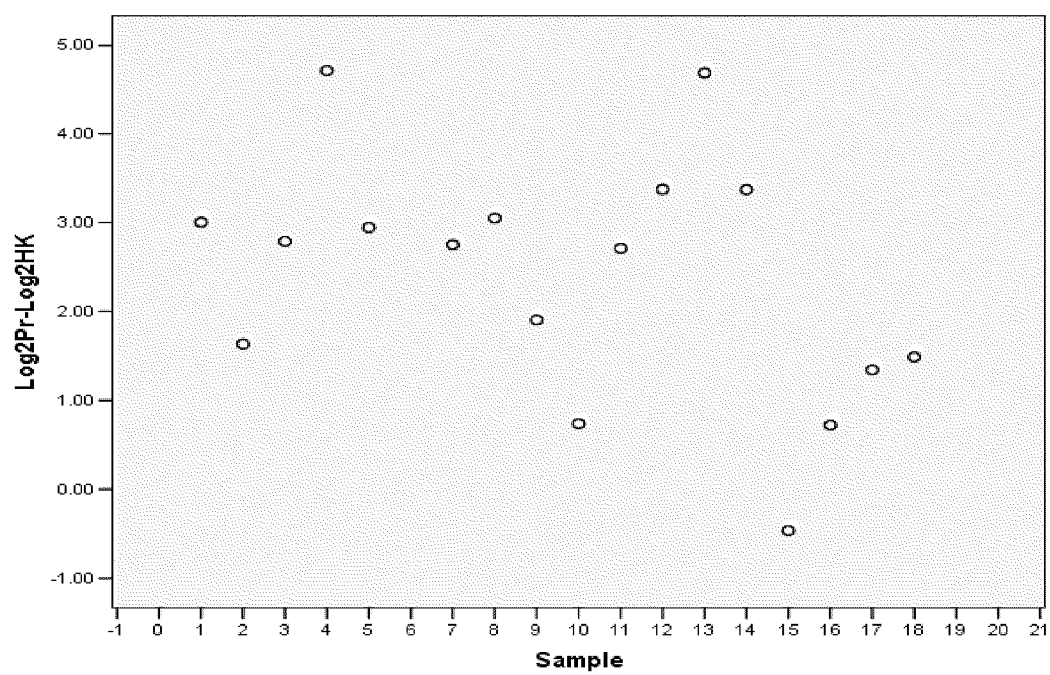


Figure 10a

39/61

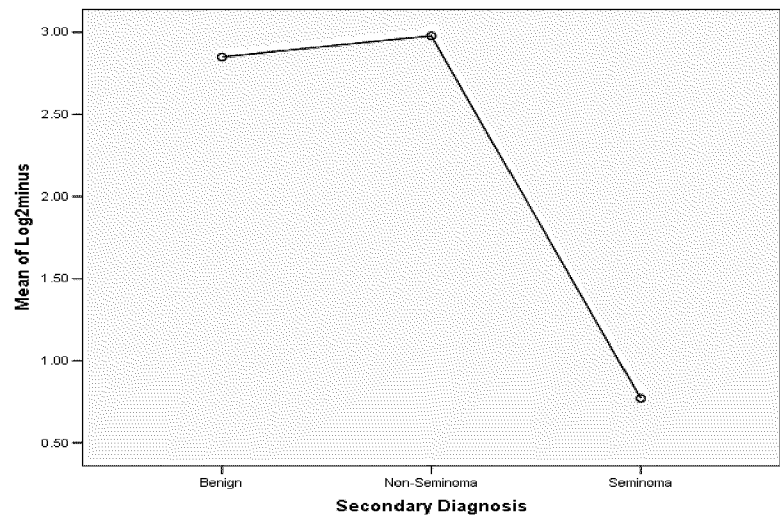


Figure 10a (cont'd)

Benign to Seminoma

ROC Curve

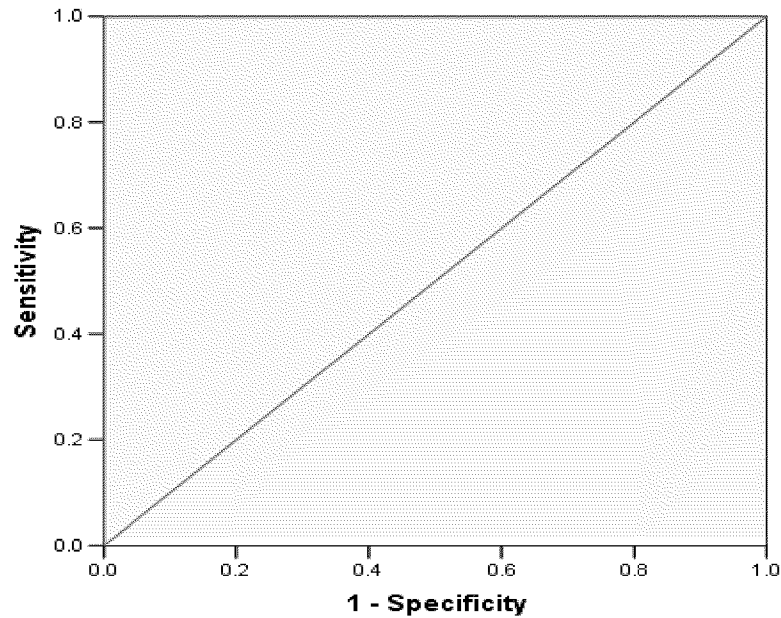


Figure 10b

40/61

Non-Seminoma to Seminoma

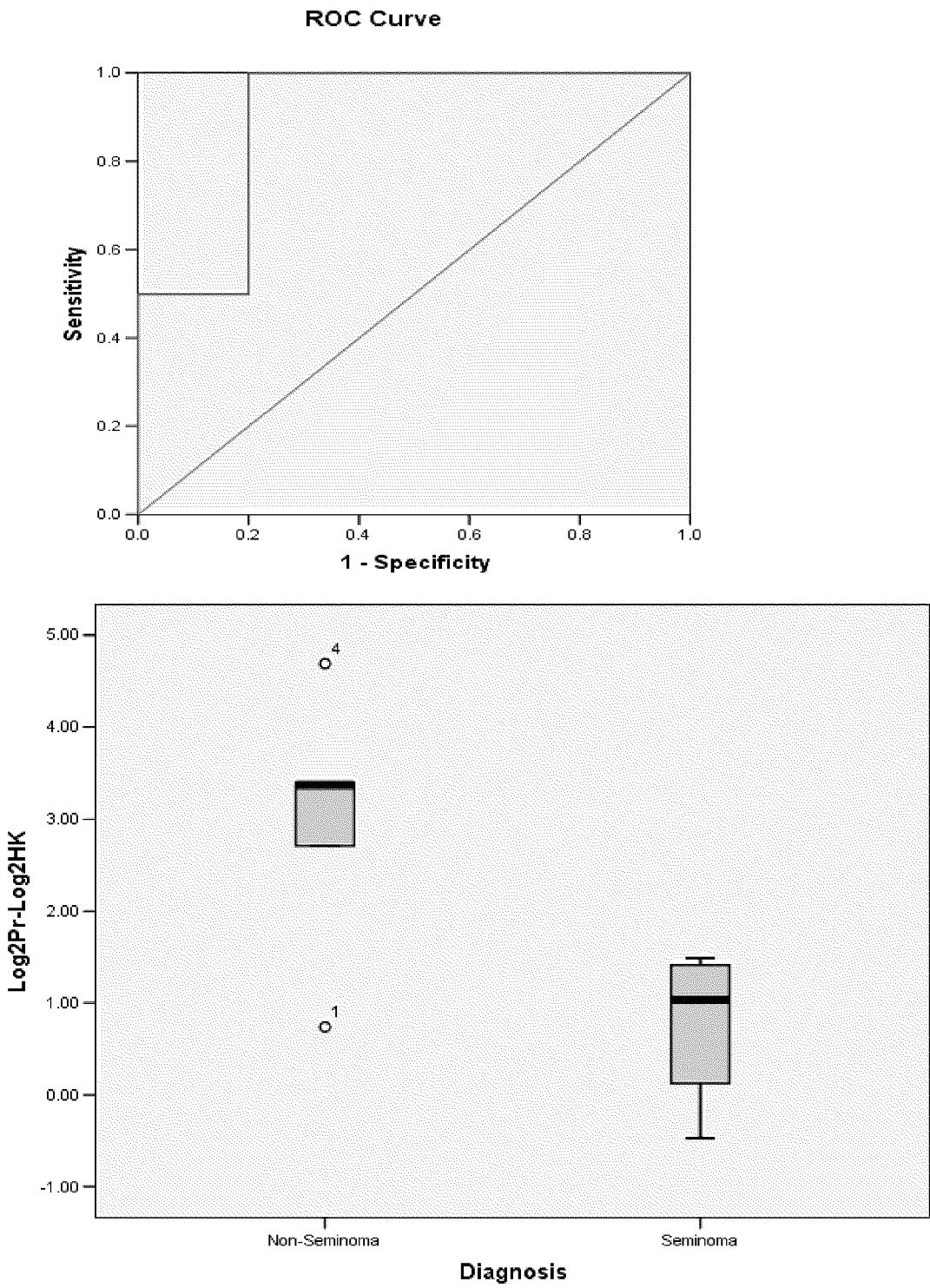


Figure 10b (cont'd)

41/61

Transcript 3

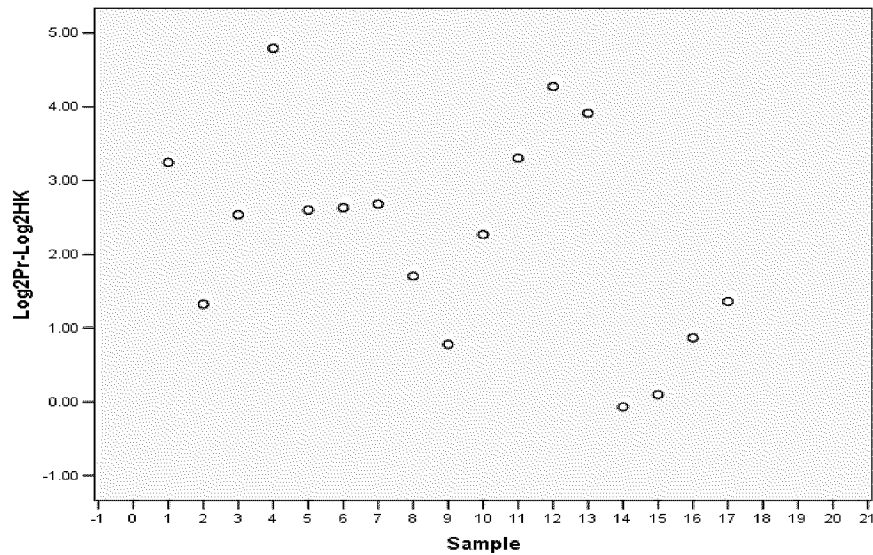


Figure 11a

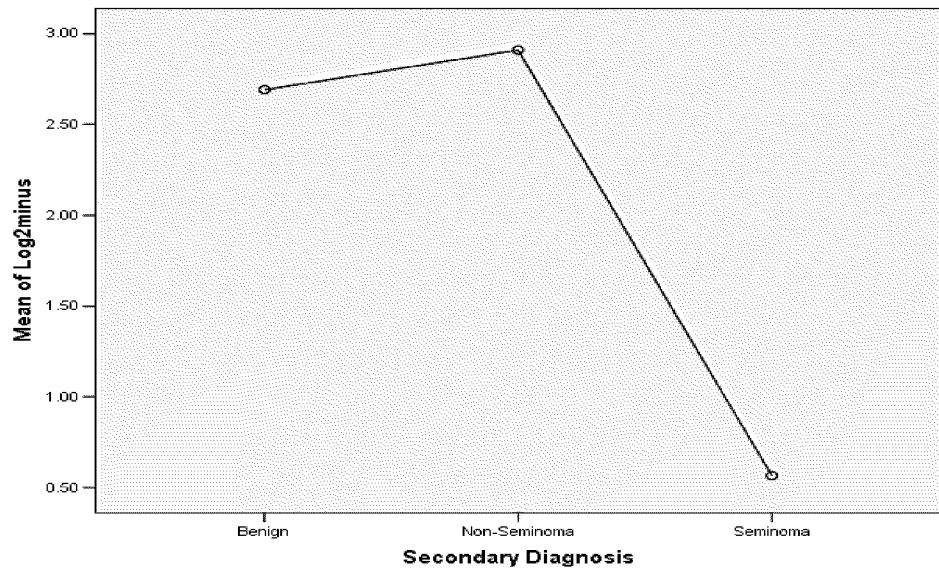


Figure 11a (cont'd)

42/61

Benign to Seminoma

ROC Curve

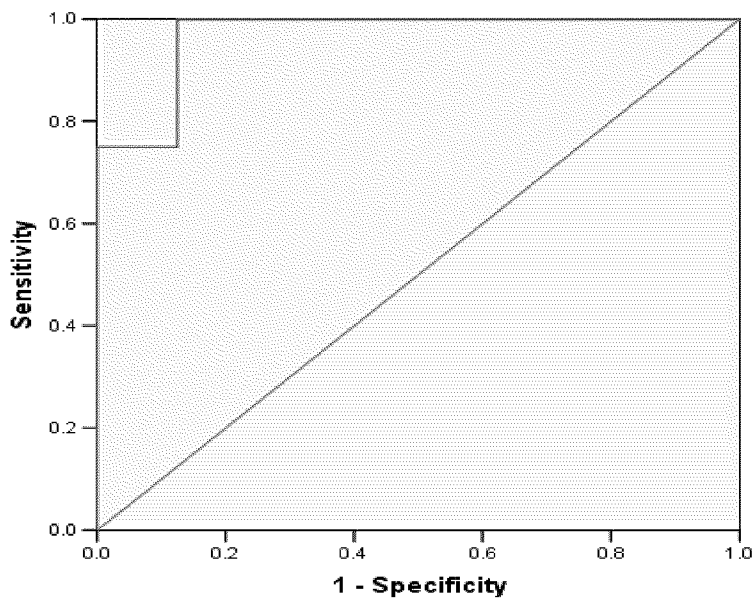


Figure 11b

Non-Seminoma to Seminoma

ROC Curve

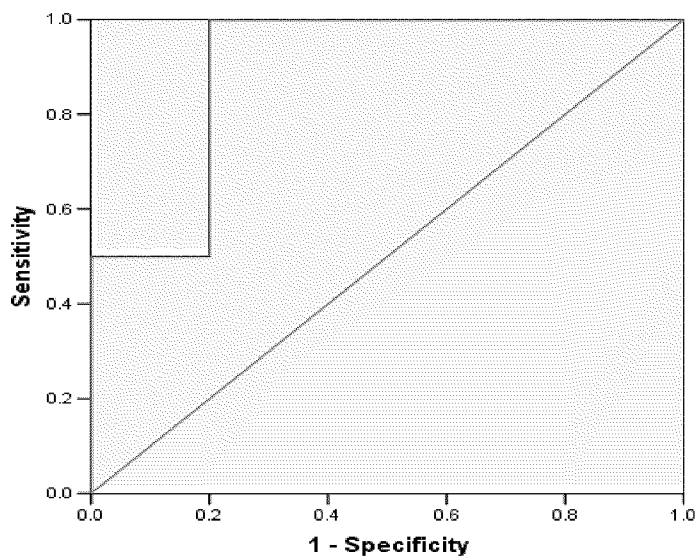


Figure 11b (cont'd)

43/61

Transcript 4

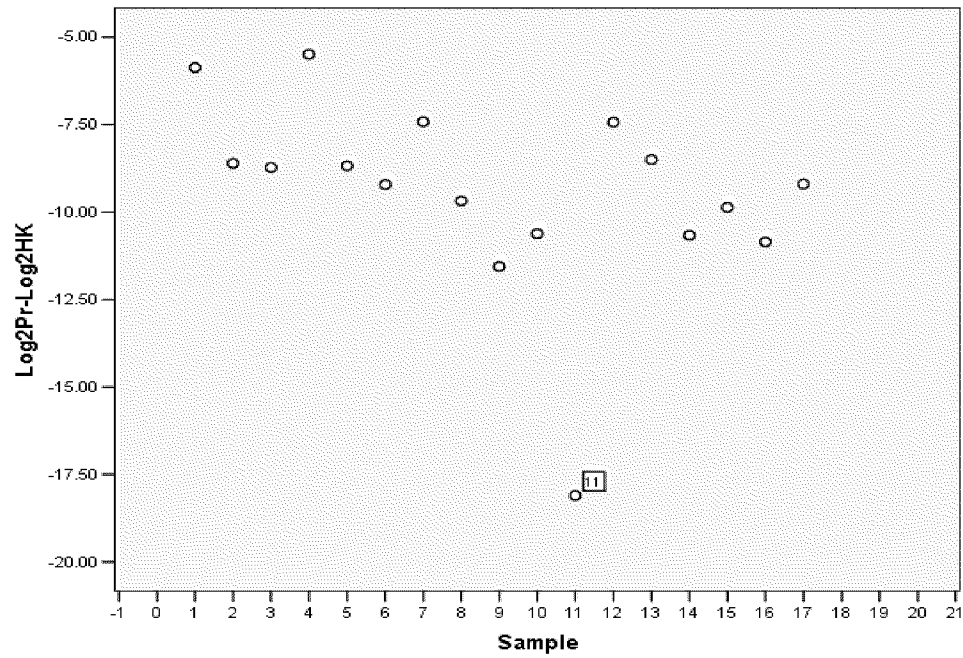


Figure 12

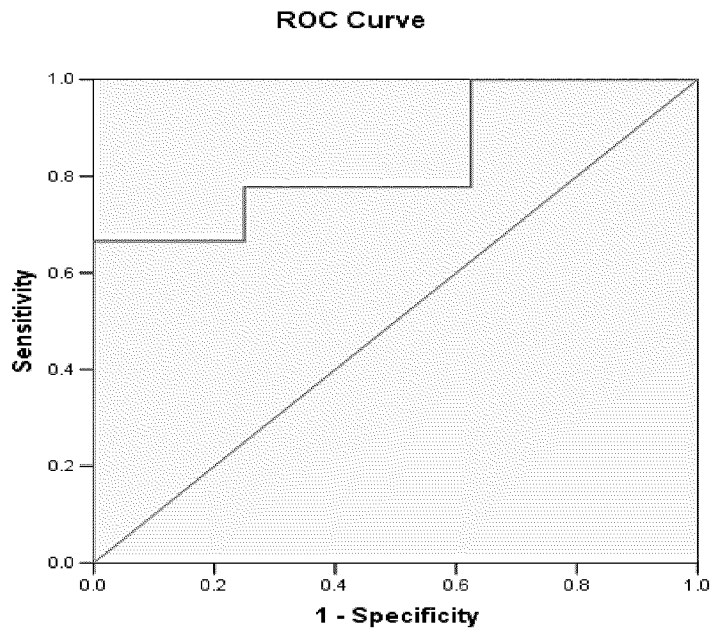


Figure 12 (cont'd)

44/61

Transcript 11

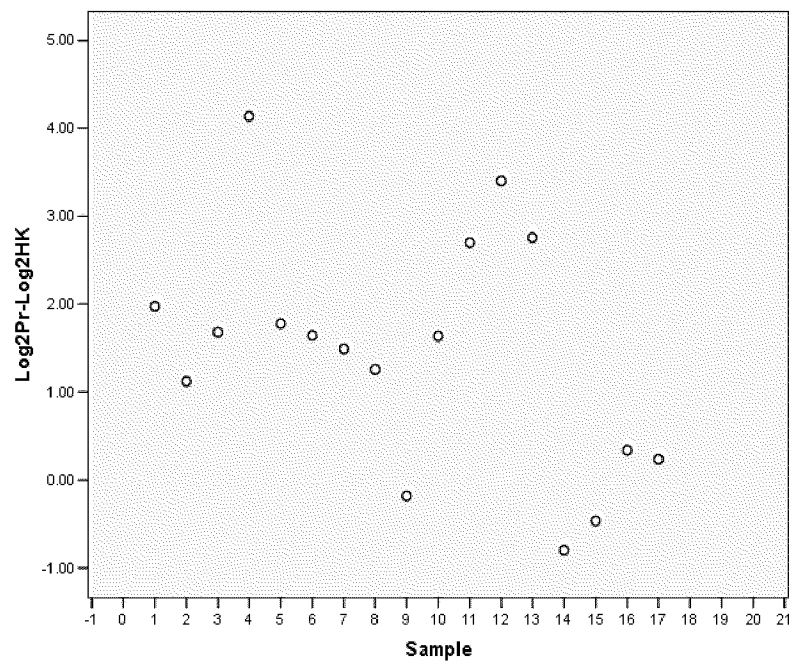


Figure 13a

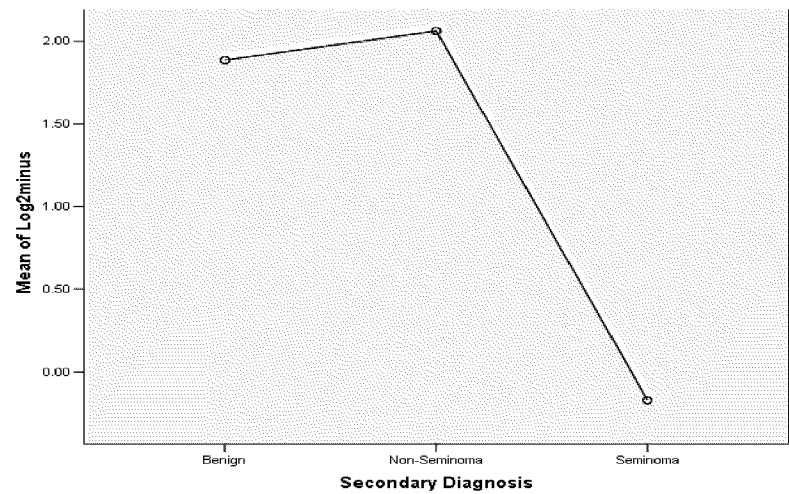


Figure 13a (cont'd)

45/61

Benign to Seminoma

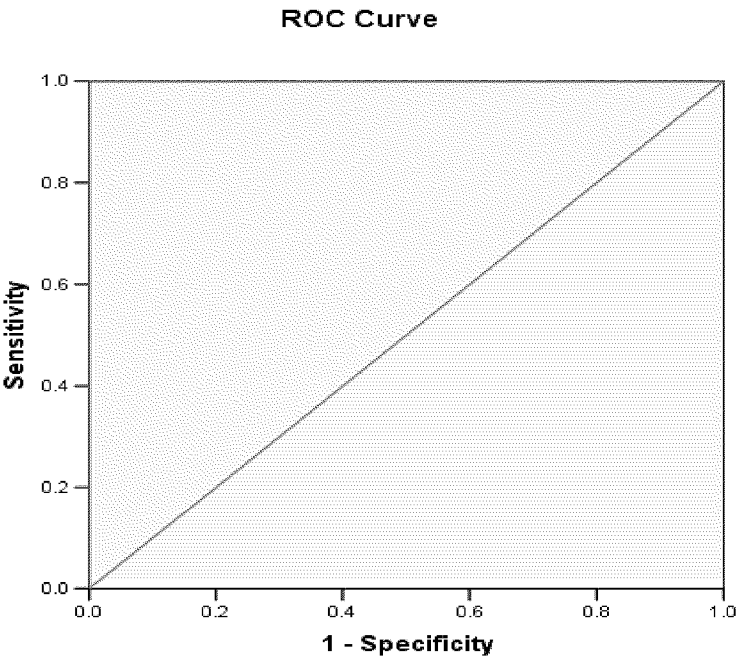


Figure 13b

Non-Seminoma to Seminoma

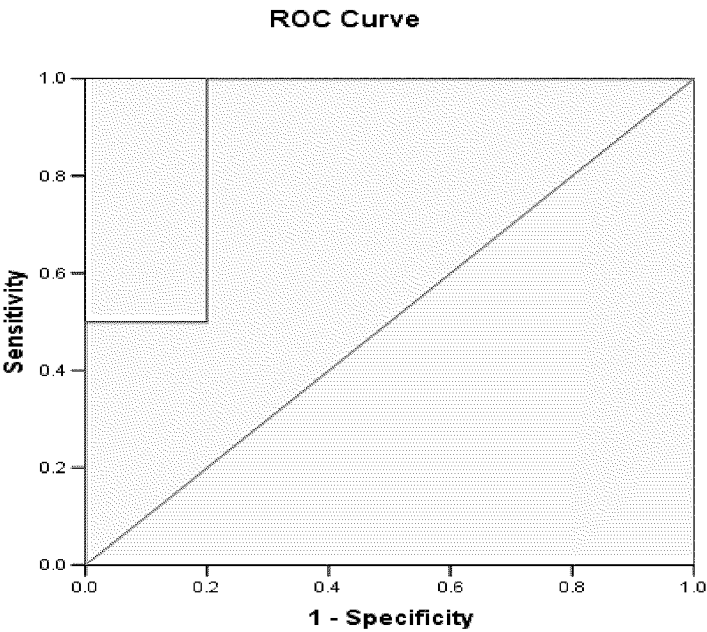


Figure 13b (cont'd)

Transcript 12

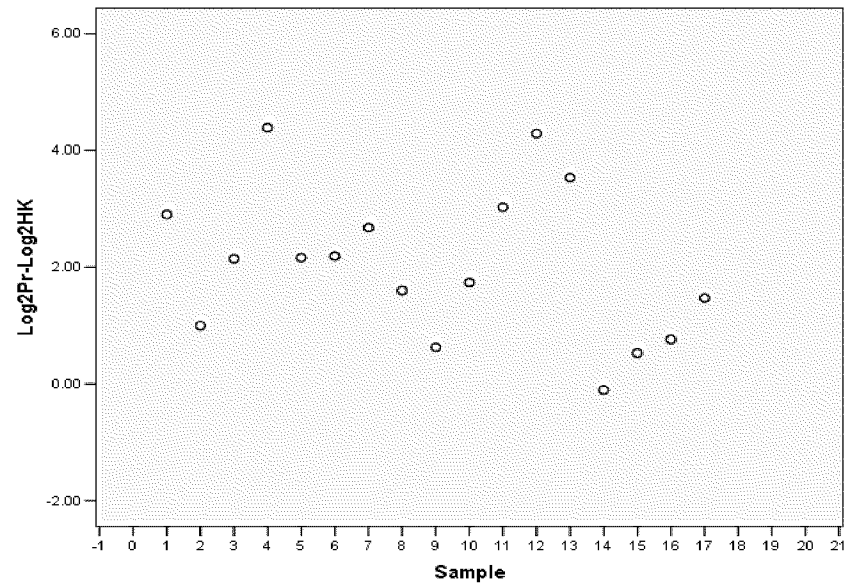


Figure 14a

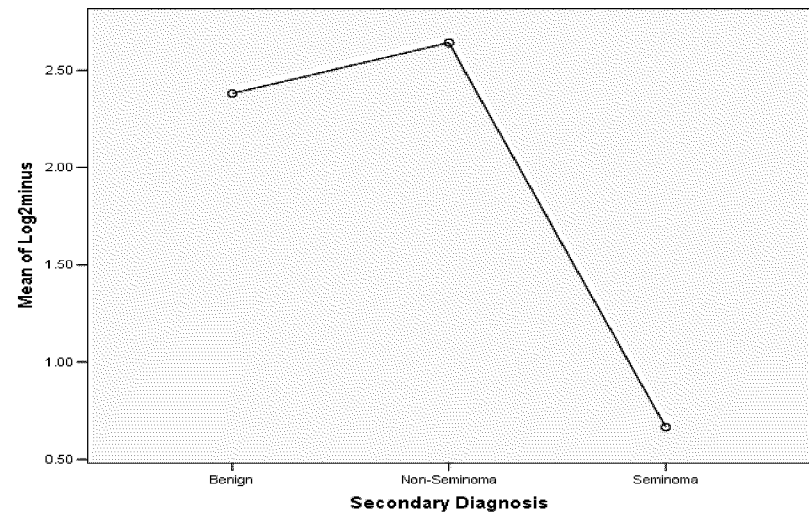


Figure 14a (cont'd)

47/61

Benign to Seminoma

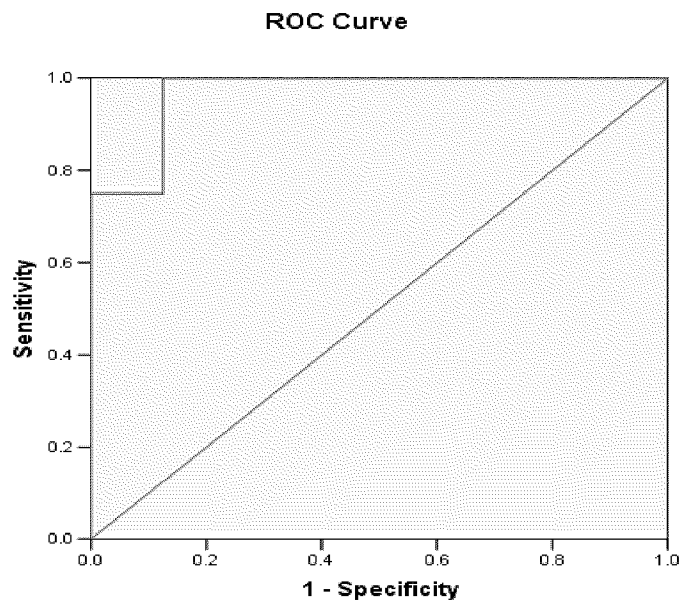


Figure 14b

Non-Seminoma to Seminoma

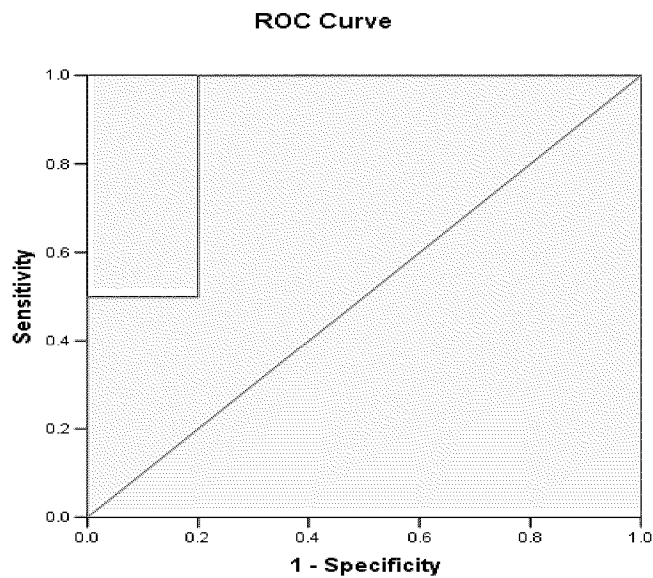


Figure 14b (cont'd)

48/61

Transcript 13

Benign to Malignant

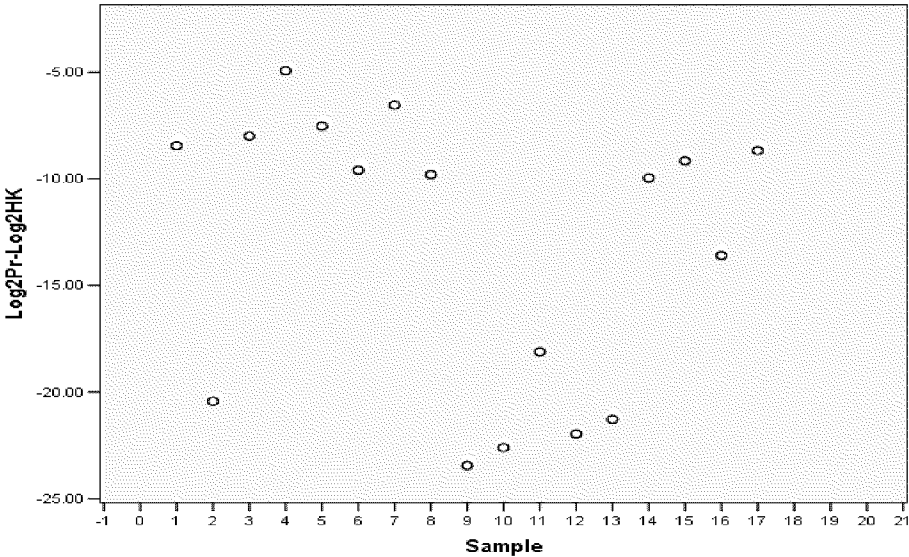


Figure 15a

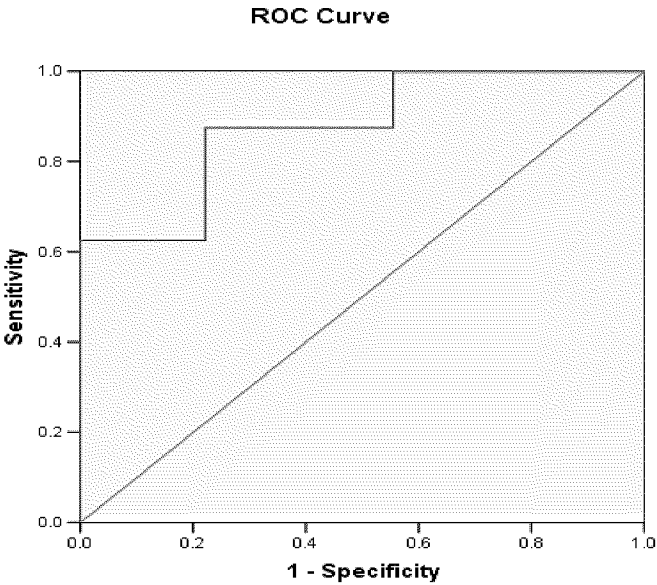


Figure 15a (cont'd)

49/61

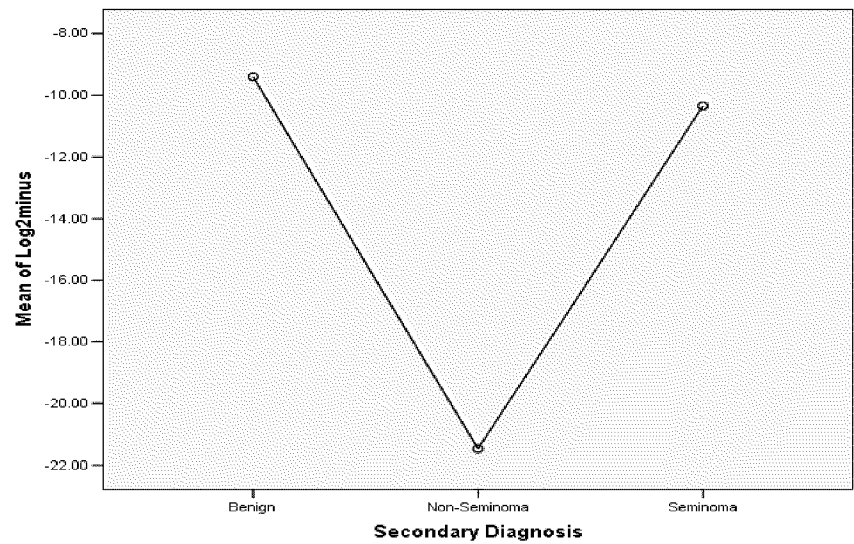


Figure 15a (cont'd)

Non-Seminoma to Seminoma

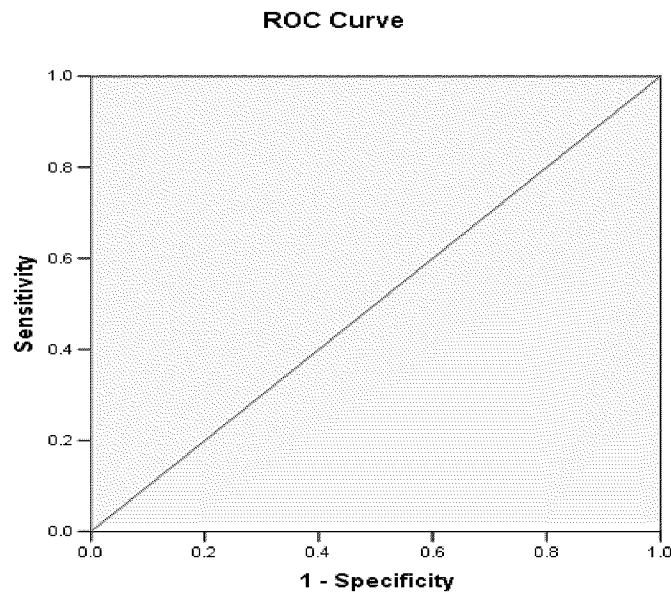


Figure 15b

50/61

Benign to Non-Seminoma

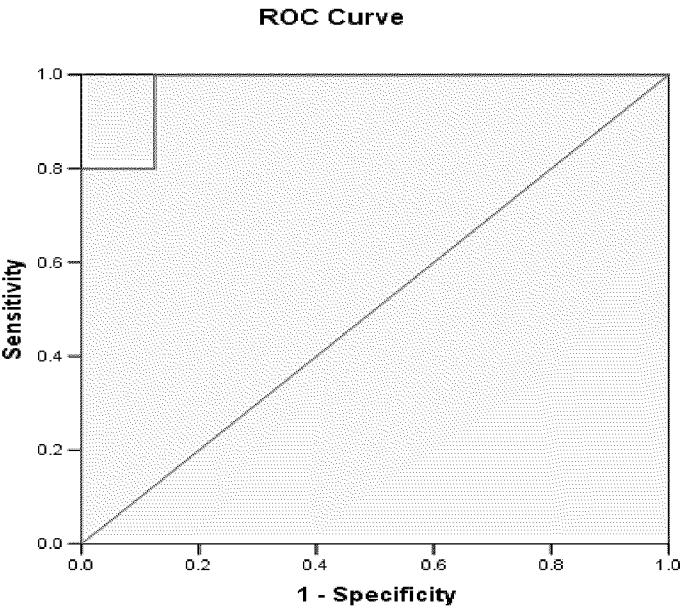


Figure 15b (cont'd)

51/61

Transcript 15

Benign to Malignant

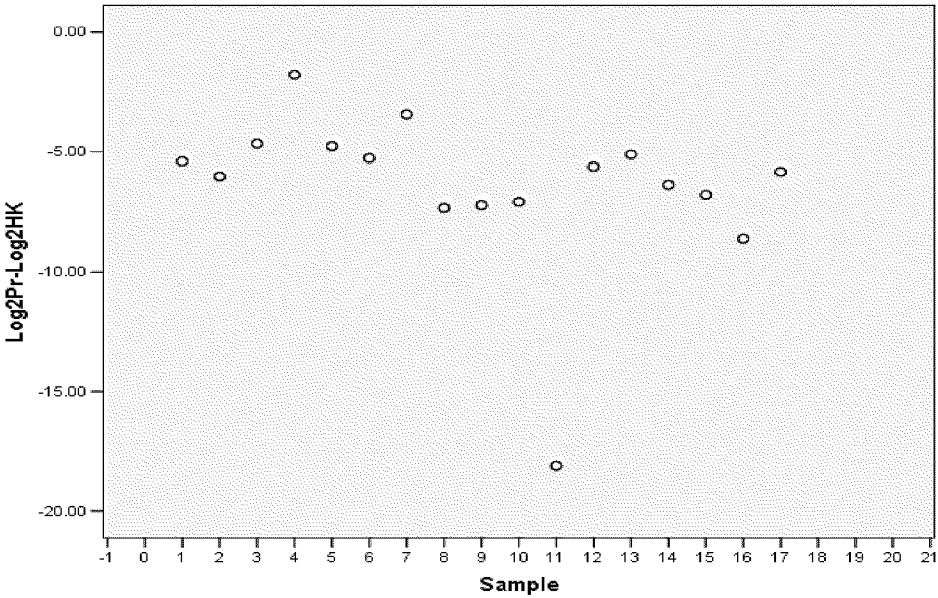


Figure 16

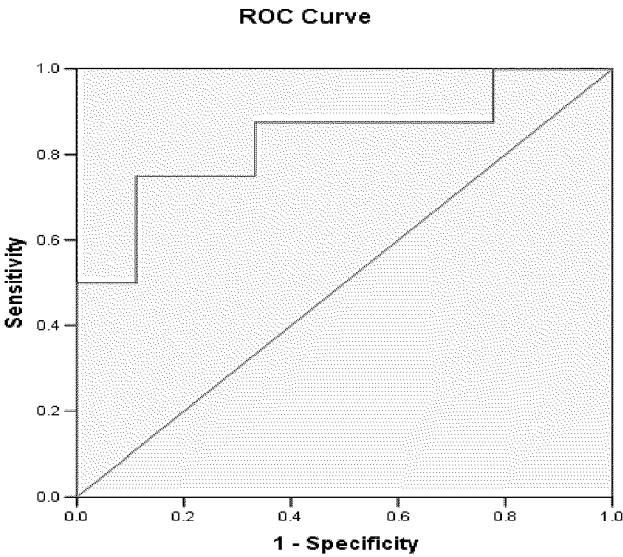


Figure 16 (cont'd)

52/61

Transcript 16

Benign to Malignant

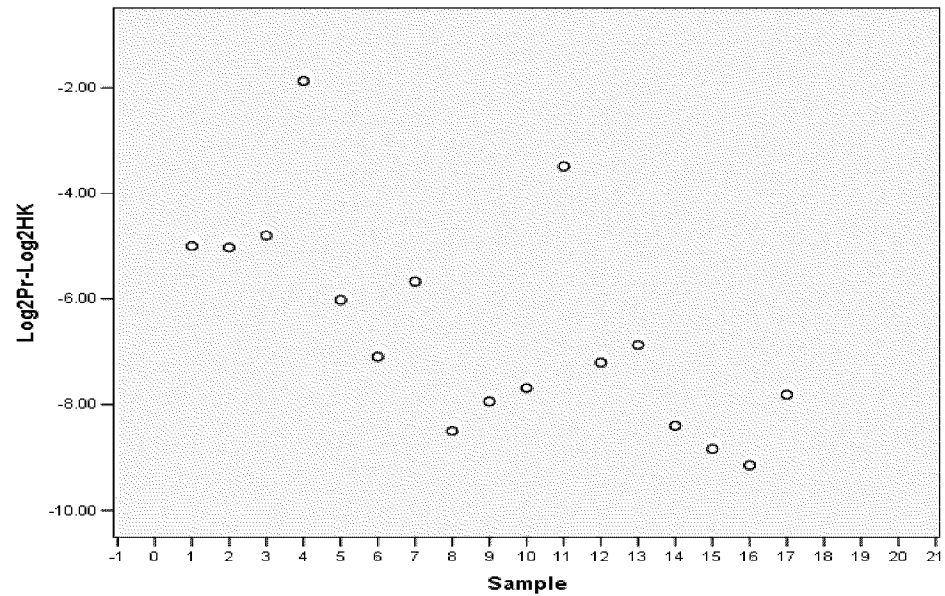


Figure 17a

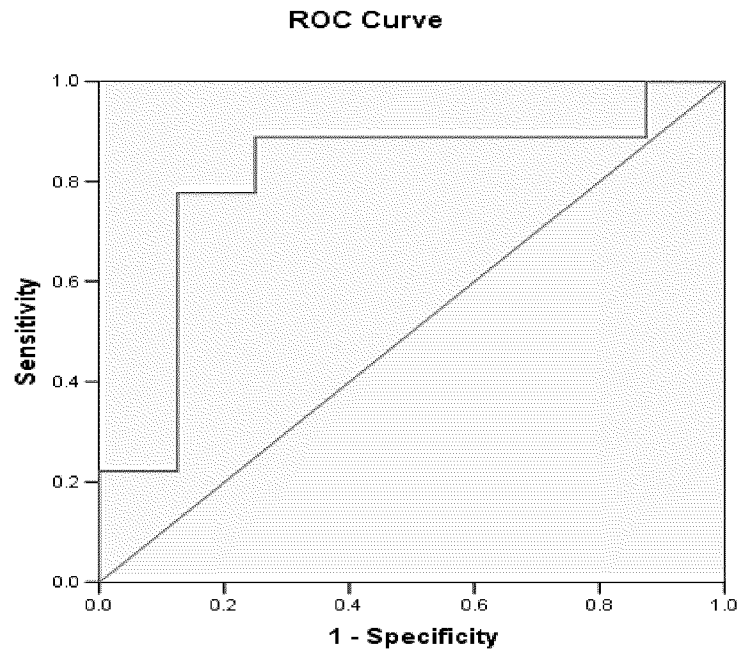


Figure 17a (cont'd)

53/61

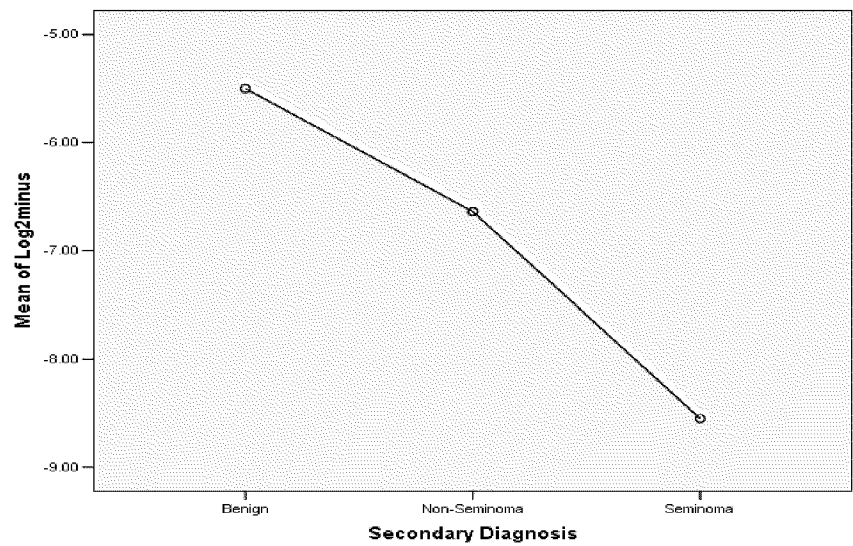


Figure 17a (cont'd)

Benign to Seminoma

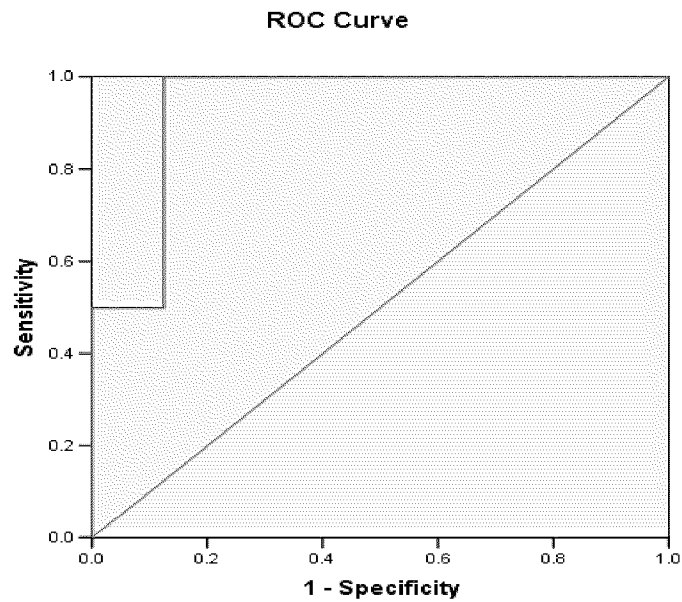


Figure 17b

54/61

Transcript 20

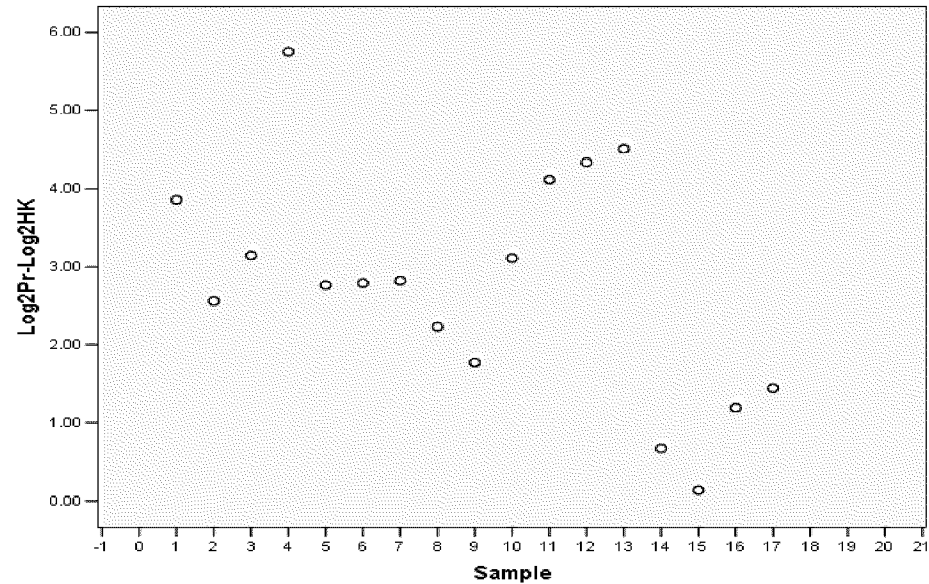


Figure 18a

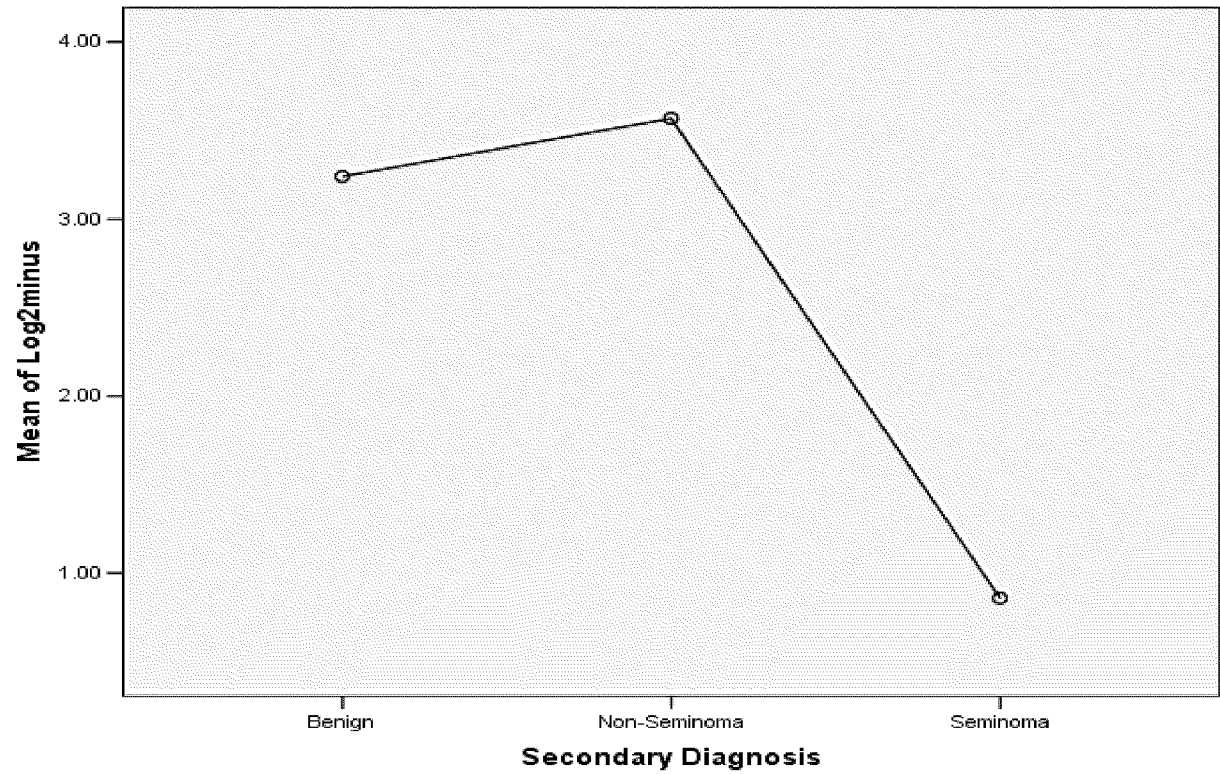


Figure 18a (cont'd)

55/61

Benign to Seminoma

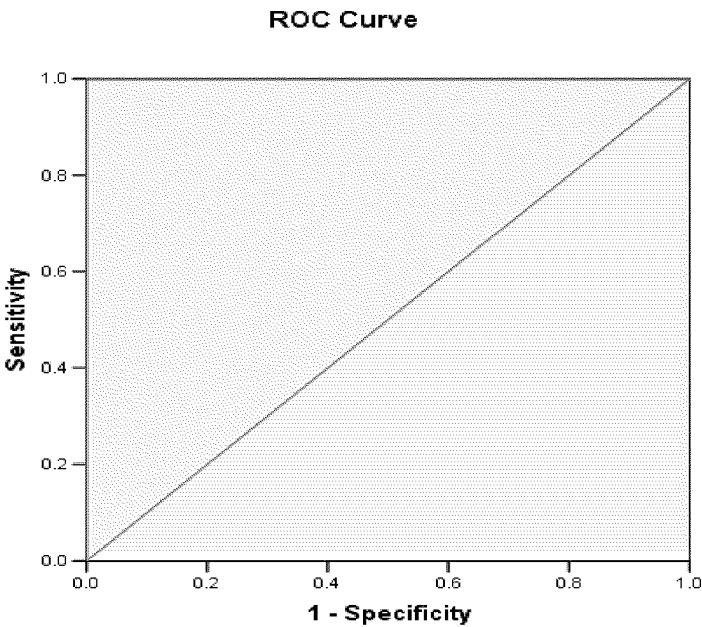


Figure 18b

Non-Seminoma to Seminoma

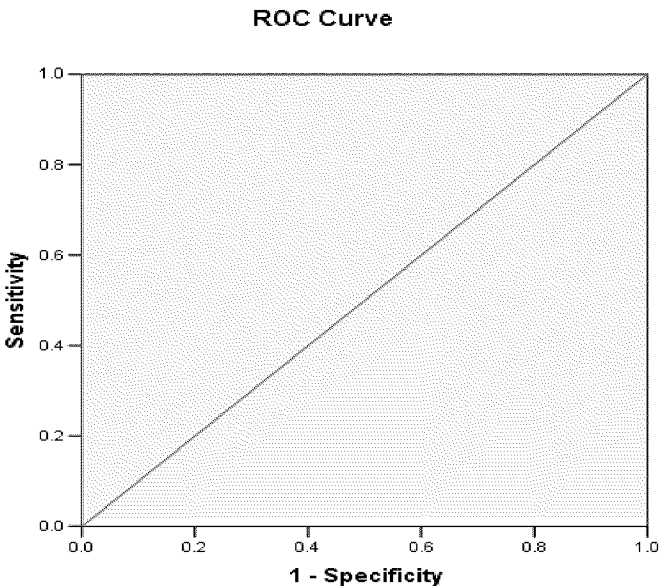


Figure 18b (cont'd)

56/61

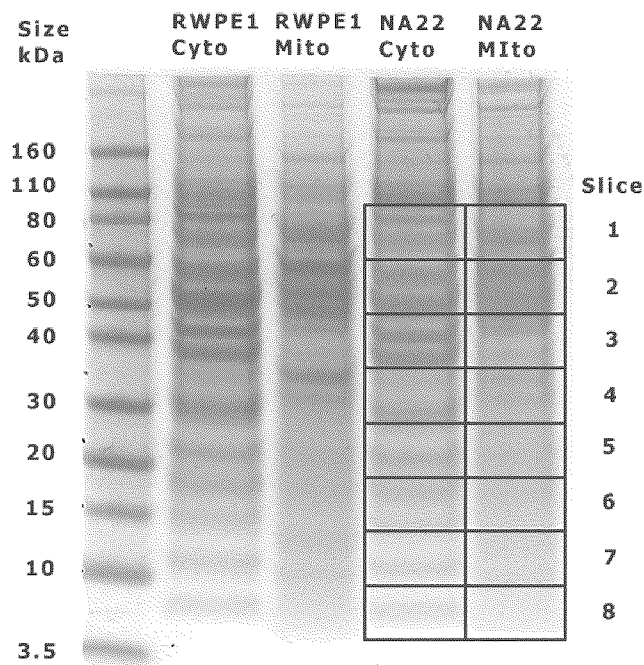


Figure 19

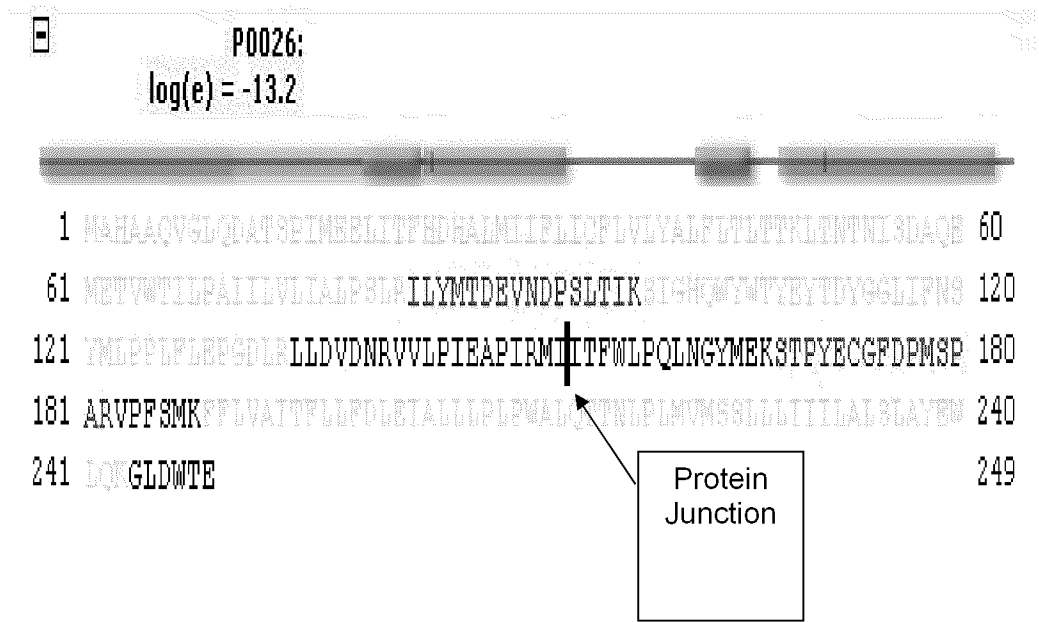


Figure 20a

57/61

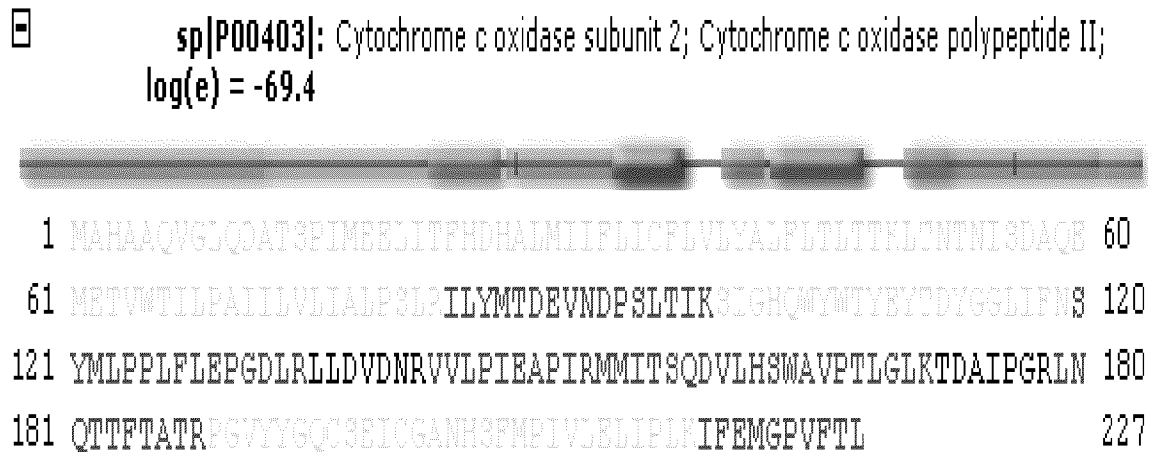


Figure 20b

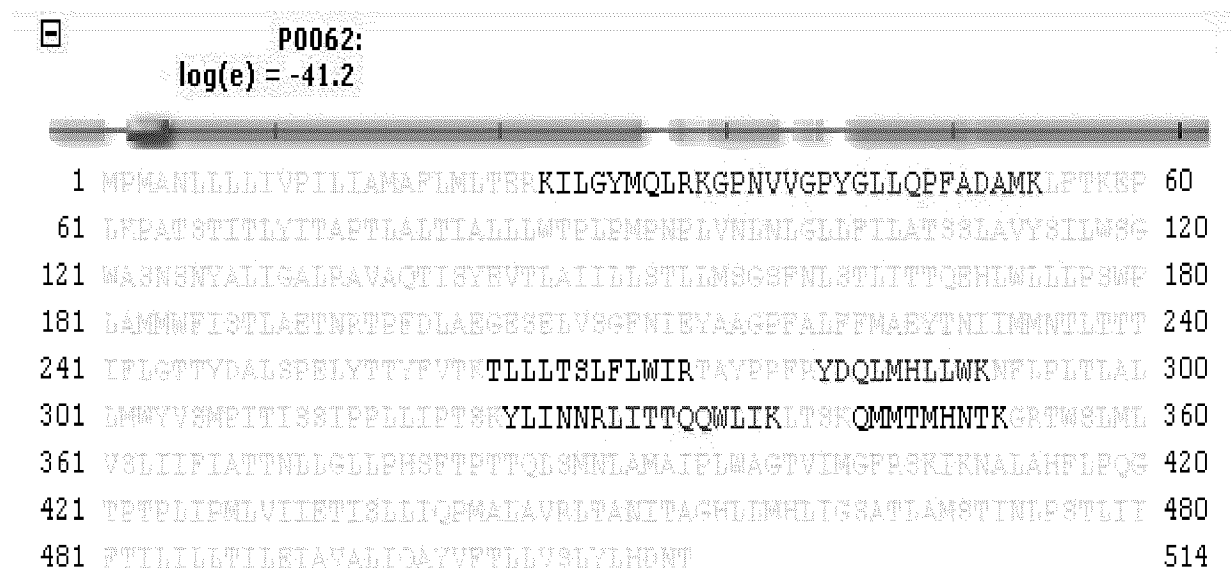


Figure 21a

58/61

sp|P03886|: NADH-ubiquinone oxidoreductase chain 1; EC 1.6.5.3; NADH dehydrogenase
log(e) = -42.9 subunit 1;

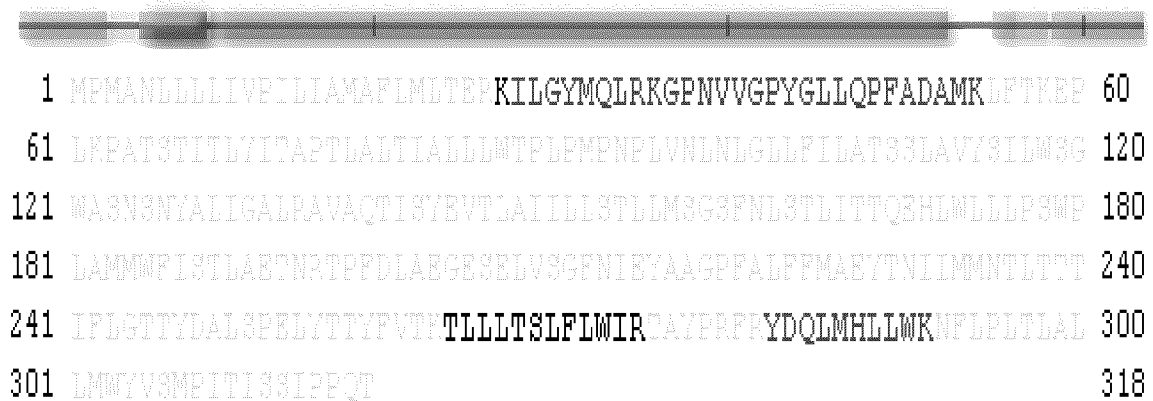


Figure 21b

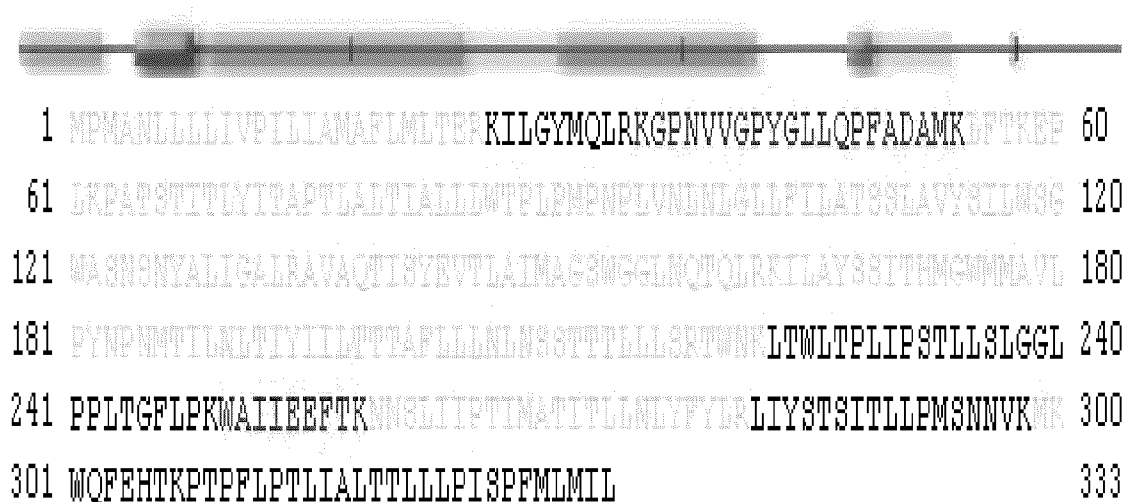


Figure 22

59/61

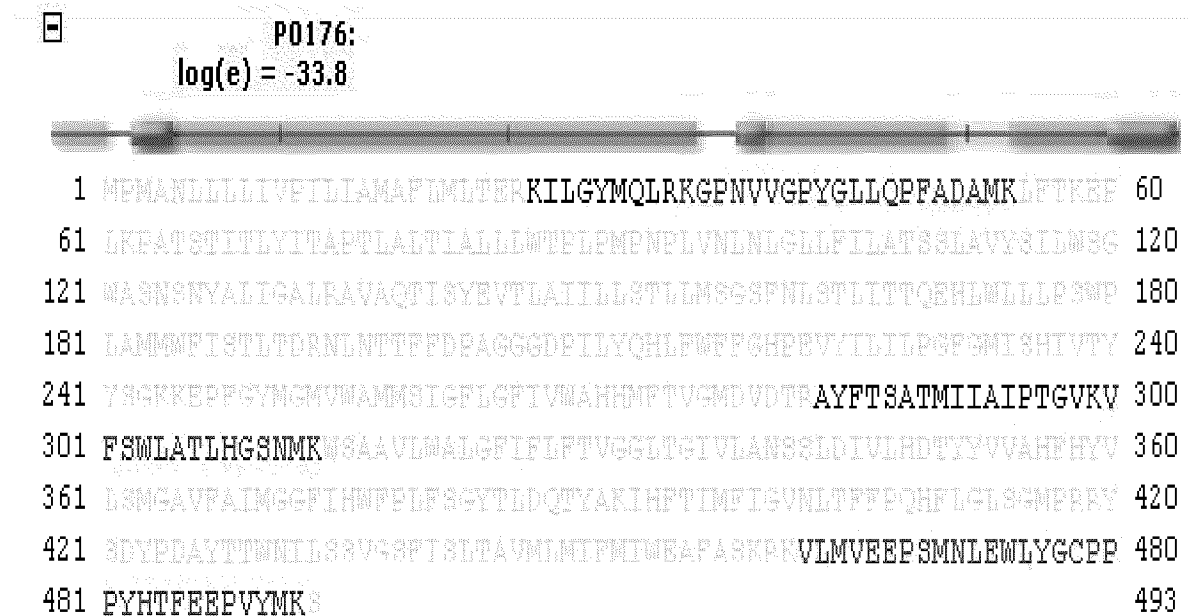


Figure 23a

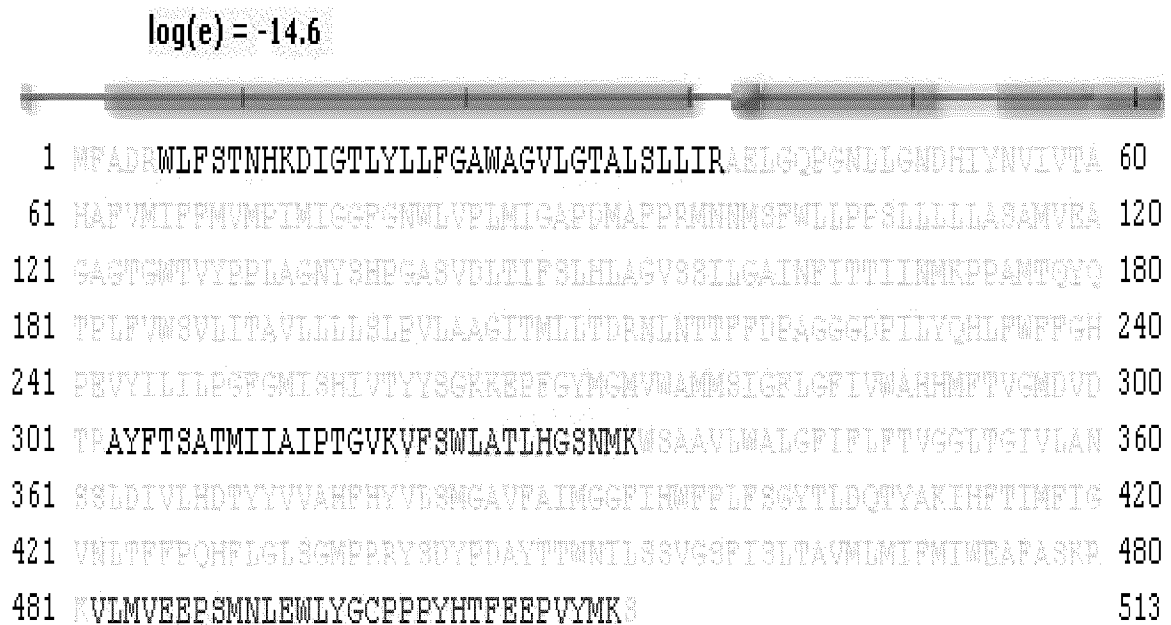


Figure 23b

60/61

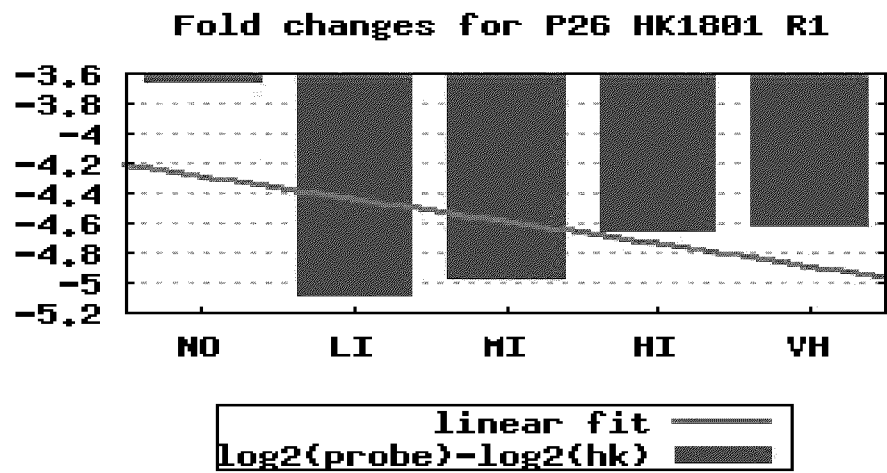


Figure 24a

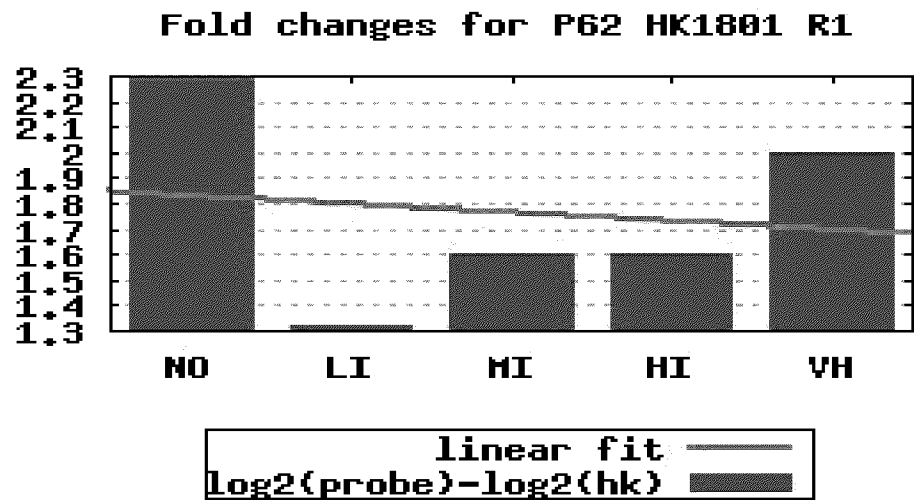


Figure 24b

61/61

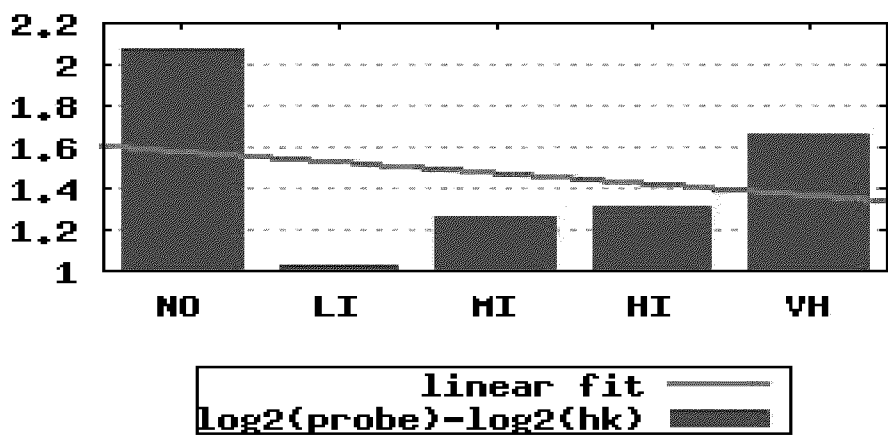


Figure 24c

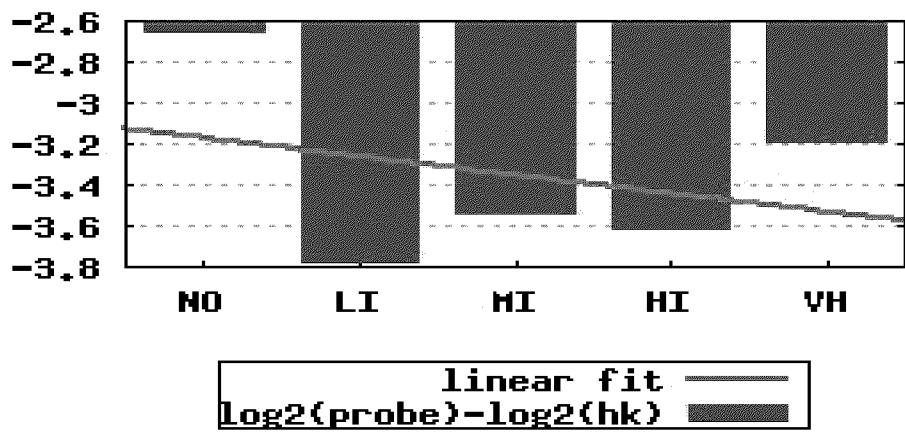


Figure 24d

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2010/000423

| <p>A. CLASSIFICATION OF SUBJECT MATTER IPC: <i>C12N 15/12</i> (2006.01), <i>C07H 21/00</i> (2006.01), <i>C07K 14/47</i> (2006.01), <i>C07K 16/18</i> (2006.01), <i>C12Q 1/68</i> (2006.01), <i>G01N 33/50</i> (2006.01), <i>G01N 33/574</i> (2006.01), <i>G01N 33/68</i> (2006.01), <i>C07K 19/00</i> (2006.01), <i>C12N 15/62</i> (2006.01). According to International Patent Classification (IPC) or to both national classification and IPC</p> | | |
|--|---|---|
| <p>B. FIELDS SEARCHED</p> | | |
| <p>Minimum documentation searched (classification system followed by classification symbols)</p> | | |
| <p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> | | |
| <p>Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)</p> | | |
| <p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p> | | |
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | Mita S. et al. Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. Nucleic Acids Research, 1990, Vol. 18, No. 3, pp. 561-567, ISSN: 0305-1048 | 1, 6 and 7 |
| X | Uchida T. et al. Down-regulation of mitochondrial gene expression by the anti-tumor arotinoid mofarotene (Ro 40-8757). International Journal of Cancer, 1994, Vol. 58, No. 6, pp. 891-897, ISSN: 0020-7136 | 14-17 |
| A | Zhu W. et al. Large-scale mitochondrial DNA deletion mutations and nuclear genome instability in human breast cancer. Cancer Detection and Prevention, 2004, Vol. 28, No. 2, pp. 119-126, ISSN: 0361-090X | |
| A | Dai J.G. et al. Mitochondrial DNA 4977 BP deletion mutations in lung carcinoma. Indian Journal of Cancer. January-March 2006, Vol. 43, No. 1, pp. 20-25, ISSN: 0019-509X | |
| <p><input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.</p> | | |
| * Special categories of cited documents : | <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> | |
| "A" document defining the general state of the art which is not considered to be of particular relevance | <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> | |
| "E" earlier application or patent but published on or after the international filing date | <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> | |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | <p>"&" document member of the same patent family</p> | |
| "O" document referring to an oral disclosure, use, exhibition or other means | | |
| "P" document published prior to the international filing date but later than the priority date claimed | | |
| <p>Date of the actual completion of the international search 10 June 2010 (10-06-2010)</p> | | <p>Date of mailing of the international search report 21 June 2010 (21-06-2010)</p> |
| <p>Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476</p> | | <p>Authorized officer Ali Abdallah (819) 994-4253</p> |

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/CA2010/000423**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:
- a. (means)
- ☒ on paper
- ☒ in electronic form
- b. (time)
- ☐ in the international application as filed
- ☒ together with the international application in electronic form
- ☐ subsequently to this Authority for the purposes of search
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments :

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/CA2010/000423**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of the first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons :

1. ☐ Claim Nos. :
 because they relate to subject matter not required to be searched by this Authority, namely :

2. ☐ Claim Nos. :
 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically :

3. ☐ Claim Nos. :
 because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows :

See supplemental Box No. III

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claim Nos. :
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim Nos. :

- Remark on Protest** ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Supplemental Box No. III

The claims are directed to a plurality of inventive concepts as follows:

- Group 1 - Claims 1-10 (all partially) are directed to an isolated mitochondrial fusion protein with an amino acid sequence of SEQ ID NO: 36, a method, an antibody and a kit all relating to said fusion protein.
- Group 2 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 37;
- Group 3 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 38;
- Group 4 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 39;
- Group 5 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 40;
- Group 6 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 41;
- Group 7 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 42;
- Group 8 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 43;
- Group 9 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 44;
- Group 10 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 4;
- Group 11 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 4;
- Group 12 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 47;
- Group 13 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 48;
- Group 14 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 49;
- Group 15 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 50;
- Group 16 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 51;
- Group 17 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 52;
- Group 18 - Claims 1-10 (all partially) are directed to same subject matter as in group 1, but relating to SEQ ID NO: 55;
- Group 19 - Claims 1-17 (all partially) are directed to an isolated mitochondrial fusion protein defined by SEQ ID NOs: 56-58, a method, an antibody, a probe and a kit all relating to said fusion protein;
- Group 20 - Claims 1-17 (all partially) are directed to same subject matter as in group 19, but relating to SEQ ID NOs: 59-61;
- Group 21 - Claims 1-17 (all partially) are directed to same subject matter as in group 19, but relating to SEQ ID NOs: 62-64; and
- Group 22 - Claims 1-17 (all partially) are directed to same subject matter as in group 19, but relating to SEQ ID NOs: 65-67.

The claims must be limited to one inventive concept as set out in Rule 13 of the PCT. The only linking feature between the different groups is a mitochondrial fusion protein.

Based on *a posteriori* considerations, mitochondrial fusion proteins are well known in the art (see Mita S. et al., Nucleic Acids Research, 1990, Vol. 18, No. 3, pp. 561-567, and Table 1 of the instant application). Therefore, every mitochondrial fusion protein may be considered as a separate invention.