



(11)

EP 2 194 145 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:
05.06.2013 Bulletin 2013/23

(51) Int Cl.:
C12Q 1/68 ^(2006.01) **G01N 33/564** ^(2006.01)

(21) Application number: **08016970.9**

(22) Date of filing: **24.04.2003**

(54) **Diagnosing and monitoring cardiac transplant rejection based on marker expression level**

Diagnose und Überwachung von Herztransplantatabstossung mittels Markerexpressionsniveau

Diagnostic et surveillance du rejet d'une greffe cardiaque basés sur le niveau d'expression d'un marqueur

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IT LI LU MC NL PT RO SE SI SK TR**

(30) Priority: **24.04.2002 US 131831**
20.12.2002 US 325899

(43) Date of publication of application:
09.06.2010 Bulletin 2010/23

(62) Document number(s) of the earlier application(s) in
accordance with Art. 76 EPC:
03799755.8 / 1 585 972

(73) Proprietor: **XDx, Inc.**
San Francisco CA 94080 (US)

(72) Inventors:

- **Wohlgemuth, Jay**
Menlo Park, CA 94303 (US)
- **Fry, Kirk**
Palo Alto, CA 94303 (US)
- **Woodward, Robert**
Pleasanton, CA 94558 (US)
- **Ly, Ngoc**
San Bruno, CA 94066 (US)
- **Prentice, James**
San Francisco, CA 94027 (US)
- **Morris, MacDonald**
Atherton, CA 94027 (US)
- **Rosenberg, Steven**
Oakland, CA 94682 (US)

(74) Representative: **Holmberg, Martin Tor**
Bergenstråhle & Lindvall AB
P.O. Box 17704
118 93 Stockholm (SE)

(56) References cited:
WO-A-01/81916 WO-A1-01/14557

- **GORCZYNSKI R M ET AL: "CORRELATION OF PERIPHERAL BLOOD LYMPHOCYTE AND INTRAGRAFT CYTOKINE MRNA EXPRESSION WITH REJECTION IN ORTHOTOPIC LIVER TRANSPLANTATION" SURGERY, C.V. MOSBY CO., ST. LOUIS,, US, vol. 120, no. 3, September 1996 (1996-09), pages 496-502, XP008005398 ISSN: 0039-6060**
- **DUGRE F J ET AL: "Cytokine and cytotoxic molecule gene expression determined in peripheral blood mononuclear cells in the diagnosis of acute renal rejection" TRANSPLANTATION, WILLIAMS AND WILKINS, BALTIMORE, MD, US, vol. 70, no. 7, 15 October 2000 (2000-10-15), pages 1074-1080, XP002434104 ISSN: 0041-1337**
- **VASCONCELLOS L M ET AL: "CYTOTOXIC LYMPHOCYTE GENE EXPRESSION IN PERIPHERAL BLOOD LEUKOCYTES CORRELATES WITH REJECTING RENAL ALLOGRAFTS" TRANSPLANTATION, WILLIAMS AND WILKINS, BALTIMORE, MD, US, vol. 66, no. 5, 15 September 1998 (1998-09-15), pages 562-566, XP009083920 ISSN: 0041-1337**
- **MORGUN A ET AL: "CYTOKINE AND TLRC7 MRNA EXPRESSION DURING ACUTE REJECTION IN CARDIAC ALLOGRAFT RECIPIENTS" TRANSPLANTATION PROCEEDINGS, ORLANDO, FL, US LNKD- DOI: 10.1016/S0041-1345(00)02613-0, vol. 1/02, no. 33, 1 February 2001 (2001-02-01), XP001073543 ISSN: 0041-1345**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 2 194 145 B1

- NISHIMURA H ET AL: "DEVELOPMENT OF LUPUS-LIKE AUTOIMMUNE DISEASES BY DISRUPTION OF THE" IMMUNITY, CELL PRESS, US LNKD- DOI:10.1016/S1074-7613(00)80089-8, vol. 11, no. 2, August 1999 (1999-08), pages 141-151, XP000971788 ISSN: 1074-7613
- DENG M C ET AL: "Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling." AMERICAN JOURNAL OF TRANSPLANTATION : OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF TRANSPLANTATION AND THE AMERICAN SOCIETY OF TRANSPLANT SURGEONS JAN 2006 LNKD- PUBMED:16433769, vol. 6, no. 1, January 2006 (2006-01), pages 150-160, XP002578025 ISSN: 1600-6135

Description**Related Applications**

- 5 **[0001]** This application claims priority to U.S. Application No. 10/131,831, filed April 24, 2002, and U.S. Application No. 10/325,899, filed December 20, 2002.

Field of the Invention

- 10 **[0002]** This invention is in the field of expression profiling following organ transplantation.

Background of the Invention

- 15 **[0003]** Many of the current shortcomings in diagnosis, prognosis, risk stratification and treatment of disease can be approached through the identification of the molecular mechanisms underlying a disease and through the discovery of nucleotide sequences (or sets of nucleotide sequences) whose expression patterns predict the occurrence or progression of disease states, or predict a patient's response to a particular therapeutic intervention. In particular, identification of nucleotide sequences and sets of nucleotide sequences with such predictive value from cells and tissues that are readily accessible would be extremely valuable. For example, peripheral blood is attainable from all patients and can easily be
20 obtained at multiple time points at low cost. This is a desirable contrast to most other cell and tissue types, which are less readily accessible, or accessible only through invasive and aversive procedures. In addition, the various cell types present in circulating blood are ideal for expression profiling experiments as the many cell types in the blood specimen can be easily separated if desired prior to analysis of gene expression. While blood provides a very attractive substrate for the study of diseases using expression profiling techniques, and for the development of diagnostic technologies and
25 the identification of therapeutic targets, the value of expression profiling in blood samples rests on the degree to which changes in gene expression in these cell types are associated with a predisposition to, and pathogenesis and progression of a disease.

- 30 **[0004]** Hematopoiesis is the development and maturation of all cell types of the blood. These include erythrocytes, platelets and leukocytes. Leukocytes are further subdivided into granulocytes (neutrophils, eosinophils, basophils) and mononuclear cells (monocytes, lymphocytes). These cells develop and mature from precursor cells to replenish the circulating pool and to respond to insults and challenges to the system. This occurs in the bone marrow, spleen, thymus, liver, lymph nodes, mucosal associated lymphoid tissue (MALT) and peripheral blood.

- 35 **[0005]** Precursor cells differentiate into immature forms of each lineage and these immature cells develop further into mature cells. This process occurs under the influence and direction of hematopoietic growth factors. When hematopoiesis is stimulated, there is an increase in the number of immature cells in the peripheral blood and in some cases, precursor cells are found at increased frequency. For example, CD34+ cells (hematopoietic stem cells) may increase in frequency in the peripheral blood with an insult to the immune system. For neutrophils, "band" forms are increased, for erythrocytes, reticulocytes or nucleated red cells are seen. Lymphocytes are preceded by lymphoblasts (immature lymphocytes).

- 40 **[0006]** It may be an important clinical goal to measure the rate of production of blood cells of a variety of lineages. Hematological disorders involving over or under production of various blood cells may be treated pharmacologically. For example, anemia (low red blood cells) may be treated with erythropoietin (a hematopoietic growth factor) and response to this therapy can be assessed by measuring RBC production rates. Low neutrophils counts can be treated by administration of G-CSF and this therapy may be monitored by measuring neutrophil production rates. Alternatively, the diagnosis of blood cell disorders is greatly facilitated by determination of lineage specific production rates. For
45 example, anemia (low RBCs) may be caused by decreased cellular production or increased destruction of cells. In the latter case, the rate of cellular production will be increased rather than decreased and the therapeutic implications are very different. Further discussion of the clinical uses of measures of blood cell production rates is given in below.

- 50 **[0007]** Assessment of blood cell production rates may be useful for diagnosis and management of non-hematological disorders. In particular, acute allograft rejection diagnosis and monitoring may benefit from such an approach. Current diagnosis and monitoring of acute allograft rejection is achieved through invasive allograft biopsy and assessment of the biopsy histology. This approach is sub-optimal because of expense of the procedure, cost, pain and discomfort of the patient, the need for trained physician operators, the risk of complications of the procedure, the lack of insight into the functioning of the immune system and variability of pathological assessment. In addition, biopsy can diagnose acute allograft rejection only after significant cellular infiltration into the allograft has occurred. At this point, the process has
55 already caused damage to the allograft. For all these reasons, a simple blood test that can diagnose and monitor acute rejection at an earlier stage in the process is needed. Allograft rejection depends on the presence of functioning cells of the immune system. In addition, the process of rejection may cause activation of hematopoiesis. Finally, effective immunosuppressive therapy to treat or prevent acute rejection may suppress hematopoiesis. For these reasons, as-

assessment of hematopoietic cellular production rates may be useful in the diagnosis and monitoring of acute rejection.

[0008] Current techniques for measuring cellular development and production rates are inadequate. The most common approach is to measure the number of mature cells of a lineage of interest over time. For example, if a patient is being treated for anemia (low red blood cell counts), then the physician will order a blood cell count to assess the number of red blood cells (RBCs) in circulation. For this to be effective, the physician must measure the cell count over time and may have to wait 2-4 weeks before being able to assess response to therapy. The same limitation is true for assessment of any cell lineage in the blood.

[0009] An alternative approach is to count the number of immature cells in the peripheral blood by counting them under the microscope. This may allow a more rapid assessment of cellular production rates, but is limited by the need for assessment by a skilled hematologist, observer variability and the inability to distinguish all precursor cells on the basis of morphology alone.

[0010] Bone marrow biopsy is the gold standard for assessment of cellular production rates. In addition to the limitations of the need for skilled physicians, reader variability and the lack of sensitivity of morphology alone, the technique is also limited by the expense, discomfort to the patient and need for a prolonged visit to a medical center. Thus there is a need for a reliable, rapid means for measuring the rate of hematopoiesis in a patient.

[0011] In addition to the relationship between hematopoiesis and variety of disease processes, there is an extensive literature supporting the role of leukocytes, e.g., T-and B-lymphocytes, monocytes and granulocytes, including neutrophils, in a wide range of disease processes, including such broad classes as cardiovascular diseases, inflammatory, autoimmune and rheumatic diseases, infectious diseases, transplant rejection, cancer and malignancy, and endocrine diseases. For example, among cardiovascular diseases, such commonly occurring diseases as atherosclerosis, restenosis, transplant vasculopathy and acute coronary syndromes all demonstrate significant T cell involvement (Smith-Norowitz et al. (1999) *Clin Immunol* 93:168-175; Jude et al. (1994) *Circulation* 90:1662-8; Belch et al. (1997) *Circulation* 95:2027-31). These diseases are now recognized as manifestations of chronic inflammatory disorders resulting from an ongoing response to an injury process in the arterial tree (Ross et al. (1999) *Ann Thorac Surg* 67:1428-33). Differential expression of lymphocyte, monocyte and neutrophil genes and their products has been demonstrated clearly in the literature. Particularly interesting are examples of differential expression in circulating cells of the immune system that demonstrate specificity for a particular disease, such as arteriosclerosis, as opposed to a generalized association with other inflammatory diseases, or for example, with unstable angina rather than quiescent coronary disease.

[0012] A number of individual genes, e.g., CD11b/CD18 (Kassirer et al. (1999) *Am Heart J* 138:555-9); leukocyte elastase (Amaro et al. (1995) *Eur Heart J* 16:615-22; and CD40L (Aukrust et al. (1999) *Circulation* 100:614-20) demonstrate some degree of sensitivity and specificity as markers of various vascular diseases. In addition, the identification of differentially expressed target and fingerprint genes isolated from purified populations of monocytes manipulated in various in vitro paradigms has been proposed for the diagnosis and monitoring of a range of cardiovascular diseases, see, e.g., US Patents Numbers 6,048,709; 6,087,477; 6,099,823; and 6,124,433 "COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE" to Falb (see also, WO 97/30065). Lockhart, in US Patent Number 6,033,860 "EXPRESSION PROFILES IN ADULT AND FETAL ORGANS" proposes the use of expression profiles for a subset of identified genes in the identification of tissue samples, and the monitoring of drug effects.

[0013] Gorczynski et al. (1996) *Surgery* 120(3):496-502 identified IL2, IL6 and interferon gamma as gene expression markers in blood correlating with rejection of liver allografts. Dugre et al. (2000) *Transplantation* 70(7): 1074-1080 identified IL4, IL5, IL6, interferon gamma, perforin and Granzyme B as gene expression markers in blood correlating with rejection of renal allografts. Vasconcellos et al. (1998) *Transplantation* 66(5):562-566 identified FasL, perforin and Granzyme B as gene expression markers in blood correlating with rejection of renal allografts. Morgun et al. (2001) *Transplantation Proceedings* 1/02(33) identified interferon gamma, TIRC7, perforin and Granzyme B as gene expression markers in blood correlating with rejection of cardiac allografts.

[0014] The accuracy of technologies based on expression profiling for the diagnosis, prognosis, and monitoring of disease would be dramatically increased if numerous differentially expressed nucleotide sequences, each with a measure of specificity for a disease in question, could be identified and assayed in a concerted manner. PCT application WO 02/057414 "LEUKOCYTE EXPRESSION PROFILING" to Wohlgemuth identifies one such set of differentially expressed nucleotides.

[0015] In order to achieve this improved accuracy, the sets of nucleotide sequences once identified need to be validated to identify those differentially expressed nucleotides within a given set that are most useful for diagnosis, prognosis, and monitoring of disease. The present invention addresses these and other needs, and applies to transplant rejection and detection of the rate of hematopoiesis for which differential regulation of genes, or other nucleotide sequences, of peripheral blood can be demonstrated.

Summary of the Invention

[0016] In order to meet these needs, the present invention is thus directed to a system for detecting differential gene expression.

[0017] In a further variation, the invention is directed to a method of diagnosing or monitoring cardiac transplant rejection in a patient by detecting the expression level of one or more genes in the patient to diagnose or monitor cardiac transplant rejection in the patient wherein the one or more genes include a nucleotide sequence according to SEQ ID NO:52

[0018] In another aspect, the methods of diagnosing or monitoring transplant rejection include detecting the expression level of at least two of the genes. In another variation, methods of diagnosing or monitoring transplant rejection include detecting the expression level of at least ten of the genes. In a further variation, the methods of diagnosing or monitoring transplant rejection include detecting the expression level of at least one hundred of the genes. In still a further variation, the methods of diagnosing or monitoring transplant rejection include detecting the expression level of all the listed genes.

[0019] In another variation, transplant rejection may be heart transplant rejection.

[0020] In another aspect, the methods of detecting transplant rejection include detecting the expression level by measuring the RNA level expressed by one or more genes. The method may further include isolating RNA from the patient prior to detecting the RNA level expressed by the one or more genes.

[0021] In one variation, the RNA level is detected by PCR. In a still further variation, the PCR uses primers consisting of nucleotide sequences selected from the group consisting of SEQ ID NO:716, SEQ ID NO:1047, SEQ ID NO:1690, SEQ ID NO:1938, SEQ ID NO:1378, SEQ ID NO:2186. The RNA level may be detected by hybridization to the probes.

In a further variation, the RNA level is detected by hybridization to an oligonucleotide. Examples of oligonucleotide include oligonucleotides having a nucleotide sequence according to SEQ ID NO:52. The oligonucleotide may be DNA, RNA, cDNA, PNA, genomic DNA, or synthetic oligonucleotides.

[0022] In another aspect, the methods of detecting transplant rejection disclose detecting the expression level by measuring one or more proteins expressed by the one or more genes. In one variation, the one or more proteins include an amino acid sequence selected from, SEQ ID NO:2448

[0023] The method of diagnosing or monitoring cardiac transplant rejection in a patient further provides for detecting the expression level of one or more genes in the patient to diagnose or monitor cardiac transplant rejection in the patient by measuring one or more proteins expressed by the one or more genes. The one or more proteins may include an amino acid sequence selected from SEQ ID NO:2448. Alternatively, the expression level of the one or more genes may be detected by measuring one or more proteins expressed by one or more genes, and one or more proteins expressed by one or more additional genes. In one variation, the one or more proteins expressed by the one or more genes include an amino acid sequence selected from SEQ ID NO:2448.

[0024] Protein detection may be accomplished by measuring serum. In another variation, the protein is a cell surface protein. In a further variation, the measuring includes using a fluorescent activated cell sorter.

[0025] In another aspect, the invention discloses a substantially purified oligonucleotide having the nucleotide sequence according to SEQ ID NO:52, and substantially purified oligonucleotides having at least 90% sequence identity to an oligonucleotide having the nucleotide sequence according to SEQ ID NO:52,

[0026] In a further aspect, the invention discloses a substantially purified oligonucleotide that hybridizes at high stringency to an oligonucleotide having the nucleotide sequence according to SEQ ID NO:52. The sequences may be used as diagnostic oligonucleotides for transplant rejection and/or cardiac transplant rejection. The oligonucleotide may have nucleotide sequence including DNA, cDNA, PNA, genomic DNA, or synthetic oligonucleotides.

[0027] In another aspect, the invention discloses a method of diagnosing or monitoring transplant rejection in a patient wherein the expression level of one or more genes in a patient's bodily fluid is detected. In a further variation, the bodily fluid is peripheral blood.

[0028] In another aspect, the invention discloses a method of diagnosing or monitoring transplant rejection in a patient, comprising detecting the expression level of four or more genes in the patient to diagnose or monitor transplant rejection in the patient wherein the four or more genes include a nucleotide sequence according to SEQ ID NO:52,

[0029] In a further aspect, the invention also discloses a system for detecting gene expression in body fluid including at least two isolated polynucleotides wherein the isolated polynucleotides detect expression of a gene wherein the gene includes a nucleotide sequence according to and the gene is differentially expressed in body fluid in an individual rejecting a transplanted organ compared to the expression of the gene in leukocytes in an individual not rejecting a transplanted organ.

[0030] In another aspect, the invention discloses a system for detecting gene expression in body fluid including at least two isolated polynucleotides wherein the isolated polynucleotides detect expression of a gene wherein the gene includes a nucleotide sequence according to SEQ ID NO:52, and the gene expression is related to the rate of hematopoiesis or the distribution of hematopoietic cells along their maturation pathway.

Brief Description of the Figure**[0031]**

Figure 1: Figure 1 is a schematic flow chart illustrating a schematic instruction set for characterization of the nucleotide sequence and/or the predicted protein sequence of novel nucleotide sequences.

Figure 2: Figure 2 depicts the components of an automated RNA preparation machine.

Figure 3 shows the results of six hybridizations on a mini array graphed (n=6 for each column). The error bars are the SEM. This experiment shows that the average signal from AP prepared RNA is 47% of the average signal from GS prepared RNA for both Cy3 and Cy5.

Figure 4 shows the average background subtracted signal for each of nine leukocyte-specific genes on a mini array. This average is for 3-6 of the above-described hybridizations for each gene. The error bars are the SEM.

Figure 5 shows the ratio of Cy3 to Cy5 signal for a number of genes. After normalization, this ratio corrects for variability among hybridizations and allows comparison between experiments done at different times. The ratio is calculated as the Cy3 background subtracted signal divided by the Cy5 background subtracted signal. Each bar is the average for 3-6 hybridizations. The error bars are SEM.

Figure 6 shows data median Cy3 background subtracted signals for control RNAs using mini arrays.

Figure 7: Cardiac Allograft rejection diagnostic genes.

A. Example of rejection and no-rejection samples expression data for 5 marker genes. For each sample, the associated rejection grades are shown as are the expression ratios for 5 differentially expressed genes. The genes are identified by the SEQ ID number for the oligonucleotide. The average fold difference between grade 0 and grade 3A samples is calculated at the bottom.

B. CART classification model. Decision tree for a 3 gene classification model for diagnosis of cardiac rejection. In the first step, expression of gene 223 is used to divide the patients to 2 branches. The remaining samples in each branch are then further divided by one remaining gene. The samples are classified as either rejection or no rejection. 1 no rejection sample is misclassified as a rejection sample.

C. Surrogates for the CART classification model. For each of the 3 splitter genes in the CART rejection model described in the example, 5 top surrogate genes are listed that were identified by the CART algorithm.

Figure 8: Validation of differential expression of a gene discovered using microarrays using real-time PCR

Figure 8A. The Ct for each patient sample on multiple assays is shown along with the Ct in the R50 control RNA. Triangles represent -RT (reverse transcriptase) controls.

Figure 8B. The fold difference between the expression of Granzyme B and an Actin reference is shown for 3 samples from patients with and without CMV disease.

Figure 9: Endpoint testing of PCR primers

Electrophoresis and microfluidics are used to assess the product of gene specific PCR primers. β -GUS gel image. Lane 3 is the image for primers F178 and R242. Lanes 2 and 1 correspond to the no-template control and -RT control, respectively.

The electropherogram of β -GUS primers F178 and R242, a graphical representation of Lane 3 from the gel image. β -Actin gel image. Lane 3 is the image for primers F75 and R178. Lanes 2 and 1 correspond to the no-template control and -RT control, respectively.

The electropherogram of β -Actin primers F75 and R178, a graphical representation of Lane 3 from the gel image.

Figure 10: PCR Primer efficiency testing. A standard curve of Ct versus log of the starting RNA amount is shown for 2 genes.

Figure 11: Real-time PCR control gene analysis

11 candidate control genes were tested using real-time PCR on 6 whole blood samples (PAX) paired with 6 mononuclear samples (CPT) from the same patient. Each sample was tested twice. For each gene, the variability of the gene across the samples is shown on the vertical axis (top graph). The average Ct value for each gene is also shown (bottom graph). 2ug RNA was used for PAX samples and 0.5 ug total RNA was used for the mononuclear samples (CPT).

Figure 12: Rejection marker discovery by co-expression with established marker Microarrays were used to measure expression of genes SEQ ID 85 and 302 in samples derived from 240 transplant recipients. For each sample, the expression measurement for 85 is plotted against 302.

Figure 13: ROC (receiver operator characteristics) curve for a 3-gene PCR assay for diagnosis of rejection (see example 17). The Sensitivity and False Positive Rate for each test cutoff is shown.

Brief Description of the Tables

[0032]

Table 1: Table 1 lists diseases or conditions amenable to study by leukocyte profiling.

Table 2: Transplant Markers

A. Transplant Genes: Genes useful for monitoring of allograft rejection are listed in this here. The gene symbol and name are given. SEQ ID 50mer is the sequence ID of a 50mer oligonucleotide that is specific for the gene and is listed for the SEQ ID 50mer oligonucleotides identified in the sequence listings. The NCBI Unigene number (HS) from (Build 160, 16 Feb 2003) is given as is an accession number (ACC) from (Genbank Release 135, 15 April 2003) for an RNA or cDNA is Genbank that corresponds to the gene.

B. Microarray Data: Gene, Gene Name and ACC are given for each gene as in A (above). The SEQ ID 50mer and SEQ ID RNA/cDNA sequences included in the sequence listing are also identified in the table. Each identified gene has a Non-Parametric Score and Median Rank in NR given from the non-parametric analysis of the data. The genes are ranked from highest to lowest scoring. Down Regulated genes are noted with a 1 in this column.

C. PCR Primers: Primers and probes for real-time PCR assays are given along with their SEQ ID #s. Each gene has 1 or 2 sets of a forward and reverse PCR primer and a hybridization probe for detection in TaqMan or similar assays.

D. PCR Data: Real-time PCR data was generated on a set of transplant samples using sybr green technology as described in the text. For each gene the number of samples (n) used in the analysis is given. An odds ratio and the p-values for a Fisher test and t-test are given for the comparison of acute rejection samples is given (see text).

E. Transplant proteins: For each gene, the corresponding protein in the RefSeq data base (Genbank Release 135, 18 April 2003) is given (RefSeq Peptide Accession #) along the the SEQ ID for certain proteins for the sequence listing.

Table 3: Viral gene for arrays. Viral genomes were used to design oligonucleotides for the microarrays. The accession numbers for the viral genomes used are given, along with the gene name and location of the region used for oligonucleotide design.

Table 4. Dependent variables for discovery of gene expression markers of cardiac allograft rejection. A stable Grade 0 is a Grade 0 biopsy in a patient who does not experience rejection with the subsequent biopsy. HG or highest grade means that the higher of the biopsy grades from the centralized and local pathologists was used for a definition of the dependent variable.

Table 5: Real-time PCR assay reporter and quencher dyes. Various combinations of reporter and quencher dyes are useful for real-time PCR assays. Reporter and quencher dyes work optimally in specific combinations defined by their spectra. For each reporter, appropriate choices for quencher dyes are given.

Table 6: Rejection marker PCR assay results

Results of real-time PCR assays are listed for the comparison of rejection samples to no rejection samples. The fold change is given for expression of each gene in rejection/no rejection samples. The p-value for the t-test comparing the rejection and no rejection classes is given.

Table 7: Summary results of array rejection significance analysis. Summary results are given for correlation analysis of leukocyte gene expression to acute rejection using significance analysis for microarrays (SAM). Five analyses are described. The ISHLT grades used to define the rejection and no rejection classes are given. In each case the highest grade from three pathology reading was taken for analysis. All samples are used for two analyses. The other analyses reduce redundancy of patients used in the analysis by using only one sample per patient ("Non-redundant") or using only one sample per patient within a given class ("Non-redundant within class"). The number of samples used in the analysis is given and the lowest false detection rate (FDR) achieved is noted.

Table 9: Rejection marker sequence analysis. For 63 of the allograft rejection markers listed in Table 2, an analysis of the gene sequence was done. The genes and proteins are identified by accession numbers. The cellular localization of each gene is described as either secreted, nuclear, mitochondrial, cytoplasmic or cellular membrane. The function of the gene is also described.

Table 10: Gene expression markers for immature cells of a variety of lineages are given in Table 10 by way of example

Table 11: Changes in the rate of hematopoiesis have been correlated to a number of disease states and other pathologies. Examples of such conditions are listed in Table 11.

Detailed Description**Definitions**

[0033] Unless defined otherwise, all scientific and technical terms are understood to have the same meaning as commonly used in the art to which they pertain. For the purpose of the present disclosure, the following terms are defined below.

[0034] In the context of the disclosure, the term "gene expression system" refers to any system, device or means to detect gene expression and includes diagnostic agents, candidate libraries, oligonucleotide sets or probe sets.

[0035] The term "monitoring" is used herein to describe the use of gene sets to provide useful information about an individual or an individual's health or disease status. "Monitoring" can include, determination of prognosis, risk-stratification, selection of drug therapy, assessment of ongoing drug therapy, prediction of outcomes, determining response to therapy, diagnosis of a disease or disease complication, following progression of a disease or providing any information relating to a patient's health status over time, selecting patients most likely to benefit from experimental therapies with known molecular mechanisms of action, selecting patients most likely to benefit from approved drugs with known molecular mechanisms where that mechanism may be important in a small subset of a disease for which the medication may not have a label, screening a patient population to help decide on a more invasive/expensive test, for example a cascade of tests from a non-invasive blood test to a more invasive option such as biopsy, or testing to assess side effects of drugs used to treat another indication..

[0036] The term "diagnostic oligonucleotide set" generally refers to a set of two or more oligonucleotides that, when evaluated for differential expression of their products, collectively yields predictive data. Such predictive data typically relates to diagnosis, prognosis, monitoring of therapeutic outcomes, and the like. In general, the components of a diagnostic oligonucleotide set are distinguished from nucleotide sequences that are evaluated by analysis of the DNA to directly determine the genotype of an individual as it correlates with a specified trait or phenotype, such as a disease, in that it is the pattern of expression of the components of the diagnostic nucleotide set, rather than mutation or polymorphism of the DNA sequence that provides predictive value. It will be understood that a particular component (or member) of a diagnostic nucleotide set can, in some cases, also present one or more mutations, or polymorphisms that are amenable to direct genotyping by any of a variety of well known analysis methods, e.g., Southern blotting, RFLP, AFLP, SSCP, SNP, and the like.

[0037] A "disease specific target oligonucleotide sequence" is a gene or other oligonucleotide that encodes a polypeptide, most typically a protein, or a subunit of a multi-subunit protein, that is a therapeutic target for a disease, or group of diseases.

[0038] A "candidate library" or a "candidate oligonucleotide library" refers to a collection of oligonucleotide sequences (or gene sequences) that by one or more criteria have an increased probability of being associated with a particular disease or group of diseases. The criteria can be, for example, a differential expression pattern in a disease state or in activated or resting leukocytes in vitro as reported in the scientific or technical literature, tissue specific expression as reported in a sequence database, differential expression in a tissue or cell type of interest, or the like. Typically, a candidate library has at least 2 members or components; more typically, the library has in excess of about 10, or about 100, or about 1000, or even more, members or components.

[0039] The term "disease criterion" is used herein to designate an indicator of a disease, such as a diagnostic factor, a prognostic factor, factor indicated by a medical or family history, a genetic factor, or a symptom, as well as an overt or confirmed diagnosis of a disease associated with several indicators such as those selected from the above list. A disease criterion includes data describing a patient's health status, including retrospective or prospective health data, e.g. in the form of the patient's medical history, laboratory test results, diagnostic test result, clinical events, medications, lists, response(s) to treatment and risk factors, etc.

[0040] The terms "molecular signature" or "expression profile" refers to the collection of expression values for a plurality (e.g., at least 2, but frequently about 10, about 100, about 1000, or more) of members of a candidate library. In many cases, the molecular signature represents the expression pattern for all of the nucleotide sequences in a library or array of candidate or diagnostic nucleotide sequences or genes. Alternatively, the molecular signature represents the expression pattern for one or more subsets of the candidate library. The term "oligonucleotide" refers to two or more nucleotides. Nucleotides may be DNA or RNA, naturally occurring or synthetic.

[0041] The term "healthy individual," as used herein, is relative to a specified disease or disease criterion. That is, the individual does not exhibit the specified disease criterion or is not diagnosed with the specified disease. It will be understood, that the individual in question, can, of course, exhibit symptoms, or possess various indicator factors for another disease.

[0042] Similarly, an "individual diagnosed with a disease" refers to an individual diagnosed with a specified disease (or disease criterion). Such an individual may, or may not, also exhibit a disease criterion associated with, or be diagnosed with another (related or unrelated) disease.

[0043] An "array" is a spatially or logically organized collection, e.g., of oligonucleotide sequences or nucleotide sequence products such as RNA or proteins encoded by an oligonucleotide sequence. In some embodiments, an array includes antibodies or other binding reagents specific for products of a candidate library.

[0044] When referring to a pattern of expression, a "qualitative" difference in gene expression refers to a difference that is not assigned a relative value. That is, such a difference is designated by an "all or nothing" valuation. Such an all or nothing variation can be, for example, expression above or below a threshold of detection (an on/off pattern of expression). Alternatively, a qualitative difference can refer to expression of different types of expression products, e.g., different alleles (e.g., a mutant or polymorphic allele), variants (including sequence variants as well as post-translationally modified variants), etc.

[0045] In contrast, a "quantitative" difference, when referring to a pattern of gene expression, refers to a difference in expression that can be assigned a value on a graduated scale, (e.g., a 0-5 or 1-10 scale, a + - +++ scale, a grade 1-grade 5 scale, or the like; it will be understood that the numbers selected for illustration are entirely arbitrary and in no way are meant to be interpreted to limit the invention).

Genes Expression Systems

[0046] A gene expression system having one or more DNA molecules wherein the one or more DNA molecules has a nucleotide sequence which detects expression of a gene corresponding to the oligonucleotides depicted in the Sequence Listing is disclosed. In one format, the oligonucleotide detects expression of a gene that is differentially expressed in leukocytes. The gene expression system may be a candidate library, a diagnostic agent, a diagnostic oligonucleotide set or a diagnostic probe set. The DNA molecules may be genomic DNA, protein nucleic acid (PNA), cDNA or synthetic oligonucleotides. Following the procedures taught herein, one can identify sequences of interest for analyzing gene expression in leukocytes. Such sequences may be predictive of a disease state.

Diagnostic oligonucleotides

[0047] A diagnostic nucleotide set(s) comprising members of the leukocyte candidate library listed in Table 2, and in the Sequence Listing, for which a correlation exists between the health status of an individual, the individual's expression of RNA or protein products corresponding to the nucleotide sequence, and the diagnosis and prognosis of transplant rejection is disclosed. In some instances, only one oligonucleotide is necessary for such detection. Members of a diagnostic oligonucleotide set may be identified by any means capable of detecting expression of RNA or protein products, including but not limited to differential expression screening, PCR, RT-PCR, SAGE analysis, high-throughput sequencing, microarrays, liquid or other arrays, protein-based methods (e.g., western blotting, proteomics, and other methods described herein), and data mining methods, as further described herein.

[0048] Also disclosed is a diagnostic oligonucleotide set that comprises at least two oligonucleotide sequences listed in Table 2, or the Sequence Listing which are differentially expressed in leukocytes in an individual with at least one disease criterion for at least one leukocyte-implicated disease relative to the expression in individual without the at least one disease criterion, wherein expression of the two or more nucleotide sequences is correlated with at least one disease criterion, as described below.

[0049] Further disclosed is a diagnostic nucleotide set that comprises at least one oligonucleotide having an oligonucleotide sequence listed in Table 2, or the Sequence Listing which is differentially expressed, and further wherein the differential expression/correlation has not previously been described. In some embodiments, the diagnostic nucleotide set is immobilized on an array.

[0050] In one embodiment, diagnostic nucleotides (or nucleotide sets) are related to the members of the leukocyte candidate library listed in Table 2, or in the Sequence Listing, for which a correlation exists between the health status, diagnosis and prognosis of transplant rejection (or disease criterion) of an individual. The diagnostic nucleotides are partially or totally contained in (or derived from) full-length gene sequences (or predicted full-length gene sequences) for the members of the candidate library listed in Table 2, and the sequence listing. In some cases, oligonucleotide sequences are designed from EST or Chromosomal sequences from a public database. In these cases the full-length gene sequences may not be known. Full-length sequences in these cases can be predicted using gene prediction algorithms. Alternatively the full-length can be determined by cloning and sequencing the full-length gene or genes that contain the sequence of interest using standard molecular biology approaches described here. The same is true for oligonucleotides designed from our sequencing of cDNA libraries where the cDNA does not match any sequence in the public databases.

[0051] The diagnostic nucleotides may also be derived from other genes that are coexpressed with the correlated sequence or full-length gene. Genes may share expression patterns because they are regulated in the same molecular pathway. Because of the similarity of expression behavior genes are identified as surrogates in that they can substitute for a diagnostic gene in a diagnostic gene set. Example 4 demonstrates the discovery of surrogates from the data and

the sequence listing identifies and gives the sequence for surrogates for cardiac diagnostic genes.

[0052] As used herein the term "gene cluster" or "cluster" refers to a group of genes related by expression pattern. In other words, a cluster of genes is a group of genes with similar regulation across different conditions, such as graft non-rejection versus graft rejection. The expression profile for each gene in a cluster should be correlated with the expression profile of at least one other gene in that cluster. Correlation may be evaluated using a variety of statistical methods. As used herein the term "surrogate" refers to a gene with an expression profile such that it can substitute for a diagnostic gene in a diagnostic assay. Such genes are often members of the same gene cluster as the diagnostic gene. For each member of a diagnostic gene set, a set of potential surrogates can be identified through identification of genes with similar expression patterns as described below.

[0053] Many statistical analyses produce a correlation coefficient to describe the relatedness between two gene expression patterns. Patterns may be considered correlated if the correlation coefficient is greater than or equal to 0.8. In preferred embodiments, the correlation coefficient should be greater than 0.85, 0.9 or 0.95. Other statistical methods produce a measure of mutual information to describe the relatedness between two gene expression patterns. Patterns may be considered correlated if the normalized mutual information value is greater than or equal to 0.7. In preferred embodiments, the normalized mutual information value should be greater than 0.8, 0.9 or 0.95. Patterns may also be considered similar if they cluster closely upon hierarchical clustering of gene expression data (Eisen et al. 1998). Similar patterns may be those genes that are among the 1, 2, 5, 10, 20, 50 or 100 nearest neighbors in a hierarchical clustering or have a similarity score (Eisen et al. 1998) of > 0.5, 0.7, 0.8, 0.9, 0.95 or 0.99. Similar patterns may also be identified as those genes found to be surrogates in a classification tree by CART (Breiman et al. 1994). Often, but not always, members of a gene cluster have similar biological functions in addition to similar gene expression patterns.

[0054] Correlated genes, clusters and surrogates are identified for the diagnostic genes of the invention. These surrogates may be used as diagnostic genes in an assay instead of, or in addition to, the diagnostic genes for which they are surrogates.

[0055] Diagnostic probe sets are also disclosed. It is understood that a probe includes any reagent capable of specifically identifying a nucleotide sequence of the diagnostic nucleotide set, including but not limited to amplified DNA, amplified RNA, cDNA, synthetic oligonucleotide, partial or full-length nucleic acid sequences. In addition, the probe may identify the protein product of a diagnostic nucleotide sequence, including, for example, antibodies and other affinity reagents.

[0056] It is also understood that each probe can correspond to one gene, or multiple probes can correspond to one gene, or both, or one probe can correspond to more than one gene.

[0057] Homologs and variants of the disclosed nucleic acid molecules may be used. Homologs and variants of these nucleic acid molecules will possess a relatively high degree of sequence identity when aligned using standard methods. The sequences encompassed by the disclosure have at least 40-50, 50-60, 70-80, 80-85, 85-90, 90-95 or 95-100% sequence identity to the sequences disclosed herein.

[0058] It is understood that for expression profiling, variations in the disclosed sequences will still permit detection of gene expression. The degree of sequence identity required to detect gene expression varies depending on the length of the oligomer. For a 60 mer, 6-8 random mutations or 6-8 random deletions in a 60 mer do not affect gene expression detection. Hughes, TR, et al. "Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nature Biotechnology, 19:343-347(2001). As the length of the DNA sequence is increased, the number of mutations or deletions permitted while still allowing gene expression detection is increased.

[0059] As will be appreciated by those skilled in the art, the sequences of the present disclosure may contain sequencing errors. That is, there may be incorrect nucleotides, frameshifts, unknown nucleotides, or other types of sequencing errors in any of the sequences; however, the correct sequences will fall within the homology and stringency definitions herein.

[0060] The minimum length of an oligonucleotide probe necessary for specific hybridization in the human genome can be estimated using two approaches. The first method uses a statistical argument that the probe will be unique in the human genome by chance. Briefly, the number of independent perfect matches (Po) expected for an oligonucleotide of length L in a genome of complexity C can be calculated from the equation (Laird CD, Chromosoma 32:378 (1971):

$$Po = (1/4)^L * 2C$$

[0061] In the case of mammalian genomes, $2C = \sim 3.6 \times 10^9$, and an oligonucleotide of 14-15 nucleotides is expected to be represented only once in the genome. However, the distribution of nucleotides in the coding sequence of mammalian genomes is nonrandom (Lathe, R. J. Mol. Biol. 183:1 (1985) and longer oligonucleotides may be preferred in order to increase the specificity of hybridization. In practical terms, this works out to probes that are 19-40 nucleotides long (Sambrook J et al., *infra*). The second method for estimating the length of a specific probe is to use a probe long enough to hybridize under the chosen conditions and use a computer to search for that sequence or close matches to the

sequence in the human genome and choose a unique match. Probe sequences are chosen based on the desired hybridization properties as described in Chapter 11 of Sambrook et al, *infra*. The PRIMER3 program is useful for designing these probes (S. Rozen and H. Skaletsky 1996, 1997; Primer3 code available at the web site located at genome.wi.mit.edu/genome_software/other/primer3.html). The sequences of these probes are then compared pair wise against a database of the human genome sequences using a program such as BLAST or MEGABLAST (Madden, T.L et al.(1996) *Meth. Enzymol.* 266:131-141). Since most of the human genome is now contained in the database, the number of matches will be determined. Probe sequences are chosen that are unique to the desired target sequence.

[0062] In some embodiments, a diagnostic probe set is immobilized on an array. The array is optionally comprises one or more of: a chip array, a plate array, a bead array, a pin array, a membrane array, a solid surface array, a liquid array, an oligonucleotide array, a polynucleotide array or a cDNA array, a microtiter plate, a pin array, a bead array, a membrane or a chip.

[0063] In some embodiments, the leukocyte-implicated disease is selected from the diseases listed in Table 1. In other embodiments, In some embodiments, the disease is atherosclerosis or cardiac allograft rejection. In other embodiments, the disease is congestive heart failure, angina, and myocardial infarction.

[0064] In some embodiments, diagnostic nucleotides of the invention are used as a diagnostic gene set in combination with genes that are know to be associated with a disease state ("known markers"). The use of the diagnostic nucleotides in combination with the known markers can provide information that is not obtainable through the known markers alone. The known markers include those identified by the prior art listing provided.

Hematopoiesis

[0065] Methods of measurement of the rate of hematopoiesis using the diagnostic oligonucleotides of the invention and measurement of the rates of hematopoiesis by any technique as a method for the monitoring and diagnosis of transplant rejection are also disclosed. Precursor and immature cells often have cell specific phenotypic markers. These are genes and/or proteins that expressed in a restricted manner in immature or precursor cells. This expression decreases with maturation. Gene expression markers for immature cells of a variety of lineages are given in Table 10 below by way of example.

Table 10:

Gene	Cell type
CD10	B-lymphoblasts
RAG1	B-lymphoblasts
RAG2	B-lymphoblasts
NF-E2	Platelets/Megakaryocyte/Erythroid
GATA-1	Platelets/Megakaryocyte
GP IIb	Platelets
pf4	Platelets
EPO-R	Erythroblast
Band 4.1	Erythrocyte
ALAS2	Erythroid specific heme biosynthesis
hemoglobin chains	Erythrocyte
2,3-BPG mutase	Erythrocyte
CD16b	Neutrophil
LAP	Neutrophil
CD16	NK cells
CD159a	NK cells

[0066] By measuring the levels of these and other genes in peripheral blood samples, an assessment of the number and proportion of immature or precursor cells can be made. Of particular use is RNA quantification in erythrocytes and platelets. These cells are anucleated in their mature forms. During development, platelets pinch off of a megakaryocyte

and take a compliment of RNA without a nucleus. This RNA is quickly consumed by the platelet. Erythrocytes start as nucleated cells, but the nucleus extrudes toward the end of the maturation process. These cells have RNA which is rapidly consumed within the first 2 days of the cells 120 day life span.

[0067] For these anucleated cell types, gene expression markers must be specific only to the cell line (and not the immature form) to be useful as measures of cellular production rates. Genes specific to the lineage vs. other blood cell types will serve as markers of cellular production rates when measured on the RNA level. This is because RNA is specific to immature forms in these cases. For example, hemoglobin is specific to erythrocytes, but hemoglobin RNA is specific to newly produced erythrocytes. Therefore, if the rate of production of erythrocytes increases, so will the level of a lineage specific RNA (e.g., hemoglobin).

[0068] Hematopoietic growth factors and cytokines have incomplete lineage specificity. G-CSF is administered to patient with low granulocyte counts and the effect is a stimulation of all lineages (granulocytes, erythrocytes, platelets, etc...). Hemolytic anemia leads to increased production of multiple cell lineages although the only lineage in increased demand is the erythrocyte. Because of this lack of specificity of hematopoietic responses, erythrocyte and platelet production rates may serve as surrogates of increased production of lymphocyte lineages. Using RBCs and platelets production rates as surrogates for lymphocyte lineages may be useful because of the lack of a nucleus in these cells and the ease of measuring cellular production rates by simply measuring lineage specific RNA levels.

[0069] Hematopoies rates can be measured using gene expression profiling of peripheral blood. RBC and platelet specific genes provide unique opportunity for this because of their lack of a nucleus and kinetics. New cells = new much more RNA from these cell types in peripheral blood. Immature lymphocytes may be even more specific for immune activation and rejection. Cell specific markers of lymphocyte precursors were identified (aka lymphoblasts) see below. Granulocyte precursors and markers of megakaryocytes or premature forms of any blood cells may be useful in this regard.

Applications for measuring the rate of hematopoiesis

[0070] Changes in the rate of hematopoiesis have been correlated to a number of disease states and other pathologies. Examples of such conditions are listed in Table 11. One of skill in the art would be aware of other such conditions. In addition, one aspect of the present disclosure is the identification of the linkage between changes in the rate of hematopoiesis. The methods of the present disclosure directed to measuring the rates of hematopoiesis can therefore be applied to the diagnosis and monitoring of a number of disease states and other pathologies. In addition, these methods can be beneficial in determining appropriate therapies for patients.

Table: 11

Disorder / condition	Cell type	Cell production	Therapy
Anemia - Iron Deficiency	Erythrocyte	Decreased	Iron
Anemia - B 12, Folate deficiency	Erythrocyte	Decreased	B12, Folate
Anemia - Aplastic	Erythrocyte	Decreased	Epogen, transfusion
Anemia - hemolytic	Erythrocyte	Increased	Immunosuppression, Splenectomy
Anemia - Renal failure	Erythrocyte	Decreased	Erythropoietin
Anemia - Chronic disease	Erythrocyte	Decreased	Treat underlying cause
Polycythemia rubra vera	Erythrocyte	Increased	
Idiophic Thrombocytopenic purpura	Platelet	Increased	Immunosuppression, Splenectomy
Thrombotic Thrombocytopenic purpura	Platelet	Increased or decreased	Immunosuppression, plasmapheresis
Essential thrombocytosis	Platelet	Increased	
Leukemia	All lineages, variable	Increase, decreased or abnormal	Chemotherapy, BMT

(continued)

Disorder / condition	Cell type	Cell production	Therapy
Cytopenias due to immunosuppression	All lineages, variable	Decreased	Epo,neupogen
Cytopenias due to Chemotherapy	All lineages, variable	Decreased	Epo, GCSF, GMCSF
GVHD	All lineages, variable	Decreased	Immunosuppression
Myelodysplasia	All lineages, variable	Decreased, increased or abnormal	Chemo?
Allograft rejection	Lymphocytes, All lineages	Increased	Immunosuppression
Autoimmune diseases (many)	Lymphocytes, All lineages	Increased	Immunosuppression

[0071] The methods disclosed are also useful for monitoring treatment regimens of diseases or other pathologies which are correlated with changes in the rate of hematopoiesis. Furthermore, the methods may be used to monitor treatment with agents that affect the rate of hematopoiesis. One of skill in the art is aware of many such agents. The following agents are examples of such.

[0072] Erythropoietin is a growth factor that is used to treat a variety of anemias that are due to decreased red cell production. Monitoring of red cell production by gene expression or other means may improve dosing and provide a means for earlier assessment of response to therapy for this expensive drug.

[0073] Neupogen (G-CSF) is used for the treatment of low neutrophil counts (neutropenia) usually related to immunosuppression or chemotherapy. Monitoring neutrophil production by gene expression testing or another means may improve dosing, patient selection, and shorten duration of therapy.

[0074] Prednisone / Immunosuppression - One of most common side effects of immunosuppression is suppression of hematopoiesis. This may occur in any cell lineage. Gene expression monitoring or other measures of hematopoietic rates could be used to monitor regularly for cytopenias in a particular cell line and the information could be used to modify dosing, modify therapy or add a specific hematologic growth factor. Following cell counts themselves is less sensitive and results in the need for prolonged trials of therapies at a given dose before efficacy and toxicity can be assessed.

[0075] Monitoring of chemotherapeutic agents -Most chemotherapy agents suppress the bone marrow for some or all lineages. Gene expression testing or other means of assessing hematopoietic rates could be used to monitor regularly for cytopenias in a particular cell line and use information to modify dosing, modify therapy or add a specific hematologic growth factor.

General Molecular Biology References

[0076] In the context of the disclosure, nucleic acids and/or proteins are manipulated according to well known molecular biology techniques. Detailed protocols for numerous such procedures are described in, e.g., in Ausubel et al. Current Protocols in Molecular Biology (supplemented through 2000) John Wiley & Sons, New York ("Ausubel"); Sambrook et al. Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook"), and Berger and Kimmel Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA ("Berger").

[0077] In addition to the above references, protocols for in vitro amplification techniques, such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), Q-replicase amplification, and other RNA polymerase mediated techniques (e.g., NASBA), useful e.g., for amplifying cDNA probes of the disclosure, are found in Mullis et al. (1987) U.S. Patent No. 4,683,202; PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) ("Innis"); Arnheim and Levinson (1990) C&EN 36; The Journal Of NIH Research (1991) 3:81; Kwok et al. (1989) Proc Natl Acad Sci USA 86, 1173; Guatelli et al. (1990) Proc Natl Acad Sci USA 87:1874; Lomell et al. (1989) J Clin Chem 35:1826; Landegren et al. (1988) Science 241:1077; Van Brunt (1990) Biotechnology 8:291; Wu and Wallace (1989) Gene 4: 560; Barringer et al. (1990) Gene 89:117, and Sooknanan and Malek (1995) Biotechnology 13:563. Additional methods, useful for cloning nucleic acids in the context of the present disclosure, include Wallace et al. U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369:684 and the references therein.

[0078] Certain polynucleotides of the disclosure, e.g., oligonucleotides can be synthesized utilizing various solid-phase strategies involving mononucleotide- and/or trinucleotide-based phosphoramidite coupling chemistry. For example, nu-

cleic acid sequences can be synthesized by the sequential addition of activated monomers and/or trimers to an elongating polynucleotide chain. See e.g., Caruthers, M.H. et al. (1992) *Meth Enzymol* 211:3.

[0079] In lieu of synthesizing the desired sequences, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company, The Great American Gene Company ExpressGen, Inc., Operon Technologies, Inc. and many others.

[0080] Similarly, commercial sources for nucleic acid and protein microarrays are available, and include, e.g., Agilent Technologies, Palo Alto, CA Affymetrix, Santa Clara, CA; and others.

[0081] One area of relevance to the present disclosure is hybridization of oligonucleotides. Those of skill in the art differentiate hybridization conditions based upon the stringency of hybridization. For example, highly stringent conditions could include hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65° C, and washing in 0.1XSSC/0.1% SDS at 68° C. (Ausubel F. M. et al., eds., 1989, *Current Protocols in Molecular Biology*, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3). Moderate stringency conditions could include, e.g., washing in 0.2XSSC/0.1% SDS at 42° C. (Ausubel et al., 1989, *supra*).

[0082] The disclosure also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore the complements of, the DNA sequences disclosed. Such hybridization conditions may be highly stringent or less highly stringent, as described above. In instances wherein the nucleic acid molecules are deoxyoligonucleotides ("oligos"), highly stringent conditions may refer, e.g., to washing in 6xSSC/0.05% sodium pyrophosphate at 37° C. (for 14-base oligos), 48° C. (for 17-base oligos), 55° C. (for 20-base oligos), and 60° C. (for 23-base oligos). These nucleic acid molecules may act as target nucleotide sequence antisense molecules, useful, for example, in target nucleotide sequence regulation and/or as antisense primers in amplification reactions of target nucleotide sequence nucleic acid sequences. Further, such sequences may be used as part of ribozyme and/or triple helix sequences, also useful for target nucleotide sequence regulation. Still further, such molecules may be used as components of diagnostic methods whereby the presence of a disease-causing allele, may be detected.

Identification of diagnostic nucleotide sets

Candidate library

[0083] Libraries of candidates that are differentially expressed in leukocytes are substrates for the identification and evaluation of diagnostic oligonucleotide sets and disease specific target nucleotide sequences.

[0084] The term leukocyte is used generically to refer to any nucleated blood cell that is not a nucleated erythrocyte. More specifically, leukocytes can be subdivided into two broad classes. The first class includes granulocytes, including, most prevalently, neutrophils, as well as eosinophils and basophils at low frequency. The second class, the non-granular or mononuclear leukocytes, includes monocytes and lymphocytes (e.g., T cells and B cells). There is an extensive literature in the art implicating leukocytes, e.g., neutrophils, monocytes and lymphocytes in a wide variety of disease processes, including inflammatory and rheumatic diseases, neurodegenerative diseases (such as Alzheimer's dementia), cardiovascular disease, endocrine diseases, transplant rejection, malignancy and infectious diseases, and other diseases listed in Table 1. Mononuclear cells are involved in the chronic immune response, while granulocytes, which make up approximately 60% of the leukocytes, have a non-specific and stereotyped response to acute inflammatory stimuli and often have a life span of only 24 hours.

[0085] In addition to their widespread involvement and/or implication in numerous disease related processes, leukocytes are particularly attractive substrates for clinical and experimental evaluation for a variety of reasons. Most importantly, they are readily accessible at low cost from essentially every potential subject. Collection is minimally invasive and associated with little pain, disability or recovery time. Collection can be performed by minimally trained personnel (e.g., phlebotomists, medical technicians, etc.) in a variety of clinical and non-clinical settings without significant technological expenditure. Additionally, leukocytes are renewable, and thus available at multiple time points for a single subject.

Assembly of an initial candidate library

[0086] The initial candidate library was assembled from a combination of "mining" publication and sequence databases and construction of a differential expression library. Candidate oligonucleotide sequences in the library may be represented by a full-length or partial nucleic acid sequence, deoxyribonucleic acid (DNA) sequence, cDNA sequence, RNA sequence, synthetic oligonucleotides, etc. The nucleic acid sequence can be at least 19 nucleotides in length, at least 25 nucleotides, at least 40 nucleotides, at least 100 nucleotides, or larger. Alternatively, the protein product of a candidate nucleotide sequence may be represented in a candidate library using standard methods, as further described below. In selecting and validating diagnostic oligonucleotides, an initial library of 8,031 candidate oligonucleotide sequences using nucleic acid sequences of 50 nucleotides in length was constructed as described below.

Candidate nucleotide library of the invention

[0087] We identified members of an initial candidate nucleotide library that are differentially expressed in activated leukocytes and resting leukocytes. From that initial candidate nucleotide library, a pool of candidates was selected as listed in Table 2, and the sequence listing. Accordingly, the invention provides the candidate leukocyte nucleotide library comprising the nucleotide sequences listed in Table 2, and in the sequence listing. In another embodiment, the invention provides an candidate library comprising at least one nucleotide sequence listed in Tables 2 and 8 and the sequence listing. In another embodiment, the invention provides an candidate library comprising at least two nucleotide sequences listed in Table 2 and the sequence listing. In another embodiment, the at least two nucleotide sequence are at least 19 nucleotides in length, at least 35 nucleotides, at least 40 nucleotides or at least 100 nucleotides. In some embodiments, the nucleotide sequences comprises deoxyribonucleic acid (DNA) sequence, ribonucleic acid (RNA) sequence, synthetic oligonucleotide sequence, or genomic DNA sequence. It is understood that the nucleotide sequences may each correspond to one gene, or that several nucleotide sequences may correspond to one gene, or both.

[0088] The invention also discloses probes to the candidate nucleotide library. The probes can comprise at least two nucleotide sequences listed in Table 2, or the sequence listing which are differentially expressed in leukocytes in an individual with a least one disease criterion for at least one leukocyte-related disease and in leukocytes in an individual without the at least one disease criterion, wherein expression of the two or more nucleotide sequences is correlated with at least one disease criterion. It is understood that a probe may detect either the RNA expression or protein product expression of the candidate nucleotide library. Alternatively, or in addition, a probe can detect a genotype associated with a candidate nucleotide sequence, as further described below. The probes for the candidate nucleotide library can also be immobilized on an array.

[0089] The candidate nucleotide library disclosed in the invention is useful in identifying diagnostic nucleotide sets of the invention. The candidate nucleotide sequences may be further characterized, and may be identified as a disease target nucleotide sequence and/or a novel nucleotide sequence, as described below. The candidate nucleotide sequences may also be suitable for use as imaging reagents, as described below.

Detection of non-leukocyte expressed genes

[0090] When measuring gene expression levels in a blood sample, RNAs may be measured that are not derived from leukocytes. Examples are viral genes, free RNAs that have been released from damaged non-leukocyte cell types or RNA from circulating non-leukocyte cell types. For example, in the process of acute allograft rejection, tissue damage may result in release of allograft cells or RNAs derived from allograft cells into the circulation. In the case of cardiac allografts, such transcripts may be specific to muscle (myoglobin) or to cardiac muscle (Troponin I, Troponin T, CK-MB). Presence of cardiac specific mRNAs in peripheral blood may indicate ongoing or recent cardiac cellular damage (resulting from acute rejection). Therefore, such genes may be excellent diagnostic markers for allograft rejection.

Generation of Expression PatternsRNA, DNA or protein sample procurement

[0091] Following identification or assembly of a library of differentially expressed candidate nucleotide sequences, leukocyte expression profiles corresponding to multiple members of the candidate library are obtained. Leukocyte samples from one or more subjects are obtained by standard methods. Most typically, these methods involve trans-cutaneous venous sampling of peripheral blood. While sampling of circulating leukocytes from whole blood from the peripheral vasculature is generally the simplest, least invasive, and lowest cost alternative, it will be appreciated that numerous alternative sampling procedures exist, and are favorably employed in some circumstances. No pertinent distinction exists, in fact, between leukocytes sampled from the peripheral vasculature, and those obtained, e.g., from a central line, from a central artery, or indeed from a cardiac catheter, or during a surgical procedure which accesses the central vasculature. In addition, other body fluids and tissues that are, at least in part, composed of leukocytes are also desirable leukocyte samples. For example, fluid samples obtained from the lung during bronchoscopy may be rich in leukocytes, and amenable to expression profiling in the context of the disclosure, e.g., for the diagnosis, prognosis, or monitoring of lung transplant rejection, inflammatory lung diseases or infectious lung disease. Fluid samples from other tissues, e.g., obtained by endoscopy of the colon, sinuses, esophagus, stomach, small bowel, pancreatic duct, biliary tree, bladder, ureter, vagina, cervix or uterus, etc., are also suitable. Samples may also be obtained other sources containing leukocytes, e.g., from urine, bile, cerebrospinal fluid, feces, gastric or intestinal secretions, semen, or solid organ or joint biopsies.

[0092] Most frequently, mixed populations of leukocytes, such as are found in whole blood are utilized in the methods of the present disclosure. A crude separation, e.g., of mixed leukocytes from red blood cells, and/or concentration, e.g.,

over a sucrose, percoll or ficoll gradient, or by other methods known in the art, can be employed to facilitate the recovery of RNA or protein expression products at sufficient concentrations, and to reduce non-specific background. In some instances, it can be desirable to purify sub-populations of leukocytes, and methods for doing so, such as density or affinity gradients, flow cytometry, fluorescence Activated Cell Sorting (FACS), immune-magnetic separation, "panning," and the like, are described in the available literature and below.

Obtaining DNA, RNA and protein samples for expression profiling

[0093] Expression patterns can be evaluated at the level of DNA, or RNA or protein products. For example, a variety of techniques are available for the isolation of RNA from whole blood. Any technique that allows isolation of mRNA from cells (in the presence or absence of rRNA and tRNA) can be utilized. In brief, one method that allows reliable isolation of total RNA suitable for subsequent gene expression analysis, is described as follows. Peripheral blood (either venous or arterial) is drawn from a subject, into one or more sterile, endotoxin free, tubes containing an anticoagulant (e.g., EDTA, citrate, heparin, etc.). Typically, the sample is divided into at least two portions. One portion, e.g., of 5-8 ml of whole blood is frozen and stored for future analysis, e.g., of DNA or protein. A second portion, e.g., of approximately 8 ml whole blood is processed for isolation of total RNA by any of a variety of techniques as described in, e.g., Sambrook, Ausubel, below, as well as U.S. Patent Numbers: 5,728,822 and 4,843,155.

[0094] Typically, a subject sample of mononuclear leukocytes obtained from about 8 ml of whole blood, a quantity readily available from an adult human subject under most circumstances, yields 5-20 μ g of total RNA. This amount is ample, e.g., for labeling and hybridization to at least two probe arrays. Labeled probes for analysis of expression patterns of nucleotides of the candidate libraries are prepared from the subject's sample of RNA using standard methods. In many cases, cDNA is synthesized from total RNA using a polyT primer and labeled, e.g., radioactive or fluorescent, nucleotides. The resulting labeled cDNA is then hybridized to probes corresponding to members of the candidate nucleotide library, and expression data is obtained for each nucleotide sequence in the library. RNA isolated from subject samples (e.g., peripheral blood leukocytes, or leukocytes obtained from other biological fluids and samples) is next used for analysis of expression patterns of nucleotides of the candidate libraries.

[0095] In some cases, however, the amount of RNA that is extracted from the leukocyte sample is limiting, and amplification of the RNA is desirable. Amplification may be accomplished by increasing the efficiency of probe labeling, or by amplifying the RNA sample prior to labeling. It is appreciated that care must be taken to select an amplification procedure that does not introduce any bias (with respect to gene expression levels) during the amplification process.

[0096] Several methods are available that increase the signal from limiting amounts of RNA, e.g. use of the Clontech (Glass Fluorescent Labeling Kit) or Stratagene (Fairplay Microarray Labeling Kit), or the Micromax kit (New England Nuclear, Inc.). Alternatively, cDNA is synthesized from RNA using a T7- polyT primer, in the absence of label, and DNA dendrimers from Genisphere (3DNA Submicro) are hybridized to the poly T sequence on the primer, or to a different "capture sequence" which is complementary to a fluorescently labeled sequence. Each 3DNA molecule has 250 fluorescent molecules and therefore can strongly label each cDNA.

[0097] Alternatively, the RNA sample is amplified prior to labeling. For example, linear amplification may be performed, as described in U.S. Patent No. 6,132,997. A T7-polyT primer is used to generate the cDNA copy of the RNA. A second DNA strand is then made to complete the substrate for amplification. The T7 promoter incorporated into the primer is used by a T7 polymerase to produce numerous antisense copies of the original RNA. Fluorescent dye labeled nucleotides are directly incorporated into the RNA. Alternatively, amino allyl labeled nucleotides are incorporated into the RNA, and then fluorescent dyes are chemically coupled to the amino allyl groups, as described in Hughes. Other exemplary methods for amplification are described below.

[0098] It is appreciated that the RNA isolated must contain RNA derived from leukocytes, but may also contain RNA from other cell types to a variable degree. Additionally, the isolated RNA may come from subsets of leukocytes, e.g. monocytes and/or T-lymphocytes, as described above. Such consideration of cell type used for the derivation of RNA depend on the method of expression profiling used. Subsets of leukocytes can be obtained by fluorescence activated cell sorting (FACS), microfluidics cell separation systems or a variety of other methods. Cell sorting may be necessary for the discovery of diagnostic gene sets, for the implementation of gene sets as products or both. Cell sorting can be achieved with a variety of technologies (See Galbraith et al. 1999, Cantor et al. 1975, see also the technology of Guava Technologies, Hayward, CA).

[0099] DNA samples may be obtained for analysis of the presence of DNA mutations, single nucleotide polymorphisms (SNPs), or other polymorphisms. DNA is isolated using standard techniques, e.g. *Maniatus, supra*.

[0100] Expression of products of candidate nucleotides may also be assessed using proteomics. Protein(s) are detected in samples of patient serum or from leukocyte cellular protein. Serum is prepared by centrifugation of whole blood, using standard methods. Proteins present in the serum may have been produced from any of a variety of leukocytes and non-leukocyte cells, and include secreted proteins from leukocytes. Alternatively, leukocytes or a desired sub-population of leukocytes are prepared as described above. Cellular protein is prepared from leukocyte samples using methods well

known in the art, e.g., Trizol (Invitrogen Life Technologies, cat # 15596108; Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156; Simms, D., Cizdziel, P.E., and Chomczynski, P. (1993) Focus® 15, 99; Chomczynski, P., Bowers-Finn, R., and Sabatini, L. (1987) J. of NIH Res. 6, 83; Chomczynski, P. (1993) Bio/Techniques 15, 532; Bracete, A.M., Fox, D.K., and Simms, D. (1998) Focus 20, 82; Sewall, A. and McRae, S. (1998) Focus 20, 36; Anal Biochem 1984 Apr;138(1):141-3, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids; Wessel D, Flugge UI. (1984) Anal Biochem. 1984 Apr;138(1):141-143.

[0101] The assay itself may be a cell sorting assay in which cells are sorted and/or counted based on cell surface expression of a protein marker. (See Cantor et al. 1975, Galbraith et al. 1999)

Obtaining expression patterns

[0102] Expression patterns, or profiles, of a plurality of nucleotides corresponding to members of the candidate library are then evaluated in one or more samples of leukocytes. Typically, the leukocytes are derived from patient peripheral blood samples, although, as indicated above, many other sample sources are also suitable. These expression patterns constitute a set of relative or absolute expression values for a some number of RNAs or protein products corresponding to the plurality of nucleotide sequences evaluated, which is referred to herein as the subject's "expression profile" for those nucleotide sequences. While expression patterns for as few as one independent member of the candidate library can be obtained, it is generally preferable to obtain expression patterns corresponding to a larger number of nucleotide sequences, e.g., about 2, about 5, about 10, about 20, about 50, about 100, about 200, about 500, or about 1000, or more. The expression pattern for each differentially expressed component member of the library provides a finite specificity and sensitivity with respect to predictive value, e.g., for diagnosis, prognosis, monitoring, and the like.

Clinical Studies, Data and Patient Groups

[0103] For the purpose of discussion, the term subject, or subject sample of leukocytes, refers to an individual regardless of health and/or disease status. A subject can be a patient, a study participant, a control subject, a screening subject, or any other class of individual from whom a leukocyte sample is obtained and assessed in the context of the disclosure. Accordingly, a subject can be diagnosed with a disease, can present with one or more symptom of a disease, or a predisposing factor, such as a family (genetic) or medical history (medical) factor, for a disease, or the like. Alternatively, a subject can be healthy with respect to any of the aforementioned factors or criteria. It will be appreciated that the term "healthy" as used herein, is relative to a specified disease, or disease factor, or disease criterion, as the term "healthy" cannot be defined to correspond to any absolute evaluation or status. Thus, an individual defined as healthy with reference to any specified disease or disease criterion, can in fact be diagnosed with any other one or more disease, or exhibit any other one or more disease criterion.

[0104] Furthermore, while the discussion of the disclosure focuses, and is exemplified using human sequences and samples, the disclosure is equally applicable, through construction or selection of appropriate candidate libraries, to non-human animals, such as laboratory animals, e.g., mice, rats, guinea pigs, rabbits; domesticated livestock, e.g., cows, horses, goats, sheep, chicken, etc.; and companion animals, e.g., dogs, cats, etc.

Methods for obtaining expression data

[0105] Numerous methods for obtaining expression data are known, and any one or more of these techniques, singly or in combination, are suitable for determining expression profiles in the context of the present disclosure. For example, expression patterns can be evaluated by northern analysis, PCR, RT-PCR, Taq Man analysis, FRET detection, monitoring one or more molecular beacon, hybridization to an oligonucleotide array, hybridization to a cDNA array, hybridization to a polynucleotide array, hybridization to a liquid microarray, hybridization to a microelectric array, molecular beacons, cDNA sequencing, clone hybridization, cDNA fragment fingerprinting, serial analysis of gene expression (SAGE), subtractive hybridization, differential display and/or differential screening (see, e.g., Lockhart and Winzeler (2000) *Nature* 405:827-836, and references cited therein).

[0106] For example, specific PCR primers are designed to a member(s) of an candidate nucleotide library. cDNA is prepared from subject sample RNA by reverse transcription from a poly-dT oligonucleotide primer, and subjected to PCR. Double stranded cDNA may be prepared using primers suitable for reverse transcription of the PCR product, followed by amplification of the cDNA using in vitro transcription. The product of in vitro transcription is a sense-RNA corresponding to the original member(s) of the candidate library. PCR product may be also be evaluated in a number of ways known in the art, including real-time assessment using detection of labeled primers, e.g. TaqMan or molecular beacon probes. Technology platforms suitable for analysis of PCR products include the ABI 7700, 5700, or 7000 Sequence Detection Systems (Applied Biosystems, Foster City, CA), the MJ Research Opticon (MJ Research, Waltham, MA), the Roche Light Cycler (Roche Diagnostics, Indianapolis, IN), the Stratagene MX4000 (Stratagene, La Jolla, CA),

and the Bio-Rad iCycler (Bio-Rad Laboratories, Hercules, CA). Alternatively, molecular beacons are used to detect presence of a nucleic acid sequence in an unamplified RNA or cDNA sample, or following amplification of the sequence using any method, e.g. IVT (In Vitro transcription) or NASBA (nucleic acid sequence based amplification). Molecular beacons are designed with sequences complementary to member(s) of an candidate nucleotide library, and are linked to fluorescent labels. Each probe has a different fluorescent label with non-overlapping emission wavelengths. For example, expression of ten genes may be assessed using ten different sequence-specific molecular beacons.

[0107] Alternatively, or in addition, molecular beacons are used to assess expression of multiple nucleotide sequences at once. Molecular beacons with sequence complimentary to the members of a diagnostic nucleotide set are designed and linked to fluorescent labels. Each fluorescent label used must have a non-overlapping emission wavelength. For example, 10 nucleotide sequences can be assessed by hybridizing 10 sequence specific molecular beacons (each labeled with a different fluorescent molecule) to an amplified or un-amplified RNA or cDNA sample. Such an assay bypasses the need for sample labeling procedures.

[0108] Alternatively, or in addition bead arrays can be used to assess expression of multiple sequences at once. See, e.g. LabMAP 100, Luminex Corp, Austin, Texas). Alternatively, or in addition electric arrays are used to assess expression of multiple sequences, as exemplified by the e-Sensor technology of Motorola (Chicago, Ill.) or Nanochip technology of Nanogen (San Diego, CA.)

[0109] Of course, the particular method elected will be dependent on such factors as quantity of RNA recovered, practitioner preference, available reagents and equipment, detectors, and the like. Typically, however, the elected method(s) will be appropriate for processing the number of samples and probes of interest. Methods for high-throughput expression analysis are discussed below.

[0110] Alternatively, expression at the level of protein products of gene expression is performed. For example, protein expression, in a sample of leukocytes, can be evaluated by one or more method selected from among: western analysis, two-dimensional gel analysis, chromatographic separation, mass spectrometric detection, protein-fusion reporter constructs, colorimetric assays, binding to a protein array and characterization of polysomal mRNA. One particularly favorable approach involves binding of labeled protein expression products to an array of antibodies specific for members of the candidate library. Methods for producing and evaluating antibodies are widespread in the art, see, e.g., Coligan, *supra*; and Harlow and Lane (1989) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, NY ("Harlow and Lane"). Additional details regarding a variety of immunological and immunoassay procedures adaptable to the present disclosure by selection of antibody reagents specific for the products of candidate nucleotide sequences can be found in, e.g., Stites and Terr (eds.)(1991) *Basic and Clinical Immunology*, 7th ed., and Paul, *supra*. Another approach uses systems for performing desorption spectrometry. Commercially available systems, e.g., from Ciphergen Biosystems, Inc. (Fremont, CA) are particularly well suited to quantitative analysis of protein expression. Indeed, Protein Chip® arrays (see, e.g., the web site ciphergen.com) used in desorption spectrometry approaches provide arrays for defection of protein expression. Alternatively, affinity reagents, e.g., antibodies, small molecules, etc.) are developed that recognize epitopes of the protein product. Affinity assays are used in protein array assays, e.g. to detect the presence or absence of particular proteins. Alternatively, affinity reagents are used to detect expression using the methods described above. In the case of a protein that is expressed on the cell surface of leukocytes, labeled affinity reagents are bound to populations of leukocytes, and leukocytes expressing the protein are identified and counted using fluorescent activated cell sorting (FACS).

[0111] It is appreciated that the methods of expression evaluation discussed herein, although discussed in the context of discovery of diagnostic nucleotide sets, are equally applicable for expression evaluation when using diagnostic nucleotide sets for, e.g. diagnosis of diseases, as further discussed below.

High Throughput Expression Assays

[0112] A number of suitable high throughput formats exist for evaluating gene expression. Typically, the term high throughput refers to a format that performs at least about 100 assays, or at least about 500 assays, or at least about 1000 assays, or at least about 5000 assays, or at least about 10,000 assays, or more per day. When enumerating assays, either the number of samples or the number of candidate nucleotide sequences evaluated can be considered. For example, a northern analysis of, e.g., about 100 samples performed in a gridded array, e.g., a dot blot, using a single probe corresponding to an candidate nucleotide sequence can be considered a high throughput assay. More typically, however, such an assay is performed as a series of duplicate blots, each evaluated with a distinct probe corresponding to a different member of the candidate library. Alternatively, methods that simultaneously evaluate expression of about 100 or more candidate nucleotide sequences in one or more samples, or in multiple samples, are considered high throughput.

[0113] Numerous technological platforms for performing high throughput expression analysis are known. Generally, such methods involve a logical or physical array of either the subject samples, or the candidate library, or both. Common array formats include both liquid and solid phase arrays. For example, assays employing liquid phase arrays, e.g., for

hybridization of nucleic acids, binding of antibodies or other receptors to ligand, etc., can be performed in multiwell, or microtiter, plates. Microtiter plates with 96, 384 or 1536 wells are widely available, and even higher numbers of wells, e.g., 3456 and 9600 can be used. In general, the choice of microtiter plates is determined by the methods and equipment, e.g., robotic handling and loading systems, used for sample preparation and analysis. Exemplary systems include, e.g., the ORCA™ system from Beckman-Coulter, Inc. (Fullerton, CA) and the Zymate systems from Zymark Corporation (Hopkinton, MA).

[0114] Alternatively, a variety of solid phase arrays can favorably be employed in to determine expression patterns in the context of the disclosure. Exemplary formats include membrane or filter arrays (e.g., nitrocellulose, nylon), pin arrays, and bead arrays (e.g., in a liquid "slurry"). Typically, probes corresponding to nucleic acid or protein reagents that specifically interact with (e.g., hybridize to or bind to) an expression product corresponding to a member of the candidate library are immobilized, for example by direct or indirect cross-linking, to the solid support. Essentially any solid support capable of withstanding the reagents and conditions necessary for performing the particular expression assay can be utilized. For example, functionalized glass, silicon, silicon dioxide, modified silicon, any of a variety of polymers, such as (poly)tetrafluoroethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, or combinations thereof can all serve as the substrate for a solid phase array.

[0115] In a preferred embodiment, the array is a "chip" composed, e.g., of one of the above specified materials. Polynucleotide probes, e.g., RNA or DNA, such as cDNA, synthetic oligonucleotides, and the like, or binding proteins such as antibodies, that specifically interact with expression products of individual components of the candidate library are affixed to the chip in a logically ordered manner, i.e., in an array. In addition, any molecule with a specific affinity for either the sense or anti-sense sequence of the marker nucleotide sequence (depending on the design of the sample labeling), can be fixed to the array surface without loss of specific affinity for the marker and can be obtained and produced for array production, for example, proteins that specifically recognize the specific nucleic acid sequence of the marker, ribozymes, peptide nucleic acids (PNA), or other chemicals or molecules with specific affinity.

[0116] Detailed discussion of methods for linking nucleic acids and proteins to a chip substrate, are found in, e.g., US Patent No. 5,143,854 "LARGE SCALE PHOTOLITHOGRAPHIC SOLID PHASE SYNTHESIS OF POLYPEPTIDES AND RECEPTOR BINDING SCREENING THEREOF" to Pirrung et al., issued, September 1, 1992; US Patent No. 5,837,832 "ARRAYS OF NUCLEIC ACID PROBES ON BIOLOGICAL CHIPS" to Chee et al., issued November 17, 1998; US Patent No. 6,087,112 "ARRAYS WITH MODIFIED OLIGONUCLEOTIDE AND POLYNUCLEOTIDE COMPOSITIONS" to Dale, issued July 11, 2000; US Patent No. 5,215,882 "METHOD OF IMMOBILIZING NUCLEIC ACID ON A SOLID SUBSTRATE FOR USE IN NUCLEIC ACID HYBRIDIZATION ASSAYS" to Bahl et al., issued June 1, 1993; US Patent No. 5,707,807 "MOLECULAR INDEXING FOR EXPRESSED GENE ANALYSIS" to Kato, issued January 13, 1998; US Patent No. 5,807,522 "METHODS FOR FABRICATING MICROARRAYS OF BIOLOGICAL SAMPLES" to Brown et al., issued September 15, 1998; US Patent No. 5,958,342 "JET DROPLET DEVICE" to Gamble et al., issued Sept. 28, 1999; US Patent 5,994,076 "METHODS OF ASSAYING DIFFERENTIAL EXPRESSION" to Chenchik et al., issued Nov. 30, 1999; US Patent No. 6,004,755 "QUANTITATIVE MICROARRAY HYBRIDIZATION ASSAYS" to Wang, issued Dec. 21, 1999; US Patent No. 6,048,695 "CHEMICALLY MODIFIED NUCLEIC ACIDS AND METHOD FOR COUPLING NUCLEIC ACIDS TO SOLID SUPPORT" to Bradley et al., issued April 11, 2000; US Patent No. 6,060,240 "METHODS FOR MEASURING RELATIVE AMOUNTS OF NUCLEIC ACIDS IN A COMPLEX MIXTURE AND RETRIEVAL OF SPECIFIC SEQUENCES THEREFROM" to Kamb et al., issued May 9, 2000; US Patent No. 6,090,556 "METHOD FOR QUANTITATIVELY DETERMINING THE EXPRESSION OF A GENE" to Kato, issued July 18, 2000; and US Patent 6,040,138 "EXPRESSION MONITORING BY HYBRIDIZATION TO HIGH DENSITY OLIGONUCLEOTIDE ARRAYS" to Lockhart et al., issued March 21, 2000.

[0117] For example, cDNA inserts corresponding to candidate nucleotide sequences, in a standard TA cloning vector are amplified by a polymerase chain reaction for approximately 30-40 cycles. The amplified PCR products are then arrayed onto a glass support by any of a variety of well known techniques, e.g., the VSLIPS™ technology described in US Patent No. 5,143,854. RNA, or cDNA corresponding to RNA, isolated from a subject sample of leukocytes is labeled, e.g., with a fluorescent tag, and a solution containing the RNA (or cDNA) is incubated under conditions favorable for hybridization, with the "probe" chip. Following incubation, and washing to eliminate non-specific hybridization, the labeled nucleic acid bound to the chip is detected qualitatively or quantitatively, and the resulting expression profile for the corresponding candidate nucleotide sequences is recorded. It is appreciated that the probe used for diagnostic purposes may be identical to the probe used during diagnostic nucleotide sequence discovery and validation. Alternatively, the probe sequence may be different than the sequence used in diagnostic nucleotide sequence discovery and validation. Multiple cDNAs from a nucleotide sequence that are non-overlapping or partially overlapping may also be used.

[0118] In another approach, oligonucleotides corresponding to members of an candidate nucleotide library are synthesized and spotted onto an array. Alternatively, oligonucleotides are synthesized onto the array using methods known in the art, e.g. Hughes, et al. *supra*. The oligonucleotide is designed to be complementary to any portion of the candidate nucleotide sequence. In addition, in the context of expression analysis for, e.g. diagnostic use of diagnostic nucleotide sets, an oligonucleotide can be designed to exhibit particular hybridization characteristics, or to exhibit a particular

specificity and/or sensitivity, as further described below.

[0119] Hybridization signal may be amplified using methods known in the art, and as described herein, for example use of the Clontech kit (Glass Fluorescent Labeling Kit), Stratagene kit (Fairplay Microarray Labeling Kit), the Micromax kit (New England Nuclear, Inc.), the Genisphere kit (3DNA Submicro), linear amplification, e.g. as described in U.S. Patent No. 6,132,997 or described in Hughes, TR, et al., Nature Biotechnology, 19:343-347 (2001) and/or Westin et al. *Nat Biotech.* 18:199-204.

[0120] Alternatively, fluorescently labeled cDNA are hybridized directly to the microarray using methods known in the art. For example, labeled cDNA are generated by reverse transcription using Cy3- and Cy5-conjugated deoxynucleotides, and the reaction products purified using standard methods. It is appreciated that the methods for signal amplification of expression data useful for identifying diagnostic nucleotide sets are also useful for amplification of expression data for diagnostic purposes.

[0121] Microarray expression may be detected by scanning the microarray with a variety of laser or CCD-based scanners, and extracting features with numerous software packages, for example, Imagen (Biodiscovery), Feature Extraction (Agilent), Scanalyze (Eisen, M. 1999. SCANALYZE User Manual; Stanford Univ., Stanford, CA. Ver 2.32.), GenePix (Axon Instruments).

[0122] In another approach, hybridization to microelectric arrays is performed, e.g. as described in Umek et al (2001) *J Mol Diagn.* 3:74-84. An affinity probe, e.g. DNA, is deposited on a metal surface. The metal surface underlying each probe is connected to a metal wire and electrical signal detection system. Unlabelled RNA or cDNA is hybridized to the array, or alternatively, RNA or cDNA sample is amplified before hybridization, e.g. by PCR. Specific hybridization of sample RNA or cDNA results in generation of an electrical signal, which is transmitted to a detector. See Westin (2000) *Nat Biotech.* 18:199-204 (describing anchored multiplex amplification of a microelectronic chip array); Edman (1997) *NAR* 25:4907-14; Vignali (2000) *J Immunol Methods* 243:243-55.

[0123] In another approach, a microfluidics chip is used for RNA sample preparation and analysis. This approach increases efficiency because sample preparation and analysis are streamlined. Briefly, microfluidics may be used to sort specific leukocyte sub-populations prior to RNA preparation and analysis. Microfluidics chips are also useful for, e.g., RNA preparation, and reactions involving RNA (reverse transcription, RT-PCR). Briefly, a small volume of whole, anti-coagulated blood is loaded onto a microfluidics chip, for example chips available from Caliper (Mountain View, CA) or Nanogen (San Diego, CA.) A microfluidics chip may contain channels and reservoirs in which cells are moved and reactions are performed. Mechanical, electrical, magnetic, gravitational, centrifugal or other forces are used to move the cells and to expose them to reagents. For example, cells of whole blood are moved into a chamber containing hypotonic saline, which results in selective lysis of red blood cells after a 20-minute incubation. Next, the remaining cells (leukocytes) are moved into a wash chamber and finally, moved into a chamber containing a lysis buffer such as guanidine isothiocyanate. The leukocyte cell lysate is further processed for RNA isolation in the chip, or is then removed for further processing, for example, RNA extraction by standard methods. Alternatively, the microfluidics chip is a circular disk containing ficoll or another density reagent. The blood sample is injected into the center of the disc, the disc is rotated at a speed that generates a centrifugal force appropriate for density gradient separation of mononuclear cells, and the separated mononuclear cells are then harvested for further analysis or processing.

[0124] It is understood that the methods of expression evaluation, above, although discussed in the context of discovery of diagnostic nucleotide sets, are also applicable for expression evaluation when using diagnostic nucleotide sets for, e.g. diagnosis of diseases, as further discussed below.

Evaluation of expression patterns

[0125] Expression patterns can be evaluated by qualitative and/or quantitative measures. Certain of the above described techniques for evaluating gene expression (as RNA or protein products) yield data that are predominantly qualitative in nature. That is, the methods detect differences in expression that classify expression into distinct modes without providing significant information regarding quantitative aspects of expression. For example, a technique can be described as a qualitative technique if it detects the presence or absence of expression of an candidate nucleotide sequence, i.e., an on/off pattern of expression. Alternatively, a qualitative technique measures the presence (and/or absence) of different alleles, or variants, of a gene product.

[0126] In contrast, some methods provide data that characterizes expression in a quantitative manner. That is, the methods relate expression on a numerical scale, e.g., a scale of 0-5, a scale of 1-10, a scale of + - +++, from grade 1 to grade 5, a grade from a to z, or the like. It will be understood that the numerical, and symbolic examples provided are arbitrary, and that any graduated scale (or any symbolic representation of a graduated scale) can be employed in the context of the present disclosure to describe quantitative differences in nucleotide sequence expression. Typically, such methods yield information corresponding to a relative increase or decrease in expression.

[0127] Any method that yields either quantitative or qualitative expression data is suitable for evaluating expression of candidate nucleotide sequence in a subject sample of leukocytes. In some cases, e.g., when multiple methods are

employed to determine expression patterns for a plurality of candidate nucleotide sequences, the recovered data, e.g., the expression profile, for the nucleotide sequences is a combination of quantitative and qualitative data.

[0128] In some applications, expression of the plurality of candidate nucleotide sequences is evaluated sequentially. This is typically the case for methods that can be characterized as low- to moderate-throughput. In contrast, as the throughput of the elected assay increases, expression for the plurality of candidate nucleotide sequences in a sample or multiple samples of leukocytes, is assayed simultaneously. Again, the methods (and throughput) are largely determined by the individual practitioner, although, typically, it is preferable to employ methods that permit rapid, e.g. automated or partially automated, preparation and detection, on a scale that is time-efficient and cost-effective.

[0129] It is understood that the preceding discussion, while directed at the assessment of expression of the members of candidate libraries, is also applies to the assessment of the expression of members of diagnostic nucleotide sets, as further discussed below.

Genotyping

[0130] In addition to, or in conjunction with the correlation of expression profiles and clinical data, it is often desirable to correlate expression patterns with the subject's genotype at one or more genetic loci. The selected loci can be, for example, chromosomal loci corresponding to one or more member of the candidate library, polymorphic alleles for marker loci, or alternative disease related loci (not contributing to the candidate library) known to be, or putatively associated with, a disease (or disease criterion). Indeed, it will be appreciated, that where a (polymorphic) allele at a locus is linked to a disease (or to a predisposition to a disease), the presence of the allele can itself be a disease criterion.

[0131] Numerous well known methods exist for evaluating the genotype of an individual, including southern analysis, restriction fragment length polymorphism (RFLP) analysis, polymerase chain reaction (PCR), amplification length polymorphism (AFLP) analysis, single stranded conformation polymorphism (SSCP) analysis, single nucleotide polymorphism (SNP) analysis (e.g., via PCR, Taqman or molecular beacons), among many other useful methods. Many such procedures are readily adaptable to high throughput and/or automated (or semi-automated) sample preparation and analysis methods. Most, can be performed on nucleic acid samples recovered via simple procedures from the same sample of leukocytes as yielded the material for expression profiling. Exemplary techniques are described in, e.g., Sambrook, and Ausubel, *supra*.

Identification of the diagnostic nucleotide sets

[0132] Identification of diagnostic nucleotide sets and disease specific target nucleotide sequence proceeds by correlating the leukocyte expression profiles with data regarding the subject's health status to produce a data set designated a "molecular signature." Examples of data regarding a patient's health status, also termed "disease criteria(ion)", is described below and in the Section titled "selected diseases," below. Methods useful for correlation analysis are further described elsewhere in the specification.

[0133] Generally, relevant data regarding the subject's health status includes retrospective or prospective health data, e.g., in the form of the subject's medical history, as provided by the subject, physician or third party, such as, medical diagnoses, laboratory test results, diagnostic test results, clinical events, or medication lists, as further described below. Such data may include information regarding a patient's response to treatment and/or a particular medication and data regarding the presence of previously characterized "risk factors." For example, cigarette smoking and obesity are previously identified risk factors for heart disease. Further examples of health status information, including diseases and disease criteria, is described in the section titled Selected diseases, below.

[0134] Typically, the data describes prior events and evaluations (i.e., retrospective data). However, it is envisioned that data collected subsequent to the sampling (i.e., prospective data) can also be correlated with the expression profile. The tissue sampled, e.g., peripheral blood, bronchial lavage, etc., can be obtained at one or more multiple time points and subject data is considered retrospective or prospective with respect to the time of sample procurement.

[0135] Data collected at multiple time points, called "longitudinal data", is often useful, and thus, the disclosure encompasses the analysis of patient data collected from the same patient at different time points. Analysis of paired samples, such as samples from a patient at different time, allows identification of differences that are specifically related to the disease state since the genetic variability specific to the patient is controlled for by the comparison. Additionally, other variables that exist between patients may be controlled for in this way, for example, the presence or absence of inflammatory diseases (e.g., rheumatoid arthritis) the use of medications that may effect leukocyte gene expression, the presence or absence of co-morbid conditions, etc. Methods for analysis of paired samples are further described below. Moreover, the analysis of a pattern of expression profiles (generated by collecting multiple expression profiles) provides information relating to changes in expression level over time, and may permit the determination of a rate of change, a trajectory, or an expression curve. Two longitudinal samples may provide information on the change in expression of a gene over time, while three longitudinal samples may be necessary to determine the "trajectory" of expression of a gene.

Such information may be relevant to the diagnosis of a disease. For example, the expression of a gene may vary from individual to individual, but a clinical event, for example, a heart attack, may cause the level of expression to double in each patient. In this example, clinically interesting information is gleaned from the change in expression level, as opposed to the absolute level of expression in each individual.

[0136] When a single patient sample is obtained, it may still be desirable to compare the expression profile of that sample to some reference expression profile. In this case, one can determine the change of expression between the patient's sample and a reference expression profile that is appropriate for that patient and the medical condition in question. For example, a reference expression profile can be determined for all patients without the disease criterion in question who have similar characteristics, such as age, sex, race, diagnoses etc.

[0137] Generally, small sample sizes of 20-100 samples are used to identify a diagnostic nucleotide set. Larger sample sizes are generally necessary to validate the diagnostic nucleotide set for use in large and varied patient populations, as further described below. For example, extension of gene expression correlations to varied ethnic groups, demographic groups, nations, peoples or races may require expression correlation experiments on the population of interest.

Expression Reference Standards

[0138] Expression profiles derived from a patient (i.e., subjects diagnosed with, or exhibiting symptoms of, or exhibiting a disease criterion, or under a doctor's care for a disease) sample are compared to a control or standard expression RNA to facilitate comparison of expression profiles (e.g. of a set of candidate nucleotide sequences) from a group of patients relative to each other (i.e., from one patient in the group to other patients in the group, or to patients in another group).

[0139] The reference RNA used should have desirable features of low cost and simplicity of production on a large scale. Additionally, the reference RNA should contain measurable amounts of as many of the genes of the candidate library as possible.

[0140] For example, in one approach to identifying diagnostic nucleotide sets, expression profiles derived from patient samples are compared to a expression reference "standard." Standard expression reference can be, for example, RNA derived from resting cultured leukocytes or commercially available reference RNA, such as Universal reference RNA from Stratagene. See Nature, V406, 8-17-00, p. 747-752. Use of an expression reference standard is particularly useful when the expression of large numbers of nucleotide sequences is assayed, e.g. in an array, and in certain other applications, e.g. qualitative PCR, RT-PCR, etc., where it is desirable to compare a sample profile to a standard profile, and/or when large numbers of expression profiles, e.g. a patient population, are to be compared. Generally, an expression reference standard should be available in large quantities, should be a good substrate for amplification and labeling reactions, and should be capable of detecting a large percentage of candidate nucleic acids using suitable expression profiling technology.

[0141] Alternatively, or in addition, the expression profile derived from a patient sample is compared with the expression of an internal reference control gene, for example, β -actin or CD4. The relative expression of the profiled genes and the internal reference control gene (from the same individual) is obtained. An internal reference control may also be used with a reference RNA. For example, an expression profile for "gene 1" and the gene encoding CD4 can be determined in a patient sample and in a reference RNA. The expression of each gene can be expressed as the "relative" ratio of expression the gene in the patient sample compared with expression of the gene in the reference RNA. The expression ratio (sample/reference) for gene 1 may be divided by the expression ratio for CD4 (sample/reference) and thus the relative expression of gene 1 to CD4 is obtained.

[0142] A buffy coat control RNA useful for expression profiling, and a method of using control RNA produced from a population of buffy coat cells, the white blood cell layer derived from the centrifugation of whole blood are also provided. Buffy coat contains all white blood cells, including granulocytes, mononuclear cells and platelets. A method of preparing control RNA from buffy coat cells for use in expression profile analysis of leukocytes is also provided. Buffy coat fractions are obtained, e.g. from a blood bank or directly from individuals, preferably from a large number of individuals such that bias from individual samples is avoided and so that the RNA sample represents an average expression of a healthy population. Buffy coat fractions from about 50 or about 100, or more individuals are preferred. 10 ml buffy coat from each individual is used. Buffy coat samples are treated with an erythrocyte lysis buffer, so that erythrocytes are selectively removed. The leukocytes of the buffy coat layer are collected by centrifugation. Alternatively, the buffy cell sample can be further enriched for a particular leukocyte sub-populations, e.g. mononuclear cells, T-lymphocytes, etc. To enrich for mononuclear cells, the buffy cell pellet, above, is diluted in PBS (phosphate buffered saline) and loaded onto a non-polystyrene tube containing a polysucrose and sodium diatrizoate solution adjusted to a density of 1.077 $\pm$ 0.001 g/ml. To enrich for T-lymphocytes, 45 ml of whole blood is treated with RosetteSep (Stem Cell Technologies), and incubated at room temperature for 20 minutes. The mixture is diluted with an equal volume of PBS plus 2% FBS and mixed by inversion. 30 ml of diluted mixture is layered on top of 15 ml DML medium (Stem Cell Technologies). The tube is centrifuged at 1200 x g, and the enriched cell layer at the plasma : medium interface is removed, washed with

PBS + 2% FBS, and cells collected by centrifugation at 1200 x g. The cell pellet is treated with 5 ml of erythrocyte lysis buffer (EL buffer, Qiagen) for 10 minutes on ice, and enriched T-lymphocytes are collected by centrifugation.

[0143] In addition or alternatively, the buffy cells (whole buffy coat or sub-population, e.g. mononuclear fraction) can be cultured *in vitro* and subjected to stimulation with cytokines or activating chemicals such as phorbol esters or ionomycin. Such stimuli may increase expression of nucleotide sequences that are expressed in activated immune cells and might be of interest for leukocyte expression profiling experiments.

[0144] Following sub-population selection and/or further treatment, e.g. stimulation as described above, RNA is prepared using standard methods. For example, cells are pelleted and lysed with a phenol/guanidinium thiocyanate and RNA is prepared. RNA can also be isolated using a silica gel-based purification column or the column method can be used on RNA isolated by the phenol/guanidinium thiocyanate method. RNA from individual buffy coat samples can be pooled during this process, so that the resulting reference RNA represents the RNA of many individuals and individual bias is minimized or eliminated. In addition, a new batch of buffy coat reference RNA can be directly compared to the last batch to ensure similar expression pattern from one batch to another, using methods of collecting and comparing expression profiles described above/below. One or more expression reference controls are used in an experiment. For example, RNA derived from one or more of the following sources can be used as controls for an experiment: stimulated or unstimulated whole buffy coat, stimulated or unstimulated peripheral mononuclear cells, or stimulated or unstimulated T-lymphocytes.

[0145] Alternatively, the expression reference standard can be derived from any subject or class of subjects including healthy subjects or subjects diagnosed with the same or a different disease or disease criterion. Expression profiles from subjects in two distinct classes are compared to determine which subset of nucleotide sequences in the candidate library best distinguish between the two subject classes, as further discussed below. It will be appreciated that in the present context, the term "distinct classes" is relevant to at least one distinguishable criterion relevant to a disease of interest, a "disease criterion." The classes can, of course, demonstrate significant overlap (or identity) with respect to other disease criteria, or with respect to disease diagnoses, prognoses, or the like. The mode of discovery involves, e.g., comparing the molecular signature of different subject classes to each other (such as patient to control, patients with a first diagnosis to patients with a second diagnosis, etc.) or by comparing the molecular signatures of a single individual taken at different time points. The disclosure can be applied to a broad range of diseases, disease criteria, conditions and other clinical and/or epidemiological questions, as further discussed above/below.

[0146] It is appreciated that while the present discussion pertains to the use of expression reference controls while identifying diagnostic nucleotide sets, expression reference controls are also useful during use of diagnostic nucleotide sets, e.g. use of a diagnostic nucleotide set for diagnosis of a disease, as further described below.

Analysis of expression profiles

[0147] In order to facilitate ready access, e.g., for comparison, review, recovery, and/or modification, the molecular signatures/expression profiles are typically recorded in a database. Most typically, the database is a relational database accessible by a computational device, although other formats, e.g., manually accessible indexed files of expression profiles as photographs, analogue or digital imaging readouts, spreadsheets, etc. can be used. Further details regarding preferred embodiments are provided below. Regardless of whether the expression patterns initially recorded are analog or digital in nature and/or whether they represent quantitative or qualitative differences in expression, the expression patterns, expression profiles (collective expression patterns), and molecular signatures (correlated expression patterns) are stored digitally and accessed via a database. Typically, the database is compiled and maintained at a central facility, with access being available locally and/or remotely.

[0148] As additional samples are obtained, and their expression profiles determined and correlated with relevant subject data, the ensuing molecular signatures are likewise recorded in the database. However, rather than each subsequent addition being added in an essentially passive manner in which the data from one sample has little relation to data from a second (prior or subsequent) sample, the algorithms optionally additionally query additional samples against the existing database to further refine the association between a molecular signature and disease criterion. Furthermore, the data set comprising the one (or more) molecular signatures is optionally queried against an expanding set of additional or other disease criteria. The use of the database in integrated systems and web embodiments is further described below.

Analysis of expression profile data from arrays

[0149] Expression data is analyzed using methods well known in the art, including the software packages Imagene (Biodiscovery, Marina del Rey, CA), Feature Extraction Software (Agilent, Palo Alto, CA), and Scanalyze (Stanford University). In the discussion that follows, a "feature" refers to an individual spot of DNA on an array. Each gene may be represented by more than one feature. For example, hybridized microarrays are scanned and analyzed on an Axon Instruments scanner using GenePix 3.0 software (Axon Instruments, Union City, CA). The data extracted by GenePix

is used for all downstream quality control and expression evaluation. The data is derived as follows. The data for all features flagged as "not found" by the software is removed from the dataset for individual hybridizations. The "not found" flag by GenePix indicates that the software was unable to discriminate the feature from the background. Each feature is examined to determine the value of its signal. The median pixel intensity of the background (B_n) is subtracted from the median pixel intensity of the feature (F_n) to produce the background-subtracted signal (hereinafter, "BGSS"). The BGSS is divided by the standard deviation of the background pixels to provide the signal-to-noise ratio (hereinafter, "S/N"). Features with a S/N of three or greater in both the Cy3 channel (corresponding to the sample RNA) and Cy5 channel (corresponding to the reference RNA) are used for further analysis (hereinafter denoted "useable features"). Alternatively, different S/Ns are used for selecting expression data for an analysis. For example, only expression data with signal to noise ratios > 3 might be used in an analysis. Alternatively, features with S/N values < 3 may be flagged as such and included in the analysis. Such flagged data sets include more values and may allow one to discover expression markers that would be missed otherwise. However, such data sets may have a higher variability than filtered data, which may decrease significance of findings or performance of correlation statistics.

[0150] For each usable feature (i), the expression level (e) is expressed as the logarithm of the ratio (R) of the Background Subtracted Signal (hereinafter "BGSS") for the Cy3 (sample RNA) channel divided by the BGSS for the Cy5 channel (reference RNA). This "log ratio" value is used for comparison to other experiments.

$$R_i = \frac{BGSS_{sample}}{BGSS_{reference}} \quad (0.1)$$

$$e_i = \log r_i \quad (0.2)$$

[0151] Variation in signal across hybridizations may be caused by a number of factors affecting hybridization, DNA spotting, wash conditions, and labeling efficiency.

[0152] A single reference RNA may be used with all of the experimental RNAs, permitting multiple comparisons in addition to individual comparisons. By comparing sample RNAs to the same reference, the gene expression levels from each sample are compared across arrays, permitting the use of a consistent denominator for our experimental ratios.

[0153] Alternative methods of analyzing the data may involve 1) using the sample channel without normalization by the reference channel, 2) using an intensity-dependent normalization based on the reference which provides a greater correction when the signal in the reference channel is large, 3) using the data without background subtraction or subtracting an empirically derived function of the background intensity rather than the background itself.

Scaling

[0154] The data may be scaled (normalized) to control for labeling and hybridization variability within the experiment, using methods known in the art. Scaling is desirable because it facilitates the comparison of data between different experiments, patients, etc. Generally the BGSS are scaled to a factor such as the median, the mean, the trimmed mean, and percentile. Additional methods of scaling include: to scale between 0 and 1, to subtract the mean, or to subtract the median.

[0155] Scaling is also performed by comparison to expression patterns obtained using a common reference RNA, as described in greater detail above. As with other scaling methods, the reference RNA facilitates multiple comparisons of the expression data, e.g., between patients, between samples, etc. Use of a reference RNA provides a consistent denominator for experimental ratios.

[0156] In addition to the use of a reference RNA, individual expression levels may be adjusted to correct for differences in labeling efficiency between different hybridization experiments, allowing direct comparison between experiments with different overall signal intensities, for example. A scaling factor (a) may be used to adjust individual expression levels as follows. The median of the scaling factor (a), for example, BGSS, is determined for the set of all features with a S/N greater than three. Next, the BGSS _{i} (the BGSS for each feature " i ") is divided by the median for all features (a), generating a scaled ratio. The scaled ratio is used to determine the expression value for the feature (e_i), or the log ratio.

$$S_i = \frac{BGSS_i}{a} \quad (0.3)$$

$$e_i = \log\left(\frac{Cy3S_i}{Cy5S_i}\right) \quad (0.4)$$

[0157] In addition, or alternatively, control features are used to normalize the data for labeling and hybridization variability within the experiment. Control feature may be cDNA for genes from the plant, *Arabidopsis thaliana*, that are included when spotting the mini-array. Equal amounts of RNA complementary to control cDNAs are added to each of the samples before they were labeled. Using the signal from these control genes, a normalization constant (L) is determined according to the following formula:

$$L_j = \frac{\frac{\sum_{i=1}^N BGSS_{j,i}}{N}}{\frac{\sum_{j=1}^K \frac{\sum_{i=1}^N BGSS_{j,i}}{N}}{K}}$$

where $BGSS_i$ is the signal for a specific feature, N is the number of *A. thaliana* control features, K is the number of hybridizations, and L_j is the normalization constant for each individual hybridization.

[0158] Using the formula above, the mean for all control features of a particular hybridization and dye (e.g., Cy3) is calculated. The control feature means for all Cy3 hybridizations are averaged, and the control feature mean in one hybridization divided by the average of all hybridizations to generate a normalization constant for that particular Cy3 hybridization (L_j), which is used as a in equation (0.3). The same normalization steps may be performed for Cy3 and Cy5 values.

[0159] An alternative scaling method can also be used. The log of the ratio of Green/Red is determined for all features. The median log ratio value for all features is determined. The feature values are then scaled using the following formula: $\text{Log_Scaled-Feature-Ratio} = \text{Log_Feature_Ratio} - \text{Median_Log_Ratio}$.

[0160] Many additional methods for normalization exist and can be applied to the data. In one method, the average ratio of Cy3 BGSS / Cy5 BGSS is determined for all features on an array. This ratio is then scaled to some arbitrary number, such as 1 or some other number. The ratio for each probe is then multiplied by the scaling factor required to bring the average ratio to the chosen level. This is performed for each array in an analysis. Alternatively, the ratios are normalized to the average ratio across all arrays in an analysis. Other methods of normalization include forcing the distribution of signal strengths of the various arrays into greater agreement by transforming them to match certain points (quartiles, or deciles, etc.) in a standard distribution, or in the most extreme case using the rank of the signal of each oligonucleotide relative to the other oligonucleotides on the array.

[0161] If multiple features are used per gene sequence or oligonucleotide, these repeats can be used to derive an average expression value for each gene. If some of the replicate features are of poor quality and don't meet requirements for analysis, the remaining features can be used to represent the gene or gene sequence.

Correlation analysis

[0162] Correlation analysis is performed to determine which array probes have expression behavior that best distinguishes or serves as markers for relevant groups of samples representing a particular clinical condition. Correlation analysis, or comparison among samples representing different disease criteria (e.g., clinical conditions), is performed using standard statistical methods. Numerous algorithms are useful for correlation analysis of expression data, and the selection of algorithms depends in part on the data analysis to be performed. For example, algorithms can be used to identify the single most informative gene with expression behavior that reliably classifies samples, or to identify all the genes useful to classify samples. Alternatively, algorithms can be applied that determine which set of 2 or more genes have collective expression behavior that accurately classifies samples. The use of multiple expression markers for diagnostics may overcome the variability in expression of a gene between individuals, or overcome the variability intrinsic to the assay. Multiple expression markers may include redundant markers (surrogates), in that two or more genes or probes may provide the same information with respect to diagnosis. This may occur, for example, when two or more

genes or gene probes are coordinately expressed. For diagnostic application, it may be appropriate to utilize a gene and one or more of its surrogates in the assay. This redundancy may overcome failures (technical or biological) of a single marker to distinguish samples. Alternatively, one or more surrogates may have properties that make them more suitable for assay development, such as a higher baseline level of expression, better cell specificity, a higher fold change between sample groups or more specific sequence for the design of PCR primers or complimentary probes. It will be appreciated that while the discussion above pertains to the analysis of RNA expression profiles the discussion is equally applicable to the analysis of profiles of proteins or other molecular markers.

[0163] Prior to analysis, expression profile data may be formatted or prepared for analysis using methods known in the art. For example, often the log ratio of scaled expression data for every array probe is calculated using the following formula:

$\log(\text{Cy 3 BGSS} / \text{Cy5 BGSS})$, where Cy 3 signal corresponds to the expression of the gene in the clinical sample, and Cy5 signal corresponds to expression of the gene in the reference RNA.

Data may be further filtered depending on the specific analysis to be done as noted below. For example, filtering may be aimed at selecting only samples with expression above a certain level, or probes with variability above a certain level between sample sets.

[0164] The following non-limiting discussion consider several statistical methods known in the art.

[0165] Briefly, the t-test and ANOVA are used to identify single genes with expression differences between or among populations, respectively. Multivariate methods are used to identify a set of two or more genes for which expression discriminates between two disease states more specifically than expression of any single gene.

t-test

[0166] The simplest measure of a difference between two groups is the Student's t test. See, e.g., Welsh et al. (2001) *Proc Natl Acad Sci USA* 98:1176-81 (demonstrating the use of an unpaired Student's t-test for the discovery of differential gene expression in ovarian cancer samples and control tissue samples). The t-test assumes equal variance and normally distributed data. This test identifies the probability that there is a difference in expression of a single gene between two groups of samples. The number of samples within each group that is required to achieve statistical significance is dependent upon the variation among the samples within each group. The standard formula for a t-test is:

$$t(e_i) = \frac{\bar{e}_{i,c} - \bar{e}_{i,t}}{\sqrt{(s_{i,c}^2 / n_c) + (s_{i,t}^2 / n_t)}}, \quad (0.5)$$

where $\bar{e}_i$ is the difference between the mean expression level of gene i in groups c and t, $s_{i,c}$ is the variance of gene x in group c and $s_{i,t}$ is the variance of gene x in group t. n_c and n_t are the numbers of samples in groups c and t.

[0167] The combination of the t statistic and the degrees of freedom [$\min(n_t, n_c) - 1$] provides a p value, the probability of rejecting the null hypothesis. A p-value of ≤ 0.01 , signifying a 99 percent probability the mean expression levels are different between the two groups (a 1% chance that the mean expression levels are in fact not different and that the observed difference occurred by statistical chance), is often considered acceptable.

[0168] When performing tests on a large scale, for example, on a large dataset of about 8000 genes, a correction factor must be included to adjust for the number of individual tests being performed. The most common and simplest correction is the Bonferroni correction for multiple tests, which divides the p-value by the number of tests run. Using this test on an 8000 member dataset indicates that a p value of ≤ 0.00000125 is required to identify genes that are likely to be truly different between the two test conditions.

Significance analysis for microarrays (SAM)

[0169] significance analysis for microarrays (SAM) (Tusher 2001) is a method through which genes with a correlation between their expression values and the response vector are statistically discovered and assigned a statistical significance. The ratio of false significant to significant genes is the False Discovery Rate (FDR). This means that for each threshold there are a set of genes which are called significant, and the FDR gives a confidence level for this claim. If a gene is called differentially expressed between 2 classes by SAM, with a FDR of 5%, there is a 95% chance that the gene is actually differentially expressed between the classes. SAM takes into account the variability and large number of variables of microarrays. SAM will identify genes that are most globally differentially expressed between the classes. Thus, important genes for identifying and classifying outlier samples or patients may not be identified by SAM.

Non-Parametric Tests

[0170] Wilcoxon's signed ranks method is one example of a non-parametric test and is utilized for paired comparisons. See e.g., Sokal and Rohlf (1987) Introduction to Biostatistics 2nd edition, WH Freeman, New York. At least 6 pairs are necessary to apply this statistic. This test is useful for analysis of paired expression data (for example, a set of patients who have cardiac transplant biopsy on 2 occasions and have a grade 0 on one occasion and a grade 3A on another). The Fisher Exact Test with a threshold and the Mann-Whitney Test are other non-parametric tests that may be used.

ANOVA

[0171] Differences in gene expression across multiple related groups may be assessed using an Analysis of Variance (ANOVA), a method well known in the art (Michelson and Schofield, 1996).

Multivariate analysis

[0172] Many algorithms suitable for multivariate analysis are known in the art. Generally, a set of two or more genes for which expression discriminates between two disease states more specifically than expression of any single gene is identified by searching through the possible combinations of genes using a criterion for discrimination, for example the expression of gene X must increase from normal 300 percent, while the expression of genes Y and Z must decrease from normal by 75 percent. Ordinarily, the search starts with a single gene, then adds the next best fit at each step of the search. Alternatively, the search starts with all of the genes and genes that do not aid in the discrimination are eliminated step-wise.

Paired samples

[0173] Paired samples, or samples collected at different time-points from the same patient, are often useful, as described above. For example, use of paired samples permits the reduction of variation due to genetic variation among individuals. In addition, the use of paired samples has a statistical significance, in that data derived from paired samples can be calculated in a different manner that recognizes the reduced variability. For example, the formula for a t-test for paired samples is:

$$t(e_x) = \frac{\bar{D}_{e_x}}{\sqrt{\frac{\sum D^2 - (\sum D)^2 / b}{b - 1}}}, \quad (0.5)$$

where D is the difference between each set of paired samples and b is the number of sample pairs. $\bar{D}$ is the mean of the differences between the members of the pairs. In this test, only the differences between the paired samples are considered, then grouped together (as opposed to taking all possible differences between groups, as would be the case with an ordinary t-test). Additional statistical tests useful with paired data, e.g., ANOVA and Wilcoxon's signed rank test, are discussed above.

Diagnostic classification

[0174] Once a discriminating set of genes is identified, the diagnostic classifier (a mathematical function that assigns samples to diagnostic categories based on expression data) is applied to unknown sample expression levels.

[0175] Methods that can be used for this analysis include the following non-limiting list:

[0176] CLEAVER is an algorithm used for classification of useful expression profile data. See Raychaudhuri et al. (2001) Trends Biotechnol 19:189-193. CLEAVER uses positive training samples (e.g., expression profiles from samples known to be derived from a particular patient or sample diagnostic category, disease or disease criteria), negative training samples (e.g., expression profiles from samples known not to be derived from a particular patient or sample diagnostic category, disease or disease criteria) and test samples (e.g., expression profiles obtained from a patient), and determines whether the test sample correlates with the particular disease or disease criteria, or does not correlate with a particular disease or disease criteria. CLEAVER also generates a list of the 20 most predictive genes for classification.

[0177] Artificial neural networks (hereinafter, "ANN") can be used to recognize patterns in complex data sets and can discover expression criteria that classify samples into more than 2 groups. The use of artificial neural networks for

discovery of gene expression diagnostics for cancers using expression data generated by oligonucleotide expression microarrays is demonstrated by Khan et al. (2001) *Nature Med.* 7:673-9. Khan found that 96 genes provided 0% error rate in classification of the tumors. The most important of these genes for classification was then determined by measuring the sensitivity of the classification to a change in expression of each gene. Hierarchical clustering using the 96 genes results in correct grouping of the cancers into diagnostic categories.

[0178] Golub uses cDNA microarrays and a distinction calculation to identify genes with expression behavior that distinguishes myeloid and lymphoid leukemias. See Golub et al. (1999) *Science* 286:531-7. Self organizing maps were used for new class discovery. Cross validation was done with a "leave one out" analysis. 50 genes were identified as useful markers. This was reduced to as few as 10 genes with equivalent diagnostic accuracy.

[0179] Hierarchical and non-hierarchical clustering methods are also useful for identifying groups of genes that correlate with a subset of clinical samples such as with transplant rejection grade. Alizadeh used hierarchical clustering as the primary tool to distinguish different types of diffuse B-cell lymphomas based on gene expression profile data. See Alizadeh et al. (2000) *Nature* 403:503-11. Alizadeh used hierarchical clustering as the primary tool to distinguish different types of diffuse B-cell lymphomas based on gene expression profile data. A cDNA array carrying 17856 probes was used for these experiments, 96 samples were assessed on 128 arrays, and a set of 380 genes was identified as being useful for sample classification.

[0180] Perou demonstrates the use of hierarchical clustering for the molecular classification of breast tumor samples based on expression profile data. See Perou et al. (2000) *Nature* 406:747-52. In this work, a cDNA array carrying 8102 gene probes was used. 1753 of these genes were found to have high variation between breast tumors and were used for the analysis.

[0181] Hastie describes the use of gene shaving for discovery of expression markers. Hastie et al. (2000) *Genome Biol.* 1(2):RESEARCH 0003.1-0003.21. The gene shaving algorithm identifies sets of genes with similar or coherent expression patterns, but large variation across conditions (RNA samples, sample classes, patient classes). In this manner, genes with a tight expression pattern within a transplant rejection grade, but also with high variability across rejection grades are grouped together. The algorithm takes advantage of both characteristics in one grouping step. For example, gene shaving can identify useful marker genes with co-regulated expression. Sets of useful marker genes can be reduced to a smaller set, with each gene providing some non-redundant value in classification. This algorithm was used on the data set described in Alizadeh et al., supra, and the set of 380 informative gene markers was reduced to 234.

[0182] Supervised harvesting of expression trees (Hastie 2001) identifies genes or clusters that best distinguish one class from all the others on the data set. The method is used to identify the genes/clusters that can best separate one class versus all the others for datasets that include two or more classes or all classes from each other. This algorithm can be used for discovery or testing of a diagnostic gene set.

[0183] CART is a decision tree classification algorithm (Breiman 1984). From gene expression and or other data, CART can develop a decision tree for the classification of samples. Each node on the decision tree involves a query about the expression level of one or more genes or variables. Samples that are above the threshold go down one branch of the decision tree and samples that are not go down the other branch. See example 4 for further description of its use in classification analysis and examples of its usefulness in discovering and implementing a diagnostic gene set. CART identifies surrogates for each splitter (genes that are the next best substitute for a useful gene in classification).

[0184] Multiple Additive Regression Trees (Friedman, JH 1999, MART) is similar to CART in that it is a classification algorithm that builds decision trees to distinguish groups. MART builds numerous trees for any classification problem and the resulting model involves a combination of the multiple trees. MART can select variables as it build models and thus can be used on large data sets, such as those derived from an 8000 gene microarray. Because MART uses a combination of many trees and does not take too much information from any one tree, it resists over training. MART identifies a set of genes and an algorithm for their use as a classifier.

[0185] A Nearest Shrunken Centroids Classifier can be applied to microarray or other data sets by the methods described by Tibshirani et al. 2002. This algorithms also identified gene sets for classification and determines their 10 fold cross validation error rates for each class of samples. The algorithm determines the error rates for models of any size, from one gene to all genes in the set. The error rates for either or both sample classes can be minimized when a particular number of genes are used. When this gene number is determined, the algorithm associated with the selected genes can be identified and employed as a classifier on prospective sample.

[0186] Once a set of genes and expression criteria for those genes have been established for classification, cross validation is done. There are many approaches, including a 10 fold cross validation analysis in which 10% of the training samples are left out of the analysis and the classification algorithm is built with the remaining 90%. The 10% are then used as a test set for the algorithm. The process is repeated 10 times with 10% of the samples being left out as a test set each time. Through this analysis, one can derive a cross validation error which helps estimate the robustness of the algorithm for use on prospective (test) samples.

[0187] Clinical data are gathered for every patient sample used for expression analysis. Clinical variables can be quantitative or non-quantitative. A clinical variable that is quantitative can be used as a variable for significance or

classification analysis. Non-quantitative clinical variables, such as the sex of the patient, can also be used in a significance analysis or classification analysis with some statistical tool. It is appreciated that the most useful diagnostic gene set for a condition may be optimal when considered along with one or more predictive clinical variables. Clinical data can also be used as supervising vectors for a correlation analysis. That is to say that the clinical data associated with each sample can be used to divide the samples into meaningful diagnostic categories for analysis. For example, samples can be divided into 2 or more groups based on the presence or absence of some diagnostic criterion (a). In addition, clinical data can be utilized to select patients for a correlation analysis or to exclude them based on some undesirable characteristic, such as an ongoing infection, a medicine or some other issue. Clinical data can also be used to assess the pre-test probability of an outcome. For example, patients who are female are much more likely to be diagnosed as having systemic lupus erythematosus than patients who are male.

[0188] Once a set of genes are identified that classify samples with acceptable accuracy. These genes are validated as a set using new samples that were not used to discover the gene set. These samples can be taken from frozen archives from the discovery clinical study or can be taken from new patients prospectively. Validation using a "test set" of samples can be done using expression profiling of the gene set with microarrays or using real-time PCR for each gene on the test set samples. Alternatively, a different expression profiling technology can be used.

Immune Monitoring

[0189] Leukocyte gene expression can be used to monitor the immune system. Immune monitoring examines both the level of gene expression for a set of genes in a given cell type and for genes which are expressed in a cell type selective manner. Gene expression monitoring will also detect the presence or absence of new cell types, progenitor cells, differentiation of cells and the like. Gene expression patterns may be associated with activation or the resting state of cells of the immune system that are responsible for or responsive to a disease state. For example, in the process of transplant rejection, cells of the immune system are activated by the presence of the foreign tissue. Genes and gene sets that monitor and diagnose this process are providing a measure of the level and type of activation of the immune system. Genes and gene sets that are useful in monitoring the immune system may be useful for diagnosis and monitoring of all diseases that involve the immune system. Some examples are transplant rejection, rheumatoid arthritis, lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS, and viral, bacterial and fungal infection. All disorders and diseases disclosed herein are contemplated. Genes and gene sets that monitor immune activation are useful for monitoring response to immunosuppressive drug therapy, which is used to decrease immune activation. Genes are found to correlate with immune activation by correlation of expression patterns to the known presence of immune activation or quiescence in a sample as determined by some other test.

Selected Diseases

[0190] In principle, diagnostic nucleotide sets of the disclosure may be developed and applied to essentially any disease, or disease criterion, as long as at least one subset of nucleotide sequences is differentially expressed in samples derived from one or more individuals with a disease criteria or disease and one or more individuals without the disease criteria or disease, wherein the individual may be the same individual sampled at different points in time, or the individuals may be different individuals (or populations of individuals). For example, the subset of nucleotide sequences may be differentially expressed in the sampled tissues of subjects with the disease or disease criterion (e.g., a patient with a disease or disease criteria) as compared to subjects without the disease or disease criterion (e.g., patients without a disease (control patients)). Alternatively, or in addition, the subset of nucleotide sequence(s) may be differentially expressed in different samples taken from the same patient, e.g. at different points in time, at different disease stages, before and after a treatment, in the presence or absence of a risk factor, etc.

[0191] Expression profiles corresponding to sets of nucleotide sequences that correlate not with a diagnosis, but rather with a particular aspect of a disease can also be used to identify the diagnostic nucleotide sets and disease specific target nucleotide sequences of the disclosure. For example, such an aspect, or disease criterion, can relate to a subject's medical or family history, e.g., childhood illness, cause of death of a parent or other relative, prior surgery or other intervention, medications, symptoms (including onset and/or duration of symptoms), etc. Alternatively, the disease criterion can relate to a diagnosis, e.g., hypertension, diabetes, atherosclerosis, or prognosis (e.g., prediction of future diagnoses, events or complications), e.g., acute myocardial infarction, restenosis following angioplasty, reperfusion injury, allograft rejection, rheumatoid arthritis or systemic lupus erythematosus disease activity or the like. In other cases, the disease criterion corresponds to a therapeutic outcome, e.g., transplant rejection, bypass surgery or response to a medication, restenosis after stent implantation, collateral vessel growth due to therapeutic angiogenesis therapy, decreased angina due to revascularization, resolution of symptoms associated with a myriad of therapies, and the like. Alternatively, the disease criteria corresponds with previously identified or classic risk factors and may correspond to prognosis or future disease diagnosis. As indicated above, a disease criterion can also correspond to genotype for one

or more loci. Disease criteria (including patient data) may be collected (and compared) from the same patient at different points in time, from different patients, between patients with a disease (criterion) and patients representing a control population, etc. Longitudinal data, i.e., data collected at different time points from an individual (or group of individuals) may be used for comparisons of samples obtained from an individual (group of individuals) at different points in time, to permit identification of differences specifically related to the disease state, and to obtain information relating to the change in expression over time, including a rate of change or trajectory of expression over time. The usefulness of longitudinal data is further discussed in the section titled "Identification of diagnostic nucleotide sets".

[0192] It is further understood that diagnostic nucleotide sets may be developed for use in diagnosing conditions for which there is no present means of diagnosis. For example, in rheumatoid arthritis, joint destruction is often well under way before a patient experiences symptoms of the condition. A diagnostic nucleotide set may be developed that diagnoses rheumatic joint destruction at an earlier stage than would be possible using present means of diagnosis, which rely in part on the presentation of symptoms by a patient. Diagnostic nucleotide sets may also be developed to replace or augment current diagnostic procedures. For example, the use of a diagnostic nucleotide set to diagnose cardiac allograft rejection may replace the current diagnostic test, a graft biopsy.

[0193] It is understood that the following discussion of diseases is exemplary and non-limiting, and further that the general criteria discussed above, e.g. use of family medical history, are generally applicable to the specific diseases discussed below.

[0194] In addition to leukocytes, as described throughout, the general method is applicable to nucleotide sequences that are differentially expressed in any subject tissue or cell type, by the collection and assessment of samples of that tissue or cell type. However, in many cases, collection of such samples presents significant technical or medical problems given the current state of the art.

Organ transplant rejection and success

[0195] A frequent complication of organ transplantation is recognition of the transplanted organ as foreign by the immune system resulting in rejection. Diagnostic nucleotide sets can be identified and validated for monitoring organ transplant success, rejection and treatment. Medications currently exist that suppress the immune system, and thereby decrease the rate of and severity of rejection. However, these drugs also suppress the physiologic immune responses, leaving the patient susceptible to a wide variety of opportunistic infections and cancers. At present there is no easy, reliable way to diagnose transplant rejection. Organ biopsy is the preferred method, but this is expensive, painful and associated with significant risk and has inadequate sensitivity for focal rejection.

[0196] Diagnostic nucleotide sets of the present disclosure can be developed and validated for use as diagnostic tests for transplant rejection and success. It is appreciated that the methods of identifying diagnostic nucleotide sets are applicable to any organ transplant population. For example, diagnostic nucleotide sets are developed for cardiac allograft rejection and success.

[0197] In some cases, disease criteria correspond to acute stage rejection diagnosis based on organ biopsy and graded using the International Society for Heart and Lung Transplantation ("ISHLT") criteria. This grading system classifies endomyocardial biopsies on the histological level as Grade 0, 1A, 1B, 2, 3A, 3B, or 4. Grade 0 biopsies have no evidence of rejection, while each successive grade has increased severity of leukocyte infiltration and/or damage to the graft myocardial cells. It is appreciated that there is variability in the Grading systems between medical centers and pathologists and between repeated readings of the same pathologist at different times. When using the biopsy grade as a disease criterion for leukocyte gene expression correlation analysis, it may be desirable to have a single pathologist read all biopsy slides or have multiple pathologists read all slides to determine the variability in this disease criterion. It is also appreciated that cardiac biopsy, in part due to variability, is not 100% sensitive or 100% specific for diagnosing acute rejection. When using the cardiac biopsy grade as a disease criterion for the discovery of diagnostic gene sets, it may be desirable to divide patient samples into diagnostic categories based on the grades. Examples of such classes are those patients with: Grade 0 vs. Grades 1A-4, Grade 0 vs. Grades 1B-4, Grade 0 vs. Grades 2-4, Grade 0-1 vs. Grade 2-4, Grade 0-1 vs. Grade 3A-4, or Grade 0 vs. Grade 3A-4.

[0198] Other disease criteria correspond to the cardiac biopsy results and other criteria, such as the results of cardiac function testing by echocardiography, hemodynamics assessment by cardiac catheterization, CMV infection, weeks post transplant, medication regimen, demographics and/or results of other diagnostic tests.

[0199] Other disease criteria correspond to information from the patient's medical history and information regarding the organ donor. Alternatively, disease criteria include the presence or absence of cytomegalovirus (CMV) infection, Epstein-Barr virus (EBV) infection, allograft dysfunction measured by physiological tests of cardiac function (e.g., hemodynamic measurements from catheterization or echocardiograph data), and symptoms of other infections. Alternatively, disease criteria correspond to therapeutic outcome, e.g. graft failure, re-transplantation, death, hospitalization, need for intravenous immunosuppression, transplant vasculopathy, response to immunosuppressive medications, etc. Disease criteria may further correspond to a rejection episode of at least moderate histologic grade, which results in

treatment of the patient with additional corticosteroids, anti-T cell antibodies, or total lymphoid irradiation; a rejection with histologic grade 2 or higher; a rejection with histologic grade <2; the absence of histologic rejection and normal or unchanged allograft function (based on hemodynamic measurements from catheterization or on echocardiographic data); the presence of severe allograft dysfunction or worsening allograft dysfunction during the study period (based on hemodynamic measurements from catheterization or on echocardiographic data).; documented CMV infection by culture, histology, or PCR, and at least one clinical sign or symptom of infection; specific graft biopsy rejection grades; rejection of mild to moderate histologic severity prompting augmentation of the patient's chronic immunosuppressive regimen; rejection of mild to moderate severity with allograft dysfunction prompting plasmaphoresis or a diagnosis of "humoral" rejection; infections other than CMV, especially infection with Epstein Barr virus (EBV); lymphoproliferative disorder (also called post-transplant lymphoma); transplant vasculopathy diagnosed by increased intimal thickness on intravascular ultrasound (IVUS), angiography, or acute myocardial infarction; graft failure or retransplantation; and all cause mortality. Further specific examples of clinical data useful as disease criteria are provided in Example 3.

[0200] In another example, diagnostic nucleotide sets are developed and validated for use in diagnosis and monitoring of kidney allograft recipients. Disease criteria correspond to, e.g., results of biopsy analysis for kidney allograft rejection, serum creatine level, creatinine clearance, radiological imaging results for the kidney and urinalysis results. Another disease criterion corresponds to the need for hemodialysis, retransplantation, death or other renal replacement therapy. Diagnostic nucleotide sets are developed and validated for use in diagnosis and treatment of bone marrow transplant and liver transplantation patients, respectively. Disease criteria for bone marrow transplant correspond to the diagnosis and monitoring of graft rejection and/or graft versus host disease, the recurrence of cancer, complications due to immunosuppression, hematologic abnormalities, infection, hospitalization and/or death. Disease criteria for liver transplant rejection include levels of serum markers for liver damage and liver function such as AST (aspartate aminotransferase), ALT (alanine aminotransferase), Alkaline phosphatase, GGT, (gamma-glutamyl transpeptidase) Bilirubin, Albumin and Prothrombin time. Further disease criteria correspond to hepatic encephalopathy, medication usage, ascites, graft failure, retransplantation, hospitalization, complications of immunosuppression, results of diagnostic tests, results of radiological testing, death and histological rejection on graft biopsy. In addition, urine can be utilized for at the target tissue for profiling in renal transplant, while biliary and intestinal secretions and feces may be used favorably for hepatic or intestinal organ allograft rejection. Diagnostic nucleotide sets can also be discovered and developed for the diagnosis and monitoring of chronic renal allograft rejection.

[0201] In the case of renal allografts, gene expression markers may be identified that are secreted proteins. These proteins may be detected in the urine of allograft recipients using standard immunoassays. Proteins are more likely to be present in the urine if they are of low molecular weight. Lower molecular weight proteins are more likely to pass through the glomerular membrane and into the urine.

[0202] In another example, diagnostic nucleotide sets are developed and validated for use in diagnosis and treatment of xenograft recipients. This can include the transplantation of any organ from a non-human animal to a human or between non-human animals. Considerations for discovery and application of diagnostics and therapeutics and for disease criterion are substantially similar to those for allograft transplantation between humans.

[0203] In another example, diagnostic nucleotide sets are developed and validated for use in diagnosis and treatment of artificial organ recipients. This includes, but is not limited to mechanical circulatory support, artificial hearts, left ventricular assist devices, renal replacement therapies, organ prostheses and the like. Disease criteria are thrombosis (blood clots), infection, death, hospitalization, and worsening measures of organ function (e.g., hemodynamics, creatinine, liver function testing, renal function testing, functional capacity).

[0204] In another example, diagnostic nucleotide sets are developed and validated for use in matching donor organs to appropriate recipients. Diagnostic gene set can be discovered that correlate with successful matching of donor organ to recipient. Disease criteria include graft failure, acute and chronic rejection, death, hospitalization, immunosuppressive drug use, and complications of immunosuppression. Gene sets may be assayed from the donor or recipient's peripheral blood, organ tissue or some other tissue.

[0205] In another example, diagnostic nucleotide sets are developed and validated for use in diagnosis and induction of patient immune tolerance (decrease rejection of an allograft by the host immune system). Disease criteria include rejection, assays of immune activation, need for immunosuppression and all disease criteria noted above for transplantation of each organ.

Viral diseases

[0206] Diagnostic leukocyte nucleotide sets may be developed and validated for use in diagnosing viral disease, as well as diagnosing and monitoring transplant rejection. In another aspect, viral nucleotide sequences may be added to a leukocyte nucleotide set for use in diagnosis of viral diseases, as well as diagnosing and monitoring transplant rejection. Alternatively, viral nucleotide sets and leukocyte nucleotides sets may be used sequentially.

Epstein-Barr virus (EBV)

[0207] EBV causes a variety of diseases such as mononucleosis, B-cell lymphoma, and pharyngeal carcinoma. It infects mononuclear cells and circulating atypical lymphocytes are a common manifestation of infection. Peripheral leukocyte gene expression is altered by infection. Transplant recipients and patients who are immunosuppressed are at increased risk for EBV-associated lymphoma.

[0208] Diagnostic nucleotide sets may be developed and validated for use in diagnosis and monitoring of EBV, as well as diagnosing and monitoring transplant rejection. In one aspect, the diagnostic nucleotide set is a leukocyte nucleotide set. Alternatively, EBV nucleotide sequences are added to a leukocyte nucleotide set, for use in diagnosing EBV. Disease criteria correspond with diagnosis of EBV, and, in patients who are EBV-sero-positive, presence (or prospective occurrence) of EBV-related illnesses such as mononucleosis, and EBV-associated lymphoma. Diagnostic nucleotide sets are useful for diagnosis of EBV, and prediction of occurrence of EBV-related illnesses.

Cytomegalovirus (CMV)

[0209] Cytomegalovirus cause inflammation and disease in almost any tissue, particularly the colon, lung, bone marrow and retina, and is a very important cause of disease in immunosuppressed patients, e.g. transplant, cancer, AIDS. Many patients are infected with or have been exposed to CMV, but not all patients develop clinical disease from the virus. Also, CMV negative recipients of allografts that come from CMV positive donors are at high risk for CMV infection. As immunosuppressive drugs are developed and used, it is increasingly important to identify patients with current or impending clinical CMV disease, because the potential benefit of immunosuppressive therapy must be balanced with the increased rate of clinical CMV infection and disease that may result from the use of immunosuppression therapy. CMV may also play a role in the occurrence of atherosclerosis or restenosis after angioplasty. CMV expression also correlates to transplant rejection, and is useful in diagnosing and monitoring transplant rejection.

[0210] Diagnostic nucleotide sets are developed for use in diagnosis and monitoring of CMV infection or re-activation of CMV infection. In one aspect, the diagnostic nucleotide set is a leukocyte nucleotide set. In another aspect, CMV nucleotide sequences are added to a leukocyte nucleotide set, for use in diagnosing CMV. Disease criteria correspond to diagnosis of CMV (e.g., sero-positive state) and presence of clinically active CMV. Disease criteria may also correspond to prospective data, e.g. the likelihood that CMV will become clinically active or impending clinical CMV infection. Antiviral medications are available and diagnostic nucleotide sets can be used to select patients for early treatment, chronic suppression or prophylaxis of CMV activity.

Hepatitis Band C

[0211] These chronic viral infections affect about 1.25 and 2.7 million patients in the US, respectively. Many patients are infected, but suffer no clinical manifestations. Some patients with infection go on to suffer from chronic liver failure, cirrhosis and hepatic carcinoma.

[0212] Diagnostic nucleotide sets are developed for use in diagnosis and monitoring of HBV or HCV infection. In one aspect, the diagnostic nucleotide set is a leukocyte nucleotide set. In another aspect, viral nucleotide sequences are added to a leukocyte nucleotide set, for use in diagnosing the virus and monitoring progression of liver disease. Disease criteria correspond to diagnosis of the virus (e.g., sero-positive state or other disease symptoms). Alternatively, disease criteria correspond to liver damage, e.g., elevated alkaline phosphatase, ALT, AST or evidence of ongoing hepatic damage on liver biopsy. Alternatively, disease criteria correspond to serum liver tests (AST, ALT, Alkaline Phosphatase, GGT, PT, bilirubin), liver biopsy, liver ultrasound, viral load by serum PCR, cirrhosis, hepatic cancer, need for hospitalization or listing for liver transplant. Diagnostic nucleotide sets are used to diagnose HBV and HCV, and to predict likelihood of disease progression. Antiviral therapeutic usage, such as Interferon gamma and Ribavirin, can also be disease criteria.

HIV

[0213] HIV infects T cells and certainly causes alterations in leukocyte expression. Diagnostic nucleotide sets are developed for diagnosis and monitoring of HIV. In one aspect, the diagnostic nucleotide set is a leukocyte nucleotide set. In another aspect, viral nucleotide sequences are added to a leukocyte nucleotide set, for use in diagnosing the virus. Disease criteria correspond to diagnosis of the virus (e.g., sero-positive state). In addition, disease criteria correspond to viral load, CD4 T cell counts, opportunistic infection, response to antiretroviral therapy, progression to AIDS, rate of progression and the occurrence of other HIV related outcomes (e.g., malignancy, CNS disturbance). Response to antiretrovirals may also be disease criteria.

Pharmacogenomics

[0214] Pharmacogenomics is the study of the individual propensity to respond to a particular drug therapy (combination of therapies). In this context, response can mean whether a particular drug will work on a particular patient, e.g. some patients respond to one drug but not to another drug. Response can also refer to the likelihood of successful treatment or the assessment of progress in treatment. Titration of drug therapy to a particular patient is also included in this description, e.g. different patients can respond to different doses of a given medication. This aspect may be important when drugs with side-effects or interactions with other drug therapies are contemplated.

[0215] Diagnostic nucleotide sets are developed and validated for use in assessing whether a patient will respond to a particular therapy and/or monitoring response of a patient to drug therapy(therapies). Disease criteria correspond to presence or absence of clinical symptoms or clinical endpoints, presence of side-effects or interaction with other drug (s). The diagnostic nucleotide set may further comprise nucleotide sequences that are targets of drug treatment or markers of active disease.

Validation and accuracy of diagnostic nucleotide sets

[0216] Prior to widespread application of the diagnostic probe sets of the disclosure the predictive value of the probe set is validated. When the diagnostic probe set is discovered by microarray based expression analysis, the differential expression of the member genes may be validated by a less variable and more quantitative and accurate technology such as real time PCR. In this type of experiment the amplification product is measured during the PCR reaction. This enables the researcher to observe the amplification before any reagent becomes rate limiting for amplification. In kinetic PCR the measurement is of C_T (threshold cycle) or C_P (crossing point). This measurement ($C_T=C_P$) is the point at which an amplification curve crosses a threshold fluorescence value. The threshold is set to a point within the area where all of the reactions were in their linear phase of amplification. When measuring C_T , a lower C_T value is indicative of a higher amount of starting material since an earlier cycle number means the threshold was crossed more quickly.

[0217] Several fluorescence methodologies are available to measure amplification product in real-time PCR. Taqman (Applied BioSystems, Foster City, CA) uses fluorescence resonance energy transfer (FRET) to inhibit signal from a probe until the probe is degraded by the sequence specific binding and Taq 3' exonuclease activity. Molecular Beacons (Stratagene, La Jolla, CA) also use FRET technology, whereby the fluorescence is measured when a hairpin structure is relaxed by the specific probe binding to the amplified DNA. The third commonly used chemistry is Sybr Green, a DNA-binding dye (Molecular Probes, Eugene, OR). The more amplified product that is produced, the higher the signal. The Sybr Green method is sensitive to non-specific amplification products, increasing the importance of primer design and selection. Other detection chemistries can also been used, such as ethidium bromide or other DNA-binding dyes and many modifications of the fluorescent dye/quencher dye Taqman chemistry, for example scorpions.

[0218] Real-time PCR validation can be done as described in Example 12.

[0219] Typically, the oligonucleotide sequence of each probe is confirmed, e.g. by DNA sequencing using an oligonucleotide-specific primer. Partial sequence obtained is generally sufficient to confirm the identity of the oligonucleotide probe. Alternatively, a complementary polynucleotide is fluorescently labeled and hybridized to the array, or to a different array containing a resynthesized version of the oligo nucleotide probe, and detection of the correct probe is confirmed.

[0220] Typically, validation is performed by statistically evaluating the accuracy of the correspondence between the molecular signature for a diagnostic probe set and a selected indicator. For example, the expression differential for a nucleotide sequence between two subject classes can be expressed as a simple ratio of relative expression. The expression of the nucleotide sequence in subjects with selected indicator can be compared to the expression of that nucleotide sequence in subjects without the indicator, as described in the following equations.

$$\sum E_x a_i / N = E_x A \quad \text{the average expression of nucleotide sequence x in the members of group A;}$$

$$\sum E_x b_i / M = E_x B \quad \text{the average expression of nucleotide sequence x in the members of group B;}$$

$$E_x A / E_x B = \Delta E_x AB \quad \text{the average differential expression of nucleotide sequence x between groups A and B:}$$

where Σ indicates a sum; E_x is the expression of nucleotide sequence x relative to a standard; a_i are the individual members of group A, group A has N members; b_i are the individual members of group B, group B has M members.

[0221] The expression of at least two nucleotide sequences, e.g., nucleotide sequence X and nucleotide sequence Y

are measured relative to a standard in at least one subject of group A (e.g., with a disease) and group B (e.g., without the disease). Ideally, for purposes of validation the indicator is independent from (i.e., not assigned based upon) the expression pattern. Alternatively, a minimum threshold of gene expression for nucleotide sequences X and Y, relative to the standard, are designated for assignment to group A. For nucleotide sequence x, this threshold is designated ΔEx , and for nucleotide sequence y, the threshold is designated ΔEy .

[0222] The following formulas are used in the calculations below:

$$\text{Sensitivity} = (\text{true positives} / (\text{true positives} + \text{false negatives}))$$

$$\text{Specificity} = (\text{true negatives} / (\text{true negatives} + \text{false positives}))$$

[0223] If, for example, expression of nucleotide sequence x above a threshold: $x > \Delta Ex$, is observed for 80/100 subjects in group A and for 10/100 subjects in group B, the sensitivity of nucleotide sequence x for the assignment to group A, at the given expression threshold ΔEx , is 80%, and the specificity is 90%.

[0224] If the expression of nucleotide sequence y is $> \Delta Ey$ in 80/100 subjects in group A, and in 10/100 subjects in group B, then, similarly the sensitivity of nucleotide sequence y for the assignment to group A at the given threshold ΔEy is 80% and the specificity is 90%. If in addition, 60 of the 80 subjects in group A that meet the expression threshold for nucleotide sequence y also meet the expression threshold ΔEx and that 5 of the 10 subjects in group B that meet the expression threshold for nucleotide sequence y also meet the expression threshold ΔEx , the sensitivity of the test ($x > \Delta Ex$ and $y > \Delta Ey$) for assignment of subjects to group A is 60% and the specificity is 95%.

[0225] Alternatively, if the criteria for assignment to group A are change to: Expression of $x > \Delta Ex$ or expression of $y > \Delta Ey$, the sensitivity approaches 100% and the specificity is 85%.

[0226] Clearly, the predictive accuracy of any diagnostic probe set is dependent on the minimum expression threshold selected. The expression of nucleotide sequence X (relative to a standard) is measured in subjects of groups A (with disease) and B (without disease). The minimum threshold of nucleotide sequence expression for x, required for assignment to group A is designated ΔEx 1.

[0227] If 90/100 patients in group A have expression of nucleotide sequence $x > \Delta Ex$ 1 and 20/100 patients in group B have expression of nucleotide sequence $x > \Delta Ex$ 1, then the sensitivity of the expression of nucleotide sequence x (using ΔEx 1 as a minimum expression threshold) for assignment of patients to group A will be 90% and the specificity will be 80%.

[0228] Altering the minimum expression threshold results in an alteration in the specificity and sensitivity of the nucleotide sequences in question. For example, if the minimum expression threshold of nucleotide sequence x for assignment of subjects to group A is lowered to ΔEx 2, such that 100/100 subjects in group A and 40/100 subjects in group B meet the threshold, then the sensitivity of the test for assignment of subjects to group A will be 100% and the specificity will be 60%.

[0229] Thus, for 2 nucleotide sequences X and Y: the expression of nucleotide sequence x and nucleotide sequence y (relative to a standard) are measured in subjects belonging to groups A (with disease) and B (without disease). Minimum thresholds of nucleotide sequence expression for nucleotide sequences X and Y (relative to common standards) are designated for assignment to group A. For nucleotide sequence x, this threshold is designated ΔEx 1 and for nucleotide sequence y, this threshold is designated ΔEy 1.

[0230] If in group A, 90/100 patients meet the minimum requirements of expression ΔEx 1 and ΔEy 1, and in group B, 10/100 subjects meet the minimum requirements of expression ΔEx 1 and ΔEy 1, then the sensitivity of the test for assignment of subjects to group A is 90% and the specificity is 90%.

[0231] Increasing the minimum expression thresholds for X and Y to ΔEx 2 and ΔEy 2, such that in group A, 70/100 subjects meet the minimum requirements of expression ΔEx 2 and ΔEy 2, and in group B, 3/100 subjects meet the minimum requirements of expression ΔEx 2 and ΔEy 2. Now the sensitivity of the test for assignment of subjects to group A is 70% and the specificity is 97%.

[0232] If the criteria for assignment to group A is that the subject in question meets either threshold, ΔEx 2 or ΔEy 2, and it is found that 100/100 subjects in group A meet the criteria and 20/100 subjects in group B meet the criteria, then the sensitivity of the test for assignment to group A is 100% and the specificity is 80%.

[0233] Individual components of a diagnostic probe set each have a defined sensitivity and specificity for distinguishing between subject groups. Such individual nucleotide sequences can be employed in concert as a diagnostic probe set to increase the sensitivity and specificity of the evaluation. The database of molecular signatures is queried by algorithms to identify the set of nucleotide sequences (i.e., corresponding to members of the probe set) with the highest average differential expression between subject groups. Typically, as the number of nucleotide sequences in the diagnostic probe set increases, so does the predictive value, that is, the sensitivity and specificity of the probe set. When the probe sets

are defined they may be used for diagnosis and patient monitoring as discussed below. The diagnostic sensitivity and specificity of the probe sets for the defined use can be determined for a given probe set with specified expression levels as demonstrated above. By altering the expression threshold required for the use of each nucleotide sequence as a diagnostic, the sensitivity and specificity of the probe set can be altered by the practitioner. For example, by lowering the magnitude of the expression differential threshold for each nucleotide sequence in the set, the sensitivity of the test will increase, but the specificity will decrease. As is apparent from the foregoing discussion, sensitivity and specificity are inversely related and the predictive accuracy of the probe set is continuous and dependent on the expression threshold set for each nucleotide sequence. Although sensitivity and specificity tend to have an inverse relationship when expression thresholds are altered, both parameters can be increased as nucleotide sequences with predictive value are added to the diagnostic nucleotide set. In addition a single or a few markers may not be reliable expression markers across a population of patients. This is because of the variability in expression and measurement of expression that exists between measurements, individuals and individuals over time. Inclusion of a large number of candidate nucleotide sequences or large numbers of nucleotide sequences in a diagnostic nucleotide set allows for this variability as not all nucleotide sequences need to meet a threshold for diagnosis. Generally, more markers are better than a single marker. If many markers are used to make a diagnosis, the likelihood that all expression markers will not meet some thresholds based upon random variability is low and thus the test will give fewer false negatives.

[0234] It is appreciated that the desired diagnostic sensitivity and specificity of the diagnostic nucleotide set may vary depending on the intended use of the set. For example, in certain uses, high specificity and high sensitivity are desired. For example, a diagnostic nucleotide set for predicting which patient population may experience side effects may require high sensitivity so as to avoid treating such patients. In other settings, high sensitivity is desired, while reduced specificity may be tolerated. For example, in the case of a beneficial treatment with few side effects, it may be important to identify as many patients as possible (high sensitivity) who will respond to the drug, and treatment of some patients who will not respond is tolerated. In other settings, high specificity is desired and reduced sensitivity may be tolerated. For example, when identifying patients for an early-phase clinical trial, it is important to identify patients who may respond to the particular treatment. Lower sensitivity is tolerated in this setting as it merely results in reduced patients who enroll in the study or requires that more patients are screened for enrollment.

Methods of using diagnostic nucleotide sets.

[0235] The disclosure also provide methods of using the diagnostic nucleotide sets to: diagnose disease; assess severity of disease; predict future occurrence of disease; predict future complications of disease; determine disease prognosis; evaluate the patient's risk, or "stratify" a group of patients; assess response to current drug therapy; assess response to current non-pharmacological therapy; determine the most appropriate medication or treatment for the patient; predict whether a patient is likely to respond to a particular drug; and determine most appropriate additional diagnostic testing for the patient, among other clinically and epidemiologically relevant applications.

[0236] The nucleotide sets of the disclosure can be utilized for a variety of purposes by physicians, healthcare workers, hospitals, laboratories, patients, companies and other institutions. As indicated previously, essentially any disease, condition, or status for which at least one nucleotide sequence is differentially expressed in leukocyte populations (or sub-populations) can be evaluated, e.g., diagnosed, monitored, etc. using the diagnostic nucleotide sets and methods of the disclosure. In addition to assessing health status at an individual level, the diagnostic nucleotide sets of the present disclosure are suitable for evaluating subjects at a "population level," e.g., for epidemiological studies, or for population screening for a condition or disease.

Collection and preparation of sample

[0237] RNA, protein and/or DNA is prepared using methods well-known in the art, as further described herein. It is appreciated that subject samples collected for use in the methods of the disclosure are generally collected in a clinical setting, where delays may be introduced before RNA samples are prepared from the subject samples of whole blood, e.g. the blood sample may not be promptly delivered to the clinical lab for further processing. Further delay may be introduced in the clinical lab setting where multiple samples are generally being processed at any given time. For this reason, methods which feature lengthy incubations of intact leukocytes at room temperature are not preferred, because the expression profile of the leukocytes may change during this extended time period. For example, RNA can be isolated from whole blood using a phenol/guanidine isothiocyanate reagent or another direct whole-blood lysis method, as described in, e.g., U.S. Patent Nos. 5,346,994 and 4,843,155. This method may be less preferred under certain circumstances because the large majority of the RNA recovered from whole blood RNA extraction comes from erythrocytes since these cells outnumber leukocytes 1000:1. Care must be taken to ensure that the presence of erythrocyte RNA and protein does not introduce bias in the RNA expression profile data or lead to inadequate sensitivity or specificity of probes.

[0238] Alternatively, intact leukocytes may be collected from whole blood using a lysis buffer that selectively lyses erythrocytes, but not leukocytes, as described, e.g., in (U.S. Patent Nos. 5,973,137, and 6,020,186). Intact leukocytes are then collected by centrifugation, and leukocyte RNA is isolated using standard protocols, as described herein. However, this method does not allow isolation of sub-populations of leukocytes, e.g. mononuclear cells, which may be desired. In addition, the expression profile may change during the lengthy incubation in lysis buffer, especially in a busy clinical lab where large numbers of samples are being prepared at any given time.

[0239] Alternatively, specific leukocyte cell types can be separated using density gradient reagents (Boyum, A, 1968.). For example, mononuclear cells may be separated from whole blood using density gradient centrifugation, as described, e.g., in U.S. Patents Nos. 4190535, 4350593, 4751001, 4818418, and 5053134. Blood is drawn directly into a tube containing an anticoagulant and a density reagent (such as Ficoll or Percoll). Centrifugation of this tube results in separation of blood into an erythrocyte and granulocyte layer, a mononuclear cell suspension, and a plasma layer. The mononuclear cell layer is easily removed and the cells can be collected by centrifugation, lysed, and frozen. Frozen samples are stable until RNA can be isolated. Density centrifugation, however, must be conducted at room temperature, and if processing is unduly lengthy, such as in a busy clinical lab, the expression profile may change.

[0240] Alternatively, cells can be separated using fluorescence activated cell sorting (FACS) or some other technique, which divides cells into subsets based on gene or protein expression. This may be desirable to enrich the sample for cells of interest, but it may also introduce cell manipulations and time delays, which result in alteration of gene expression profiles (Cantor et al. 1975; Galbraith et al. 1999).

[0241] The quality and quantity of each clinical RNA sample is desirably checked before amplification and labeling for array hybridization, using methods known in the art. For example, one microliter of each sample may be analyzed on a Bioanalyzer (Agilent 2100 Palo Alto, CA. USA) using an RNA 6000 nano LabChip (Caliper, Mountain View, CA. USA). Degraded RNA is identified by the reduction of the 28S to 18S ribosomal RNA ratio and/or the presence of large quantities of RNA in the 25-100 nucleotide range.

[0242] It is appreciated that the RNA sample for use with a diagnostic nucleotide set may be produced from the same or a different cell population, sub-population and/or cell type as used to identify the diagnostic nucleotide set. For example, a diagnostic nucleotide set identified using RNA extracted from mononuclear cells may be suitable for analysis of RNA extracted from whole blood or mononuclear cells, depending on the particular characteristics of the members of the diagnostic nucleotide set. Generally, diagnostic nucleotide sets must be tested and validated when used with RNA derived from a different cell population, sub-population or cell type than that used when obtaining the diagnostic gene set. Factors such as the cell-specific gene expression of diagnostic nucleotide set members, redundancy of the information provided by members of the diagnostic nucleotide set, expression level of the member of the diagnostic nucleotide set, and cell-specific alteration of expression of a member of the diagnostic nucleotide set will contribute to the usefulness of using a different RNA source than that used when identifying the members of the diagnostic nucleotide set. It is appreciated that it may be desirable to assay RNA derived from whole blood, obviating the need to isolate particular cell types from the blood.

Rapid method of RNA extraction suitable for production in a clinical setting of high quality RNA for expression profiling

[0243] In a clinical setting, obtaining high quality RNA preparations suitable for expression profiling, from a desired population of leukocytes poses certain technical challenges, including: the lack of capacity for rapid, high-throughput sample processing in the clinical setting, and the possibility that delay in processing (in a busy lab or in the clinical setting) may adversely affect RNA quality, e.g. by permitting the expression profile of certain nucleotide sequences to shift. Also, use of toxic and expensive reagents, such as phenol, may be disfavored in the clinical setting due to the added expense associated with shipping and handling such reagents.

[0244] A useful method for RNA isolation for leukocyte expression profiling would allow the isolation of monocyte and lymphocyte RNA in a timely manner, while preserving the expression profiles of the cells, and allowing inexpensive production of reproducible high-quality RNA samples. Accordingly, the invention provides a method of adding inhibitor (s) of RNA transcription and/or inhibitor(s) of protein synthesis, such that the expression profile is "frozen" and RNA degradation is reduced. A desired leukocyte population or sub-population is then isolated, and the sample may be frozen or lysed before further processing to extract the RNA. Blood is drawn from subject population and exposed to ActinomycinD (to a final concentration of 10 ug/ml) to inhibit transcription, and cycloheximide (to a final concentration of 10 ug/ml) to inhibit protein synthesis. The inhibitor(s) can be injected into the blood collection tube in liquid form as soon as the blood is drawn, or the tube can be manufactured to contain either lyophilized inhibitors or inhibitors that are in solution with the anticoagulant. At this point, the blood sample can be stored at room temperature until the desired leukocyte population or sub-population is isolated, as described elsewhere. RNA is isolated using standard methods, e.g., as described above, or a cell pellet or extract can be frozen until further processing of RNA is convenient.

[0245] Also disclosed is a method of using a low-temperature density gradient for separation of a desired leukocyte sampler as well as the combination of use of a low-temperature density gradient and the use of transcriptional and/or

protein synthesis inhibitor(s). A desired leukocyte population is separated using a density gradient solution for cell separation that maintains the required density and viscosity for cell separation at 0-4°C. Blood is drawn into a tube containing this solution and may be refrigerated before and during processing as the low temperatures slow cellular processes and minimize expression profile changes. Leukocytes are separated, and RNA is isolated using standard methods. Alternately, a cell pellet or extract is frozen until further processing of RNA is convenient. Care must be taken to avoid rewarming the sample during further processing steps.

[0246] Alternatively, a method of using low-temperature density gradient separation, combined with the use of actinomycin A and cyclohexamide, as described above can be used.

Assessing expression for diagnostics

[0247] Expression profiles for the set of diagnostic nucleotide sequences in a subject sample can be evaluated by any technique that determines the expression of each component nucleotide sequence. Methods suitable for expression analysis are known in the art, and numerous examples are discussed in the Sections titled "Methods of obtaining expression data" and "high throughput expression Assays", above.

[0248] In many cases, evaluation of expression profiles is most efficiently, and cost effectively, performed by analyzing RNA expression. Alternatively, the proteins encoded by each component of the diagnostic nucleotide set are detected for diagnostic purposes by any technique capable of determining protein expression, e.g., as described above. Expression profiles can be assessed in subject leukocyte sample using the same or different techniques as those used to identify and validate the diagnostic nucleotide set. For example, a diagnostic nucleotide set identified as a subset of sequences on a cDNA microarray can be utilized for diagnostic (or prognostic, or monitoring, etc.) purposes on the same array from which they were identified. Alternatively, the diagnostic nucleotide sets for a given disease or condition can be organized onto a dedicated sub-array for the indicated purpose. It is important to note that if diagnostic nucleotide sets are discovered using one technology, e.g. RNA expression profiling, but applied as a diagnostic using another technology, e.g. protein expression profiling, the nucleotide sets must generally be validated for diagnostic purposes with the new technology. In addition, it is appreciated that diagnostic nucleotide sets that are developed for one use, e.g. to diagnose a particular disease, may later be found to be useful for a different application, e.g. to predict the likelihood that the particular disease will occur. Generally, the diagnostic nucleotide set will need to be validated for use in the second circumstance. As discussed herein, the sequence of diagnostic nucleotide set members may be amplified from RNA or cDNA using methods known in the art providing specific amplification of the nucleotide sequences.

General Protein Methods

[0249] Protein products of the nucleotide sequences disclosed in the invention may include proteins that represent functionally equivalent gene products. Such an equivalent gene product may contain deletions, additions or substitutions of amino acid residues within the amino acid sequence encoded by the nucleotide sequences described, above, but which result in a silent change, thus producing a functionally equivalent nucleotide sequence product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

[0250] For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Functionally equivalent", as utilized herein, refers to a protein capable of exhibiting a substantially similar in vivo activity as the endogenous gene products encoded by the nucleotide described, above.

[0251] The gene products (protein products of the nucleotide sequences) may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the gene polypeptides and peptides of the invention by expressing nucleic acid encoding nucleotide sequences are described herein. Methods which are well known to those skilled in the art can be used to construct expression vectors containing nucleotide sequence protein coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination/genetic recombination. See, for example, the techniques described in Sambrook et al., 1989, supra, and Ausubel et al., 1989, supra. Alternatively, RNA capable of encoding nucleotide sequence protein sequences may be chemically synthesized using, for example, synthesizers. See, for example, the techniques described in "Oligonucleotide Synthesis", 1984, Gait, M. J. ed., IRL Press, Oxford.

[0252] A variety of host-expression vector systems may be utilized to express the nucleotide sequence coding sequences of the disclosure. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells which may, when transformed or transfected with

the appropriate nucleotide coding sequences, exhibit the protein encoded by the nucleotide sequence of the disclosure in situ. These include but are not limited to microorganisms such as bacteria (e.g., *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing nucleotide sequence protein coding sequences; yeast (e.g. *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing the nucleotide sequence protein coding sequences; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the nucleotide sequence protein coding sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing nucleotide sequence protein coding sequences; or mammalian cell systems (e.g. COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5 K promoter).

[0253] In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the nucleotide sequence protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, EMBO J. 2:1791), in which the nucleotide sequence protein coding sequence may be ligated individually into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the likes of pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target nucleotide sequence protein can be released from the GST moiety. Other systems useful in the disclosure include use of the FLAG epitope or the 6-HIS systems.

[0254] In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign nucleotide sequences. The virus grows in *Spodoptera frugiperda* cells. The nucleotide sequence coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of nucleotide sequence coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted nucleotide sequence is expressed. (E.g., see Smith et al., 1983, J. Virol. 46: 584; Smith, U.S. Pat. No. 4,215,051).

[0255] In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the nucleotide sequence coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric nucleotide sequence may then be inserted in the adenovirus genome by in vitro or in vivo recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing nucleotide sequence encoded protein in infected hosts. (E.g., See Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted nucleotide sequence coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire nucleotide sequence, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the nucleotide sequence coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner et al., 1987, Methods in Enzymol. 153:516-544).

[0256] In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the product of the nucleotide sequence in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

[0257] For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the nucleotide sequence encoded protein may be engineered. Rather than using expression

vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express nucleotide sequence encoded protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the nucleotide sequence encoded protein.

[0258] A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, et al., 1977, Cell 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc. Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy, et al., 1980, Cell 22:817) genes can be employed in tk-, hgp^rt- or ap^rt- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, et al., 1980, Natl. Acad. Sci. USA 77:3567; O'Hare, et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, et al., 1981, J. Mol. Biol. 150:1); and hyg^r, which confers resistance to hygromycin (Santerre, et al., 1984, Gene 30:147) genes.

[0259] An alternative fusion protein system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, et al., 1991, Proc. Natl. Acad. Sci. USA 88: 8972-8976). In this system, the nucleotide sequence of interest is subcloned into a vaccinia recombination plasmid such that the nucleotide sequence's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni.sup.2 +-nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

[0260] Where recombinant DNA technology is used to produce the protein encoded by the nucleotide sequence for such assay systems, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection.

Antibodies

[0261] Indirect labeling involves the use of a protein, such as a labeled antibody, which specifically binds to the protein encoded by the nucleotide sequence. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by an Fab expression library.

[0262] Antibodies to the protein encoded by the nucleotide sequences are also disclosed. Described herein are methods for the production of antibodies capable of specifically recognizing one or more nucleotide sequence epitopes. Such antibodies may include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. Such antibodies may be used, for example, in the detection of a nucleotide sequence in a biological sample, or, alternatively, as a method for the inhibition of abnormal gene activity, for example, the inhibition of a disease target nucleotide sequence, as further described below. Thus, such antibodies may be utilized as part of cardiovascular or other disease treatment method, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels of nucleotide sequence encoded proteins, or for the presence of abnormal forms of the such proteins.

[0263] For the production of antibodies to a nucleotide sequence, various host animals may be immunized by injection with a protein encoded by the nucleotide sequence, or a portion thereof. Such host animals may include but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and Corynebacterium parvum.

[0264] Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

[0265] Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to the hybridoma technique of Kohler and Milstein, (1975, Nature 256:495-497; and U.S. Pat. No. 4,376,110), the human B-cell hybridoma technique (Kosbor et al., 1983, Immunology Today 4:72; Cole et

al., 1983, Proc. Natl. Acad. Sci. USA 80:2026-2030), and the EBV-hybridoma technique (Cole et al., 1985, Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb may be cultivated in vitro or in vivo.

[0266] In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci., 81:6851-6855; Neuberger et al., 1984, Nature, 312:604-608; Takeda et al., 1985, Nature, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

[0267] Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, 1988, Science 242:423-426; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward et al., 1989, Nature 334:544-546) can be adapted to produce nucleotide sequence-single chain antibodies. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

[0268] Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')₂ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., 1989, Science, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Disease Specific target nucleotide sequences

[0269] Disease specific target nucleotide sequences, and sets of disease specific target nucleotide sequences are also disclosed. The diagnostic nucleotide sets, subsets thereof, novel nucleotide sequences, and individual members of the diagnostic nucleotide sets identified as described above are also disease specific target nucleotide sequences. In particular, individual nucleotide sequences that are differentially regulated or have predictive value that is strongly correlated with a disease or disease criterion are especially favorable as disease specific target nucleotide sequences. Sets of genes that are co-regulated may also be identified as disease specific target nucleotide sets. Such nucleotide sequences and/or nucleotide sequence products are targets for modulation by a variety of agents and techniques. For example, disease specific target nucleotide sequences (or the products of such nucleotide sequences, or sets of disease specific target nucleotide sequences) can be inhibited or activated by, e.g., target specific monoclonal antibodies or small molecule inhibitors, or delivery of the nucleotide sequence or gene product of the nucleotide sequence to patients. Also, sets of genes can be inhibited or activated by a variety of agents and techniques. The specific usefulness of the target nucleotide sequence(s) depends on the subject groups from which they were discovered, and the disease or disease criterion with which they correlate.

Imaging

[0270] Further disclosed are imaging reagents. The differentially expressed leukocyte nucleotide sequences, diagnostic nucleotide sets, or portions thereof, and novel nucleotide sequences of the invention are nucleotide sequences expressed in cells with or without disease. Leukocytes expressing a nucleotide sequence(s) that is differentially expressed in a disease condition may localize within the body to sites that are of interest for imaging purposes. For example, a leukocyte expressing a nucleotide sequence(s) that are differentially expressed in an individual having atherosclerosis may localize or accumulate at the site of an atherosclerotic plaque. Such leukocytes, when labeled, may provide a detection reagent for use in imaging regions of the body where labeled leukocyte accumulate or localize, for example, at the atherosclerotic plaque in the case of atherosclerosis. For example, leukocytes are collected from a subject, labeled in vitro, and reintroduced into a subject. Alternatively, the labeled reagent is introduced into the subject individual, and leukocyte labeling occurs within the patient.

[0271] Imaging agents that detect the imaging targets of the invention are produced by well-known molecular and immunological methods (for exemplary protocols, see, e.g., Ausubel, Berger, and Sambrook, as well as Harlow and Lane, *supra*).

[0272] For example, a full-length nucleic acid sequence, or alternatively, a gene fragment encoding an immunogenic peptide or polypeptide fragments, is cloned into a convenient expression vector, for example, a vector including an in-frame epitope or substrate binding tag to facilitate subsequent purification. Protein is then expressed from the cloned cDNA sequence and used to generate antibodies, or other specific binding molecules, to one or more antigens of the imaging target protein. Alternatively, a natural or synthetic polypeptide (or peptide) or small molecule that specifically binds (or is specifically bound to) the expressed imaging target can be identified through well established techniques

(see, e.g., Mendel et al. (2000) *Anticancer Drug Des* 15:29-41; Wilson (2000) *Curr Med Chem* 7:73-98; Hamby and Showalter (1999) *Pharmacol Ther* 82:169-93; and Shimazawa et al. (1998) *Curr Opin Struct Biol* 8:451-8). The binding molecule, e.g., antibody, small molecule ligand, etc., is labeled with a contrast agent or other detectable label, e.g., gadolinium, iodine, or a gamma-emitting source. For in-vivo imaging of a disease process that involved leukocytes, the labeled antibody is infused into a subject, e.g., a human patient or animal subject, and a sufficient period of time is passed to permit binding of the antibody to target cells. The subject is then imaged with appropriate technology such as MRI (when the label is gadolinium) or with a gamma counter (when the label is a gamma emitter).

Identification of nucleotide sequence involved in leukocyte adhesion

[0273] A method of identifying nucleotide sequences involved in leukocyte adhesion is described. The interaction between the endothelial cell and leukocyte is a fundamental mechanism of all inflammatory disorders, including the diagnosis and prognosis of allograft rejection the diseases listed in Table 1. For example, the first visible abnormality in atherosclerosis is the adhesion to the endothelium and diapedesis of mononuclear cells (e.g., T-cell and monocyte). Insults to the endothelium (for example, cytokines, tobacco, diabetes, hypertension and many more) lead to endothelial cell activation. The endothelium then expresses adhesion molecules, which have counter receptors on mononuclear cells. Once the leukocyte receptors have bound the endothelial adhesion molecules, they stick to the endothelium, roll a short distance, stop and transmigrate across the endothelium. A similar set of events occurs in both acute and chronic inflammation. When the leukocyte binds the endothelial adhesion molecule, or to soluble cytokines secreted by endothelial or other cells, a program of gene expression is activated in the leukocyte. This program of expression leads to leukocyte rolling, firm adhesion and transmigration into the vessel wall or tissue parenchyma. Inhibition of this process is highly desirable goal in anti-inflammatory drug development. In addition, leukocyte nucleotide sequences and epithelial cell nucleotide sequences, that are differentially expressed during this process may be disease-specific target nucleotide sequences.

[0274] Human endothelial cells, e.g. derived from human coronary arteries, human aorta, human pulmonary artery, human umbilical vein or microvascular endothelial cells, are cultured as a confluent monolayer, using standard methods. Some of the endothelial cells are then exposed to cytokines or another activating stimuli such as oxidized LDL, hyperglycemia, shear stress, or hypoxia (Moser et al. 1992). Some endothelial cells are not exposed to such stimuli and serve as controls. For example, the endothelial cell monolayer is incubated with culture medium containing 5 U/ml of human recombinant IL-1alpha or 10 ng/ml TNF (tumor necrosis factor), for a period of minutes to overnight. The culture medium composition is changed or the flask is sealed to induce hypoxia. In addition, tissue culture plate is rotated to induce sheer stress.

[0275] Human T-cells and/or monocytes are cultured in tissue culture flasks or plates, with LGM-3 media from Clonetics. Cells are incubated at 37 degree C, 5% CO₂ and 95% humidity. These leukocytes are exposed to the activated or control endothelial layer by adding a suspension of leukocytes on to the endothelial cell monolayer. The endothelial cell monolayer is cultured on a tissue culture treated plate/ flask or on a microporous membrane. After a variable duration of exposures, the endothelial cells and leukocytes are harvested separately by treating all cells with trypsin and then sorting the endothelial cells from the leukocytes by magnetic affinity reagents to an endothelial cell specific marker such as PECAM-1 (Stem Cell Technologies). RNA is extracted from the isolated cells by standard techniques. Leukocyte RNA is labeled as described above, and hybridized to leukocyte candidate nucleotide library. Epithelial cell RNA is also labeled and hybridized to the leukocyte candidate nucleotide library. Alternatively, the epithelial cell RNA is hybridized to a epithelial cell candidate nucleotide library, prepared according to the methods described for leukocyte candidate libraries, above.

[0276] Hybridization to candidate nucleotide libraries will reveal nucleotide sequences that are upregulated or down-regulated in leukocyte and/or epithelial cells undergoing adhesion. The differentially regulated nucleotide sequences are further characterized, e.g. by isolating and sequencing the full-length sequence, analysis of the DNA and predicted protein sequence, and functional characterization of the protein product of the nucleotide sequence, as described above. Further characterization may result in the identification of leukocyte adhesion specific target nucleotide sequences, which may be candidate targets for regulation of the inflammatory process. Small molecule or antibody inhibitors can be developed to inhibit the target nucleotide sequence function. Such inhibitors are tested for their ability to inhibit leukocyte adhesion in the in vitro test described above.

Integrated systems

[0277] Integrated systems for the collection and analysis of expression profiles, and molecular signatures, as well as for the compilation, storage and access of the databases of the disclosure, typically include a digital computer with software including an instruction set for sequence searching and analysis, and, optionally, high-throughput liquid control software, image analysis software, data interpretation software, a robotic control armature for transferring solutions from a source to a destination (such as a detection device) operably linked to the digital computer, an input device (e.g., a

computer keyboard) for entering subject data to the digital computer, or to control analysis operations or high throughput sample transfer by the robotic control armature. Optionally, the integrated system further comprises an image scanner for digitizing label signals from labeled assay components, e.g., labeled nucleic acid hybridized to a candidate library microarray. The image scanner can interface with image analysis software to provide a measurement of the presence or intensity of the hybridized label, i.e., indicative of an on/off expression pattern or an increase or decrease in expression.

[0278] Readily available computational hardware resources using standard operating systems are fully adequate, e.g., a PC (Intel x86 or Pentium chip- compatible DOS,TM OS2,TM WINDOWS,TM WINDOWS NT,TM WINDOWS95,TM WINDOWS98,TM LINUX, or even Macintosh, Sun or PCs will suffice) for use in the integrated systems of the disclosure. Current art in software technology is similarly adequate (i.e., there are a multitude of mature programming languages and source code suppliers) for design, e.g., of an upgradeable open-architecture object-oriented heuristic algorithm, or instruction set for expression analysis, as described herein. For example, software for aligning or otherwise manipulating , molecular signatures can be constructed by one of skill using a standard programming language such as Visual basic, Fortran, Basic, Java, or the like, according to the methods herein.

[0279] Various methods and algorithms, including genetic algorithms and neural networks, can be used to perform the data collection, correlation, and storage functions, as well as other desirable functions, as described herein. In addition, digital or analog systems such as digital or analog computer systems can control a variety of other functions such as the display and/or control of input and output files.

[0280] For example, standard desktop applications such as word processing software (e.g., Corel WordPerfectTM or Microsoft WordTM) and database software (e.g., spreadsheet software such as Corel Quattro ProTM, Microsoft ExcelTM, or database programs such as Microsoft AccessTM or ParadoxTM) can be adapted to the present invention by inputting one or more character string corresponding, e.g., to an expression pattern or profile, subject medical or historical data, molecular signature, or the like, into the software which is loaded into the memory of a digital system, and carrying out the operations indicated in an instruction set, e.g., as exemplified in Figure 2. For example, systems can include the foregoing software having the appropriate character string information, e.g., used in conjunction with a user interface in conjunction with a standard operating system such as a Windows, Macintosh or LINUX system. For example, an instruction set for manipulating strings of characters, either by programming the required operations into the applications or with the required operations performed manually by a user (or both). For example, specialized sequence alignment programs such as PILEUP or BLAST can also be incorporated into the systems, e.g., for alignment of nucleic acids or proteins (or corresponding character strings).

[0281] Software for performing the statistical methods required for the invention, e.g., to determine correlations between expression profiles and subsets of members of the diagnostic nucleotide libraries, such as programmed embodiments of the statistical methods described above, are also included in the computer systems. Alternatively, programming elements for performing such methods as principle component analysis (PCA) or least squares analysis can also be included in the digital system to identify relationships between data. Exemplary software for such methods is provided by Partek, Inc., St. Peter, Mo; at the web site partek.com.

[0282] Any controller or computer optionally includes a monitor which can include, e.g., a flat panel display (e.g., active matrix liquid crystal display, liquid crystal display), a cathode ray tube ("CRT") display, or another display system which serves as a user interface, e.g., to output predictive data. Computer circuitry, including numerous integrated circuit chips, such as a microprocessor, memory, interface circuits, and the like, is often placed in a casing or box which optionally also includes a hard disk drive, a floppy disk drive, a high capacity removable drive such as a writeable CD-ROM, and other common peripheral elements.

[0283] Inputting devices such as a keyboard, mouse, or touch sensitive screen, optionally provide for input from a user and for user selection, e.g., of sequences or data sets to be compared or otherwise manipulated in the relevant computer system. The computer typically includes appropriate software for receiving user instructions, either in the form of user input into a set parameter or data fields (e.g., to input relevant subject data), or in the form of preprogrammed instructions, e.g., preprogrammed for a variety of different specific operations. The software then converts these instructions to appropriate language for instructing the system to carry out any desired operation.

[0284] The integrated system may also be embodied within the circuitry of an application specific integrated circuit (ASIC) or programmable logic device (PLD). In such a case, the invention is embodied in a computer readable descriptor language that can be used to create an ASIC or PLD. The integrated system can also be embodied within the circuitry or logic processors of a variety of other digital apparatus, such as PDAs, laptop computer systems, displays, image editing equipment, etc.

[0285] The digital system can comprise a learning component where expression profiles, and relevant subject data are compiled and monitored in conjunction with physical assays, and where correlations, e.g., molecular signatures with predictive value for a disease, are established or refined. Successful and unsuccessful combinations are optionally documented in a database to provide justification/preferences for user-base or digital system based selection of diagnostic nucleotide sets with high predictive accuracy for a specified disease or condition.

[0286] The integrated systems can also include an automated workstation. For example, such a workstation can

prepare and analyze leukocyte RNA samples by performing a sequence of events including: preparing RNA from a human blood sample; labeling the RNA with an isotopic or non-isotopic label; hybridizing the labeled RNA to at least one array comprising all or part of the candidate library; and detecting the hybridization pattern. The hybridization pattern is digitized and recorded in the appropriate database.

Automated RNA preparation tool

[0287] An automated RNA preparation tool for the preparation of mononuclear cells from whole blood samples, and preparation of RNA from the mononuclear cell is also disclosed. In a preferred embodiment, the use of the RNA preparation tool is fully automated, so that the cell separation and RNA isolation would require no human manipulations. Full automation is advantageous because it minimizes delay, and standardizes sample preparation across different laboratories. This standardization increases the reproducibility of the results.

[0288] Figure 2 depicts the processes performed by the RNA preparation tool. A primary component of the device is a centrifuge (A). Tubes of whole blood containing a density gradient solution, transcription/translation inhibitors, and a gel barrier that separates erythrocytes from mononuclear cells and serum after centrifugation are placed in the centrifuge (B). The barrier is permeable to erythrocytes and granulocytes during centrifugation, but does not allow mononuclear cells to pass through (or the barrier substance has a density such that mononuclear cells remain above the level of the barrier during the centrifugation). After centrifugation, the erythrocytes and granulocytes are trapped beneath the barrier, facilitating isolation of the mononuclear cell and serum layers. A mechanical arm removes the tube and inverts it to mix the mononuclear cell layer and the serum (C). The arm next pours the supernatant into a fresh tube (D), while the erythrocytes and granulocytes remained below the barrier. Alternatively, a needle is used to aspirate the supernatant and transfer it to a fresh tube. The mechanical arms of the device opens and closes lids, dispenses PBS to aid in the collection of the mononuclear cells by centrifugation, and moves the tubes in and out of the centrifuge. Following centrifugation, the supernatant is poured off or removed by a vacuum device (E), leaving an isolated mononuclear cell pellet. Purification of the RNA from the cells is performed automatically, with lysis buffer and other purification solutions (F) automatically dispensed and removed before and after centrifugation steps. The result is a purified RNA solution. In another embodiment, RNA isolation is performed using a column or filter method. In yet another embodiment, the disclosure includes an on-board homogenizer for use in cell lysis.

Other automated systems

[0289] Automated and/or semi-automated methods for solid and liquid phase high-throughput sample preparation and evaluation are available, and supported by commercially available devices. For example, robotic devices for preparation of nucleic acids from bacterial colonies, e.g., to facilitate production and characterization of the candidate library include, for example, an automated colony picker (e.g., the Q-bot, Genetix, U.K.) capable of identifying, sampling, and inoculating up to 10,000/4 hrs different clones into 96 well microtiter dishes. Alternatively, or in addition, robotic systems for liquid handling are available from a variety of sources, e.g., automated workstations like the automated synthesis apparatus developed by Takeda Chemical Industries, LTD. (Osaka, Japan) and many robotic systems utilizing robotic arms (Zymate II, Zymark Corporation, Hopkinton, Mass.; Orca, Beckman Coulter, Inc. (Fullerton, CA)) which mimic the manual operations performed by a scientist. Any of the above devices are suitable for use with the present disclosure, e.g., for high-throughput analysis of library components or subject leukocyte samples. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

[0290] High throughput screening systems that automate entire procedures, e.g., sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the relevant assay are commercially available. (see, e.g., Zymark Corp., Hopkinton, MA; Air Technical Industries, Mentor, OH; Beckman Instruments, Inc. Fullerton, CA; Precision Systems, Inc., Natick, MA, etc.). These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, arrays and array readers are available, e.g., from Affymetrix, PE Biosystems, and others.

[0291] The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like.

[0292] A variety of commercially available peripheral equipment, including, e.g., optical and fluorescent detectors, optical and fluorescent microscopes, plate readers, CCD arrays, phosphorimagers, scintillation counters, phototubes, photodiodes, and the like, and software is available for digitizing, storing and analyzing a digitized video or digitized optical or other assay results, e.g., using PC (Intel x86 or pentium chip- compatible DOS™, OS2™ WINDOWS™, WINDOWS NT™ or WINDOWS95™ based machines), MACINTOSH™, or UNIX based (e.g., SUN™ work station) computers.

Embodiment in a web site.

[0293] The methods described above can be implemented in a localized or distributed computing environment. For example, if a localized computing environment is used, an array comprising a candidate nucleotide library, or diagnostic nucleotide set, is configured in proximity to a detector, which is, in turn, linked to a computational device equipped with user input and output features.

[0294] In a distributed environment, the methods can be implemented on a single computer with multiple processors or, alternatively, on multiple computers. The computers can be linked, e.g. through a shared bus, but more commonly, the computer(s) are nodes on a network. The network can be generalized or dedicated, at a local level or distributed over a wide geographic area. In certain embodiments, the computers are components of an intra-net or an internet.

[0295] The predictive data corresponding to subject molecular signatures (e.g., expression profiles, and related diagnostic, prognostic, or monitoring results) can be shared by a variety of parties. In particular, such information can be utilized by the subject, the subject's health care practitioner or provider, a company or other institution, or a scientist. An individual subject's data, a subset of the database or the entire database recorded in a computer readable medium can be accessed directly by a user by any method of communication, including, but not limited to, the internet. With appropriate computational devices, integrated systems, communications networks, users at remote locations, as well as users located in proximity to, e.g., at the same physical facility, the database can access the recorded information. Optionally, access to the database can be controlled using unique alphanumeric passwords that provide access to a subset of the data. Such provisions can be used, e.g., to ensure privacy, anonymity, etc.

[0296] Typically, a client (e.g., a patient, practitioner, provider, scientist, or the like) executes a Web browser and is linked to a server computer executing a Web server. The Web browser is, for example, a program such as IBM's Web Explorer, Internet explorer, NetScape or Mosaic, or the like. The Web server is typically, but not necessarily, a program such as IBM's HTTP Daemon or other WWW daemon (e.g., LINUX-based forms of the program). The client computer is bi-directionally coupled with the server computer over a line or via a wireless system. In turn, the server computer is bi-directionally coupled with a website (server hosting the website) providing access to software implementing the methods of this disclosure.

[0297] A user of a client connected to the Intranet or Internet may cause the client to request resources that are part of the web site(s) hosting the application(s) providing an implementation of the methods described herein. Server program(s) then process the request to return the specified resources (assuming they are currently available). A standard naming convention has been adopted, known as a Uniform Resource Locator ("URL"). This convention encompasses several types of location names, presently including subclasses such as Hypertext Transport Protocol ("http"), File Transport Protocol ("ftp"), gopher, and Wide Area Information Service ("WAIS"). When a resource is downloaded, it may include the URLs of additional resources. Thus, the user of the client can easily learn of the existence of new resources that he or she had not specifically requested.

[0298] Methods of implementing Intranet and/or Intranet embodiments of computational and/or data access processes are well known to those of skill in the art and are documented, e.g., in ACM Press, pp. 383-392; ISO-ANSI, Working Draft, "Information Technology-Database Language SQL", Jim Melton, Editor, International Organization for Standardization and American National Standards Institute, Jul. 1992; ISO Working Draft, "Database Language SQL-Part 2: Foundation (SQL/Foundation)", CD9075-2:199.chi.SQL, Sep. 11, 1997; and Cluer et al. (1992) A General Framework for the Optimization of Object-Oriented Queries, Proc SIGMOD International Conference on Management of Data, San Diego, California, Jun. 2-5, 1992, SIGMOD Record, vol. 21, Issue 2, Jun., 1992; Stonebraker, M., Editor;. Other resources are available, e.g., from Microsoft, IBM, Sun and other software development companies.

[0299] Using the tools described above, users of the reagents, methods and database as discovery or diagnostic tools can query a centrally located database with expression and subject data. Each submission of data adds to the sum of expression and subject information in the database. As data is added, a new correlation statistical analysis is automatically run that incorporates the added clinical and expression data. Accordingly, the predictive accuracy and the types of correlations of the recorded molecular signatures increases as the database grows.

[0300] For example, subjects, such as patients, can access the results of the expression analysis of their leukocyte samples and any accrued knowledge regarding the likelihood of the patient's belonging to any specified diagnostic (or prognostic, or monitoring, or risk group), i.e., their expression profiles, and/or molecular signatures. Optionally, subjects can add to the predictive accuracy of the database by providing additional information to the database regarding diagnoses, test results, clinical or other related events that have occurred since the time of the expression profiling. Such information can be provided to the database via any form of communication, including, but not limited to, the internet. Such data can be used to continually define (and redefine) diagnostic groups. For example, if 1000 patients submit data regarding the occurrence of myocardial infarction over the 5 years since their expression profiling, and 300 of these patients report that they have experienced a myocardial infarction and 700 report that they have not, then the 300 patients define a new "group A." As the algorithm is used to continually query and revise the database, a new diagnostic nucleotide set that differentiates groups A and B (i.e., with and without myocardial infarction within a five year period) is identified.

This newly defined nucleotide set is then be used (in the manner described above) as a test that predicts the occurrence of myocardial infarction over a five-year period. While submission directly by the patient is exemplified above, any individual with access and authority to submit the relevant data e.g., the patient's physician, a laboratory technician, a health care or study administrator, or the like, can do so.

[0301] As will be apparent from the above examples, transmission of information via the internet (or via an intranet) is optionally bi-directional. That is, for example, data regarding expression profiles, subject data, and the like are transmitted via a communication system to the database, while information regarding molecular signatures, predictive analysis, and the like, are transmitted from the database to the user. For example, using appropriate configurations of an integrated system including a microarray comprising a diagnostic nucleotide set, a detector linked to a computational device can directly transmit (locally or from a remote workstation at great distance, e.g., hundreds or thousands of miles distant from the database) expression profiles and a corresponding individual identifier to a central database for analysis according to the methods of the disclosure. According to, e.g., the algorithms described above, the individual identifier is assigned to one or more diagnostic (or prognostic, or monitoring, etc.) categories. The results of this classification are then relayed back, via, e.g., the same mode of communication, to a recipient at the same or different internet (or intranet) address.

Kits

[0302] The present disclosure is optionally provided to a user as a kit. Typically, a kit contains one or more diagnostic nucleotide sets of the disclosure. Alternatively, the kit contains the candidate nucleotide library of the disclosure. Most often, the kit contains a diagnostic nucleotide probe set, or other subset of a candidate library, e.g., as a cDNA or antibody microarray packaged in a suitable container. The kit may further comprise, one or more additional reagents, e.g., substrates, labels, primers, for labeling expression products, tubes and/or other accessories, reagents for collecting blood samples, buffers, e.g., erythrocyte lysis buffer, leukocyte lysis buffer, hybridization chambers, cover slips, etc., as well as a software package, e.g., including the statistical methods of the disclosure, e.g., as described above, and a password and/or account number for accessing the compiled database. The kit optionally further comprises an instruction set or user manual detailing preferred methods of using the diagnostic nucleotide sets in the methods of the disclosure. In one embodiment, the kit may include contents useful for the discovery of diagnostic nucleotide sets using microarrays. The kit may include sterile, endotoxin and RNase free blood collection tubes. The kit may also include alcohol swabs, tourniquet, blood collection set, and/or PBS (phosphate buffer saline; needed when method of example 2 is used to derived mononuclear RNA). The kit may also include cell lysis buffer. The kit may include RNA isolation kit, substrates for labeling of RNA (may vary for various expression profiling techniques). The kit may also include materials for fluorescence microarray expression profiling, including one or more of the following: reverse transcriptase and 10x RT buffer, T7(dT)24 primer (primer with T7 promoter at 5' end), DTT, deoxynucleotides, optionally 100mM each, RNase inhibitor, second strand cDNA buffer, DNA polymerase, Rnase H, T7 RNA polymerase ribonucleotides, in vitro transcription buffer, and/or Cy3 and Cy5 labeled ribonucleotides. The kit may also include microarrays containing candidate gene libraries, cover slips for slides, and/or hybridization chambers. The kit may further include software package for identification of diagnostic gene set from data, that contains statistical methods, and/or allows alteration in desired sensitivity and specificity of gene set. The software may further facilitate access to and data analysis by centrally a located database server. The software may further include a password and account number to access central database server. In addition, the kit may include a kit user manual.

[0303] In another embodiment, the kit may include contents useful for the application of diagnostic nucleotide sets using microarrays. The kit may include sterile, endotoxin and/or RNase free blood collection tubes. The kit may also include, alcohol swabs, tourniquet, and/or a blood collection set. The kit may further include PBS (phosphate buffer saline; needed when method of example 2 is used to derived mononuclear RNA), cell lysis buffer, and/or an RNA isolation kit. In addition, the kit may include substrates for labeling of RNA (may vary for various expression profiling techniques). For fluorescence microarray expression profiling, components may include reverse transcriptase and 10x RT buffer, T7 (dT)24 primer (primer with T7 promoter at 5' end), DTT, deoxynucleotides (optionally 100mM each), RNase inhibitor, second strand cDNA buffer, DNA polymerase, Rnase H, T7 RNA polymerase, ribonucleotides, in vitro transcription buffer, and/or Cy3 and Cy5 labeled ribonucleotides. The kit may further include microarrays containing candidate gene libraries. The kit may also include cover slips for slides, and/or hybridization chambers. The kit may include a software package for identification of diagnostic gene set from data. The software package may contain statistical methods, allow alteration in desired sensitivity and specificity of gene set, and/or facilitate access to and data analysis by centrally located database server. The software package may include a password and account number to access central database server. In addition, the kit may include a kit user manual.

[0304] In another embodiment, the kit may include contents useful for the application of diagnostic nucleotide sets using real-time PCR. This kit may include sterile, endotoxin and/or RNase free blood collection tubes. The kit may further include alcohol swabs, tourniquet, and/or a blood collection set. The kit may also include PBS (phosphate buffer saline;

needed when method of example 2 is used to derived mononuclear RNA). In addition, the kit may include cell lysis buffer and/or an RNA isolation kit. The kit may also include substrates for real time RT-PCR, which may vary for various real-time PCR techniques, including poly dT primers, random hexamer primers, reverse Transcriptase and RT buffer, DTT, deoxynucleotides 100 mM, RNase H, primer pairs for diagnostic and control gene set, 10x PCR reaction buffer, and/or Taq DNA polymerase. The kit may also include fluorescent probes for diagnostic and control gene set (alternatively, fluorescent dye that binds to only double stranded DNA). The kit may further include reaction tubes with or without barcode for sample tracking, 96-well plates with barcode for sample identification, one barcode for entire set, or individual barcode per reaction tube in plate. The kit may also include a software package for identification of diagnostic gene set from data, and /or statistical methods. The software package may allow alteration in desired sensitivity and specificity of gene set, and/or facilitate access to and data analysis by centrally located database server. The kit may include a password and account number to access central database server. Finally, the kit may include a kit user manual.

[0305] This disclosure will be better understood by reference to the following non-limiting Examples:

LIST OF EXAMPLE TITLES

[0306]

Example 1: Preparation of a leukocyte cDNA array comprising a candidate gene library

Example 2: Preparation of RNA from mononuclear cells for expression profiling

Example 3: Preparation of Universal Control RNA for use in leukocyte expression profiling

Example 4: RNA Labeling and hybridization to a leukocyte cDNA array of candidate nucleotide sequences.

Example 5: Clinical study for the Identification of diagnostic gene sets useful in diagnosis and treatment of Cardiac allograft rejection

Example 6: Identification of diagnostic nucleotide sets for kidney and liver allograft rejection

Example 7: Identification of diagnostic nucleotide sets for diagnosis of cytomegalovirus

Example 8: Design of oligonucleotide probes

Example 9: Production of an array of 8,000 spotted 50 mer oligonucleotides.

Example 10: Identification of diagnostic nucleotide sets for diagnosis of Cardiac Allograft Rejection using microarrays

Example 11: Amplification, labeling, and hybridization of total RNA to an oligonucleotide microarray

Example 12: Real-time PCR validation of array expression results

Example 13: Real-time PCR expression markers of acute allograft rejection

Example 14: Identification of diagnostic nucleotide sets for diagnosis of Cardiac Allograft Rejection using microarrays

Example 15: Correlation and Classification Analysis

Example 16: Acute allograft rejection: biopsy tissue gene expression profiling

Example 17: Microarray and PCR gene expression panels for diagnosis and monitoring or acute allograft rejection

Example 18: Assay sample preparation

Example 19: Allograft rejection diagnostic gene sequence analysis

Example 20: Detection of proteins expressed by diagnostic gene sequences

Example 21: Detecting changes in the rate of hematopoiesis

Examples

Example 1: Preparation of a leukocyte cDNA array comprising a candidate gene library

[0307] Candidate genes and gene sequences for leukocyte expression profiling are identified through methods described elsewhere in this document. Candidate genes are used to obtain or design probes for peripheral leukocyte expression profiling in a variety of ways.

[0308] A cDNA microarray carrying 384 probes was constructed using sequences selected from the initial candidate library. cDNAs is selected from T-cell libraries, PBMC libraries and buffy coat libraries.

96-Well PCR

[0309] Plasmids are isolated in 96-well format and PCR was performed in 96-well format. A master mix is made that contain the reaction buffer, dNTPs, forward and reverse primer and DNA polymerase was made. 99 ul of the master mix was aliquoted into 96-well plate. 1 ul of plasmid (1-2 ng/ul) of plasmid was added to the plate. The final reaction concentration was 10 mM Tris pH 8.3, 3.5 mM MgCl₂, 25 mM KCl, 0.4 mM dNTPs, 0.4 uM M 13 forward primer, 0.4 M 13 reverse primer, and 10 U of Taq Gold (Applied Biosystems). The PCR conditions were:

Step 1 95C for 10 min
 Step 2 95C for 15 sec
 Step 3 56C for 30 sec
 Step 4 72C for 2 min 15 seconds
 Step 5 go to Step 2 39 times
 Step 6 72C for 10 minutes
 Step 7 4C for ever.

PCR Purification

[0310] PCR purification is done in a 96-well format. The ArrayIt (Telechem International, Inc.) PCR purification kit is used and the provided protocol was followed without modification. Before the sample is evaporated to dryness, the concentration of PCR products was determined using a spectrophotometer. After evaporation, the samples are re-suspended in 1x Micro Spotting Solution (ArrayIt) so that the majority of the samples were between 0.2-1.0 ug/ul.

Array Fabrication

[0311] Spotted cDNA microarrays are then made from these PCR products by ArrayIt using their protocols, which may be found at the ArrayIt website. Each fragment was spotted 3 times onto each array. Candidate genes and gene sequences for leukocyte expression profiling are identified through methods described elsewhere in this document. Those candidate genes are used for peripheral leukocyte expression profiling. The candidate libraries can be used to obtain or design probes for expression profiling in a variety of ways.

[0312] Oligonucleotide probes are prepared using the gene sequences of Table 2, and the sequence listing. Oligo probes are designed on a contract basis by various companies (for example, Compugen, Mergen, Affymetrix, Telechem), or designed from the candidate sequences using a variety of parameters and algorithms as indicated at located at the MIT web site. Briefly, the length of the oligonucleotide to be synthesized is determined, preferably greater than 18 nucleotides, generally 18-24 nucleotides, 24-70 nucleotides and, in some circumstances, more than 70 nucleotides. The sequence analysis algorithms and tools described above are applied to the sequences to mask repetitive elements, vector sequences and low complexity sequences. Oligonucleotides are selected that are specific to the candidate nucleotide sequence (based on a Blast n search of the oligonucleotide sequence in question against gene sequences databases, such as the Human Genome Sequence, UniGene, dbEST or the non-redundant database at NCBI), and have <50% G content and 25-70% G+C content. Desired oligonucleotides are synthesized using well-known methods and apparatus, or ordered from a company (for example Sigma). Oligonucleotides are spotted onto microarrays. Alternatively, oligonucleotides are synthesized directly on the array surface, using a variety of techniques (Hughes et al. 2001, Yershov et al. 1996, Lockhart et al 1996).

Example 2: Preparation of RNA from mononuclear cells for expression profiling

[0313] Blood was isolated from the subject for leukocyte expression profiling using the following methods: Two tubes were drawn per patient. Blood was drawn from either a standard peripheral venous blood draw or directly from a large-bore intra-arterial or intravenous catheter inserted in the femoral artery, femoral vein, subclavian vein or internal jugular vein. Care was taken to avoid sample contamination with heparin from the intravascular catheters, as heparin can interfere with subsequent RNA reactions. For each tube, 8 ml of whole blood was drawn into a tube (CPT, Becton-Dickinson order #362753) containing the anticoagulant Citrate, 25°C density gradient solution (e.g. Ficoll, Percoll) and a polyester gel barrier that upon centrifugation was permeable to RBCs and granulocytes but not to mononuclear cells. The tube was inverted several times to mix the blood with the anticoagulant. The tubes were centrifuged at 1750xg in a swing-out rotor at room temperature for 20 minutes. The tubes were removed from the centrifuge and inverted 5-10 times to mix the plasma with the mononuclear cells, while trapping the RBCs and the granulocytes beneath the gel barrier. The plasma/mononuclear cell mix was decanted into a 15ml tube and 5ml of phosphate-buffered saline (PBS) is added. The 15ml tubes were spun for 5 minutes at 1750xg to pellet the cells. The supernatant was discarded and 1.8 ml of RLT lysis buffer is added to the mononuclear cell pellet. The buffer and cells were pipetted up and down to ensure complete lysis of the pellet. The cell lysate was frozen and stored until it is convenient to proceed with isolation of total RNA.

[0314] Total RNA was purified from the lysed mononuclear cells using the Qiagen Rneasy Miniprep kit, as directed by the manufacturer (10/99 version) for total RNA isolation, including homogenization (Qias shredder columns) and on-column DNase treatment. The purified RNA was eluted in 50ul of water. The further use of RNA prepared by this method is described in Examples 10 and 11.

[0315] Some samples were prepared by a different protocol, as follows:
 Two 8 ml blood samples were drawn from a peripheral vein into a tube (CPT, Becton-Dickinson order #362753) containing

anticoagulant (Citrate), 25°C density gradient solution (Ficoll) and a polyester gel barrier that upon centrifugation is permeable to RBCs and granulocytes but not to mononuclear cells. The mononuclear cells and plasma remained above the barrier while the RBCs and granulocytes were trapped below. The tube was inverted several times to mix the blood with the anticoagulant, and the tubes were subjected to centrifugation at 1750xg in a swing-out rotor at room temperature for 20 min. The tubes were removed from the centrifuge, and the clear plasma layer above the cloudy mononuclear cell layer was aspirated and discarded. The cloudy mononuclear cell layer was aspirated, with care taken to rinse all of the mononuclear cells from the surface of the gel barrier with PBS (phosphate buffered saline). Approximately 2 mls of mononuclear cell suspension was transferred to a 2ml microcentrifuge tube, and centrifuged for 3min. at 16,000 rpm in a microcentrifuge to pellet the cells. The supernatant was discarded and 1.8 ml of RLT lysis buffer (Qiagen) were added to the mononuclear cell pellet, which lysed the cells and inactivated Rnases. The cells and lysis buffer were pipetted up and down to ensure complete lysis of the pellet. Cell lysate was frozen and stored until it was convenient to proceed with isolation of total RNA.

[0316] RNA samples were isolated from 8 mL of whole blood. Yields ranged from 2 ug to 20ug total RNA for 8mL blood. A260/A280 spectrophotometric ratios were between 1.6 and 2.0, indicating purity of sample. 2ul of each sample were run on an agarose gel in the presence of ethidium bromide. No degradation of the RNA sample and no DNA contamination was visible.

[0317] In some cases, specific subsets of mononuclear cells were isolated from peripheral blood of human subjects. When this was done, the StemSep cell separation kits (manual version 6.0.0) were used from StemCell Technologies (Vancouver, Canada). This same protocol can be applied to the isolation of T cells, CD4 T cells, CD8 T cells, B cells, monocytes, NK cells and other cells. Isolation of cell types using negative selection with antibodies may be desirable to avoid activation of target cells by antibodies.

Example 3: RepARATION of Universal Control RNA for use in leukocyte expression profiling

[0318] Control RNA was prepared using total RNA from Buffy coats and/or total RNA from enriched mononuclear cells isolated from Buffy coats, both with and without stimulation with ionomycin and PMA. The following control RNAs were prepared:

Control 1: Buffy Coat Total RNA

Control 2: Mononuclear cell Total RNA

Control 3: Stimulated buffy coat Total RNA

Control 4: Stimulated mononuclear Total RNA

Control 5: 50% Buffy coat Total RNA / 50% Stimulated buffy coat Total RNA

Control 6: 50% Mononuclear cell Total RNA / 50% Stimulated Mononuclear Total RNA

[0319] Some samples were prepared using the following protocol: Buffy coats from 38 individuals were obtained from Stanford Blood Center. Each buffy coat is derived from ~350 mL whole blood from one individual. 10 ml buffy coat was removed from the bag, and placed into a 50 ml tube. 40 ml of Buffer EL (Qiagen) was added, the tube was mixed and placed on ice for 15 minutes, then cells were pelleted by centrifugation at 2000xg for 10 minutes at 4°C. The supernatant was decanted and the cell pellet was re-suspended in 10 ml of Qiagen Buffer EL. The tube was then centrifuged at 2000xg for 10 minutes at 4°C. The cell pellet was then re-suspended in 20 ml TRIZOL (GibcoBRL) per Buffy coat sample, the mixture was shredded using a rotary homogenizer, and the lysate was then frozen at -80°C prior to proceeding to RNA isolation.

[0320] Other control RNAs were prepared from enriched mononuclear cells prepared from Buffy coats. Buffy coats from Stanford Blood Center were obtained, as described above. 10 ml buffy coat was added to a 50 ml polypropylene tube, and 10 ml of phosphate buffer saline (PBS) was added to each tube. A polysucrose (5.7 g/dL) and sodium diatrizoate (9.0 g/dL) solution at a 1.077 +/-0.0001 g/ml density solution of equal volume to diluted sample was prepared (Histopaque 1077, Sigma cat. no 1077-1). This and all subsequent steps were performed at room temperature. 15 ml of diluted buffy coat/PBS was layered on top of 15 ml of the histopaque solution in a 50 ml tube. The tube was centrifuged at 400xg for 30 minutes at room temperature. After centrifugation, the upper layer of the solution to within 0.5 cm of the opaque interface containing the mononuclear cells was discarded. The opaque interface was transferred into a clean centrifuge tube. An equal volume of PBS was added to each tube and centrifuged at 350xg for 10 minutes at room temperature. The supernatant was discarded. 5 ml of Buffer EL (Qiagen) was used to resuspend the remaining cell pellet and the tube was centrifuged at 2000xg for 10 minutes at room temperature. The supernatant was discarded. The pellet was resuspended in 20 ml of TRIZOL (GibcoBRL) for each individual buffy coat that was processed. The sample was homogenized using a rotary homogenizer and frozen at -80C until RNA was isolated. RNA was isolated from frozen lysed Buffy coat samples as follows: frozen samples were thawed, and 4 ml of chloroform was added to each buffy coat sample. The sample was mixed by vortexing and centrifuged at 2000xg for 5 minutes. The aqueous layer was moved

to new tube and then repurified by using the RNeasy Maxi RNA clean up kit, according to the manufacturer's instruction (Qiagen, PN 75162). The yield, purity and integrity were assessed by spectrophotometer and gel electrophoresis. Some samples were prepared by a different protocol, as follows. The further use of RNA prepared using this protocol is described in Example 11.

[0321] 50 whole blood samples were randomly selected from consented blood donors at the Stanford Medical School Blood Center. Each buffy coat sample was produced from ~350 mL of an individual's donated blood. The whole blood sample was centrifuged at ~4,400 x g for 8 minutes at room temperature, resulting in three distinct layers: a top layer of plasma, a second layer of buffy coat, and a third layer of red blood cells. 25 ml of the buffy coat fraction was obtained and diluted with an equal volume of PBS (phosphate buffered saline). 30 ml of diluted buffy coat was layered onto 15 ml of sodium diatrizoate solution adjusted to a density of 1.077+/-0.001 g/ml (Histopaque 1077, Sigma) in a 50mL plastic tube. The tube was spun at 800 g for 10 minutes at room temperature. The plasma layer was removed to the 30 ml mark on the tube, and the mononuclear cell layer removed into a new tube and washed with an equal volume of PBS, and collected by centrifugation at 2000 g for 10 minutes at room temperature. The cell pellet was resuspended in 10 ml of Buffer EL (Qiagen) by vortexing and incubated on ice for 10 minutes to remove any remaining erythrocytes. The mononuclear cells were spun at 2000 g for 10 minutes at 4 degrees Celsius. The cell pellet was lysed in 25 ml of a phenol/guanidinium thiocyanate solution (TRIZOL Reagent, Invitrogen). The sample was homogenized using a Power-Gen 5 rotary homogenizer (Fisher Scientific) and Omini disposable generator probes (Fisher Scientific). The Trizol lysate was frozen at -80 degrees C until the next step.

[0322] The samples were thawed out and incubated at room temperature for 5 minutes. 5 ml chloroform was added to each sample, mixed by vortexing, and incubated at room temperature for 3 minutes. The aqueous layers were transferred to new 50 ml tubes. The aqueous layer containing total RNA was further purified using the Qiagen RNeasy Maxi kit (PN 75162), per the manufacturer's protocol (October 1999). The columns were eluted twice with 1 ml Rnase-free water, with a minute incubation before each spin. Quantity and quality of RNA was assessed using standard methods. Generally, RNA was isolated from batches of 10 buffy coats at a time, with an average yield per buffy coat of 870 µg, and an estimated total yield of 43.5 mg total RNA with a 260/280 ratio of 1.56 and a 28S/18S ratio of 1.78.

[0323] Quality of the RNA was tested using the Agilent 2100 Bioanalyzer using RNA 6000 microfluidics chips. Analysis of the electrophorograms from the Bioanalyzer for five different batches demonstrated the reproducibility in quality between the batches.

[0324] Total RNA from all five batches were combined and mixed in a 50 ml tube, then aliquoted as follows: 2 x 10 ml aliquots in 15 ml tubes, and the rest in 100 µl aliquots in 1.5 ml microcentrifuge tubes. The aliquots gave highly reproducible results with respect to RNA purity, size and integrity. The RNA was stored at -80°C.

Test hybridization of Reference RNA.

[0325] When compared with BC38 and Stimulated mononuclear reference samples, the R50 performed as well, if not better than the other reference samples as shown in Figure 3. In an analysis of hybridizations, where the R50 targets were fluorescently labeled with Cy-5 using methods described herein and the amplified and labeled aRNA was hybridized (as in example 11) to the oligonucleotide array described in example 9. The R50 detected 97.3% of probes with a Signal to Noise ratio (S/N) of greater than three and 99.9 % of probes with S/N greater than one.

Example 4. RNA Labeling and hybridization to a leukocyte cDNA array of candidate nucleotide sequences.

Comparison of Guanine-Silica to Acid-Phenol RNA Purification (GSvsAP)

[0326] These data are from a set of 12 hybridizations designed to identify differences between the signal strength from two different RNA purification methods. The two RNA methods used were guanidine-silica (GS, Qiagen) and acid-phenol (AP, Trizol, Gibco BRL). Ten tubes of blood were drawn from each of four people. Two were used for the AP prep, the other eight were used for the GS prep. The protocols for the leukocyte RNA preps using the AP and GS techniques were completed as described here:

Guanidine-silica (GS) method:

[0327] For each tube, 8ml blood was drawn into a tube containing the anticoagulant Citrate, 25°C density gradient solution and a polyester gel barrier that upon centrifugation is permeable to RBCs and granulocytes but not to mononuclear cells. The mononuclear cells and plasma remained above the barrier while the RBCs and granulocytes were trapped below. CPT tubes from Becton-Dickinson (#362753) were used for this purpose. The tube was inverted several times to mix the blood with the anticoagulant. The tubes were immediately centrifuged @1750xg in a swinging bucket rotor at room temperature for 20 min. The tubes were removed from the centrifuge and inverted 5-10 times. This mixed the

plasma with the mononuclear cells, while the RBCs and the granulocytes remained trapped beneath the gel barrier. The plasma/mononuclear cell mix was decanted into a 15ml tube and 5ml of phosphate-buffered saline (PBS) was added. The 15ml tubes are spun for 5 minutes at 1750xg to pellet the cells. The supernatant was discarded and 1.8 ml of RLT lysis buffer (guanidine isothiocyanate) was added to the mononuclear cell pellet. The buffer and cells were pipetted up and down to ensure complete lysis of the pellet. The cell lysate was then processed exactly as described in the Qiagen RNeasy Miniprep kit protocol (10/99 version) for total RNA isolation (including steps for homogenization (QiaShredder columns) and on-column DNase treatment. The purified RNA was eluted in 50ul of water.

Acid-phenol (AP) method:

[0328] For each tube, 8ml blood was drawn into a tube containing the anticoagulant Citrate, 25°C density gradient solution and a polyester gel barrier that upon centrifugation is permeable to RBCs and granulocytes but not to mononuclear cells. The mononuclear cells and plasma remained above the barrier while the RBCs and granulocytes were trapped below. CPT tubes from Becton-Dickinson (#362753) were used for this purpose. The tube was inverted several times to mix the blood with the anticoagulant. The tubes were immediately centrifuged @1750xg in a swinging bucket rotor at room temperature for 20 min. The tubes were removed from the centrifuge and inverted 5-10 times. This mixed the plasma with the mononuclear cells, while the RBCs and the granulocytes remained trapped beneath the gel barrier. The plasma/mononuclear cell mix was decanted into a 15ml tube and sol of phosphate-buffered saline (PBS) was added. The 15ml tubes are spun for 5 minutes @1750xg to pellet the cells. The supernatant was discarded and the cell pellet was lysed using 0.6 mL Phenol/guanidine isothiocyanate (e.g. Trizol reagent, GibcoBRL). Subsequent total RNA isolation proceeded using the manufacturers protocol.

[0329] RNA from each person was labeled with either Cy3 or Cy5, and then hybridized in pairs to the mini-array. For instance, the first array was hybridized with GS RNA from one person (Cy3) and GS RNA from a second person (Cy5).

[0330] Techniques for labeling and hybridization for all experiments discussed here were completed as detailed above.

Arrays were prepared as described in example 1.

[0331] RNA isolated from subject samples, or control Buffy coat RNA, were labeled for hybridization to a cDNA array. Total RNA (up to 100 μ g) was combined with 2 μ l of 100 μ M solution of an Oligo (dT)12-18 (GibcoBRL) and heated to 70°C for 10 minutes and place on ice. Reaction buffer was added to the tube, to a final concentration of 1xRT buffer (GibcoBRL), 10 mM DTT (GibcoBRL), 0.1 mM unlabeled dATP, dTTP, and dGTP, and 0.025 mM unlabeled dCTP, 200 pg of CAB (A. thaliana photosystem I chlorophyll a/b binding protein), 200 pg of RCA (A. thaliana RUBISCO activase), 0.25 mM of Cy-3 or Cy-5 dCTP, and 400 U Superscript II RT (GibcoBRL).

[0332] The volumes of each component of the labeling reaction were as follows: 20 μ l of 5xRT buffer; 10 μ l of 100 mM DTT; 1 μ l of 10 mM dNTPs without dCTP; 0.5 μ l of 5 mM CTP; 13 μ l of H₂O; 0.02 μ l of 10 ng/ μ l CAB and RCA; 1 μ l of 40 Units/ μ l RNaseOUT Recombinant Ribonuclease Inhibitor (GibcoBRL); 2.5 μ l of 1.0 mM Cy-3 or Cy-5 dCTP; and 2.0 μ l of 200 Units/ μ l of Superscript II RT.

[0333] The sample was vortexed and centrifuged. The sample was incubated at 4°C for 1 hour for first strand cDNA synthesis, then heated at 70°C for 10 minutes to quench enzymatic activity. 1 μ l of 10 mg/ml of RNase A was added to degrade the RNA strand, and the sample was incubated at 37°C for 30 minutes. Next, the Cy-3 and Cy-5 cDNA samples were combined into one tube. Unincorporated nucleotides were removed using QIAquick RCR purification protocol (Qiagen), as directed by the manufacturer. The sample was evaporated to dryness and resuspended in 5 μ l of water. The sample was mixed with hybridization buffer containing 5xSSC, 0.2% SDS, 2 mg/ml Cot-1 DNA (GibcoBRL), 1 mg/ml yeast tRNA (GibcoBRL), and 1.6 ng/ μ l poly dA40-60 (Pharmacia). This mixture was placed on the microarray surface and a glass cover slip was placed on the array (Coming). The microarray glass slide was placed into a hybridization chamber (Amaylt). The chamber was then submerged in a water bath overnight at 62° C. The microarray was removed from the cassette and the cover slip was removed by repeatedly submerging it to a wash buffer containing 1xSSC, and 0.1% SDS. The microarray slide was washed in 1xSSC/0.1% SDS for 5 minutes. The slide was then washed in 0.1% SSC/0.1% SDS for 5 minutes. The slide was finally washed in 0.1xSSC for 2 minutes. The slide was spun at 1000 rpm for 2 minutes to dry out the slide, then scanned on a microarray scanner (Axon Instruments, Union City, CA.).

[0334] Six hybridizations with 20 μ g of RNA were performed for each type of RNA preparation (GS or AP). Since both the Cy3 and the Cy5 labeled RNA are from test preparations, there are six data points for each GS prepped, Cy3-labeled RNA and six for each GS-prepped, Cy5-labeled RNA. The mini array hybridizations were scanned on an Axon Instruments scanner using GenPix 3.0 software. The data presented were derived as follows. First, all features flagged as "not found" by the software were removed from the dataset for individual hybridizations. These features are usually due to high local background or other processing artifacts. Second, the median fluorescence intensity minus the background fluorescence intensity was used to calculate the mean background subtracted signal for each dye for each hybridization. In Figure 3, the mean of these means across all six hybridizations is graphed (n=6 for each column). The error bars are the SEM. This experiment shows that the average signal from AP prepared RNA is 47% of the average signal from GS prepared RNA for both Cy3 and Cy5.

Generation of expression data for leukocyte genes from peripheral leukocyte samples

[0335] Six hybridizations were performed with RNA purified from human blood leukocytes using the protocols given above. Four of the six were prepared using the GS method and 2 were prepared using the AP method. Each preparation of leukocyte RNA was labeled with Cy3 and 10 µg hybridized to the mini-array. A control RNA was batch labeled with Cy5 and 10 µg hybridized to each mini-array together with the Cy3-labeled experimental RNA.

[0336] The control RNA used for these experiments was Control 1: Buffy Coat RNA, as described above. The protocol for the preparation of that RNA is reproduced here:

Buffy Coat RNA Isolation:

[0337] Buffy coats were obtained from Stanford Blood Center (in total 38 individual buffy coats were used. Each buffy coat is derived from ~350 mL whole blood from one individual. 10 ml buffy coat was taken and placed into a 50 ml tube and 40 ml of a hypochlorous acid (HOCl) solution (Buffer EL from Qiagen) was added. The tube was mixed and placed on ice for 15 minutes. The tube was then centrifuged at 2000xg for 10 minutes at 4°C. The supernatant was decanted and the cell pellet was re-suspended in 10 ml of hypochlorous acid solution (Qiagen Buffer EL). The tube was then centrifuged at 2000xg for 10 minutes at 4°C. The cell pellet was then re-suspended in 20 ml phenol/guanidine thiocyanate solution (TRIZOL from GibcoBRL) for each individual buffy coat that was processed.

[0338] The mixture was then shredded using a rotary homogenizer. The lysate was then frozen at -80°C prior to proceeding to RNA isolation.

[0339] The arrays were then scanned and analyzed on an Axon Instruments scanner using GenePix 3.0 software. The data presented were derived as follows. First, all features flagged as "not found" by the software were removed from the dataset for individual hybridizations. Second, control features were used to normalize the data for labeling and hybridization variability within the experiment. The control features are cDNA for genes from the plant, *Arabidopsis thaliana*, that were included when spotting the mini-array. Equal amounts of RNA complementary to two of these cDNAs were added to each of the samples before they were labeled. A third was pre-labeled and equal amounts were added to each hybridization solution before hybridization. Using the signal from these genes, we derived a normalization constant (L_j) according to the following formula:

$$L_j = \frac{\sum_{i=1}^N BGSS_{j,i}}{\sum_{j=1}^K \frac{\sum_{i=1}^N BGSS_{j,i}}{N}}$$

where BGSS_j is the signal for a specific feature as identified in the GenePix software as the median background subtracted signal for that feature, N is the number of *A. thaliana* control features, K is the number of hybridizations, and L is the normalization constant for each individual hybridization. Using the formula above, the mean over all control features of a particular hybridization and dye (eg Cy3) was calculated. Then these control feature means for all Cy3 hybridizations were averaged. The control feature mean in one hybridization divided by the average of all hybridizations gives a normalization constant for that particular Cy3 hybridization.

[0340] The same normalization steps were performed for Cy3 and Cy5 values, both fluorescence and background. Once normalized, the background Cy3 fluorescence was subtracted from the Cy3 fluorescence for each feature. Values less than 100 were eliminated from further calculations since low values caused spurious results.

[0341] Figure 4 shows the average background subtracted signal for each of nine leukocyte-specific genes on the mini array. This average is for 3-6 of the above-described hybridizations for each gene. The error bars are the SEM.

[0342] The ratio of Cy3 to Cy5 signal is shown for a number of genes. This ratio corrects for variability among hybridizations and allows comparison between experiments done at different times. The ratio is calculated as the Cy3 background subtracted signal divided by the Cy5 background subtracted signal. Each bar is the average for 3-6 hybridizations. The error bars are SEM.

[0343] Together, these results show that we can measure expression levels for genes that are expressed specifically in sub-populations of leukocytes. These expression measurements were made with only 10 µg of leukocyte total RNA

that was labeled directly by reverse transcription. The signal strength can be increased by improved labeling techniques that amplify either the starting RNA or the signal fluorescence. In addition, scanning techniques with higher sensitivity can be used.

[0344] Genes in Figures 4 and 5:

Gene Name/Description	GenBank Accession Number	Gene Name Abbreviation
T cell-specific tyrosine kinase Mrna	L10717	TKTCS
Interleukin 1 alpha (IL 1) mRNA, complete cds	NM_00057S	IL1A
T-cell surface antigen CD2 (T11) mRNA, complete cds	M14362	CD2
Interleukin-13 (IL-13) precursor gene, complete cds	U31120	IL-13
Thymocyte antigen CD1a mRNA, complete cds	M28825	CD1a
CD6 mRNA for T cell glycoprotein CDS	NM_006725	CD6
MHC class II HLA-DQA1 mRNA, complete cds	U77589	HLA-DQA1
Granulocyte colony-stimulating factor	M28170	CD19
Homo sapiens CD69 antigen	NM_001781	CD69

Example 5: Clinical study to identify diagnostic gene sets useful in diagnosis and treatment of cardiac allograft recipients

[0345] An observational study was conducted in which a prospective cohort of cardiac transplant recipients were analyzed for associations between clinical events or rejection grades and expression of a leukocyte candidate nucleotide sequence library. Patients were identified at 4 cardiac transplantation centers while on the transplant waiting list or during their routing post-transplant care. All adult cardiac transplant recipients (new or re-transplants) who received an organ at the study center during the study period or within 3 months of the start of the study period were eligible. The first year after transplantation is the time when most acute rejection occurs and it is thus important to study patients during this period. Patients provided informed consent prior to study procedures.

[0346] Peripheral blood leukocyte samples were obtained from all patients at the following time points: prior to transplant surgery (when able), the same day as routinely scheduled screening biopsies, upon evaluation for suspected acute rejection (urgent biopsies), on hospitalization for an acute complication of transplantation or immunosuppression, and when Cytomegalovirus (CMV) infection was suspected or confirmed. Samples were obtained through a standard peripheral vein blood draw or through a catheter placed for patient care (for example, a central venous catheter placed for endocardial biopsy). When blood was drawn from an intravenous line, care was taken to avoid obtaining heparin with the sample as it can interfere with downstream reactions involving the RNA. Mononuclear cells were prepared from whole blood samples as described in Example 2. Samples were processed within 2 hours of the blood draw and DNA and serum were saved in addition to RNA. Samples were stored at -80° C or on dry ice and sent to the site of RNA preparation in a sealed container with ample dry ice. RNA was isolated from subject samples as described in Example 2 and hybridized to a candidate library of differentially expressed leukocyte nucleotide sequences, as further described in Examples 9-10. Methods used for amplification, labeling, hybridization and scanning are described in Example 11. Analysis of human transplant patient mononuclear cell RNA hybridized to a microarray and identification of diagnostic gene sets is shown in Example 10.

[0347] From each patient, clinical information was obtained at the following time points: prior to transplant surgery (when available), the same day as routinely scheduled screening biopsies, upon evaluation for suspected acute rejection (e.g., urgent biopsies), on hospitalization for an acute complication of transplantation or immunosuppression, and when Cytomegalovirus (CMV) infection was suspected or confirmed. Data was collected directly from the patient, from the patient's medical record, from diagnostic test reports or from computerized hospital databases. It was important to collect all information pertaining to the study clinical correlates (diagnoses and patient events and states to which expression data is correlated) and confounding variables (diagnoses and patient events and states that may result in altered leukocyte gene expression. Examples of clinical data collected are: patient sex, date of birth, date of transplant, race, requirement for prospective cross match, occurrence of pre-transplant diagnoses and complications, indication for transplantation, severity and type of heart disease, history of left ventricular assist devices, all known medical diagnoses, blood type, HLA type, viral serologies (including CMV, Hepatitis B and C, HIV and others), serum chemistries, white and red blood cell counts and differentials, CMV infections (clinical manifestations and methods of diagnosis), occurrence of new cancer, hemodynamic parameters measured by catheterization of the right or left heart (measures of graft function), results of echocardiography, results of coronary angiograms, results of intravascular ultrasound studies (diagnosis of

transplant vasculopathy), medications, changes in medications, treatments for rejection, and medication levels. Information was also collected regarding the organ donor, including demographics, blood type, HLA type, results of screening cultures, results of viral serologies, primary cause of brain death, the need for inotropic support, and the organ cold ischemia time.

[0348] Of great importance was the collection of the results of endocardial biopsy for each of the patients at each visit. Biopsy results were all interpreted and recorded using the international society for heart and lung transplantation (ISHLT) criteria, described below. Biopsy pathological grades were determined by experienced pathologists at each center.

ISHLT Criteria

Grade	Finding	Rejection Severity
0	No lymphocytic infiltrates	None
1A	Focal (perivascular or interstitial lymphocytic infiltrates without necrosis)	Borderline mild
1B	Diffuse but sparse lymphocytic infiltrates without necrosis	Mild
2	One focus only with aggressive lymphocytic infiltrate and/or myocyte damage	Mild, focal moderate
3A	Multifocal aggressive lymphocytic infiltrates and/or myocardial damage	Moderate
3B	Diffuse inflammatory lymphocytic infiltrates with necrosis	Borderline Severe
4	Diffuse aggressive polymorphous lymphocytic infiltrates with edema hemorrhage and vasculitis, with necrosis	Severe

[0349] Because variability exists in the assignment of ISHLT grades, it was important to have a centralized and blinded reading of the biopsy slides by a single pathologist. This was arranged for all biopsy slides associated with samples in the analysis. Slides were obtained and assigned an encoded number. A single pathologist then read all slides from all centers and assigned an ISHLT grade. Grades from the single pathologist were then compared to the original grades derived from the pathologists at the study centers. For the purposes of correlation analysis of leukocyte gene expression to biopsy grades, the centralized reading information was used in a variety of ways (see Example 10 for more detail). In some analyses, only the original reading was used as an outcome. In other analyses, the result from the centralized reader was used as an outcome. In other analyses, the highest of the 2 grades was used. For example, if the original assigned grade was 0 and the centralized reader assigned a 1A, then 1A was the grade used as an outcome. In some analyses, the highest grade was used and then samples associated with a Grade 1A reading were excluded from the analysis. In some analyses, only grades with no disagreement between the 2 readings were used as outcomes for correlation analysis.

[0350] Clinical data was entered and stored in a database. The database was queried to identify all patients and patient visits that meet desired criteria (for example, patients with > grade II biopsy results, no CMV infection and time since transplant < 12 weeks).

[0351] The collected clinical data (disease criteria) is used to define patient or sample groups for correlation of expression data. Patient groups are identified for comparison, for example, a patient group that possesses a useful or interesting clinical distinction, versus a patient group that does not possess the distinction. Examples of useful and interesting patient distinctions that can be made on the basis of collected clinical data are listed here:

1. Rejection episode of at least moderate histologic grade, which results in treatment of the patient with additional corticosteroids, anti-T cell antibodies, or total lymphoid irradiation.
2. Rejection with histologic grade 2 or higher.
3. Rejection with histologic grade <2.
4. The absence of histologic rejection and normal or unchanged allograft function (based on hemodynamic measurements from catheterization or on echocardiographic data).
5. The presence of severe allograft dysfunction or worsening allograft dysfunction during the study period (based on hemodynamic measurements from catheterization or on echocardiographic data).
6. Documented CMV infection by culture, histology, or PCR, and at least one clinical sign or symptom of infection.
7. Specific graft biopsy rejection grades
8. Rejection of mild to moderate histologic severity prompting augmentation of the patient's chronic immunosuppressive regimen
9. Rejection of mild to moderate severity with allograft dysfunction prompting plasmaphoresis or a diagnosis of "humoral" rejection

EP 2 194 145 B1

10. Infections other than CMV, esp. Epstein Barr virus (EBV)
11. Lymphoproliferative disorder (also called, post-transplant lymphoma)
12. Transplant vasculopathy diagnosed by increased intimal thickness on intravascular ultrasound (IVUS), angiography, or acute myocardial infarction.
13. Graft Failure or Retransplantation
14. All cause mortality
15. Grade 1A or higher rejection as defined by the initial biopsy reading.
16. Grade 1B or higher rejection as defined by the initial biopsy reading.
17. Grade 1A or higher rejection as defined by the centralized biopsy reading.
18. Grade 1B or higher rejection as defined by the centralized biopsy reading.
19. Grade 1A or higher rejection as defined by the highest of the initial and centralized biopsy reading.
20. Grade 1B or higher rejection as defined by the highest of the initial and centralized biopsy reading.
21. Any rejection > Grade 2 occurring in patient at any time in the post-transplant course.

[0352] Expression profiles of subject samples are examined to discover sets of nucleotide sequences with differential expression between patient groups, for example, by methods describes above and below. Non-limiting examples of patient leukocyte samples to obtain for discovery of various diagnostic nucleotide sets are as follows:

Leukocyte set to avoid biopsy or select for biopsy:

Samples : Grade 0 vs. Grades 1-4

Leukocyte set to monitor therapeutic response:

Examine successful vs. unsuccessful drug treatment.

Samples:

Successful: Time 1: rejection, Time 2: drug therapy Time 3: no rejection

Unsuccessful: Time 1: rejection, Time 2: drug therapy; Time 3: rejection

Leukocyte set to predict subsequent acute rejection. Biopsy may show no rejection, but the patient may develop rejection shortly thereafter. Look at profiles of patients who subsequently do and do not develop rejection.

Samples:

Group 1 (Subsequent rejection): Time 1: Grade 0; Time 2: Grade>0

Group 2 (No subsequent rejection): Time 1: Grade 0; Time 2: Grade 0

Focal rejection may be missed by biopsy. When this occurs the patient may have a Grade 0, but actually has rejection. These patients may go on to have damage to the graft etc.

Samples:

Non-rejectors: no rejection over some period of time

Rejectors: an episode of rejection over same period

Leukocyte set to diagnose subsequent or current graft failure:

Samples:

Echocardiographic or catheterization data to define worsening function over time and correlate to profiles.

Leukocyte set to diagnose impending active CMV:

Samples:

Look at patients who are CMV IgG positive. Compare patients with subsequent (to a sample) clinical CMV

infection verses no subsequent clinical CMV infection.

Leukocyte set to diagnose current active CMV:

Samples:

Analyze patients who are CMV IgG positive. Compare patients with active current clinical CMV infection vs. no active current CMV infection.

[0353] Upon identification of a nucleotide sequence or set of nucleotide sequences that distinguish patient groups with a high degree of accuracy, that nucleotide sequence or set of nucleotide sequences is Validated, and implemented as a diagnostic test. The use of the test depends on the patient groups that are used to discover the nucleotide set. For example, if a set of nucleotide sequences is discovered that have collective expression behavior that reliably distinguishes patients with no histological rejection or graft dysfunction from all others, a diagnostic is developed that is used to screen patients for the need for biopsy. Patients identified as having no rejection do not need biopsy, while others are subjected to a biopsy to further define the extent of disease. In another example, a diagnostic nucleotide set that determines continuing graft rejection associated with myocyte necrosis (> grade I) is used to determine that a patient is not receiving adequate treatment under the current treatment regimen. After increased or altered immunosuppressive therapy, diagnostic profiling is conducted to determine whether continuing graft rejection is progressing. In yet another example, a diagnostic nucleotide set(s) that determine a patient's rejection status and diagnose cytomegalovirus infection is used to balance immunosuppressive and anti-viral therapy.

[0354] The methods of this example are also applicable to cardiac xenograft monitoring.

Example 6: Identification of diagnostic nucleotide sets for kidney and liver allograft rejection

[0355] Diagnostic tests for rejection are identified using patient leukocyte expression profiles to identify a molecular signature correlated with rejection of a transplanted kidney or liver. Blood, or other leukocyte source, samples are obtained from patients undergoing kidney or liver biopsy following liver or kidney transplantation, respectively. Such results reveal the histological grade, i.e., the state and severity of allograft rejection. Expression profiles are obtained from the samples as described above, and the expression profile is correlated with biopsy results. In the case of kidney rejection, clinical data is collected corresponding to urine output, level of creatine clearance, and level of serum creatine (and other markers of renal function). Clinical data collected for monitoring liver transplant rejection includes, biochemical characterization of serum markers of liver damage and function such as SGOT, SGPT, Alkaline phosphatase, GGT, Bilirubin, Albumin and Prothrombin time.

[0356] Leukocyte nucleotide sequence expression profiles are collected and correlated with important clinical states and outcomes in renal or hepatic transplantation. Examples of useful clinical correlates are given here:

1. Rejection episode of at least moderate histologic grade, which results in treatment of the patient with additional corticosteroids, anti-T cell antibodies, or total lymphoid irradiation.
2. The absence of histologic rejection and normal or unchanged allograft function (based on tests of renal or liver function listed above).
3. The presence of severe allograft dysfunction or worsening allograft dysfunction during the study period (based on tests of renal and hepatic function listed above).
4. Documented CMV infection by culture, histology, or PCR, and at least one clinical sign or symptom of infection.
5. Specific graft biopsy rejection grades
6. Rejection of mild to moderate histologic severity prompting augmentation of the patient's chronic immunosuppressive regimen
7. Infections other than CMV, esp. Epstein Barr virus (EBV)
8. Lymphoproliferative disorder (also called, post-transplant lymphoma)
9. Graft Failure or Retransplantation
10. Need for hemodialysis or other renal replacement therapy for renal transplant patients.
11. Hepatic encephalopathy for liver transplant recipients.
12. All cause mortality

[0357] Subsets of the candidate library (or of a previously identified diagnostic nucleotide set), are identified, according to the above procedures, that have predictive and/or diagnostic value for kidney or liver allograft rejection.

Example 7: Identification of a diagnostic nucleotide set for diagnosis of cytomegalovirus

[0358] Cytomegalovirus is a very important cause of disease in immunocompromised patients, for example, transplant patients, cancer patients, and AIDS patients. The virus can cause inflammation and disease in almost any tissue (particularly the colon, lung, bone marrow and retina). It is increasingly important to identify patients with current or impending clinical CMV disease, particularly when immunosuppressive drugs are to be used in a patient, e.g. for preventing transplant rejection. Leukocytes are profiled in patients with active CMV, impending CMV, or no CMV. Expression profiles correlating with diagnosis of active or impending CMV are identified. Subsets of the candidate library (or a previously identified diagnostic nucleotide set) are identified, according to the above procedures that have predictive value for the diagnosis of active or impending CMV. Diagnostic nucleotide set(s) identified with predictive value for the diagnosis of active or impending CMV may be combined, or used in conjunction with, cardiac, liver and/or kidney allograft-related diagnostic gene set(s) (described in Examples 6 and 10).

[0359] In addition, or alternatively, CMV nucleotide sequences are obtained, and a diagnostic nucleotide set is designed using CMV nucleotide sequence. The entire sequence of the organism is known and all CMV nucleotide sequences can be isolated and added to the library using the sequence information and the approach described below. Known expressed genes are preferred. Alternatively, nucleotide sequences are selected to represent groups of CMV genes that are coordinately expressed (immediate early genes, early genes, and late genes) (Spector et al. 1990, Stamminger et al. 1990).

[0360] Oligonucleotides were designed for CMV genes using the oligo design procedures of Example 8. Probes were designed using the 14 gene sequences shown here and were included on the array described in example 9:

Cytomegalovirus (CMV) Accession #X17403	HCMVTRL2 (IRL2)	1893..2240
	HCMVTRL7 (IRL7)	complement(6595..6843)
	HCMVUL21	complement(26497..27024)
	HCMVUL27	complement(32831..34657)
	HCMVUL33	43251..44423
	HCMVUL54	complement(76903..80631)
	HCMVUL75	complement(107901..110132)
	HCMVUL83	complement(119352..121037)
	HCMVUL106	complement(154947..155324)
	HCMVUL109	complement(157514..157810)
	HCMVUL113	161503..162800
	HCMVUL122	complement(169364..170599)
	HCMVUL123 (last exon at 3'-end)	complement(171006..172225)
	HCMVUS28	219200..220171

[0361] Diagnostic nucleotide set(s) for expression of CMV genes is used in combination with diagnostic leukocyte nucleotide sets for diagnosis of other conditions, e.g. organ allograft rejection.

[0362] Using the techniques described in example 2 mononuclear samples from 180 cardiac transplant recipients (enrolled in the study described in Example 5) were used for expression profiling with the leukocyte arrays. Of these samples 15 were associated with patients who had a diagnosis of primary or reactivation CMV made by culture, PCR or any specific diagnostic test.

[0363] After preparation of RNA, amplification, labeling, hybridization, scanning, feature extraction and data processing were done as described in Example 11 using the oligonucleotide microarrays described in

Example 9

[0364] The resulting log ratio of expression of Cy3 (patient sample)/ Cy5 (R50 reference RNA) was used for analysis. Significance analysis for microarrays (SAM, Tusher 2001, see Example 15) was applied to determine which genes were most significantly differentially expressed between these 15 CMV patients and the 165 non-CMV patients. 12 genes were identified with a 0% FDR and 6 with a 0.1% FDR and are listed in Table 2. Some genes are represented by more than one oligonucleotide on the array and for 2 genes, multiple oligonucleotides from the same gene are called significant (SEQ IDs 3061, 3064 eomesodermin and 3031, 3040, 104, 2736 small inducible cytokine A4).

[0365] Clinical variables were also included in the significance analysis. For example, the white blood cell count and the number of weeks post transplant (for the patient at the time the sample was obtained) were available for most of the

180 samples The log of these variables was taken and the variables were then used in the significance analysis described above with the gene expression data Both the white blood cell count (01% FDR) and the weeks post transplant (0% FDR) appeared to correlate with CMV status CMV patients were more likely to have samples associated with later post transplant data and the lower white blood cell counts

[0366] These genes and variables can be used alone or in association with other genes or variables or with other genes to build a diagnostic gene set or a classification algorithm using the approaches described herein

[0367] Primers for real-time PCR validation were designed for some of these genes as described in Example 13 and listed in Table 2C and the sequence listing. Using the methods described in example 13, primers for Granzyme B were designed and used to validate expression findings from the arrays. 6 samples were tested (3 from patients with CMV and 3 from patients without CMV) The gene was found to be differentially expressed between the patients with and without CMV (see example 13 for full description) This same approach can be used to validate other diagnostic genes by real-time PCR Diagnostic nucleotide sets can also be identified for a variety of other viral diseases (Table 1) using this same approach

[0368] cDNA microarrays may be used to monitor viral expression In addition, these methods may be used to monitor other viruses, such as Epstein-Barr virus, Herpes Simplex 1 and vesicular stomatitis virus

Example 8- Design of oligonucleotide probes

[0369] By way of example, this section describes the design of four oligonucleotide probes using Array Designer Ver 11 (Premier Biosoft International, Palo Alto, CA) The major steps in the process are given first

[0370] Obtain best possible sequence of mRNA from GenBank If a full-length sequence reference sequence is not available, a partial sequence is used, with preference for the 3' end over the 5' end When the sequence is known to represent the antisense strand, the reverse complement of the sequence is used for probe design. For sequences represented in the subtracted leukocyte expression library that have no significant match in GenBank at the time of probe design, our sequence is used.

[0371] Mask low complexity regions and repetitive elements in the sequence using an algorithm such as RepeatMasker.

[0372] Use probe design software, such as Array Designer, version 1.1, to select a sequence of 50 residues with specified physical and chemical properties. The 50 residues nearest the 3' end constitute a search frame. The residues it contains are tested for suitability. If they don't meet the specified criteria, the search frame is moved one residue closer to the 5' end, and the 50 residues it now contains are tested. The process is repeated until a suitable 50-mer is found.

[0373] If no such 50-mer occurs in the sequence, the physical and chemical criteria are adjusted until a suitable 50-mer is found.

[0374] Compare the probe to dbEST, the UniGene cluster set, and the assembled human genome using the BLASTn search tool at NCBI to obtain the pertinent identifying information and to verify that the probe does not have significant similarity to more than one known gene.

Clone 40H 12

[0375] Clone 40H12 was sequenced and compared to the nr, dbEST, and UniGene databases at NCBI using the BLAST search tool. The sequence matched accession number NM_002310, a 'curated RefSeq project' sequence, see Pruitt et al. (2000) Trends Genet. 16:44-47, encoding leukemia inhibitory factor receptor (LIFR) mRNA with a reported E value of zero. An E value of zero indicates there is, for all practical purposes, no chance that the similarity was random based on the length of the sequence and the composition and size of the database. This sequence, cataloged by accession number NM_002310, is much longer than the sequence of clone 40H 12 and has a poly-A tail. This indicated that the sequence cataloged by accession number NM_002310 is the sense strand and a more complete representation of the mRNA than the sequence of clone 40H12, especially at the 3' end. Accession number "NM_002310" was included in a text file of accession numbers representing sense strand mRNAs, and sequences for the sense strand mRNAs were obtained by uploading a text file containing desired accession numbers as an Entrez search query using the Batch Entrez web interface and saving the results locally as a FASTA file. The following sequence was obtained, and the region of alignment of clone 40H12 is outlined:

CTCTCTCCCAGAACGTGTCTCTGCTGCAAGGCACCGGGCCCTTTCGCTCTGCAGAACTGCACTTGCAAGA
CCATTATCAACTCCTAATCCCAGCTCAGAAAGGGAGCCTCTGCGACTCATTTCATCGCCCTCCAGGACTGA
CTGCATTGCACAGATGATGGATATTTACGTATGTTTGAAACGACCATCCTGGATGGTGGACAATAAAAGA
ATGAGGACTGCTTCAAATTTCCAGTGGCTGTTATCAACATTTATTCTTCTATATCTAATGAATCAAGTAA
ATAGCCAGAAAAAGGGGGCTCCTCATGATTTGAAGTGTGTAACATAACAATTTGCAAGTGTGGAAGTGTTC
TTGGAAAGCACCCCTCTGGAACAGGCCGTGGTACTGATTATGAAGTTTGCATTGAAAACAGGTCCCGTTCT
TGTTATCAGTTGGAGAAAACCAGTATTAAAATTCCAGCTCTTTCACATGGTGATTATGAAATAACAATAA
ATTCTCTACATGATTTTGGAAAGTTCTACAAGTAAATTCACACTAAATGAACAAAACGTTTCCTTAATTCC
AGATACTCCAGAGATCTTGAATTTGTCTGCTGATTTCTCAACCTCTACATTATACCTAAAGTGGAACGAC
AGGGGTTTCAGTTTTTCCACACCGCTCAAATGTTATCTGGGAAATTAAAGTTCTACGTAAAGAGAGTATGG
AGCTCGTAAATTAGTGACCCACAACACAACCTCTGAATGGCAAAGATACACTTCATCACTGGAGTTGGGC
CTCAGATATGCCCTTGGAATGTGCCATTCAATTTGTGGAAATTAGATGCTACATTGACAATCTTCATTTT
TCTGGTCTCGAAGAGTGGAGTGACTGGAGCCCTGTGAAGAACATTTCTTGGATACCTGATTCTCAGACTA
AGGTTTTTCCTCAAGATAAAGTGATACTTGTAGGCTCAGACATAACATTTTGTGTGTGAGTCAAGAAAA

AGTGTATATCAGCACTGATTGGCCATACAACTGCCCTTGATCCATCTTGATGGGGAAAATGTTGCAATC
 AAGATTCGTAATATTTCTGTTTCTGCAAGTAGTGGAACAAATGTAGTTTTTACAACCGAAGATAACATAT
 TTGGAACCGTTATTTTTGCTGGATATCCACCAGATACTCCTCAACAACGAATTTGTGAGACACATGATTT
 5 AAAAGAAATTATATGTAGTTGGAATCCAGGAAGGGTGACAGCGTTGGTGGGCCACGTGCTACAAGCTAC
 ACTTTAGTTGAAAGTTTTTCAGGAAAATATGTTAGACTTAAAAGAGCTGAAGCACCTACAAACGAAAGCT
 ATCAATTATTATTTCAAATGCTTCCAAATCAAGAAATATATAATTTTACTTTGAATGCTCACAAATCCGCT
 GGGTCGATCACAAATCAACAATTTTAGTTAATATAACTGAAAAAGTTTATCCCCATACCTCTACTTTCATT
 AAAGTGAAGGATATTAATTCAACAGCTGTTAACTTTCTTGGCATTACCAGGCAACTTTGCAAAGATTA
 ATTTTTTATGTGAAATTGAAATTAAGAAATCTAATTCAGTACAAGAGCAGCGGAATGTCACAATCAAAGG
 10 AGTAGAAAATTCAAGTTATCTTGTTGCTCTGGACAAGTTAAATCCATACACTCTATATACTTTTCGGATT
 CGTTGTTCTACTGAAACTTTCTGGAAATGGAGCAAATGGAGCAATAAAAAACAACATTTAACAACAGAAG
 CCAGTCCTTCAAAGGGGCTGATACTTGGAGAGAGTGGAGTTCTGATGGAAAAATTTAATAATCTATTG
 GAAGCCTTTACCCATTAAATGAAGCTAATGGAAAAATACTTTCTTACAATGTATCGTGTTTCATCAGATGAG
 GAAACACAGTCCCTTTCTGAAATCCCTGATCCTCAGCACAAAGCAGAGATACGACTTGATAAGAATGACT
 ACATCATCAGCGTAGTGGCTAAAAATCTGTGGGCTCATCACCACCTTCCAAAATACCGAGTATGGAAT
 15 TCCAAATGATGATCTCAAAATAGAACAAAGTTGTTGGGATGGGAAAGGGGATTCTCCTCACCCTGGCATTAC
 GACCCCAACATGACTTGGGACTACGTCATTAAAGTGGTGAATCTGTAATAGAACTCGTCTCGGTCCGAACCATGCCCTTATGG
 ACTGGAGAAAAAGTTCCCTCAAAACAGCACTGAAACTGTAATAGAACTGATGAGTTTCGACCAGGTATAAG
 ATATAATTTTTTCTGTATGGATGCAGAAATCAAGGATATCAATTATTACGCTCCATGATTGGATATATA
 GAAGAATTGGCTCCCATTTGTGCACCAAATTTTACTGTTGAGGATACTTCTGCAGATTCGATATTAGTAA
 20 AATGGGAAGACATTCTGTGGAAGAACTTAGAGGCTTTTTAAGAGGATATTTGTTTTACTTTGGAAAAGG
 AGAAAGAGACACATCTAAGATGAGGGTTTTAGAAATCAGGTCGTTCTGACATAAAAGTTAAGAATATTACT
 GACATATCCCAGAGACACTGAGAATTGCTGATCTTCAAGGTAAACAAGTTACCACCTGGTCTTGCGAG
 CCTATACAGATGGTGGAGTGGGCCCCGAGAAGAGTATGTATGTGGTGACAAAGGAAAATTCGTGGGATT
 AATTATTGCCATTCTCATCCAGTGGCAGTGGCTGTCTATTGTTGGAGTGGTGACAAGTATCCTTTGCTAT
 CGGAAACGAGAATGGATTAAAGAAACCTTCTACCTGTATATTCCAAATCCAGAAAACGTAAAGCATTAC
 25 AGTTTCAAAAGAGTGCTGTGTGAGGGAAGCAGTGCTCTTAAACATTGGAAATGAATCCTTTGTACCCCAA
 TAATGTTGAGGTTCTGGAAGTCTGATCAGCATTCTCTAAATAGAAAGATACAGAAATAATTTCCCCAGTA
 GCTGAGCGTCTGAAGATCGCTCTGATGCAGAGCCTGAAAACCATGTGGTTGTGTCTATTGTCCACCCA
 TCATTGAGGAAGAAATACCAAACCCAGCCGAGATGAAGCTGGAGGGACTGCACAGGTTATTTACATTGA
 TGTTCACTCGATGTATCAGCCTCAAGCAAAACCAGAAGAACAAGAAAATGACCCTGTAGGAGGGGCA
 30 GGCTATAAGCCACAGATGCACCTCCCCATTAATTCTACTGTGGAAGATATAGCTGCAGAAGAGGACTTAG
 ATAAACTGCGGGTTACAGACCTCAGGCCAATGTAAATACATGGAATTTAGTGTCTCCAGACTCTCCTAG
 ATCCATAGACAGCAACAGTGAGATTGTCTCATTTGGAAGTCCATGCTCCATTAATTTCCCGACAATTTTTG
 ATTCCTCTAAAGATGAAGACTCTCCTAAATCTAATGGAGGAGGGTGGTCTTTACAAACTTTTTTTCAGA
 ACAAAACCAACGATTAACAGTGTACCGTGTCACTTCACTCAGCCATCTCAATAAGCTCTTACTGTCTAGT
 35 GTTGCTACATCAGCACTGGGCATTCTTGGAGGGATCCTGTGAAGTATTGTTAGGAGGTGAACCTTCACTAC
 ATGTTAAGTTACACTGAAAGTTCATGTGCTTTTAAATGTAGTCTAAAAGCCAAAGTATAGTGACTCAGAAT
 CCTCAATCCACAAAACCTCAAGATTGGGAGCTCTTTGTGATCAAGCCAAAGAATCTCATGTACTCTACCT
 TCAAGAAGCATTTCAGGCTAATACCTACTTGTACGTACATGTAAAACAAATCCCGCCGCAACTGTTTTTC
 TGTTCTGTGTTGTGTTTCTCATATGTATACTTGGTGGAAATTGTAAGTGGATTGTCAGGCCAGGGAG
 40 AAAATGTCCAAGTAACAGGTGAAGTTTATTTGCCTGACGTTTACTCCTTTCTAGATGAAAACCAAGCACA
 GATTTTAAACTTCTAAGATTATTTCTCTCTATCCACAGCATTACAAAAATTAATATAATTTTTAATGT
 AGTGACAGCGATTTAGTGTTTTGTTTGATAAAGTATGCTTATTTCTGTGCCTACTGTATAATGGTTATCA
 AACAGTTGTCTCAGGGGTACAACTTTGAAAAAAGTGTGACACTGACCAGCCCAAATCATAATCATGTT
 TTCTTGCTGTGATAGGTTTTGCTTGCCTTTTCATTATTTTTTAGCTTTTATGCTTGCTTCCATTATTTCA
 GTTGGTTGCCCTAATATTTAAAAATTTACACTTCTAAGACTAGAGACCCACATTTTTTAAAAATCATTTTA
 45 TTTTGTGATACAGTGACAGCTTTATATGAGCAAATTCATATATTATTCATAAGCATGTAATTCAGTGACT
 TACTATGTGAGATGACTACTAAGCAATATCTAGCAGCGTTAGTTCCATATAGTTCTGATTGGATTTCGTT
 CCTCCTGAGGAGACCATGCCGTTGAGCTTGGCTACCCAGGCAGTGGTGATCTTTGACACCTTCTGGTGGA
 TGTTCTCTCCACTCATGAGTCTTTTCATCATGCCACATTATCTGATCCAGTCTCACATTTTAAATATA
 AAATAAAGAGAGAATGCTTCTTACAGGAACAGTTACCCAAGGGCTGTTTCTTAGTAAGTGTGATAAAT
 50 GATCTGGATCCATGGGCATACCTGTGTTGAGGTGCAGCAATTGCTTGGTGAGCTGTGCAGAATTGATTG
 CTTTCAGCACAGCATCCTCTGCCCACCTTGTCTCTATAAGCGATGTCTGGAGTGATTGTGGTTCTTGG
 AAAAGCAGAAGGAAAACTAAAAAGTGTATCTGTATTTTCCCTGCTCAGGTTGCCATGTATTTTAC
 CTTTTCATATTTAAGGCAAAAGTACTGAAAAATTTAAGTGTCCGAATAAGATATGTCTTTTTTGTGTTGT
 55 TTTTTTGGTTGGTTGTTTGTGTTTTATCATCTGAGATTCTGTAATGTATTGTGAAATAATGGATCAATT
 AATTTTTTTTGAAGCTCATATTGTATCTTTTTAAAAAACCATGTTGTGGAAGGAGGAGTGACAAGTG
 ACAAATCTATTTAGGAACCTGTGTATGAATCCTGATTTTAACTGCTAGGATTGAGCTAAATTTCTGAG
 CTTTATGATCTGTGGAATTTGGAATGAAATCGAATTCATTTGTACATACATAGTATATTAAACTATA

TAATAGTTCATAGAAATGTTTCAGTAATGAAAAATATATCCAATCAGAGCCATCCCGAAAAAAAAAAAAAA
AA (SEQ ID NO: 3101)

- 5 The FASTA file, including the sequence of NM_002310, was masked using the RepeatMasker web interface (Smit, AFA & Green, P RepeatMasker at <http://ftp.genome.washington.edu/RM/RepeatMasker.html>, Smit and Green). Specifically, during masking, the following types of sequences were replaced with "N's": SINE/MIR & LINE/L2, LINE/L1, LTR/MaLR, LTR/Retroviral, Alu, and other low informational content sequences such as simple repeats. Below is the sequence following masking:

10

15

20

25

30

35

40

45

50

55

CTCTCTCCAGAACGTGTCTCTGCTGCAAGGCACCGGGCCCTTTGCTCTGCAGAACTGCACTTGCAAG
ACCATTATCAACTCCTAATCCCAGCTCAGAAAGGGAGCCTCTGCGACTCATTCATCGCCCTCCAGGACT
GACTGCATTGCACAGATGATGGATATTTACGTATGTTTGAACGACCATCCTGGATGGTGGACAATAAA
AGAATGAGGACTGCTTCAAATTTCCAGTGGCTGTTATCAACATTTATTCTTCTATATCTAATGAATCAA
GTAAATAGCCAGAAAAAGGGGGCTCCTCATGATTTGAAGTGTGTAACAAATTTGCAAGTGTGGAAC
TGTTCTTGGAAGCACCCCTCTGGAACAGGCCGTGGTACTGATTATGAAGTTTGCAATTGAAAACAGGTCC
CGTTCTTGTTATCAGTTGGAGAAAACCAGTATTAATAATTCAGCTCTTTCACATGGTGATTATGAAATA
ACAATAAATTCTCTACATGATTTTGGAAGTTCTACAAGTAAATTCACACTAAATGAACAAAACGTTTCC
TTAATTCCAGATACTCCAGAGATCTTGAATTTGTCTGCTGATTTCTCAACCTCTACATTATACCTAAAG
TGGAACGACAGGGGTTTCAGTTTTTCCACACCGCTCAAATGTTATCTGGGAAATTTAAAGTTCTACGTAAA
GAGAGTATGGAGCTCGTAAAATTAGTGACCCACAACAACCTCTGAATGGCAAAGATACACTTCATCAC
TGGAGTTGGGCCTCAGATATGCCCTTGGAATGTGCCATTCATTTTGTGGAAATTAGATGCTACATTGAC
AATCTTCATTTTTCTGGTCTCGAAGAGTGGAGTGACTGGAGCCCTGTGAAGAACATTTCTTGATACCT
GATTCTCAGACTAAGGTTTTTCCTCAAGATAAAGTGATACTTGTAGGCTCAGACATAACATTTTGTGT
GTGAGTCAAGAAAAAGTGTATCAGCACTGATTGGCCATACAACTGCCCCCTTGATCCATCTTGATGGG
GAAAATGTTGCAATCAAGATTCGTAATATTTCTGTTTCTGCAAGTAGTGGAACAAATGTAGTTTTTACA
ACCGAAGATAACATATTTGGAACCGTTATTTTTGCTGGATATCCACCAGATACTCCTCAACAACCTGAAT
TGTGAGACACATGATTTAAAAGAAATTATATGTAGTTGGAATCCAGGAAGGGTGACAGCGTTGGTGGGC
CCACGTGCTACAAGCTACACTTTAGTTGAAAGTTTTTCAGGAAAATATGTTAGACTTAAAAGAGCTGAA
GCACCTACAAACGAAAGCTATCAATTATTATTTCAAATGCTTCAAATCAAGAAATATATAATTTTACT
TTGAATGCTCACAATCCGCTGGGTCGATCACAATCAACAATTTTAGTTAATATAACTGAAAAAGTTTAT
CCCCATACTCCTACTTCATTCAAAGTGAAGGATATTAATTCAACAGCTGTTAACTTTCTTGGCATTTA
CCAGGCAACTTTGCAAAGATTAATTTTTTATGTGAAATTGAAATTAAGAAATCTAATTCAGTACAAGAG
CAGCGGAATGTCACAATCAAAGGAGTAGAAAATTCAAGTTATCTTGTTGCTCTGGACAAGTTAAATCCA
TACACTCTATATACTTTTCGGATTCTGTTGTTCTACTGAACTTTCTGGAAATGGAGCAAATGGAGCAAT
AAAAACAACATTTAACAACAGAAGCCAGTCCTTCAAAGGGGCCTGATACTTGGAGAGAGTGGAGTTCT
GATGGAAAAAATTTAATAATCTATTGGAAGCCTTTACCCATTAATGAAGCTAATGGAAAAATACTTTCC
TACAATGTATCGTGTTTCATCAGATGAGGAAACACAGTCCCTTTCTGAAATCCCTGATCCTCAGCACAAA
GCAGAGATACGACTTGATAAGAATGACTACATCATCAGCGTAGTGGCTAAAAATCTGTGGGCTCATCA
CCACCTTCCAAAATAGCGAGTATGGAAATTCAAATGATGATCTCAAAATAGAACAAGTTGTTGGGATG
GGAAAGGGGATTCTCCTCACCTGGCATTACGACCCCAACATGACTTGGGACTACGTCATTAAGTGGTGT
AACTCGTCTCGGTCCGAACCATGCCTTATGGACTGGAGAAAAGTTCCCTCAAACAGCACTGAAACTGTA

ATAGAATCTGATGAGTTTCGACCAGGTATAAGATATAATTTTTCTCTGTATGGATGCAGAAATCAAGGA
TATCAATTATTACGCTCCATGATTGGATATATAGAAGAATTGGCTCCCATTTGTTGCACCAAATTTTACT
GTTGAGGATACTTCTGCAGATTTCGATATTAGTAAATGGAAGACATTCTCTGTGGAAGAACTTAGAGGC
TTTTTAAGAGGATATTTGTTTTACTTTGGAAAAGGAGAAAGAGACACATCTAAGATGAGGGTTTTAGAA
TCAGGTCGTTCTGACATAAAAGTTAAGAATATTACTGACATATCCAGAAGACACTGAGAATTGCTGTAT
CTTCAAGGTAAAACAAGTTACCACCTGGTCTTGCGAGCCTATACAGATGGTGGAGTGGGCCCCGAGAAG
AGTATGTATGTGGTGACAAAGGAAAATTCTGTGGGATTAATTATTGCCATTCTCATCCCAGTGGCAGTG
GCTGTCAATTGTTGGAGTGGTGACAAGTATCCTTTGCTATCGGAAACGAGAATGGATTAAAGAAACCTTC
TACCCTGATATTCCAAATCCAGAAAAGTGTAAAGCATTACAGTTTCAAAGAGTGTCTGTGAGGGAAGC
AGTGCTCTTAAACATTGGAAATGAATCCTTGTAACCCCAAATAATGTTGAGGTTCTGGAAACTCGATCA
GCATTTCTTAAATAGAAAGATACAGAAATAATTTCCCCAGTAGCTGAGCGTCTCTGAAGATCGCTCTGAT
GCAGAGCCTGAAAACCATGTGGTTGTGTCTATTGTCCACCCATCATTGAGGAAGAAATACCAAACCCA
GCCGCAGATGAAGCTGGAGGGACTGCACAGGTTATTTACATTGATGTTTCAGTCGATGTATCAGCCTCAA
GCAAAACCAGAAGAAGAAACAAGAAAATGACCTGTAGGAGGGGCAGGCTATAAGCCACAGATGCACCTC
CCCATTAAATTCTACTGTGGAAGATATAGCTGCAGAAGAGGACTTAGATAAAAAGTGGGGTTACAGACCT
CAGGCCAATGTAAATACATGGAATTTAGTGTCTCCAGACTCTCCTAGATCCATAGACAGCAACAGTGAG
ATTGTCTCATTTTGAAGTCCATGCTCCATTAATTTCCGACAATTTTTTGATTCTCTCTAAAGATGAAGAC
TCTCTAAATCTAATGGAGGAGGGTGGTCTTTACAAACTTTTTTTCAGAACAAACCAAACGATTAACAG
TGTCACCGTGTCACTTCAGTCAGCCATCTCAATAAGCTCTTACTGCTAGTGTGCTACATCAGCACTGG
GCATTTCTGGAGGGATCCTGTGAAGTATTGTTAGGAGGTGAACTTCACTACATGTTAAGTTACACTGAA
AGTTCATGTGCTTTTAATGTAGTCTAAAAGCCAAAGTATAGTGACTCAGAATCCTCAATCCACAAAACCT
CAAGATTGGGAGCTCTTTGTGATCAAGCCAAAGAAATTCTCATGTACTCTACCTTCAAGAAGCATTTCAA
GGCTAATACCTACTTGTACGTACATGTAAACAAATCCCGCCGCAACTGTTTTCTGTTCTGTTGTTTGT
GGTTTTCTCATATGTATACTTGGTGGAATTGTAAGTGGATTGTCAGGCCAGGGAGAAAATGTCCAAGTA
ACAGGTGAAGTTTATTTGCCTGACGTTTACTCCTTTCTAGATGAAAACCAAGCACAGATTTTAAACCTT
CTAAGATTATTCTCCTCTATCCACAGCATTACNNNNNNNNNNNNNNNNNNNNNNNGTAGTGACAGCGAT
TTAGTGTTTTGTTTGATAAAGTATGCTTATTTCTGTGCCTACTGTATAATGGTTATCAAACAGTTGTCT
CAGGGGTACAAAACCTTTGAAAACAAGTGTGACACTGACCAGCCCCAAATCATAATCATGTTTTCTTGCTGT
GATAGGTTTTGCTTGCCTTTTCATTATTTTTTAGCTTTTATGCTTGCTTCCATTATTTTCAGTTGGTTGC
CCTAATATTTAAATTTACACTTCTAAGACTAGAGACCCACATTTTTTAAAAATCATTTTATTTTGTGA
TACAGTGACAGCTTTATATGAGCAAATTCATATTTATTCATAAGCATGTAATTCCAGTGACTTACTATG
TGAGATGACTACTAAGCAATATCTAGCAGCGTTAGTTCCATATAGTTCTGATTGGATTTCGTTCTCTCT
GAGGAGACCATGCCGTTGAGCTTGGCTACCCAGGCAGTGGTGATCTTTGACACCTTCTGGTGGATGTTCT
CTCCCACTCATGAGTCTTTTCATCATGCCACATTATCTGATCCAGTCTCTACATTTTTAAATATAAAAAC
TAAAGAGAGAATGCTTCTTACAGGAACAGTTACCCAAGGGCTGTTTCTTAGTAACTGTCAAACTGAT
CTGGATCCATGGGCATACCTGTGTTTCGAGGTGCAGCAATTGCTTGGTGAGCTGTGCAGAATTGATTGCC
TTCAGCACAGCATCCTCTGCCACCCTTGTTTTCTCATAAGCGATGTCTGGAGTGATTGTGGTTCTTGGA
AAAGCAGAAGGAAAAACTAAAAAGTGTATCTTGATTTTCCCTGCCCTCAGGTTGCCTATGTATTTTAC
CTTTTCATATTTAAGGCAAAAGTACTTGAAAATTTAAGTGTCCGAATAAGATATGTCTTTTTTGTGTTG
TTTTTTTTGGTTGGTTGGTTGGTTTTTTTTATCATCTGAGATTCTGTAATGTATTTGCAAATAATGGATCA

TTAATTTTTTTTGAAGCTCATATTGTATCTTTTTAAAAACCATGTTGTGGAAAAAGCCAGAGTGACAA
 GTGACAAAATCTATTTAGGAACTCTGTGTATGAATCCTGATTTTAACTGCTAGGATTCAGCTAAATTTTC
 5 TGAGCTTTATGATCTGTGGAAATTTGGAATGAAATCGAATTCATTTTGTACATACATAGTATATTAAAA
 CTATATAATAGTTTCATAGAAATGTTTCAGTAATGAAAAATATATCCAATCAGAGCCATCCCGAAAAAAA
 AAAAAAA (SEQ ID NO: 3102).

[0376] The length of this sequence was determined using batch, automated computational methods and the sequence, as sense strand, its length, and the desired location of the probe sequence near the 3' end of the mRNA was submitted to Array Designer Ver 1.1 (Premier Biosoft International, Palo Alto, CA). Search quality was set at 100%, number of best probes set at 1, length range set at 50 base pairs, Target T_m set at 75 C. degrees plus or minus 5 degrees, Hairpin max deltaG at 6.0 -kcal/mol., Self dimmer max deltaG at 6.0 -kcal/mol, Run/repeat (dinucleotide) max length set at 5, and Probe site minimum overlap set at 1. When none of the 49 possible probes met the criteria, the probe site would be moved 50 base pairs closer to the 5' end of the sequence and resubmitted to Array Designer for analysis. When no possible probes met the criteria, the variation on melting temperature was raised to plus and minus 8 degrees and the number of identical basepairs in a run increased to 6 so that a probe sequence was produced.

[0377] In the sequence above, using the criteria noted above, Array Designer Ver 1.1 designed a probe corresponding to oligonucleotide number 3037 and is indicated by underlining in the sequence above. It has a melting temperature of 68.4 degrees Celsius and a max run of 6 nucleotides and represents one of the cases where the criteria for probe design in Array Designer Ver 1.1 were relaxed in order to obtain an oligonucleotide near the 3' end of the mRNA (Low melting temperature was allowed).

Clone 463D12

[0378] Clone 463D12 was sequenced and compared to the nr, dbEST, and UniGene databases at NCBI using the BLAST search tool. The sequence matched accession number AI184553, an EST sequence with the definition line "qd60a05.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1733840 3' similar to gb:M29550 PROTEIN PHOSPHATASE 2B CATALYTIC SUBUNIT 1 (HUMAN);, mRNA sequence." The E value of the alignment was 1.00×10^{-118} . The GenBank sequence begins with a poly-T region, suggesting that it is the antisense strand, read 5' to 3'. The beginning of this sequence is complementary to the 3' end of the mRNA sense strand. The accession number for this sequence was included in a text file of accession numbers representing antisense sequences. Sequences for antisense strand mRNAs were obtained by uploading a text file containing desired accession numbers as an Entrez search query using the Batch Entrez web interface and saving the results locally as a FASTA file. The following sequence was obtained, and the region of alignment of clone 463D12 is outlined:

TTTTTTTTTTTTTCTTAAATAGCATTTATTTTCTCTCAAAAAGCCTATTATGTACTAACAAGTGTTCC
 40 TCTAAATTAGAAAGGCATCACTACTAAAATTTTATACATATTTTTTATATAAGAGAAGGAATATTGGGT
 TACAATCTGAATTTCTCTTTATGATTTCTCTTAAAGTATAGAACAGCTATTAATAATGACTAATATTGCT
 AAAATGAAGGCTACTAAATTTCCCAAGAATTTCCGGTGAATGCCCAAAATGGTGTTAAGATATGCAG
 AAGGGCCCATTTCAGCAAAGCAATCTCTCCACCCCTTCATAAAAGATTTAAGCTAAAAAAAAAAAAA
 45
 AAGAAAGAAATCCAACAGCTGAAGACATTGGGCTATTTATAAATCTTCTCCAGTCCCCAGACAGCCT
 CACATGGGGGCTGTAAACAGCTAACTAAAAATATCTTTGAGACTCTTATGTCCACACCCACTGACACAAG
 50 GAGAGCTGTAACCACAGTGAACTAGACTTTGCTTTCCTTTAGCAAGTATGTGCCTATGATAGTAAACT
 GGAGTAAATGTAACAAGTAATAAAACAAATTTTTTTTAAAAATAAAATTATACCTTTTTTCTCCAACAAA
 CGGTAAAGACCACGTGAAGACATCCATAAAATTAGGCAACCAGTAAAGATGTGGAGAACCAGTAAACTG
 TCGAAATTCATCACATTATTTTCATACTTTAATACAGCAGCTTTAATTATTGGAGAACATCAAAGTAAT
 55 TAGGTGCCGAAAAACATTGTTATTAATGAAGGGAACCCCTGACGTTTGACCTTTTCTGTACCATCTATA
 GCCCTGGACTTGA (SEQ ID NO: 3103)

EP 2 194 145 B1

The FASTA file, including the sequence of AA184553, was then masked using the RepeatMasker web interface, as shown below. The region of alignment of clone 463D12 is outlined.

```

5      TTTTTTTTTTTTTTCTTAAATAGCATTTATTTTCTCTCAAAAAGCCTATTATGTACTAACAAGTGTTC
      TCTAAATTAGAAAGGCATCACTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGAGAAGGAATATTGGGT
      TACAATCTGAATTTCTCTTTATGATTTCTCTTAAAGTATAGAACAGCTATTAAAATGACTAATATTGCT
      AAAATGAAGGCTACTAAATTTCCCAAGAATTTTCGGTGGAATGCCCAAAAATGGTGTTAAGATATGCAG
10     AAGGGCCCATTTCAAGCAAAGCAATCTCTCCACCCCTTCATAAAAGATTTAAGCTAAAAAAAAAAAAAA
      AAGAAAGAAAATCCAACAGCTGAAGACATTGGGCTATTTATAAATCTTCTCCAGTCCCCCAGACAGCCT
      CACATGGGGGCTGTAAACAGCTAACTAAAATATCTTTGAGACTCTTATGTCCACACCCACTGACACAAG
      GAGAGCTGTAAACCACAGTGAACTAGACTTTGCTTTTCCTTTAGCAAGTATGTGCCTATGATAGTAACT
15     GGAGTAAATGTAACA)GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCTTTTCTCCAACAAA
      CGGTAAAGACCACGTGAAGACATCCATAAAATTAGGCAACCAGTAAAGATGTGGAGAACCAGTAAACTG
      TCGAAATTCATCACATTATTTTCATACTTTAATACAGCAGCTTTAATTATTGGAGAACATCAAAGTAAT
20     TAGGTGCCGAAAAACATTGTTATTAATGAAGGGAACCCCTGACGTTTGACCTTTTCTGTACCATCTATA
      GCCCTGGACTTGA Masked version of 463D12 sequence. (SEQ ID NO:3104)

```

[0379] The sequence was submitted to Array Designer as described above, however, the desired location of the probe was indicated at base pair 50 and if no probe met the criteria, moved in the 3' direction. The complementary sequence from Array Designer was used, because the original sequence was antisense. The oligonucleotide designed by Array Designer corresponds to oligonucleotide number 3054 and is complementary to the underlined sequence above. The probe has a melting temperature of 72.7 degrees centigrade and a max run of 4 nucleotides.

Clone 72D4

[0380] Clone 72D4 was sequenced and compared to the nr, dbEST, and UniGene databases at NCBI using the BLAST search tool. No significant matches were found in any of these databases. When compared to the human genome draft, significant alignments were found to three consecutive regions of the reference sequence NT_008060, as depicted below, suggesting that the insert contains three spliced exons of an unidentified gene.

	Residue numbers on clone 72D4 sequence	Matching residue numbers on NT 008060
	1 - 198	478646 - 478843
	197 - 489	479876 - 480168
40	491 - 585	489271 - 489365

[0381] Because the reference sequence contains introns and may represent either the coding or noncoding strand for this gene, BioCardia's own sequence file was used to design the oligonucleotide. Two complementary probes were designed to ensure that the sense strand was represented. The sequence of the insert in clone 72D4 is shown below, with the three putative exons outlined.

50

55

CAGGTCACACAGCACATCAGTGGCTACATGTGAGCTCAGACCTGGGTCTGCTGCTGTCTGT
 CTCCCAATATCCATGACCTTGACTGATGCAGGTGTCTAGGGATACGTCCATCCCCGTCTCT
 5 GCTGGAGCCCAGAGCACGGAAGCCTGGCCCTCCGAGGAGACAGAAGGGAGTGTCCGACA
 CCATGACGAGAGCTTGGCAGAATAAATAACTTCTTTAAACAATTTACGGCATGAAGAAA
 TCTGGACCAGTTTATTAAATGGGATTTCTGCCACAAACCTTGGAAGAATCACATCATCTTA
 10 NNCCCAAGTGAAAACCTGTGTTGCGTAACAAAGAACATGACTGCGCTCCACACATACATCA
 TTGCCCGGCGAGGCGGGACACAAGTCAACGACGGAACACTTGAGACAGGCCTACAACCTG
 TGCACGGGTCAGAAGCAAGTTTAAGCCATACTTGCTGCAGTGAGACTACATTTCTGTCTAT
 15 AGAAGATACTGACTTGATCTGTTTTTCAGCTCCAGTTCACAGATGTGCGTGTGTGGTCC
 CCAAGTATCACCTTCCAATTTCTGGGAGCAGTGCTCTGGCCG GATCCTTGCCGCGCGGAT

AAAAAC (SEQ ID NO: 3106)

[0382] The sequence was submitted to RepeatMasker, but no repetitive sequences were found. The sequence shown above was used to design the two 50-mer probes using Array Designer as described above. The probes are shown in bold typeface in the sequence depicted below. The probe in the sequence is oligonucleotide number 3020 (SEQ ID NO: 3020) and the complementary probe is oligonucleotide number 318 (SEQ ID NO: 318). A portion of the target sequence is listed below (SEQ ID: 3106).

CAGGTCACACAGCACATCAGTGGCTACATGTGAGCTCAGACCTGGGTCTGCTGCTGTCTGTCTTCCCAA
 TATCCATGACCTTGACTGATGCAGGTGTCTAGGGATACGTCCATCCCCGTCTGCTGGAGCCCAGAGCA
 30 CGGAAGCCTGGCCCTCCGAGGAGACAGAAGGGAGTGTCCGACACCATGACGAGAGCTTGGCAGAATAAA
 TAACTTCTTTAAACAATTTTACGGCATGAAGAAATCTGGACCAGTTTATTAAATGGGATTTCTGCCACA
 AACCTTGGAAGAATCACATCATCTTANNCCCAAGTGAAAACCTGTGTTGCGTAACAAAGAACATGACTGC
 GCTCCACACATACATCATTGCCCGGCGAGGCGGGACACAAGTCAACGACGGAACACTTGAGACAGGCCT
 35 ACAACTGTGCACGGGTCAGAAGCAAGTTTAAGCCATACTTGCTGCAGTGAGACTACATTTCTGTCTATA
 GAAGATACCTGACTTGATCTGTTTTTCAGCTCCAGTTCACAGATGTGC

← --- GTCAAGGGTCTACACG

GTGTTGTGGTCCCCAAGTATCACCTTCCAATTTCTGGGAG -->
 CACAACACCAGGGGTTTCATAGTGAAGGTTAAAG - 5'

CAGTGCTCTGGCCGGATCCTTGCCGCGCGGATAAAAACT --->

Confirmation of probe sequence

[0383] Following probe design, each probe sequence was confirmed by comparing the sequence against dbEST, the UniGene cluster set, and the assembled human genome using BLASTn at NCBI. Alignments, accession numbers, gi numbers, UniGene cluster numbers and names were examined and the most common sequence used for the probe.

Example 9 - Production of an array of 8000 spotted 50mer oligonucleotides

[0384] We produced an array of 8000 spotted initial candidate 50mer oligonucleotides. Example 8 exemplifies the design and selection of probes for this array.

[0385] Sigma-Genosys (The Woodlands, TX) synthesized un-modified 50-mer oligonucleotides using standard phosphoramidite chemistry, with a starting scale of synthesis of 0.05 μ mole (see, e.g., R. Meyers, ed. (1995) Molecular Biology and Biotechnology: A Comprehensive Desk Reference). Briefly, to begin synthesis, a 3' hydroxyl nucleoside

with a dimethoxytrityl (DMT) group at the 5' end was attached to a solid support. The DMT group was removed with trichloroacetic acid (TCA) in order to free the 5'-hydroxyl for the coupling reaction. Next, tetrazole and a phosphoramidite derivative of the next nucleotide were added. The tetrazole protonates the nitrogen of the phosphoramidite, making it susceptible to nucleophilic attack. The DMT group at the 5'-end of the hydroxyl group blocks further addition of nucleotides in excess. Next, the inter-nucleotide linkage was converted to a phosphotriester bond in an oxidation step using an oxidizing agent and water as the oxygen donor. Excess nucleotides were filtered out and the cycle for the next nucleotide was started by the removal of the DMT protecting group. Following the synthesis, the oligo was cleaved from the solid support. The oligonucleotides were desalted, resuspended in water at a concentration of 100 or 200 μ M, and placed in 96-deep well format. The oligonucleotides were re-arrayed into Whatman Uniplate 384-well polypropylene V bottom plates. The oligonucleotides were diluted to a final concentration 30 μ M in 1X Micro Spotting Solution Plus (Telechem/arrayit.com, Sunnyvale, CA) in a total volume of 15 μ l. In total, 8,031 oligonucleotides were arrayed into twenty-one 384-well plates.

[0386] Arrays were produced on Telechem/arrayit.com Super amine glass substrates (Telechem/arrayit.com), which were manufactured in 0.1 mm filtered clean room with exact dimensions of 25x76x0.96 mm. The arrays were printed using the Virtek Chipwriter with a Telechem 48 pin Micro Spotting Printhead. The Printhead was loaded with 48 Stealth SMP3B TeleChem Micro Spotting Pins, which were used to print oligonucleotides onto the slide with the spot size being 110-115 microns in diameter.

Example 10: Identification of diagnostic nucleotide sets for diagnosis of Cardiac Allograft Rejection

[0387] Genes were identified which have expression patterns useful for the diagnosis and monitoring of cardiac allograft rejection. Further, sets of genes that work together in a diagnostic algorithm for allograft rejection were identified. Patients, patient clinical data and patient samples used in the discovery of markers below were derived from a clinical study described in example 5.

[0388] The collected clinical data is used to define patient or sample groups for correlation of expression data. Patient groups are identified for comparison, for example, a patient group that possesses a useful or interesting clinical distinction, verses a patient group that does not possess the distinction. Measures of cardiac allograft rejection are derived from the clinical data described above to divide patients (and patient samples) into groups with higher and lower rejection activity over some period of time or at any one point in time. Such data are rejection grade as determined from pathologist reading of the cardiac biopsies and data measuring progression of end-organ damage, including depressed left ventricular dysfunction (decreased cardiac output, decreased ejection fraction, clinical signs of low cardiac output) and usage of inotropic agents (Kobashigawa 1998).

[0389] Expression profiles correlating with occurrence of allograft rejection are identified, including expression profiles corresponding to end-organ damage and progression of end-organ damage. Expression profiles are identified predicting allograft rejection, and response to treatment or likelihood of response to treatment. Subsets of the candidate library (or a previously identified diagnostic nucleotide set) are identified, that have predictive value for the presence of allograft rejection or prediction of allograft rejection or end organ damage.

[0390] Mononuclear RNA samples were collected from patients who had recently undergone a cardiac allograft transplantation using the protocol described in example 2. The allograft rejection status at the time of sample collection was determined by examination of cardiac biopsies as described in example 5.

[0391] 180 samples were included in the analysis. Each patient sample was associated with a biopsy and clinical data collected at the time of the sample. The cardiac biopsies were graded by a pathologist at the local center and by a centralized pathologist who read the biopsy slides from all four local centers in a blinded manner. Biopsy grades included 0, 1A, 1B, 2, 3A, and 3B. No grade 4 rejection was identified. Dependent variables were developed based on these grades using either the local center pathology reading or the higher of the two readings, local or centralized. The dependent variables used for correlation of gene expression profiles with cardiac allograft rejection are shown in Table 4. Dependent variables are used to create classes of samples corresponding to the presence or absence of rejection.

[0392] Clinical data were also used to determine criteria for including samples in the analysis. The strictest inclusion criteria required that samples be from patients who did not have a bacterial or viral infection, were at least two weeks post cardiac transplant and were not currently admitted to the hospital. A second inclusion criteria (inclusion 2) reduced the post-transplant criteria to 1 week and eliminated the hospital admission criteria.

[0393] After preparation of RNA (example 2), amplification, labeling, hybridization, scanning, feature extraction and data processing were done as described in Example 11, using the oligonucleotide microarrays described in Example 9. The resulting log ratio of expression of Cy3 (patient sample)/ Cy5 (R50 reference RNA) was used for analysis. This dataset is called the "static" data. A second type of dataset, referenced, was derived from the first. These datasets compared the gene expression log ratio in each sample to a baseline sample from the same patient using the formula:

$$\text{ref log ratio} = (\log \text{ratio}_{\text{sample}}) - (\log \text{ratio}_{\text{baseline}})$$

[0394] Two referenced datasets were used, named "0 HG" and "Best 0". The baseline for 0 HG was a Grade 0 sample from the same patient as the sample, using the highest grade between the centralized and local pathologists. The baseline for Best 0 was a Grade 0 sample from the same patient as the sample, using both the local and centralized reader biopsy grade data. When possible a Grade 0 prior to the sample was used as the baseline in both referenced datasets.

[0395] The datasets were also divided into subsets to compare analysis between two subsets of roughly half of the data. The types of subsets constructed were as follows. First half/second half subsets were the first half of the samples and the second half of the samples from a dataset ordered by sample number. Odd/even subsets used the same source, a dataset ordered by sample number, but the odd subset consisted of every 2nd sample starting with the first and the even subset consisted of every 2nd sample starting with the second sample. Center 14/other subsets were the same datasets, divided by transplant hospital. The center 14 subset consisted of all samples from patients at center 14, while the other subset consisted of all samples from the other three centers (12,13, and 15).

[0396] Initially, significance analysis for microarrays (SAM, Tusher 2001, Example 15) was used to discover genes that were differentially expressed between the rejection and no-rejection groups. Ninety-six different combinations of dependent variables, inclusion criteria, static/referenced, and data subsets were used in SAM analysis to develop the primary lists of genes significantly differentially expressed between rejection and no-rejection. The most significant of these genes were chosen based on the following criteria. Tier 1 genes were those which appeared with an FDR of less than 20% in identical analyses in two independent subsets. Tier 2 genes were those which appeared in the top 20 genes on the list with an FDR less than 20% more than 50% of the time over all dependent variables with the inclusion criteria, and static/referenced constant. Tier 3 genes were those that appeared more than 50% of the time with an FDR less than 20% more than 50% of the time over all dependent variables with the inclusion criteria, and static/referenced constant. The genes that were identified by the analysis as statistically differentially expressed between rejection and no rejection are shown in Table 2.

[0397] SAM chooses genes as significantly different based on the magnitude of the difference between the groups and the variation among the samples within each group. An example of the difference between some Grade 0 and some Grade 3A samples for 9 genes is shown in Figure 7A.

[0398] Additionally, many of these same combinations were used in the Supervised Harvesting of Expression Trees (SHET, Hastie et al. 2001) algorithm (see example 15) to identify markers that the algorithm chose as the best to distinguish between the rejection and no rejection classes using a bias factor of 0.01. The top 20 or 30 terms were taken from the SHET output and among all comparisons in either the static or referenced data the results were grouped. Any gene found in the top 5 terms in more than 50% of the analyses was selected to be in group B1 (Table 2). The occurrences of each gene were tabulated over all SHET analysis (for either static or referenced data) and the 10 genes that occurred the most were selected to be in group B2 (Table 2).

[0399] An additional classification method used was CART (Salford Systems, San Diego, example 15). Either the static or referenced dataset was reduced to only the genes for which expression values (log ratios) were present in at least 80% of the samples. These data were used in CART with the default settings, using the Symmetric Gini algorithm. Each of the dependent variables was used with both the full sample set and the strict inclusion criteria. Two groups of genes were identified. Group C1 were those genes that were a primary splitter (1st decision node). Group C2 genes were the 10 genes that occurred as splitters the most often over all these analyses.

[0400] Two other classification models were developed and their best genes identified as markers of cardiac allograft rejection. Group D genes were identified from a set of 59 samples, referenced data, local biopsy reading grade, using logistic regression. Group E genes were identified from the primary static dataset using a K-nearest neighbor classification algorithm.

[0401] Both hierarchical clustering (Eisen et al. 1998) and CART were used to identify surrogates for each identified marker. Hierarchical clustering surrogates are genes co-expressed in these and were chosen from the nearest branches of the dendrogram. CART surrogates were identified by CART as the surrogates for those genes chosen as primary splitters at decision nodes.

[0402] Primers for real-time PCR validation were designed for each of the marker genes as described in Example 13.

[0403] CART was used to build a decision tree for classification of samples as rejection or no-rejection using the gene expression data from the arrays. The analysis identified sets of genes that can be used together to accurately identify samples derived from cardiac allograft transplant patients. The set of genes and the identified threshold expression levels for the decision tree are referred to as a "models". This model can be used to predict the rejection state of an unknown sample. The input data were the static expression data (log ratio) and the referenced expression data (log ratio referenced to the best available grade 0 from either the centralized reader or the local reader) for 139 of our top

marker genes. These two types of expression data were entered into the CART software as independent variables. The dependent variable was rejection state, defined for this model as no rejection = grade 0 and rejection = grade 3A. Samples were eliminated from consideration in the training set if they were from patients with either bacterial or viral infection or were from patients who were less than two weeks post-transplant. The method used was Symmetric Gini, allowing linear combinations of independent variables. The costs were set to 1 for both false negatives and false positives and the priors were set equal for the two states. No penalties were assessed for missing data, however the marker genes selected have strong representation across the dataset. 10-fold cross validation was used to test the model. Settings not specified remained at the default values.

[0404] The model shown in Figure 7B is based on decisions about expression values at three nodes, each a different marker gene. The cost assigned to this model is 0.292, based on the priors being equal, the costs set to 1 for each type of error, and the results from the 10-fold cross validation.

[0405] In the training set, no rejection samples were misclassified (sensitivity = 100%) and only 1 no-rejection sample was misclassified (specificity = 94.4%). Following 10-fold cross validation, 2 rejection samples were misclassified (sensitivity = 87.5%) and 3 no-rejection samples were misclassified (specificity = 83.3%). The CART software assigns surrogate markers for each decision node.

[0406] These genes can be used alone or in association with other genes or variables to build a diagnostic gene set or a classification algorithm. These genes can be used in association with known gene markers for rejection (such as those identified in the prior art) to provide a diagnostic algorithm.

Example 11 - Amplification, labeling, and hybridization of total RNA to an oligonucleotide microarray

Amplification, labeling, hybridization and scanning

[0407] Samples consisting of at least 0.5 to 2 μ g of intact total RNA were further processed for array hybridization. When available, 2 μ g of intact total RNA is used for amplification. Amplification and labeling of total RNA samples was performed in three successive enzymatic reactions. First, a single-stranded DNA copy of the RNA was made (hereinafter, "ss-cDNA"). Second, the ss-cDNA was used as a template for the complementary DNA strand, producing double-stranded cDNA (hereinafter, "ds-cDNA, or cDNA"). Third, linear amplification was performed by in vitro transcription from a bacterial T₇ promoter. During this step, fluorescent-conjugated nucleotides were incorporated into the amplified RNA (hereinafter, "aRNA").

[0408] The first strand cDNA was produced using the Invitrogen kit (Superscript II). The first strand cDNA was produced in a reaction composed of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, and 3 mM MgCl₂ (1x First Strand Buffer, Invitrogen), 0.5 mM dGTP, 0.5 mM dATP, 0.5 mM dTTP, 0.5 mM dCTP, 10 mM DTT, 200 U reverse transcriptase (Superscript II, Invitrogen, #18064014), 15 U RNase inhibitor (RNAGuard, Amersham Pharmacia, #27-0815-01), 5 μ M T7T24 primer (5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGGTTTTTTTTTTTTTTTTTTTTTTTTTTT TTT-3'), (SEQ ID NO:3105) and 0.5 to 2 μ g of selected sample total RNA. Several purified, recombinant control mRNAs from the plant *Arabidopsis thaliana* were added to the reaction mixture: 2-20 pg of the following genes CAB, RCA, LTP4, NAC1, RCP1, XCP2, RBCL, LTP6, TIM, and PRKase (Stratagene, #252201, #252202, #252204, #252208, #252207, #252206, #252203, #252205, #252209, #252210 respectively). The control RNAs allow the estimate of copy numbers for individual mRNAs in the clinical sample because corresponding sense oligonucleotide probes for each of these plant genes are present on the microarray. The final reaction volume of 20 μ l was incubated at 42°C for 90 min. For synthesis of the second cDNA strand, DNA polymerase and RNase were added to the previous reaction, bringing the final volume to 150 μ l. The previous contents were diluted and new substrates were added to a final concentration of 20 mM Tris-HCl (pH 7.0) (Fisher Scientific, Pittsburgh, PA #BP1756-100), 90 mM KCl (Teknova, Half Moon Bay, CA, #0313-500), 4.6 mM MgCl₂ (Teknova, Half Moon Bay, CA, #0304-500), 10 mM (NH₄)₂SO₄ (Fisher Scientific #A702-500) (1x Second Strand buffer, Invitrogen), 0.266 mM dGTP, 0.266 mM dATP, 0.266 mM dTTP, 0.266 mM dCTP, 40 U *E. coli* DNA polymerase (Invitrogen, #18010-025), and 2 U RNaseH (Invitrogen, #18021-014). The second strand synthesis took place at 16°C for 150 minutes.

[0409] Following second-strand synthesis, the ds-cDNA was purified from the enzymes, dNTPs, and buffers before proceeding to amplification, using phenol-chloroform extraction followed by ethanol precipitation of the cDNA in the presence of glycogen.

[0410] Alternatively, a silica-gel column is used to purify the cDNA (e.g. Qiaquick PCR cleanup from Qiagen, #28104). The volume of the column purified cDNA was reduced by ethanol precipitation in the presence of glycogen in which the cDNA was collected by centrifugation at > 10,000 xg for 30 minutes, the supernatant is aspirated, and 150 μ l of 70% ethanol, 30% water was added to wash the DNA pellet. Following centrifugation, the supernatant was removed, and residual ethanol was evaporated at room temperature. Alternatively, the volume of the column purified cDNA is reduced in a vacuum evaporator where the supernatant is reduced to a final volume of 7.4 μ l.

[0411] Linear amplification of the cDNA was performed by in vitro transcription of the cDNA. The cDNA pellet from

the step described above was resuspended in 7.4 μ l of water, and in vitro transcription reaction buffer was added to a final volume of 20 μ l containing 7.5 mM GTP, 7.5 mM ATP, 7.5 mM TTP, 2.25 mM CTP, 1.025 mM Cy3-conjugated CTP (Perkin Elmer; Boston, MA, #NEL-580), 1x reaction buffer (Ambion, Megascript Kit, Austin, TX and #1334) and 1 % T₇ polymerase enzyme mix (Ambion, Megascript Kit, Austin, TX and #1334). This reaction was incubated at 37°C overnight. Following in vitro transcription, the RNA was purified from the enzyme, buffers, and excess NTPs using the RNeasy kit from Qiagen (Valencia, CA; # 74106) as described in the vendor's protocol. A second elution step was performed and the two eluates were combined for a final volume of 60 μ l. RNA is quantified using an Agilent 2100 bioanalyzer with the RNA 6000 nano LabChip.

[0412] Reference RNA was prepared as described above, except Cy5-CTP was incorporated instead of Cy3CTP. Reference RNA from five reactions, each reaction started with 2 ug total RNA, was pooled together and quantitated as described above.

Hybridization to an array

[0413] RNA was prepared for hybridization as follows: for an 18mm $\times$ 55mm array, 20 μ g of amplified RNA (aRNA) was combined with 20 μ g of reference aRNA. The combined sample and reference aRNA was concentrated by evaporating the water to 10 μ l in a vacuum evaporator. The sample was fragmented by heating the sample at 95°C for 30 minutes to fragment the RNA into 50-200 bp pieces. Alternatively, the combined sample and reference aRNA was concentrated by evaporating the water to 5 μ l in a vacuum evaporator. Five μ l of 20 mM zinc acetate was added to the aRNA and the mix incubated at 60°C for 10 minutes. Following fragmentation, 40 μ l of hybridization buffer was added to achieve final concentrations of 5 $\times$ SSC and 0.20 %SDS with 0.1 μ g/ μ l of Cot-1 DNA (Invitrogen) as a competitor DNA. The final hybridization mix was heated to 98°C, and then reduced to 50°C at 0.1°C per second.

[0414] Alternatively, formamide is included in the hybridization mixture to lower the hybridization temperature.

[0415] The hybridization mixture was applied to a pre-heated 65°C microarray, surface, covered with a glass coverslip (Coming, #2935-246), and placed on a pre-heated 65°C hybridization chamber (Telechem, AHC-10). 15 μ l of 5 $\times$ SSC was placed in each of the reservoir in the hybridization chamber and the chamber was sealed and placed in a water bath at 62°C for overnight (16-20 hrs). Following incubation, the slides were washed in 2 $\times$ SSC, 0.1% SDS for five minutes at 30°C, then in 2 $\times$ SSC for five minutes at 30°C, then in 2 $\times$ SSC for another five minutes at 30°C, then in 0.2 $\times$ SSC for two minutes at room temperature. The arrays were spun at 1000 $\times$ g for 2 minutes to dry them. The dry microarrays are then scanned by methods described above.

[0416] The microarrays were imaged on the Agilent (Palo Alto, CA) scanner G2565AA. The scan settings using the Agilent software were as follows: for the PMT Sensitivity (100% Red and 100% Green); Scan Resolution (10 microns); red and green dye channels; used the default scan region for all slides in the carousel; using the largest scan region; scan date for Instrument ID; and barcode for Slide ID. The full image produced by the Agilent scanner was flipped, rotated, and split into two images (one for each signal channel) using TIFFSplitter (Agilent, Palo Alto, CA). The two channels are the output at 532 nm (Cy3-labeled sample) and 633 nm (Cy5-labeled R50). The individual images were loaded into GenePix 3.0 (Axon Instruments, Union City, CA) for feature extraction, each image was assigned an excitation wavelength corresponding the file opened; Red equals 633 nm and Green equals 532 532 nm. The setting file (gal) was opened and the grid was laid onto the image so that each spot in the grid overlapped with >50% of the feature. Then the GenePix software was used to find the features without setting minimum threshold value for a feature. For features with low signal intensity, GenePix reports "not found". For all features, the diameter setting was adjusted to include only the feature if necessary.

[0417] The GenePix software determined the median pixel intensity for each feature (F_i) and the median pixel intensity of the local background for each feature (B_i) in both channels. The standard deviation (SDF_i and SDB_i) for each is also determined. Features for which GenePix could not discriminate the feature from the background were "flagged" as described below.

[0418] Following feature extraction into a ".gpr" file, the header information of the .gpr file was changed to carry accurate information into the database. An Excel macro was written to include the following information: Name of the original .tif image file, SlideID, Version of the feature extraction software, GenePix Array List file, GenePix Settings file, ScanID, Name of person who scanned the slide, Green PMT setting, Red PMT setting, ExtractID (date .gpr file was created, formatted as yyyy.mm.dd-hh.mm.ss), Results file name (same as the .gpr file name), StorageCD, and Extraction comments.

Pre-processing with Excel Templates

[0419] Following analysis of the image and extraction of the data, the data from each hybridization was preprocessed to extract data that was entered into the database and subsequently used for analysis. The complete GPR file produced by the feature extraction in GenePix was imported into an excel file pre-processing template or processed using a AWK

script. Both programs used the same processing logic and produce identical results. The same excel template or AWK script was used to process each GPR file. The template performs a series of calculations on the data to differentiate poor features from others and to combine duplicate or triplicate feature data into a single data point for each probe.

[0420] The data columns used in the pre-processing were: Oligo ID, F633 Median (median value from all the pixels in the feature for the Cy5 dye), B633 Median (the median value of all the pixels in the local background of the selected feature for Cy5), B633 SD (the standard deviation of the values for the pixels in the local background of the selected feature for Cy5), F532 Median (median value from all the pixels in the feature for the Cy3 dye), B532 Median (the median value of all the pixels in the local background of the selected feature for Cy3), B532 SD (the standard deviation of the values for the pixels in the local background of the selected feature for Cy3), and Flags. The GenePix Flags column contains the flags set during feature extraction. "-75" indicates there were no features printed on the array in that position, "-50" indicates that GenePix could not differentiate the feature signal from the local background, and "-100" indicates that the user marked the feature as bad.

[0421] Once imported, the data associated with features with -75 flags was not used. Then the median of B633 SD and B532 SD were calculated over all features with a flag value of "0". The minimum values of B633 Median and B532 Median were identified, considering only those values associated with a flag value of "0". For each feature, the signal to noise ratio (S/N) was calculated for both dyes by taking the fluorescence signal minus the local background (BGSS) and dividing it by the standard deviation of the local background:

$$S/N = \frac{F_i - B_i}{SDB_i}$$

[0422] If the S/N was less than 3, then an adjusted background-subtracted signal was calculated as the fluorescence minus the minimum local background on the slide. An adjusted S/N was then calculated as the adjusted background subtracted signal divided by the median noise over all features for that channel. If the adjusted S/N was greater than three and the original S/N were less than three, a flag of 25 was set for the Cy5 channel, a flag of 23 was set for the Cy3 channel, and if both met these criteria, then a flag of 28 was set. If both the adjusted S/N and the original S/N were less than three, then a flag of 65 was set for Cy5, 63 set for Cy3, and 68 set if both dye channels had an adjusted S/N less than three. All signal to noise calculations, adjusted background-subtracted signal, and adjusted S/N were calculated for each dye channel. If the BGSS value was greater than or equal to 64000, a flag was set to indicate saturation; 55 for Cy5, 53 for Cy3, 58 for both.

[0423] The BGSS used for further calculations was the original BGSS if the original S/N was greater than or equal to three. If the original S/N ratio was less than three and the adjusted S/N ratio was greater than or equal to three, then the adjusted BGSS was used. If the adjusted S/N ratio was less than three, then the adjusted BGSS was used, but with knowledge of the flag status.

[0424] To facilitate comparison among arrays, the Cy3 and Cy5 data were scaled. The log of the ratio of Green/Red was determined for all features. The median log ratio value for good features (Flags 0, 23, 25, 28, 63) was determined. The feature values were scaled using the following formula: $\text{Log_Scaled_Feature_Ratio} = \text{Log_Feature_Ratio} - \text{Median_Log_Ratio}$.

[0425] The flag setting for each feature was used to determine the expression ratio for each probe, a choice of one, two or three features. If all features had flag settings in the same category (categories=negatives, 0 to 28, 53-58, and 63-68), then the average of the three scaled, anti log feature ratios was calculated. If the three features did not have flags in the same category, then the feature or features with the best quality flags were used (0>25>23>28>55>53>58>65>63>68). Features with negative flags were never used. When the best flags were two or three features in the same category, the anti log average was used. If a single feature had a better flag category than the other two then the anti log of that feature ratio was used.

[0426] Once the probe expression ratios were calculated from the one, two, or three features, the log of the scaled, averaged ratios was taken as described below and stored for use in analyzing the data. Whichever features were used to calculate the probe value, the flag from those features was carried forward and stored as the flag value for that probe. 2 different data sets can be used for analysis. Flagged data uses all values, including those with flags. Filtered data sets are created by removing flagged data from the set before analysis.

Example 12: Real-time PCR validation of array expression results

[0427] Leukocyte microarray gene expression was used to discover expression markers and diagnostic gene sets for clinical outcomes. It is desirable to validate the gene expression results for each gene using a more sensitive and quantitative technology such as real-time PCR. Further, it is possible for the diagnostic nucleotide sets to be implemented

as a diagnostic test as a real-time PCR panel. Alternatively, the quantitative information provided by real-time PCR validation can be used to design a diagnostic test using any alternative quantitative or semi-quantitative gene expression technology. To validate the results of the microarray experiments we used real-time, or kinetic, PCR. In this type of experiment the amplification product is measured during the PCR reaction. This enables the researcher to observe the amplification before any reagent becomes rate limiting for amplification. In kinetic PCR the measurement is of C_T (threshold cycle) or C_P (crossing point). This measurement ($C_T=C_P$) is the point at which an amplification curve crosses a threshold fluorescence value. The threshold is set to a point within the area where all of the reactions were in their linear phase of amplification. When measuring C_T , a lower C_T value is indicative of a higher amount of starting material since an earlier cycle number means the threshold was crossed more quickly.

[0428] Several fluorescence methodologies are available to measure amplification product in real-time PCR. Taqman (Applied BioSystems, Foster City, CA) uses fluorescence resonance energy transfer (FRET) to inhibit signal from a probe until the probe is degraded by the sequence specific binding and Taq 3' exonuclease activity. Molecular Beacons (Stratagene, La Jolla, CA) also use FRET technology, whereby the fluorescence is measured when a hairpin structure is relaxed by the specific probe binding to the amplified DNA. The third commonly used chemistry is Sybr Green, a DNA-binding dye (Molecular Probes, Eugene, OR). The more amplified product that is produced, the higher the signal. The Sybr Green method is sensitive to non-specific amplification products, increasing the importance of primer design and selection. Other detection chemistries can also been used, such as ethidium bromide or other DNA-binding dyes and many modifications of the fluorescent dye/quencher dye Taqman chemistry.

Sample prep and cDNA synthesis

[0429] The inputs for real time PCR reaction are gene-specific primers, cDNA from specific patient samples, and standard reagents. The cDNA was produced from mononuclear RNA (prepared as in example 2) or whole blood RNA by reverse transcription using Oligo dT primers (Invitrogen, 18418-012) and random hexamers (Invitrogen, 48190-011) at a final concentration of 0.5ng/ μ l and 3ng/ μ l respectively. For the first strand reaction mix, 0.5 μ g of mononuclear total RNA or 2 μ g of whole blood RNA and 1 μ l of the Oligo dT/ Random Hexamer Mix, were added to water to a final volume of 11.5 μ l. The sample mix was then placed at 70°C for 10 minutes. Following the 70°C incubation, the samples were chilled on ice, spun down, and 88.5 μ l of first strand buffer mix dispensed into the reaction tube. The final first strand buffer mix produced final concentrations of 1X first strand buffer (Invitrogen, Y00146, Carlsbad, CA), 10 mM DTT (Invitrogen, Y00147), 0.5 mM dATP (NEB, N0440S, Beverly, MA), 0.5 mM dGTP (NEB, N0442S), 0.5mM dTTP (NEB, N0443S), 0.5 mM dCTP (NEB, N0441S), 200U of reverse transcriptase (Superscript II, Invitrogen, 18064-014), and 18U of RNase inhibitor (RNAgaard Amersham Pharmacia, 27-0815-01, Piscataway, NJ). The reaction was incubated at 42°C for 90 minutes. After incubation the enzyme was heat inactivated at 70°C for 15 minutes, 2 U of RNase H added to the reaction tube, and incubated at 37°C for 20 minutes.

PRIMER DESIGN

[0430] Two methods were used to design primers. The first was to use the software, Primer Expresstm and recommendations for primer design that are provided with the GeneAmp® 7700 Sequence Detection System supplied by Applied BioSystems (Foster City, CA). The second method used to design primers was the PRIMER3 ver 0.9 program that is available from the Whitehead Research Institute, Cambridge, Massachusetts at the Whitehead Research web site. The program can also be accessed on the World Wide Web at the web site at the Massachusetts Institute of Technology website. Primers and Taqman/hybridization probes were designed as described below using both programs.

[0431] The Primer Express literature explains that primers should be designed with a melting temperature between 58 and 60 degrees C. while the Taqman probes should have a melting temperature of 68 to 70 under the salt conditions of the supplied reagents. The salt concentration is fixed in the software. Primers should be between 15 and 30 basepairs long. The primers should produce an amplicon in size between 50 and 150 base pairs, have a C-G content between 20% and 80%, have no more than 4 identical base pairs next to one another, and no more than 2 C's and G's in the last 5 bases of the 3' end. The probe cannot have a G on the 5' end and the strand with the fewest G's should be used for the probe.

[0432] Primer3 has a large number of parameters. The defaults were used for all except for melting temperature and the optimal size of the amplicon was set at 100 bases. One of the most critical is salt concentration as it affects the melting temperature of the probes and primers. In order to produce primers and probes with melting temperatures equivalent to Primer Express, a number of primers and probes designed by Primer Express were examined using PRIMER3. Using a salt concentration of 50 mM these primers had an average melting temperature of 3.7 degrees higher than predicted by Primer Express. In order to design primers and probes with equivalent melting temperatures as Primer Express using PRIMER3, a melting temperature of 62.7 plus/minus 1.0 degree was used in PRIMER3 for primers and 72.7 plus/minus 1.0 degrees for probes with a salt concentration of 50 mM.

[0433] The C source code for Primer3 was downloaded and compiled on a Sun Enterprise 250 server using the GCC compiler. The program was then used from the command line using a input file that contained the sequence for which we wanted to design primers and probes along with the input parameters as described by help files that accompany the software. Using scripting it was possible to input a number of sequences and automatically generate a number of possible probes and primers.

[0434] Primers for β -Actin (Beta Actin, Genbank Locus: NM_001101) and β -GUS: glucuronidase, beta, (GUSB, Genbank Locus: NM_000181), two reference genes, were designed using both methods and are shown here as examples:

The first step was to mask out repetitive sequences found in the mRNA sequences using RepeatMasker program that can be accessed at: the web site University of Washington Genome Repeatmasker website. (Smit, A.F.A. & Green, P.).

[0435] The last 500 basepairs on the last 3' end of masked sequence was then submitted to PRIMER3 using the following exemplary input sequences:

```
PRIMER_SEQUENCE_ID=>ACTB Beta Actin (SEQID 3083)
SEQUENCE=TTGGCTTGACTCAGGATTTAAAACTGGAACGGTGAAGGTGACAGCAGTCGGTTGGACGA
GCATCCCCCAAAGTTCACAATGTGGCCGAGGACTTTGATTGCACATTGTTGTTTTTAATAGTCATTCC
AAATATGAGATGCATTGTTACAGGAAGTCCCTTGCCATCCTAAAAGCACCCCACTTCTCTCTAAGGAGA
ATGGCCCAAGTCCTCTCCCAAGTCCACACAGGGGAGGGATAGCATTGCTTTTCGTGTAAATTATGTAATGC
AAAATTTTTTTAATCTTCGCCTTAATCTTTTTTATTTTGTGTTTTATTTTGAATGATGAGCCTTCGTGCC
CCCCTTCCCCCTTTTTTCCCCCAACTTGAGATGTATGAAGGCTTTTGGTCTCCCTGGGAGTGGGTGGAG
GCAGCCGGGCTTACCTGTACTGACTTGAGACCAGTTGAATAAAAGTGCACACCTTA
```

```
PRIMER_SEQUENCE_ID=>GUSB (SEQID 3084)
SEQUENCE=GAAGAGTACCAGAAAAGTCTGCTAGAGCAGTACCATCTGGGTCTGGATCAAAAACGCAGA
AAATATGTGGTTGGAGAGCTCATTTGGAATTTTGCCGATTTTCATGACTGAACAGTCACCGACGAGAGTG
CTGGGGAATAAAAAGGGGATCTTCACTCGGCAGAGACAACCAAAAAGTGCAGCGTTCCCTTTTGGAGAG
AGATACTGGAAGATTGCAATGAAACCAGGTATCCCCACTCAGTAGCCAAGTCACAATGTTTGGAAAAC
AGCCCGTTTACTTGAGCAAGACTGATACCACCTGCGTGTCCCTTCCCTCCCGAGTCAGGGCGACTTCCA
CAGCAGCAGAACAAAGTGCCTCCTGGACTGTTACGGCAGACCAGAACGTTTCTGGCCTGGGTTTTGTGG
TCATCTATTCTAGCAGGGAACACTAAAGGTGGAAATAAAAGATTTTCTATTATGGAAATAAAGAGTTGG
CATGAAAGTCGCTACTG
```

[0436] After running PRIMER3, 100 sets of primers and probes were generated for ACTB and GUSB. From this set, nested primers were chosen based on whether both left primers could be paired with both right primers and a single Taqman probe could be used on an insert of the correct size. With more experience we have decided not use the mix and match approach to primer selection and just use several of the top pairs of predicted primers.

[0437] For ACTB this turned out to be:

```
Forward 75 CACAATGTGGCCGAGGACTT(SEQID 3085),
Forward 80 TGTGGCCGAGGACTTTGATT(SEQID 3086),
Reverse 178 TGGCTTTTAGGATGGCAAGG(SEQID 3087), and
Reverse 168 GGGGGCTTAGTTTGCTTCCT(SEQID 3088).
```

[0438] Upon testing, the F75 and R178 pair worked best.

[0439] For GUSB the following primers were chosen:

```
Forward 59 AAGTGCAG CGTTCCTTTTGC(SEQID 3089),
Forward 65 AGCGTTCCTTTTGCAGAGA (SEQID 3090),
Reverse 158 CGGGCTGTT7TCCAAACATT (SEQID 3091), and
Reverse 197 GAAGGGACACGCAGGTGGTA (SEQID 3092).
```

[0440] No combination of these GUSB pairs worked well.

[0441] In addition to the primer pairs above, Primer Express predicted the following primers for GUSB: Forward 178 TACCACCTGCGTGTCCCTTC (SEQID 3093) and Reverse 242 GAGGCACTTGTTCTGCTGCTG (SEQID 3094). This pair of primers worked to amplify the GUSB mRNA.

[0442] The parameters used to predict these primers in Primer Express were:

EP 2 194 145 B1

Primer Tm: min 58, Max=60, opt 59, max difference=2 degrees
Primer GC: min=20% Max =80% no 3' G/C clamp
Primer: Length: min=9 max=40 opt=20
Amplicon: min Tm=0 max Tm=85
min = 50 bp max =150 bp
Probe: Tm 10 degrees > primers, do not begin with a G on 5' end
Other: max base pair repeat = 3
max number of ambiguous residues = 0
secondary structure: max consecutive bp = 4, max total bp = 8
Uniqueness: max consecutive match = 9
max % match = 75
max 3' consecutive match = 7

Granzyme B is a marker of transplant rejection.

[0443] For Granzyme B the following sequence (NM_004131) (SEQID 3096) was used as input for Primer3 :

```
GGGGACTCTGGAGGCCCTCTTGTGTGTAACAAGGTGGCCCAGGGCATTGT
CTCCTATGGACGAAACAATGGCATGCCTCCACGAGCCTGCACCAAAGTCT
CAAGCTTTGTACACTGGATAAAGAAAACCATGAAACGCTACTAACTACAG
GAAGCAAACCTAAGCCCCCGCTGTAATGAAACACCTTCTCTGGAGCCAAGT
CCAGATTTACACTGGGAGAGGTGCCAGCAACTGAATAAATACCTCTCCCA
GTGTAAATCTGGAGCCAAGTCCAGATTTACACTGGGAGAGGTGCCAGCAA
CTGAATAAATACCTCTTAGCTGAGTGG
```

[0444] For Granzyme B the following primers were chosen for testing:

Forward 81 ACGAGCGTGCACCAAAGTCT (SEQID 3097)
Forward 63 AAACAATGGCATGCCTCCAC (SEQID 3098)
Reverse 178 TCATTACAGCGGGGGCTTAG (SEQID 3099)
Reverse 168 CGGGGCTTAGTTTGCTTCCT (SEQID 3100)

[0445] Testing demonstrated that F81 and R178 worked well,

[0446] Using this approach, primers were designed for all the genes that were shown to have expression patterns that correlated with allograft rejection. These primer pairs are shown in Table 2, and are added to the sequence listing. Primers can be designed from any region of a target gene using this approach.

PRIMER ENDPOINT TESTING

[0447] Primers were first tested to examine whether they would produce the correct size product without non-specific amplification. The standard real-time PCR protocol was used without the Rox and Sybr green dyes. Each primer pair was tested on cDNA made from universal mononuclear leukocyte reference RNA that was produced from 50 individuals as described in Example 3 (R50).

[0448] The PCR reaction consisted of 1X RealTime PCR Buffer (Ambion, Austin, TX), 2mM MgCl₂ (Applied Biosystems, B02953), 0.2mM dATP (NEB), 0.2mM dTTP (NEB), 0.2mM dCTP (NEB), 0.2mM dGTP (NEB), .625U AmpliTaq Gold (Applied Biosystems, Foster City, CA), 0.3μM of each primer to be used (Sigma Genosys, The Woodlands, TX), 5μl of the R50 reverse-transcription reaction and water to a final volume of 19μl.

[0449] Following 40 cycles of PCR, 10 microliters of each product was combined with Sybr green at a final dilution of 1:72,000. Melt curves for each PCR product were determined on an ABI 7900 (Applied Biosystems, Foster City, CA), and primer pairs yielding a product with one clean peak were chosen for further analysis. One microliter of the product from these primer pairs was examined by agarose gel electrophoresis on an Agilent Bioanalyzer, DNA1000 chip (Palo Alto, CA). Results for 2 genes are shown in Figure 9. From the primer design and the sequence of the target gene, one can calculate the expected size of the amplified DNA product. Only primer pairs with amplification of the desired product and minimal amplification of contaminants were used for real-time PCR. Primers that produced multiple products of different sizes are likely not specific for the gene of interest and may amplify multiple genes or chromosomal loci.

PRIMER OPTIMIZATION/EFFICIENCY

[0450] Once primers passed the end-point PCR, the primers were tested to determine the efficiency of the reaction in a real-time PCR reaction. cDNA was synthesized from starting total RNA as described above. A set of 5 serial dilutions of the R50 reverse-transcribed cDNA (as described above) were made in water: 1:10, 1:20, 1:40, 1:80, and 1:160.

[0451] The Sybr Green real-time PCR reaction was performed using the Taqman PCR Reagent kit (Applied Biosystems, Foster City, CA, N808-0228). A master mix was made that consisted of all reagents except the primers and template. The final concentration of all ingredients in the reaction was 1X Taqman Buffer A (Applied Biosystems), 2mM MgCl₂ (Applied Biosystems), 200μM dATP (Applied Biosystems), 200μM dCTP (Applied Biosystems), 200μM dGTP (Applied Biosystems), 400μM dUTP (Applied Biosystems), 1:400,000 diluted Sybr Green dye (Molecular Probes), 1.25U AmpliTaq Gold (Applied Biosystems). The PCR master mix was dispensed into two, light-tight tubes. Each β-Actin primer F75 and R178 (Sigma-Genosys, The Woodlands, TX), was added to one tube of PCR master mix and Each β-GUS primer F178 and R242 (Sigma-Genosys), was added to the other tube of PCR master mix to a final primer concentration of 300nM. 45μl of the β-Actin or β-GUS master mix was dispensed into wells, in a 96-well plate (Applied Biosystems). 5μl of the template dilution series was dispensed into triplicate wells for each primer. The reaction was run on an ABI 7900 Sequence Detection System (Applied Biosystems) with the following conditions: 10 min. at 95°C; 40 cycles of 95°C for 15 sec, 60°C for 1 min; followed by a dissociation curve starting at 50°C and ending at 95°C.

[0452] The Sequence Detection System v2.0 software was used to analyze the fluorescent signal from each well. The high end of the baseline was adjusted to between 8 and 20 cycles to reduce the impact on any data curves, yet be as high as possible to reduce baseline drift. A threshold value was selected that allowed the majority of the amplification curves to cross the threshold during the linear phase of amplification. The dissociation curve for each well was compared to other wells for that marker. This comparison allowed identification of "bad" wells, those that did not amplify, that amplified the wrong size product, or that amplified multiple products. The cycle number at which each amplification curve crossed the threshold (C_T) was recorded and the file transferred to MS Excel for further analysis. The C_T values for triplicate wells were averaged. The data were plotted as a function of the log₁₀ of the calculated starting concentration of RNA. The starting RNA concentration for each cDNA dilution was determined based on the original amount of RNA used in the RT reaction, the dilution of the RT reaction, and the amount used (5 μl) in the real-time PCR reaction. For each gene, a linear regression line was plotted through all of the dilutions series points. The slope of the line was used to calculate the efficiency of the reaction for each primer set using the equation:

$$E = 10^{\left(\frac{-1}{\text{slope}}\right)} - 1$$

[0453] Using this equation (Pfaffl 2001, Applied Biosystems User Bulletin #2), the efficiency for these β-actin primers is 1.28 and the efficiency for these β-GUS primers is 1.14 (Figure 10). This efficiency was used when comparing the expression levels among multiple genes and multiple samples. This same method was used to calculate reaction efficiency for primer pairs for each gene studied. A primer pair was considered successful if the efficiency was reproducibly determined to be between 0.7 and 2.4.

SYBR-GREEN ASSAYS

[0454] Once markers passed the Primer Efficiency QPCR (as stated above), they were used in real-time PCR assays. Patient RNA samples were reverse-transcribed to cDNA (as described above) and 1:10 dilutions made in water. In addition to the patient samples, a no template control (NTC) and a pooled reference RNA (see example 3) described in were included on every plate.

[0455] The Sybr Green real-time PCR reaction was performed using the Taqman Core PCR Reagent kit (Applied Biosystems, Foster City, CA, N808-0228). A master mix was made that consisted of all reagents except the primers and template. The final concentration of all ingredients in the reaction was 1X Taqman Buffer A (Applied Biosystems), 2mM MgCl₂ (Applied Biosystems), 200μM dATP (Applied Biosystems), 200μM dCTP (Applied Biosystems), 200μM dGTP (Applied Biosystems), 400μM dUTP (Applied Biosystems), 1:400,000 diluted Sybr Green dye (Molecular Probes), 1.25U AmpliTaq Gold (Applied Biosystems). The PCR master mix was aliquotted into eight light-tight tubes, one for each marker to be examined across a set of samples. The optimized primer pair for each marker was then added to the PCR master mix to a final primer concentration of 300nM. 18μl of the each marker master mix was dispensed into wells in a 384well plate (Applied Biosystems). 2μl of the 1:10 diluted control or patient cDNA sample was dispensed into triplicate wells for each primer pair. The reaction was run on an ABI 7900 Sequence Detection System (Applied Biosystems) using the cycling conditions described above.

[0456] The Sequence Detection System v2.0 software (Applied Biosystems) was used to analyze the fluorescent

signal from each well. The high end of the baseline was adjusted to between 8 and 20 cycles to reduce the impact on any data curves, yet be as high as possible to reduce baseline drift. A threshold value was selected that allowed the majority of the amplification curves to cross the threshold during the linear phase of amplification. The disassociation curve for each well was compared to other wells for that marker. This comparison allowed identification of "bad" wells, those that did not amplify, that amplified the wrong size product, or that amplified multiple products. The cycle number at which each amplification curve crossed the threshold (C_T) was recorded and the file transferred to MS Excel for further analysis. The C_T value representing any well identified as bad by analysis of disassociation curves was deleted. The C_T values for triplicate wells were averaged. A standard deviation (Stdev) and a coefficient of variation (CV) were calculated for the triplicate wells. If the CV was greater than 2, an outlier among the three wells was identified and deleted. Then the average was re-calculated. In each plate, ΔC_T was calculated for each marker-control combination by subtracting the average C_T of the target marker from the average C_T of the control β -Actin or β -GUS). The expression relative to the control marker was calculated by taking two to the power of the ΔC_T of the target marker. For example, expression relative to β -Actin was calculated by the equation:

$$ErA = 2^{(C_{T, Actin} - C_{T, target})}$$

[0457] All plates were run in duplicate and analyzed in the same manner. The percent variation was determined for each sample-marker combination (relative expression) by taking the absolute value of the value of the RE for the second plate from the RE for the first plate, and dividing that by the average. If more than 25% of the variation calculations on a plate are greater than 50%, then a third plate was run.

TAQMAN PROTOCOL

Real-time PCR assays were also done using Taqman PCR chemistry.

[0458] The Taqman real-time PCR reaction was performed using the Taqman Universal PCR Master Mix (Applied BioSystems, Foster City, CA, #4324018). The master mix was aliquoted into eight, light-tight tubes, one for each marker. The optimized primer pair for each marker was then added to the correctly labeled tube of PCR master mix. A FAM/TAMRA dual-labeled Taqman probe (Biosearch Technologies, Navoto, CA, DLO-FT-2) was then added to the correctly labeled tube of PCR master mix. Alternatively, different combinations of fluorescent reporter dyes and quenchers can be used such that the absorption wavelength for the quencher matches the emission wavelength for the reporter, as shown in Table 5. 18 μ l of the each marker master mix was dispensed into a 384well plate (Applied BioSystems). 2 μ l of the template sample was dispensed into triplicate wells for each primer pair. The final concentration of each reagent was: 1X TaqMan Universal PCR Master Mix, 300nM each primer, 0.25nM probe, 2 μ l 1:10 diluted template. The reaction was run on an ABI 7900 Sequence Detection System (Applied Biosystems) using standard conditions (95°C for 10 min., 40 cycles of 95°C for 15 sec, 60°C for 1 min.).

[0459] The Sequence Detector v2.0 software (Applied BioSystems) was used to analyze the fluorescent signal from each well. The high end of the baseline was adjusted to between 8 and 20 cycles to reduce the impact on any data curves, yet be as high as possible to reduce baseline drift. A threshold value was selected that allowed most of the amplification curves to cross the threshold during the linear phase of amplification. The cycle number at which each amplification curve crossed the threshold (C_T) was recorded and the file transferred to MS Excel for further analysis. The C_T values for triplicate wells were averaged. The C_T values for triplicate wells were averaged. A standard deviation (Stdev) and a coefficient of variation (CV) were calculated for the triplicate wells. If the CV was greater than 2, an outlier among the three wells was identified and deleted. Then the average was re-calculated. In each plate, ΔC_T was calculated for each marker-control combination by subtracting the average C_T of the target marker from the average C_T of the control (β -Actin or β -GUS). The expression relative to the control marker was calculated by taking two to the power of the ΔC_T of the target marker. All plates were run in duplicate and analyzed in the same manner. The percent variation was determined for each sample-marker combination (relative expression) by taking the absolute value of the value of the RE for the second plate from the RE for the first plate, and dividing that by the average. If more than 25% of the variation calculations on a plate are greater than 50%, then a third plate was run.

BI-PLEXING

[0460] Variation of real-time PCR assays can arise from unequal amounts of RNA starting material between reactions. In some assays, to reduce variation, the control gene amplification was included in the same reaction well as the target gene. To differentiate the signal from the two genes, different fluorescent dyes were used for the control gene. β -Actin

was used as the control gene and the TaqMan probe used was labeled with the fluorescent dye VIC and the quencher TAMRA (Biosearch Technologies, Navoto, CA, DLO-FT-2). Alternatively, other combinations of fluorescent reporter dyes and quenchers (Table 5) can be used as long as the emission wavelength of the reporter for the control gene is sufficiently different from the wavelength of the reporter dye used for the target. The control gene primers and probe were used at limiting concentrations in the reaction (150 nM primers and 0.125 nM probe) to ensure that there were enough reagents to amplify the target marker. The plates were run under the same protocol and the data are analyzed in the same way, but with a separate baseline and threshold for the VIC signal. Outliers were removed as above from both the FAM and VIC signal channels. The expression relative to control was calculated as above, using the VIC signal from the control gene.

ABSOLUTE QUANTITATION

[0461] Instead of calculating the expression relative to a reference marker, an absolute quantitation can be performed using real-time PCR. To determine the absolute quantity of each marker, a standard curve is constructed using serial dilutions from a known amount of template for each marker on the plate. The standard curve may be made using cloned genes purified from bacteria or using synthetic complimentary oligonucleotides. In either case, a dilution series that covers the expected range of expression is used as template in a series of wells in the plate. From the average C_T values for these known amounts of template a standard curve can be plotted. From this curve the C_T values for the unknowns are used to identify the starting concentration of cDNA. These absolute quantities can be compared between disease classes (i.e. rejection vs. no-rejection) or can be taken as expression relative to a control gene to correct for variation among samples in sample collection, RNA purification and quantification, cDNA synthesis, and the PCR amplification.

CELL TYPE SPECIFIC EXPRESSION

[0462] Some markers are expressed only in specific types of cells. These markers may be useful markers for differentiation of rejection samples from no-rejection samples or may be used to identify differential expression of other markers in a single cell type. A specific marker for cytotoxic T-lymphocytes (such as CD8) can be used to identify differences in cell proportions in the sample. Other markers that are known to be expressed in this cell type can be compared to the level of CD8 to indicate differential gene expression within CD8 T-cells.

Control genes for PCR

[0463] As discussed above, PCR expression measurements can be made as either absolute quantification of gene expression using a standard curve or relative expression of a gene of interest compared to a control gene. In the latter case, the gene of interest and the control gene are measured in the same sample. This can be done in separate reactions or in the same reaction (biplex format, see above). In either case, the final measurement for expression of a gene is expressed as a ratio of gene expression to control gene expression. It is important for a control gene to be constitutively expressed in the target tissue of interest and have minimal variation in expression on a per cell basis between individuals or between samples derived from an individual. If the gene has this type of expression behavior, the relative expression ratio will help correct for variability in the amount of sample RNA used in an assay. In addition, an ideal control gene has a high level of expression in the sample of interest compared to the genes being assayed. This is important if the gene of interest and control gene are used in a biplex format. The assay is set up so that the control gene reaches its threshold C_t value early and its amplification is limited by primers so that it does not compete for limiting reagents with the gene of interest.

[0464] To identify an ideal control gene for an assay, a number of genes were tested for variability between samples and expression in both mononuclear RNA samples and whole blood RNA samples using the RNA procurement and preparation methods and real-time PCR assays described above. 6 whole-blood and 6 mononuclear RNA samples from transplant recipients were tested. The intensity levels and variability of each gene in duplicate experiments on both sample types are shown in Figure 11.

[0465] Based on criteria of low variability and high expression across samples, β -actin, 18s, GAPDH, b2microglobulin were found to be good examples of control genes for the PAX samples. A single control gene may be incorporated as an internal biplex control in assays.

Controlling for variation in real time PCR

[0466] Due to differences in reagents, experimenters, and preparation methods, and the variability of pipetting steps, there is significant plate-to-plate variation in real-time PCR experiments. This variation can be reduced by automation (to reduce variability and error), reagent lot quality control, and optimal data handling. However, the results on replicate

plates are still likely to be different since they are run in the machine at different times.

[0467] Variation can also enter in data extraction and analysis. Real-time PCR results are measured as the time (measured in PCR cycles) at which the fluorescence intensity (ΔR_n in Applied Biosystems SDS v2.1 software) crosses a user-determined threshold (CT). When performing relative quantification, the CT value for the target gene is subtracted from the CT value for a control gene. This difference, called ACT, is the value compared among experiments to determine whether there is a difference between samples. Variation in setting the threshold can introduce additional error. This is especially true in the duplexed experimental format, where both the target gene and the control gene are measured in the same reaction tube. Duplexing is performed using dyes specific to each of the two genes. Since two different fluorescent dyes are used on the plate, two different thresholds are set. Both of these thresholds contribute to each ACT. Slight differences in the each dye's threshold settings (relative to the other dye) from one plate to the next can have significant effects on the ACT.

[0468] There are several methods for setting the threshold for a PCR plate. Older versions of SDS software (Applied Biosystems) determine the average baseline fluorescence for the plate and the standard deviation of the baseline. The threshold is set to 10x the standard deviation of the baseline. In SDS 2.0 the users must set the baseline by themselves. Software from other machine manufacturers either requires the user to set the threshold themselves or uses different algorithms. The latest version of the SDS software (SDS 2.1) contains Automatic baseline and threshold setting. The software sets the baseline separately for each well on the plate using the ΔR_n at cycles preceding detectable levels. Variability among plates is dependent on reproducible threshold setting. This requires a mathematical or experimental data driven threshold setting protocol. Reproducibly setting the threshold according to a standard formula will minimize variation that might be introduced in the threshold setting process. Additionally, there may be experimental variation among plates that can be reduced by setting the threshold to a component of the data. We have developed a system that uses a set of reactions on each plate that are called the threshold calibrator (TCb). The TCb wells are used to set the threshold on all plates.

1. The TCb wells contain a template, primers, and probes that are common among all plates within an experiment.
2. The threshold is set within the minimum threshold and maximum threshold determined above.
3. The threshold is set to a value in this range that results in the average CT value for the TCb wells to be the same on all plates.

[0469] These methods were used to derive the primers depicted in Table 2C.

Example 13: Real-time PCR expression markers of acute allograft rejection

[0470] In examples 14 and 16, genes were identified as useful markers of cardiac and renal allograft rejection using microarrays. Some genes identified through these studies are listed in Table 2. In order to validate these findings, obtain a more precise measurement of expression levels and develop PCR reagents for diagnostic testing, real-time PCR assays were performed on samples from allograft recipients using primers to the identified genes. Some gene specific PCR primers were developed and tested for all genes in Table 2A as described in example 12. Some primers are listed in Table 2C and the sequence listing. These primers were used to measure expression of the genes relative to β -actin or β -gus in 69 mononuclear RNA samples obtained from cardiac allograft recipients using Sybr green real-time PCR assays as described in example 12. Each sample was associated with an ISHLT cardiac rejection biopsy grade. The samples were tested in 2 phases. In phase I, 14 Grade 0, 1 Grade 1A, 3 Grade 2 and 9 Grade 3A samples were tested. In phase II, 19 Grade 2, 4 Grade 1B, 4 Grade 2 and 15 Grade 3A samples were tested. Data was analyzed for each phase individually and for the combined phase I + II sample set. These data are summarized in Table 6.

[0471] The average fold change in expression between rejection (3A) and no rejection (0) samples was calculated. A t-test was done to determine the significance with which each gene was differentially expressed between rejection and no rejection and a p-value was calculated. Genes with high average fold changes and low p-values are considered best candidates for further development as rejection markers. However, it is important to note that a gene with a low average fold change and a high p-value may still be a useful marker for rejection in some patients and may work as part of a gene expression panel to diagnose rejection. These same PCR data were used to create PCR gene expression panels for diagnosis of acute rejection as discussed in example 17.

[0472] Non-parametric tests such as the Fisher Exact Test and Mann-Whitney U test are useful for choosing useful markers. They assess the ability of markers to discriminate between different classes as well as their significance. For example, one could use the median of all samples (including both non-rejector and rejector samples) as a threshold and apply the Fisher Exact test to the numbers of rejectors and non-rejectors above and below the threshold.

[0473] These methods were used to generate the data in Table 2D.

Example 14: Identification of diagnostic nucleotide sets for diagnosis of Cardiac Allograft Rejection using microarrays

[0474] Genes were identified which have expression patterns useful for the diagnosis and monitoring of acute cardiac allograft rejection. Further, sets of genes that work together in a diagnostic algorithm for allograft rejection were identified. Acute allograft rejection is a process that occurs in all solid organ transplantation including, heart, lung, liver, kidney, pancreas, pancreatic islet cell, intestine and others. Gene expression markers of acute cardiac rejection may be useful for diagnosis and monitoring of all allograft recipients. Patients, patient clinical data and patient samples used in the discovery of markers below were derived from a clinical study described in example 5.

[0475] The collected clinical data was used to define patient or sample groups for correlation of expression data. Patient groups were identified for comparison. For example, a patient group that possesses a useful or interesting clinical distinction, versus a patient group that does not possess the distinction. Measures of cardiac allograft rejection were derived from the clinical data to divide patients (and patient samples) into groups with higher and lower rejection activity over some period of time or at any one point in time. Such data were rejection grades as determined from histological reading of the cardiac biopsy specimens by a pathologist and data measuring progression of end-organ damage, including depressed left ventricular dysfunction (decreased cardiac output, decreased ejection fraction, clinical signs of low cardiac output) and usage of inotropic agents (Kobashigawa 1998).

[0476] Mononuclear RNA samples were collected and prepared from patients who had recently undergone a cardiac allograft transplantation using the protocol described in example 2. The allograft rejection status at the time of sample collection was determined by examination of cardiac biopsies as described in example 5 and as summarized here.

[0477] 300 patient samples were included in the analysis. Each patient sample was associated with a biopsy and other clinical data collected at the time of the sample. The cardiac biopsies were graded by a pathologist at the local center and by three centralized pathologists who read the biopsy slides from all four local centers in a blinded manner. Biopsy grades included 0, 1A, 1B, 2, 3A, and 3B. No grade 4 rejection was identified. Dependent variables were developed based on these grades using the local center pathology reading, the reading of a centralized and blinded pathologist, the highest of the readings, local or centralized and a consensus grade derived from all pathological readings. Samples were classified as no rejection or rejection in the following ways: Grade 0 vs. Grades 1-4, Grades 0 and 1A vs. Grades 1B-4, Grade 0 vs. Grade 3A, Grade 0 vs. Grades 1B-4, and Grade 0 vs. Grades 1B and 3A-4. Grade 0 samples were selected such that they were not immediately followed by an episode of acute rejection in the same patient. Comparing Grade 0 samples to Grade 3A samples gives the greatest difference between the rejection and no rejection groups on average.

[0478] Taking the highest of all pathologist readings has the effect of removing any sample from the no rejection class that was not a unanimous Grade 0. It also results in an increase in the number of rejection samples used in an analysis with the assumption that if a pathologist saw features of rejection, the call was likely correct and the other pathologists may have missed the finding. Many leading cardiac pathologists and clinicians believe that ISHLT grade 2 rejection does not represent significant acute rejection. Thus, for correlation analysis, exclusion of Grade 2 samples may be warranted. Clinical data were also used to determine criteria for including samples in the analysis. For example, a patient with an active infection or in the early post-transplant period (ongoing surgical inflammation) might have immune activation unrelated to rejection and thus be difficult to identify as patients without rejection. The strictest inclusion criteria required that samples be from patients who did not have a bacterial or viral infection, were at least two weeks post cardiac transplant, were asymptomatic and were not currently admitted to the hospital.

[0479] After preparation of RNA (example 2), amplification, labeling, hybridization, scanning, feature extraction and data processing were done as described in Example 11, using the oligonucleotide microarrays described in Example 9. The resulting log ratio of expression of Cy3 (patient sample)/ Cy5 (R50 reference RNA) was used for analysis.

[0480] Significance analysis for microarrays (SAM, Tusher 2001, Example 15) was used to discover genes that were differentially expressed between the rejection and no-rejection groups. Many different combinations of dependent variables, inclusion criteria, static/referenced, and data subsets were used in SAM analysis to develop the primary lists of genes significantly differentially expressed between rejection and no-rejection. As described in example 15, SAM assigns a false detection rate to each gene identified as differentially expressed. The most significant of these genes were identified.

[0481] An exemplary analysis was the comparison of Grade 0 samples to Grade 3A-4 samples using SAM. Data from the all the pathological readings was used to identify consensus Grade 0 samples and samples with at least one reading of Grade 3A or above. Using this definition of rejection and no rejection, expression profiles from rejection samples were compared to no rejection samples using SAM. The analysis identified 7 genes with a FDR of 1%, 15 genes @ 1.4%, 35 genes @ 3.9%. Many more genes were identified at higher FDR levels.

[0482] In Table 7, a number of SAM analyses are summarized. In each case the highest grade from the 3 pathologists was taken for analysis. No rejection and rejection classes are defined. Samples are either used regardless of redundancy with respect to patients or a requirement is made that only one sample is used per patient or per patient per class. The number of samples used in the analysis is given and the lowest FDR achieved is noted.

[0483] Some of the genes identified by SAM as candidate rejection markers are noted in Table 2A and B. SAM chooses genes as significantly different based on the magnitude of the difference between the groups and the variation among the samples within each group. It is important to note that a gene which is not identified by SAM as differentially expressed between rejection and no rejection may still be a useful rejection marker because: 1. The microarray technology is not adequately sensitive to detect all genes expressed at low levels. 2. A gene might be a useful member of a gene expression panel in that it is a useful rejection marker only in a subset of patients. This gene may not be significantly differentially expressed between all rejection and no rejection samples.

[0484] For the purposes of cross-validation of the results, the datasets were also divided into subsets to compare analysis between two subsets of roughly half of the data. The types of subsets constructed were as follows. First half/second half subsets were the first half of the samples and the second half of the samples from a dataset ordered by sample number. Odd/even subsets used the same source, a dataset ordered by sample number, but the odd subset consisted of every 2nd sample starting with the first and the even subset consisted of every 2nd sample starting with the second sample. Center 14/other subsets were the same datasets, divided by transplant hospital. The center 14 subset consisted of all samples from patients at center 14, while the other subset consisted of all samples from the other three centers (12, 13, and 15). When a gene was found to be significantly differentially expressed in both sets of data, a higher priority was put on that gene for development of a diagnostic test. This was reflected in a "Array Score" value (Table 2B) that also considered the false detection rate for the gene and the importance of the gene in classification models (see example 17).

[0485] Alternatively one can divide samples into 10 equal parts and do 10-fold cross validation of the results of SAM.

[0486] Microarray data was also used to generate classification models for diagnosis of rejection as described in example 17. Genes identified through classification models as useful in the diagnosis of rejection are noted in Table 2B in the column "models".

[0487] As genes were identified as useful rejection markers by microarray significance analysis, classification models, PCR analysis, or through searching the prior art, a variety of approaches were employed to discover genes that had similar expression behavior (coexpression) to the gene of interest. If a gene is a useful rejection marker, then a gene that is identified as having similar expression behavior is also likely to be a useful rejection marker. Hierarchical clustering (Eisen et al. 1998, see example 15) was used to identify co-expressed genes for established rejection markers. Genes were identified from the nearest branches of the clustering dendrogram. Gene expression profiles generated from 240 samples derived from transplant recipients were generated as described above. Hierarchical clustering was performed and co-expressed genes of rejection markers were identified. An example is shown in Figure 12. SEQ ID NO:85 was shown to be significantly differentially expressed between rejection and no rejection using both microarrays and PCR. Gene SEQ ID NO:3020 was identified by hierarchical clustering as closely co-expressed with SEQ ID NO:85.

[0488] Some of the primers for real-time PCR validation were designed for each of the marker genes as described in Example 12 and are listed in Table 2C and the sequence listing. PCR expression measurements using these primers were used to validate array findings, more accurately measure differential gene expression and create PCR gene expression panels for diagnosis of rejection as described in example 17.

[0489] Alternative methods of analyzing the data may involve 1) using the sample channel without normalization by the reference channel, 2) using an intensity-dependent normalization based on the reference which provides a greater correction when the signal in the reference channel is large, 3) using the data without background subtraction or subtracting an empirically, derived function of the background intensity rather than the background itself.

[0490] These methods were used to identify genes listed in Table 2B.

Example 15: Correlation and Classification Analysis

[0491] After generation and processing of expression data sets from microarrays as described in Example 11, a log ratio value is used for most subsequent analysis. This is the logarithm of the expression ratio for each gene between sample and universal reference. The processing algorithm assigns a number of flags to data that are of low signal to noise, saturated signal or are in some other way of low or uncertain quality. Correlation analysis can proceed with all the data (including the flagged data) or can be done on filtered data sets where the flagged data is removed from the set. Filtered data should have less variability and noise and may result in more significant or predictive results. Flagged data contains all information available and may allow discovery of genes that are missed with the filtered data set. After filtering the data for quality as described above and in example 11, missing data are common in microarray data sets. Some algorithms don't require complete data sets and can thus tolerate missing values. Other algorithms are optimal with or require imputed values for missing data. Analysis of data sets with missing values can proceed by filtering all genes from the analysis that have more than 5%, 10%, 20%, 40%, 50%, 60% or other % of values missing across all samples in the analysis. Imputation of data for missing values can be done by a variety of methods such as using the row mean, the column mean, the nearest neighbor or some other calculated number. Except when noted, default settings for filtering and imputation were used to prepare the data for all analytical software packages.

[0492] In addition to expression data, clinical data are included in the analysis. Continuous variables, such as the ejection fraction of the heart measured by echocardiography or the white blood cell count can be used for correlation analysis. Any piece of clinical data collected on study subjects can be used in a correlation or classification analysis. In some cases, it may be desirable to take the logarithm of the values before analysis. These variables can be included in an analysis along with gene expression values, in which case they are treated as another "gene". Sets of markers can be discovered that work to diagnose a patient condition and these can include both genes and clinical parameters. Categorical variables such as male or female can also be used as variables for correlation analysis. For example, the sex of a patient may be an important splitter for a classification tree.

[0493] Clinical data are used as supervising vectors (dependent variables) for the significance or classification analysis of expression data. In this case, clinical data associated with the samples are used to divide samples into clinically meaningful diagnostic categories for correlation or classification analysis. For example, pathologic specimens from kidney biopsies can be used to divide lupus patients into groups with and without kidney disease. A third or more categories can also be included (for example "unknown" or "not reported"). After generation of expression data and definition of supervising vectors, correlation, significance and classification analysis are used to determine which set of genes and set of genes are most appropriate for diagnosis and classification of patients and patient samples. Two main types of expression data analyses are commonly performed on the expression data with differing results and purposes. The first is significance analyses or analyses of difference. In this case, the goal of the analysis is to identify genes that are differentially expressed between sample groups and to assign a statistical confidence to those genes that are identified. These genes may be markers of the disease process in question and are further studied and developed as diagnostic tools for the indication. The second major type of analysis is classification analysis. While significance analysis identifies individual genes that are differentially expressed between sample groups, classification analysis identifies gene sets and an algorithm for their gene expression values that best distinguish sample (patient) groups. The resulting gene expression panel and algorithm can be used to create and implement a diagnostic test. The set of genes and the algorithm for their use as a diagnostic tool are often referred to herein as a "model". Individual markers can also be used to create a gene expression diagnostic model. However, multiple genes (or gene sets) are often more useful and accurate diagnostic tools.

Significance analysis for microarrays (SAM)

[0494] Significance analysis for microarrays (SAM) (Tusher 2001) is a method through which genes with a correlation between their expression values and the response vector are statistically discovered and assigned a statistical significance. The ratio of false significant to significant genes is the False Discovery Rate (FDR). This means that for each threshold there are some number of genes that are called significant, and the FDR gives a confidence level for this claim. If a gene is called differentially expressed between two classes by SAM, with a FDR of 5%, there is a 95% chance that the gene is actually differentially expressed between the classes. SAM will identify genes that are differentially expressed between the classes. The algorithm selects genes with low variance within a class and large variance between classes. The algorithm may not identify genes that are useful in classification, but are not differentially expressed in many of the samples. For example, a gene that is a useful marker for disease in women and not men, may not be a highly significant marker in a SAM analysis, but may be useful as part of a gene set for diagnosis of a multi-gene algorithm

[0495] After generation of data from patient samples and definition of categories using clinical data as supervising vectors, SAM is used to detect genes that are likely to be differentially expressed between the groupings. Those genes with the highest significance can be validated by real-time PCR (Example 13) or can be used to build a classification algorithm as described here.

Classification

[0496] Classification algorithms are used to identify sets of genes and formulas for the expression levels of those genes that can be applied as diagnostic and disease monitoring tests. The same classification algorithms can be applied to all types of expression and proteomic data, including microarray and PCR based expression data. Examples of classification models are given in example 17. The discussion below describes the algorithms that were used and how they were used.

[0497] Classification and Regression Trees (CART) is a decision tree classification algorithm (Breiman 1984). From gene expression and or other data, CART can develop a decision tree for the classification of samples. Each node on the decision tree involves a query about the expression level of one or more genes or variables. Samples that are above the threshold go down one branch of the decision tree and samples that are not go down the other branch. Genes from expression data sets can be selected for classification building with CART by significant differential expression in SAM analysis (or other significance test), identification by supervised tree-harvesting analysis, high fold change between sample groups, or known relevance to classification of the target diseases. In addition, clinical data can be used as

independent variables for CART that are of known importance to the clinical question or are found to be significant predictors by multivariate analysis or some other technique. CART identifies predictive variables and their associated decision rules for classification (diagnosis). CART also identifies surrogates for each splitter (genes that are the next best substitute for a useful gene in classification). Analysis is performed in CART by weighting misclassification costs to optimize desired performance of the assay. For example, it may be most important that the sensitivity of a test for a given diagnosis be > 90%. CART models can be built and tested using 10 fold cross-validation or v-fold cross validation (see below). CART works best with a smaller number of variables (5-50). Multiple Additive Regression Trees (Friedman, JH 1999, MART) is similar to CART in that it is a classification algorithm that builds decision trees to distinguish groups. MART builds numerous trees for any classification problem and the resulting model involves a combination of the multiple trees. MART can select variables as it build models and thus can be used on large data sets, such as those derived from an 8000 gene microarray. Because MART uses a combination of many trees and does not take too much information from any one tree, it resists over training. MART identifies a set of genes and an algorithm for their use as a classifier.

[0498] A Nearest Shrunken Centroids Classifier can be applied to microarray or other data sets by the methods described by Tibshirani et al. 2002. This algorithm also identified gene sets for classification and determines their 10 fold cross validation error rates for each class of samples. The algorithm determines the error rates for models of any size, from one gene to all genes in the set. The error rates for either or both sample classes can be minimized when a particular number of genes are used. When this gene number is determined, the algorithm associated with the selected genes can be identified and employed as a classifier on prospective sample.

[0499] For each classification algorithm and for significance analysis, gene sets and diagnostic algorithms that are built are tested by cross validation and prospective validation. Validation of the algorithm by these means yields an estimate of the predictive value of the algorithm on the target population. There are many approaches, including a 10 fold cross validation analysis in which 10% of the training samples are left out of the analysis and the classification algorithm is built with the remaining 90%. The 10% are then used as a test set for the algorithm. The process is repeated 10 times with 10% of the samples being left out as a test set each time. Through this analysis, one can derive a cross validation error which helps estimate the robustness of the algorithm for use on prospective (test) samples. Any % of the samples can be left out for cross validation (v-fold cross validation, LOOCV). When a gene set is established for a diagnosis with an acceptable cross validation error, this set of genes is tested using samples that were not included in the initial analysis (test samples). These samples may be taken from archives generated during the clinical study. Alternatively, a new prospective clinical study can be initiated, where samples are obtained and the gene set is used to predict patient diagnoses.

Example 16: Acute allograft rejection: biopsy tissue gene expression profiling

[0500] Acute allograft rejection involves activation of recipient leukocytes and infiltration into the rejecting organ. For example, CD8 T-cells are activated by CD4 T-cells and enter the allograft where they destroy graft tissue. These activated, graft-associated leukocytes may reside in the graft, die or exit the graft. Upon exiting, the cells can find their way into the urine or blood (in the case of renal allografts), bile or blood (liver allografts) or blood (cardiac allografts). These activated cells have specific gene expression patterns that can be measured using microarrays, PCR or other methods. These gene expression patterns can be measured in the graft tissue (graft associated leukocytes), blood leukocytes, urine leukocytes or stool/biliary leukocytes. Thus graft associated leukocyte gene expression patterns are used to discover markers of activated leukocytes that can be measured outside the graft for diagnostic testing.

[0501] Renal biopsy and cardiac biopsy tissue specimens were obtained for gene expression profiling. The specimens were obtained at the time of allograft biopsy and were preserved by flash freezing in liquid nitrogen using standard approaches or immersion in an RNA stabilization reagent as per the manufacturers recommendation (RNAlater, Qiagen, Valencia, CA). Biopsy allograft pathological evaluation was also obtained and samples were classified as having a particular ISHLT rejection grade (for cardiac) or acute rejection, chronic rejection, acute tubular necrosis or no disease (for renal).

[0502] 28 renal biopsy tissue samples were transferred to RLT buffer, homogenized and RNA was prepared using RNeasy preparation kits (Qiagen, Valencia, CA). Average total RNA yield was 1.3 ug. Samples were subjected to on column DNase digestion. 18 samples were derived from patients with ongoing acute allograft rejection and 10 were from controls with chronic rejection or acute renal failure.

[0503] RNA from the samples was used for amplification, labeling and hybridization to leukocyte arrays (example 11). Significance analysis for microarrays (SAM, Tusher 2001, Example 15) was used to identify genes that were differentially expressed between the acute rejection samples and controls. Leukocyte markers of acute rejection that are associated with the graft should be genes that are expressed at some level in activated leukocytes. Since leukocytes appear in graft tissue with some frequency with acute rejection, leukocyte genes associated with rejection are identified by SAM as upregulated in acute rejection in this experiment. 35 genes were identified as upregulated in acute rejection by SAM with less than a 5% false detection rate and 139 were detected with < 10.0% FDR.

[0504] For each of these genes, to 50mer oligonucleotide sequence was used to search NCBI databases including Unigene and OMIM. Genes were identified by sequence analysis to be either known leukocyte specific markers, known leukocyte expressed markers, known not to be leukocyte expressed or expression unknown. This information helped selected candidate leukocyte markers from all upregulated genes. This is necessary because some of the upregulated genes may have been expressed by renal tissue. Those genes that are leukocyte specific or leukocyte expressed were selected for evaluation by PCR in urine and blood samples from patients with and without acute allograft rejection (cardiac and renal). These genes are useful expression markers of acute rejection in allograft tissue specimens and may also be useful gene expression markers for the process in circulating leukocytes, or urine leukocytes. In addition, some of the leukocyte expressed genes from this analysis were selected for PCR validation and development for diagnosis of acute cardiac rejection and are noted in Table 2.

[0505] Five cardiac rejection markers in the peripheral blood were assayed using real-time PCR in renal biopsy specimens. The average fold change for these genes between acute rejection (n = 6) and controls (n = 6) is given below. Work is ongoing to increase the number of samples tested and the significance of the results.

PCR assays of cardiac rejection peripheral blood markers in renal allograft tissue. R = rejection, NR = No rejection.

Gene	Fold change (R/NR)
Granzyme B	2.16
CD20	1.42
NK cell receptor	1.72
T-box 21	1.74
IL4	1.3

[0506] Markers of renal rejection that are secreted from cells may be measured in the urine or serum of patients as a diagnostic or screening assay for rejection. Genes with lower molecular weight are most likely to be filtered into the urine to be measured in this way. Standard immunoassays may be used to measure these proteins.

Example 17: Microarray and PCR gene expression panels for diagnosis and monitoring of acute allograft rejection

Array panels / classification models

[0507] Using the methods disclosed, gene expression panels were discovered for screening and diagnosis of acute allograft rejection. Gene expression panels can be implemented for diagnostic testing using any one of a variety of technologies, including, but not limited to, microarrays and real-time PCR.

[0508] Using peripheral blood mononuclear cell RNA that was collected and prepared from cardiac allograft recipients as described in examples 2 and 5, leukocyte gene expression profiles were generated and analyzed using microarrays as described in examples 11,13, and 15. 300 samples were analyzed. ISHLT rejection grades were used to divide patients into classes of rejection and no rejection. Multiple Additive Regression Trees (MART, Friedman, JH 1999, example 15) was used to build a gene expression panel and algorithm for the diagnosis of rejection with high sensitivity. Default settings for the implementation of MART called TreeNet 1.0 (Salford Systems, San Diego, CA) were used except where noted.

[0509] 82 Grade 0 (rejection) samples and 76 Grade 1B-4 (no rejection) samples were divided into training (80% of each class) and testing (20% of each class) sets. A MART algorithm was then developed on the training set to distinguish rejection from no rejection samples using a cost of 1.02:1 for misclassification of rejection as no rejection. The resulting algorithm was then used to classify the test samples. The algorithm correctly classified 51 of 66 (77%) no rejection samples in the training set and 9 of 16 (56%) no rejection samples in the test set. For rejection samples 64 of 64 (100%) were correctly classified in the training set and 12 of 12 were correctly classified in the test set. The algorithm used 37 genes. MART ranks genes by order of importance to the model. In order, the 37 genes were: SEQ IDs: 3058, 3030, 3034, 3069, 3081, 3072, 3041, 3052, 3048, 3045, 3059, 3075, 3024, 279, 3023, 3053, 3022, 3067, 3020, 3047, 3033, 3068, 3060, 3063, 3028, 3032, 3025, 3046, 3065, 3080,3039,3055,49,3080,3038,3071.

[0510] Another MART model was built by excluding samples derived from patients in the first month post transplant and from patients with known CMV infection. 20 Grade 0 (rejection) samples and 25 Grade 1B-4 (no rejection) samples were divided into training (80% of each class) and testing (20% of each class) sets. A MART algorithm was then developed on the training set to distinguish rejection from no rejection samples using default settings. The resulting algorithm was

then used to classify the test samples. The algorithm correctly classified 100% of samples of both classes in the training and testing sets. However, this model required 169 genes. The sample analysis was done a second time with the only difference being requirement that all decision trees in the algorithm be composed of two nodes (single decision, "stump model"). In this case 15/16 no rejection samples were correctly identified in the training set and 4/4 no rejection samples were correctly identified in the test set. For the rejection samples, 17/19 were correctly identified in the training set and 5/6 were correctly classified in the test set. This model required 23 genes. In order of importance, they were: SEQ IDs: 3042, 2783, 3076, 3029, 3026, 2751, 3036, 3073, 3035, 3050, 3051, 3027, 3074, 3062, 3044, 3077, 2772, 3049, 3043, 3079, 3070, 3057, 3078.

Real-time PCR panels / classification models

[0511] PCR primers were developed for top rejection markers and used in real-time PCR assays on transplant patient samples as described in examples 12 and 13. This data was used to build PCR gene expression panels for diagnosis of rejection. Using MART (example 15) a 10-fold cross validated model was created to diagnose rejection using 12 no rejection samples (grade 0) and 10 rejection samples (grade 3A). Default settings were used with the exception of assigning a 1.02:1 cost for misclassification of rejection as no rejection and requirement that all decision trees be limited to 2 nodes ("stump model"). 20 genes were used in the model, including: SEQ IDs: 101, 3021, 102, 2781, 78, 87, 86, 36, 77, 2766, 3018, 80, 3019, 2752, 79, 99, 3016, 2790, 3020, 3056, 88. The 10-fold cross-validated sensitivity for rejection was 100% and the specificity was 85%. Some PCR primers for the genes are listed in Table 2C and the sequence listing.

[0512] A different analysis of the PCR data was performed using the nearest shrunken centroids classifier (Tibshirani et al. 2002; PAM version 1.01, see example 15). A 10-fold cross validated model was created to diagnose rejection using 13 no rejection samples (grade 0) and 10 rejection samples (grade 3A). Default settings were used with the exception of using a prior probability setting of (0.5, 0.5). The algorithm derives algorithms using any number of the genes. A 3-gene model was highly accurate with a 10 fold cross-validated sensitivity for rejection of 90%, and a specificity of 85%.

[0513] The 3 genes used in this model were: SEQ IDs 2784, 79, and 2794. Some of the PCR primers used are given in Table 2C and the sequence listing. An ROC curve was plotted for the 3-gene model and is shown in Figure 13.

Example 18: Assay sample preparation

[0514] In order to show that XDX's leukocyte-specific markers can be detected in whole blood, we collected whole blood RNA using the PAXgene whole blood collection, stabilization, and RNA isolation kit (PreAnalytix). Varying amounts of the whole blood RNA were used in the initial RT reaction (1, 2, 4, and 8ug), and varying dilutions of the different RT reactions were tested (1:5, 1:10, 1:20, 1:40, 1:80, 1:160). We did real-time PCR assays with primers specific to XDX's markers and showed that we can reliably detect these markers in whole blood.

[0515] Total RNA was prepared from 14 mononuclear samples (CPT, BD) paired with 14 whole blood samples (PAXgene, PreAnalytix) from transplant recipients. cDNA was prepared from each sample using 2ug total RNA as starting material. Resulting cDNA was diluted 1:10 and Sybr green real-time PCR assays were performed.

[0516] For real-time PCR assays, Ct values of 15-30 are desired for each gene. If a gene's Ct value is much above 30, the result may be variable and non-linear. For PAX sample, target RNA will be more dilute than in CPT samples. cDNA dilutions must be appropriate to bring Ct values to less than 30. Ct values for the first 5 genes tested in this way are shown in the table below for both whole blood RNA (PAX) and mononuclear RNA (CPT).

Gene	Ct PAX	Ct CPT
CD20	27.41512	26.70474
4761	28.45656	26.52635
3096	29.09821	27.83281
GranzymeB	31.18779	30.56954
IL4	33.11774	34.8002
Actin	19.17622	18.32966
B-GUS	26.89142	26.92735

[0517] With one exception, the genes have higher Ct values in whole blood. Using this protocol, all genes can be

detected with Cts <35. For genes found to have Ct values above 30 in target samples, less diluted cDNA may be needed.

Example 19: Allograft rejection diagnostic gene sequence analysis

[0518] Gene products that are secreted from cells or expressed as surface proteins have special diagnostic utility in that an assay may be developed to detect relative quantities of proteins in blood plasma or serum. Secreted proteins may also be detectable in urine, which may be a useful sample for the detection of rejection in renal allograft recipients. Cell surface markers may be detected using antigen specific antibodies in ELISA assays or using flow sorting techniques such as FACS.

[0519] Each gene that is found to be differentially regulated in one population of patients has several potential applications. It may be a target for new pharmaceuticals, a diagnostic marker for a condition, a benchmark for titrating drug delivery and clearance, or used in screening small molecules for new therapeutics. Any of these applications may be improved by an understanding of the physiologic function and localization of the gene product in vivo and by relating those functions to known diseases and disorders. Identifying the basic function of each candidate gene helps identify the signaling or metabolic pathways the gene is a part of, leading us to investigate other members of those pathways as potential diagnostic markers or targets of interest to drug developers.

[0520] For each of the markers in table 2, we attempted to identify the basic function and subcellular localization of the gene. These results are summarized in Table 9. In addition to initial DNA sequencing and processing, sequence analysis, and analysis of novel clones, information was obtained from the following public resources: Online Mendelian Inheritance in Man at the NCBI, LocusLink at the NCBI, the SWISS-PROT database, and Protein Reviews on the Web. For each marker represented by a curated reference mRNA from the RefSeq project, the corresponding reference protein accession number is listed. Curated sequences are those that have been manually processed by NCBI staff to represent the best estimate of the mRNA sequence as it is transcribed, based on alignments of draft DNA sequence, predicted initiation, termination and splice sites, and submissions of EST and full-length mRNA sequences from the scientific community.

[0521] These methods were used to derive the data in Table 2E.

Example 20: Detection of proteins expressed by diagnostic gene sequences

[0522] One of ordinary skill in the art is aware of many possible methods of protein detection. The following example illustrates one possible method.

[0523] The designated coding region of the sequence is amplified by PCR with adapter sequences at either end for subcloning. An epitope or other affinity "tag" such as a "His-tag" may be added to facilitate purification and/or detection of the protein. The amplified sequence is inserted into an appropriate expression vector, most typically a shuttle vector which can replicate in either bacteria, most typically *E. coli*, and the organism/cell of choice for expression such as a yeast or mammalian cell. Such shuttle vectors typically contain origins of replication for bacteria and an antibiotic resistance marker for selection in bacteria, as well as the relevant replication and selection sequences for transformation/transfection into the ultimate expression cell type. In addition, the sequence of interest is inserted into the vector so that the signals necessary for transcription (a promoter) and translation operably linked to the coding region. Said expression could be accomplished in bacteria, fungi, or mammalian cells, or by in vitro translation.

[0524] The expression vector would then typically be used to transform bacteria and clones analyzed to ensure that the proper sequence had been inserted into the expression vector in the productive orientation for expression. Said verified expression vector is then transfected into a host cell and transformants selected by a variety of methods including antibiotic resistance or nutritional complementation of an auxotrophic marker. Said transformed cells are then grown under conditions conducive to expression of the protein of interest, the cells and conditioned media harvested, and the protein of interest isolated from the most enriched source, either the cell pellet or media.

[0525] The protein is then be isolated by standard of chromatographic or other methods, including immunoaffinity chromatography using the affinity "tag" sequence or other methods, including cell fractionation, ion exchange, size exclusion chromatography, or selective precipitation. The isolated and purified protein is then be used as an antigen to generate specific antibodies. This is accomplished by standard methods including injection into heterologous species with an adjuvant, isolation of monoclonal antibodies from mice, or in vitro selection of antibodies from bacteriophage display antibody libraries. These antibodies are then used to detect the presence of the indicated protein of interest in a complex bodily fluid using standard methods such as ELISA or RIA.

Example 21: Detecting changes in the rate of hematopoiesis

[0526] Gene expression profiling of blood cells from cardiac allograft recipients was done using microarrays and real-time PCR as described in other examples herein.

EP 2 194 145 B1

[0527] Two of the genes in that were most correlated with cardiac transplant acute rejection with both microarrays and PCR were hemoglobin Beta and 2,3 DPGM. These genes are well know to be specific markers of erythrocyte lineages. This correlation was found using both purified peripheral mononuclear cells and whole blood RNA preparations.

[0528] Analysis of the five genes from the PCR data most strongly correlated with rejection showed that their expression levels were extremely highly correlated within each other ($R^2 > 0.85$).

Gene	Hs	Acc	SEQ ID No
hemoglobin, beta (HBB)	Hs.155376	NM_000518	86
2,3-bisphosphoglycerate mutase (BP	Hs.198365	X04327	87
cDNA FLJ20347	Hs.102669	AK000354	94
602620663F1cDNA	Hs.34549	AI123826	107
HA1247cDNA	Hs.33757	AI114652	91

[0529] This suggested that they were all elevated as part of a single response or process. When the microarray data was used to cluster these genes with each other and the other genes on the microarray, we found that these five genes clustered reasonably near each and of the other array genes which clustered tightly with them, four of the top 40 or so were platelet related genes. In addition, these a number of these genes clustered closely with CD34. CD34 is a marker of hematopoietic stem cells and is seen in the peripheral blood with increased hematopoiesis.

[0530] CD34, platelet RNA and erythrocyte RNA all mark immature or progenitor blood cells and it is clear that theses marker of acute rejection are part of a coordinated hematopoietic response. A small increase in the rate of production of RBCs and platelets may result in large fold changes in RNA levels. Immune activation from acute rejection may lead to increased hamatopoiesis in the bone marrow and non-marrow sites. This leads to an increase in many lineages because of the lack of complete specificity of the marrow response. Alternatively, increased hematopoiesis may occur in a transplant recipient due to an infection (viral or other), allergy or other stimulus to the system. This results in production of cells or a critical mass of immune cells that can cause rejection. In this scenario, monitoring for markers of immune activation would provide an opportunity for early diagnosis.

Table 1

Disease Classification	Disease/Patient Group
Cardiovascular Disease	Atherosclerosis Unstable angina Myocardial Infarction Restenosis after angioplasty Congestive Heart Failure Myocarditis Endocarditis Endothelial Dysfunction Cardiomyopathy Cardiovascular drug use
Infectious Disease	Hepatitis A, B, C, D, E, G Malaria Tuberculosis HIV Pneumocystis Carinii Giardia Toxoplasmosis Lyme Disease Rocky Mountain Spotted Fever Cytomegalovirus Epstein Barr Virus Herpes Simplex Virus

EP 2 194 145 B1

(continued)

Disease Classification	Disease/Patient Group
	Clostridium Dificile Colitis Meningitis (all organisms) Pneumonia (all organisms) Urinary Tract Infection (all organisms) Infectious Diarrhea (all organisms) Anti-infectious drug use
Angiogenesis	Pathologic angiogenesis Physiologic angiogenesis Treatment induced angiogenesis Pro or anti-angiogenic drug use
Transplant Rejection	Heart Lung Liver Pancreas Bowel Bone Marrow Stem Cell Graft versus host disease Transplant vasculopathy Skin Cornea Islet Cells Kidney Xenotransplants Mechanical Organ Immunosuppressive drug use
Hematological Disorders	Anemia - Iron Deficiency Anemia - B12, Folate deficiency Anemia - Aplastic Anemia - hemolytic Anemia - Renal failure Anemia - Chronic disease Polycythemia rubra vera Pernicious anemia Idiopathic Thrombocytopenic purpura Thrombotic Thrombocytopenic purpura Essential thrombocytosis Leukemia Cytopenias due to immunosuppression Cytopenias due to Chemotherapy Myelodysplasia

Table 2A.

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
HSRRN1 8S	18S ribosomal RNA		NA	X03205	
ACTB	Actin, beta		Hs.288061	NM_001101	
GUSB	Glucuronidase, beta		Hs.183868	NM_000181	
B2M	beta 2 microglobulin		Hs.75415	NM_004048	
TSN	Translin		Hs.75066	NM_004622	
CCR7	1707		Hs.1652	NM_001838	
IL1R2	4685-IL1R		Hs.25333	NM_004633	
AIF-1	Allograft inflammatory factor 1, all variants		Hs.76364	NM_004847	
ALAS2	ALAS2		Hs.323383	NM_000032 .1	
APELIN	APELIN		Hs.303084	NM_017413	
CD80	B7-1, CD80		Hs.838	NM_005191	
EPB41	Band 4.1		Hs.37427	NM_004437	
CBLB	c-cbl-B		Hs.3144	NM_004351	
CCR5	CCR5		Hs.54443	NM_000579	
MMB	CD10		Hs.1298	NM_000902	
KLRC1	CD159a		Hs.74082	NM_002259	
FCGR3A	CD16		Hs.176663	NM_000569	
FCGR3B	CD16b		Hs.372679	NM_000570	
LAG3	CD223		Hs.74011	NM_002286	
PECAMI	CD31		Hs.78146	NM_000442	
CD34	CD34		Hs.374990	NM_001773	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
FCGR1A	CD64		Hs.77424	NM_000566	
TFRC	CD71 = T9, transferrin receptor		Hs.77356	NM_003234	
CMA1	chymase		Hs.135626	NM_001836	
KIT	c-Kit		Hs.81665	NM_000222	
MPL	c-mpl		Hs.84171	NM_005373	
EphB6	FphB6		Hs.3796	NM_004445	
EPOR	EPO-R		Hs.127826	NM_000121 .2	
Foxp3	Foxp3		Hs.247700	NM_014009	
GATA1	GATA1		Hs.765	NM_002049	
ITGA2B	GP IIb		NM_000419.2	NM_000419	
GNLY	granulysin		Hs.105806	NM_006433	
GZMA	GZMA		Hs.90708	NM_006144	
HBA	hemoglobin, alpha I		Hs.398636	NM_000558 .3	
HBZ	hemoglobin, zeta		Hs.272003	NM_005332 .2	
HBB	hemoglobin, beta	36	Hs.155376	NM_000518 .4	368
HBD	hemoglobin, delta		Hs.36977	NM_000519 .2	
HBE	hemoglobin, epsilon I		Hs.117848	NM_005330	
HBG	hemoglobin, gamma A		Hs.283108	NM_000559 .2	
HBQ	hemoglobin, theta 1		Hs.247921	NM_005331	
HLA-DP	MH/c, class II, DP alpha 1		Hs.198253	NM_033554	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
HLA-DQ	MHC, class II, DQ alpha 1		Hs.198253	NM_002122	
HLA-DRB	MHC, class II, DR beta 1		Hs.375570	NM_002124.1	
ICOS	ICOS		Hs.56247	NM8012092	
IL18	IL18		Hs.83077	NM_001562	
IL3	interleukin 3 (colony-stimulating factor, multiple)		Hs.694	NM_000588	
ITGA4	Integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)		Hs.40034	NM_000885	
ITGAM	integrin, alpha M (complement component receptor 3, alpha; also known as CD11b(p170), macrophage antigen alpha polypeptide)		Hs.172631	NM_000632	
ITGB7	integrin, beta 7		Hs.1741	NM_000889	
CEBPB	LAP, CCAAT/enhancer binding protein (C/EBP), beta		Hs.99029	NM_005194	
NF-E2	NF-E2		Hs.75643	NM_006163	
PDCD1	programmed cell death 1, PD-1	52	Hs.158297	NM_005018	384
PF4	platelet factor 4 (chemokine (C-X-C motif) ligand 4)		Hs.81564	NM_002619	
PRKCC	protein kinase C, theta		Hs.211593	NM_006257.1	
PPARGC1	PPARgamma		Hs.198468	NM_013261	
RAG1	recombination activating gene 1		Hs.73958	NM_000448	
RAG2	recombination activating gene 2			NM_000536	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1) (SDF1)		Hs.237356	NM_000609	
TNFRSF4	tumor necrosis factor receptor superfamily, member 4		Hs.129780	NM_003327	
TNFSF4	tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally activated glycoprotein 1,34kDa)		Hs.181097	NM_003326	
TPS1	tryptase, alpha		Hs.334455	NM_003293	
ADA	adenosine deaminase		Hs.1217	NM_000022	
ADA	Carboxypeptidase M		Hs.334873	NM_001874 .1	
CSF2	colony stimulating factor, GM-CSF		Hs.1349	NM_000758 .2	
CSF3	colony stimulating factor 3, G-CSF		Hs.2233	NM_172219	
CRP	C-reactive protein, pentraxin-related (CRP),		Hs.76452	NM_000567 .1	
	Tyrosine Kinase 3		Hs.385	NM_004119	
FLT3 GATA3	FMS-Related GATA binding protein 3		Hs.169946	NM_002051 .	
IL7R	Interleukin 7 receptor		Hs.362807	NM 002185.1	
KLF1	Kruppel-like factor I (erythroid), EKL F		Hs.37860	NM_006563 .1	
LCK	lymphocyte-specific protein tyrosine kinase		Hs.1765	NM_005356 .2	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
LEF1	lymphoid enhancer-binding factor 1		Hs.44865	NM_016269 .2	
PLAUR	Urokinase-type Plasminogen Activator Receptor, CD87, uPAR		Hs.179657	NM_002659 .1	
TNFSF13 B	Tumor necrosis factor (ligand) superfamily, member 13b, BlyS/TALL-1/BAFF		Hs.270737	NM_006573 .3	
IL8	Interleukin 8		Hs.624	NM_000584	
GZMB	Granzyme B (granzyme 2, cytotoxic T-lymphocyte- associated serine esterase 1)		Hs.1051	NM_004131	
TNFSF6	Tumor necrosis factor (ligand) superfamily, member 6	77	Hs.2007	NM_000639	409
TCIRG1	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 prrotein a isoform 3	78	Hs.46465	NM_00601g	410
PRF1	Perforin 1 (pore forming protein)	79	Hs.2200	NM_005041	411
ILA	Interleukin 4	80	Hs.73917	NM_000589	412
IL13	Interleukin 13		Hs.845	NM_002188	
CTLA4	Cytotoxic T-lymphocyte- associated protein 4		Hs.247824	NM_005214	
CD8A	CD8 antigen, alpha polypeptide (p32)		Hs.85258	NM_001768	
BY55	Natural killer cell receptor, immunoglobulin superfamily member		Hs.8 1743	NM_007053	
OID 4460	EST	85	Hs.205159	AF150295	417
HBB	Hemoglobin, beta	86	Hs.155376	NM_000518	418

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
BPGM	2,3-bisphosphoglymate mutase	87	Hs.198365	NM_001724	419
MTHFD2	Methylene tetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase	88	Hs.154672	NM_006636	420
TAP1	Transporter 1, ATP-binding cassette, sub-family B (MDR1/TAP)		Hs.352018	NM_000593	
KPNA6	Karyopherin alpha 6 (importin alpha 7)	90	Hs.301553	AW021037	422
OID 4365	Mitochondrial solute carrier		Hs300496	A1114652	
IGHM	Immunoglobulin heavy constant mu		Hs.300697	BC032249	
OID 573	KIAA1486 protein		Hs.210958	AB040919	
OID 873	KIAA1892 protein	94	Hs.102669	AK000354	426
OID 3	EST		Hs.104157	AW968823	
CXCR4	Chemokine (C-X-C motif) receptor 4		Hs.89414	NM_003467	
CD69	CD69 antigen (p60, early T-cell activation antigen)		Hs.82401	NM_001781	
CCL5	Chemokine (C-C motif) ligand 5 (RANTES, SCYA5)		Hs.241392	NM_002985	
IL6	Interleukin 6		Hs.93913	NM_000600	
IL2	Interleukin 2		Hs.89679	NM_000586	
KLRF1	Killer cell lectin-like receptor subfamily F, member 1	101	Hs.183125	NM_016523	433

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	102	Hs.80887	NM_002350	434
IL2RA	Interleukin 2 receptor, alpha		Hs.1724	NM_000417	
CCL4	Chemokine (C-C motif) ligand 4, SCYA4		Hs.75703	NM_002984	
OID 6207	EST		Hs.92440	D20522	439
ChGn	Chondroitin beta 1,4 N-acetylgalactosaminyltransfera se		Hs.11260	NM_018371	
OID 4281	EST	107	Hs.34549	AA053887	
CXCL9	Chemokine (C-X-C motif) ligand 9 (MIG)		Hs.77367	NM_002416	
CXCL10	Chemokine (C-X-C motif) ligand 10, SCYB10		Hs.2248	NM_001565	
IL17	Interleukin 17 (cytotoxic T-lymphocyte-associated serine esterase 8)		Hs.41724	NM_002190	
IL15	Interleukin 15		Hs.168132	NM_000585	
IL10	Interleukin 10		Hs.193717	NM_000572	
IFNG	Interferon, gamma		Hs.856	NM_000619	
HLA-DRB 1	Major histocompatibility complex, class II, DR beta 1		Hs.308026	NM_002124	
CD8B1	CD8 antigen, beta polypeptide 1 (p37)		Hs.2299	NM_004931	
CD4	CD4 antigen (p55)		Hs.17483	NM_000616	
CXCR3	Chemokine (C-X-C motif) receptor GPR9		Hs.198252	NM_001504	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
OID 7094	XDX EST 479G12		NA	NA	
OID 7605	EST		Hs. 109302	AA808018	
CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)		Hs. 789	NM_001511	
OID 253	EST		Hs. 83086	AK091125	
GPI	Glucose phosphate isomerase		Hs. 409162	NM_000175	
CD47	CD47 antigen (Rh-related antigen, integrin-associated signal transducer)		Hs. 82685	NM_001777	
HLA-F	Major histocompatibility complex, class I, F		Hs. 377850	NM_018950	
OID 5350	EST		Hs. 4283	AK055687	
TCRGC2	T cell receptor gamma constant 2		Hs. 112259	M17323	
OID 7016	EST		NA	BI018696	
PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)		Hs. 196384	NM_000963	
OID 5847	Hypothetical protein FLJ32919		Hs. 293224	NM_144588	
PRDM1	PR domain containing 1, with ZNF domain		Hs. 388346	NM_001198	
CKB	Creatine kinase, Brain		Hs. 173724	NM_001823	
TNNI3	Troponin I, cardiac		Hs. 351382	NM_000363	
TNNT2	Troponin T2, cardiac		Hs. 296865	NM_000364	
MB	Myoglobin		Hs. 118836	NM_005368	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
SLC7A11	Solute carrier family 7, (cationic amino acid transporter, y+ system) member 11		Hs.6682	NM_014331	
TNFRSF5	tumor necrosis factor receptor superfamily, member 5; CD40		Hs.25648	NM_001250	
TNFRSF7	tumor necrosis factor receptor superfamily, member 7; CD27		Hs.355307	NM_001242	
CD86	CD86 antigen (CD28 antigen ligand 2, B7-2 antigen)		Hs.27954	NM_175862	
AIF1v2	Allograft inflammatory factor 1, splice variant 2		Hs.76364	NM_004847	
EBV BCLF-1	BCLF-1 major capsid		NA	AJ507799	
EBV EBV	EDNA repetitive sequence		NA	AJ507799	
CMV pp67	pp67		NA	X17403	
CMV TRL7	c6843-6595		NA	X17403	
CMV IE1e3	IE1 exon 3		NA	X17403	
CMV IE1e4	IE1 exon 4 (40 variants)		NA	X17403	
EBV EBNA-1	EBNA-1 coding region		NA	AJ507799	
EBV BZLF-1	Zebra gene		NA	AJ507799	
EBV EBN	EBNA repetitive sequence		NA	AJ507799	
EBV EBNA-LP	Short EBNA leader peptide exon		NA	AJ507799	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
CMV IE1	IE1S		NA	X17403	
CMV IE1	IE1-MC (exon 3)		NA	X17403	
CLC	Charot-Leyden crystal protein		Hs.889	NM_001828	
TERF2IP	telomeric repeat binding factor 2, interacting protein		Hs.274428	NM_018975	
HLA-A	Major histocompatibility complex, class I, A		Hs.181244	NM_002116	
OID 5891	EST 3' end		None	AW297949	
MSCP	mitochondrial solute carrier protein		Hs.283716	NM_018579	
DUSP5	dual specificity phosphatase 5		Hs.2128	NM_004419	
PRO1853	Hypothetical protein PRO1853		Hs.433466	NM_018607	
OID 6420	73A7, FLJ00290 protein		Hs.98531	AK090404	
CDSN	Corneodesmosin		Hs.507	NM_001264	
OID 4269	EST		Hs.44628	BM727677	
RPS25	Ribosomal protein S25		Hs.409158	NM_001028	
GAPD	Glyceraldehyde-3-phosphate dehydrogenase		Hs.169476	NM_002046	
RPLP1	Ribosomal protein, large, P1		Hs.424299	NM_001003	
OID_5115	qz23b07.x1 cDNA, 3' end /clone=IMAGE:2027701		NA	AI364926	
SLC9A8	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 8		Hs.380978	AB023156	
OID_1512	IMAGE:3865861 5 clone 5'		Hs.381302	BE618004	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
POLR2D	Polymerase (RNA) II (DNA directed) polypeptide D		Hs.194638	NM_004805	
ARPC3	Actin related protein 2/3 complex, subunit 3, 21kDa		Hs.293750	NM_005719	
OID_6282	EST 3' end		Hs.17132	BC041913	
PRO1073	PRO1073 protein		Hs.356442	AF001542	
OID_7222	EST, weakly similar to A43932 mucin 2 precursor, intestinal		Hs.28310	BG260891	
FPRL1	Formyl peptide receptor-like 1		Hs.99855	NM_001462	
FKBP1	FK506 binding protein like		Hs.99134	NM_022110	
PREB	Prolactin regulatory element binding		Hs.279784	NM_013388	
OID_1551	Hypothetical protein LOC200227		Hs.250824	BE887646	
OID_7595	DKPZP566F0546 protein		Hs.144505	NM_015653	
RNF19	Ring finger protein 19		Hs.48320	NM_015435	
SMCY	SMC (mouse) homolog, Y chromosome (SMCY)		Hs.80358	NM_004653	
OID_4184	CMVHCMVUL109		NA	X17403	
OID_7504	Hypothetical protein FLJ35207		Hs.86543	NM_152312	
DNAJC3	DnaJ (Hsp40) homolog, subfamily C, member 3		Hs.9683	NM_006260	
ARHU	Ras homolog gene family, member U		Hs.20252	NM_021205	
OID_7200	Hypothetical protein FLJ22059		Hs.13323	NM_022752	
SERPINB 2	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2		Hs.75716	NM_002575	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
ENO1	Enolase 1, alpha		Hs.254105	NM_001428	
OID_7696	EST 3' end		Hs.438092	AW217325	
OID_4173	CMV HCMVTRL2 (IRL2)		NA	X17403	
CSF2RB	Upstream variant mRNA of colony stimulating factor 2 receptor, beta, low-affinity (granulocyte- macrophage)		Hs.285401	AL540399	
OID_7410	CM2-LT0042-281299-062-e11 LT0042 cDNA, mRNA sequence		Hs.375145	AW837717	
OID_4180	CMV HCMVUS28		NA	X17403	
OID_5101	EST		Hs.144814	BG461987	
MOP3	MOP-3		Hs.380419	NM_018183	
RPL18A	Ribosomal protein L18a		Hs.337766	NM_000980	
INPP5A	Inositol polyphosphate-5- phosphatase, 40kDa		Hs.124029	NM_005539	
hIAN7	Immune associated nucleotide		Hs.124675	BG772661	
RPS29	Ribosomal protein S29		Hs.534	NM_001032	
OID_6008	EST 3' end		Hs.352323	AW592876	
OID_4186	CMV HCMVUL122		NA	X17403	
VNN2	vanin 2		Hs.121102	NM_004665	
OID_7703	KIAA0907 protein		Hs.24656	NM_014949	
OID_7057	480F8		NA	480F8	
OID_4291	EST		Hs.355841	BC038439	
OID_1366	EST		Hs.165695	AW850041	
EEF1A1	Eukaryotic translation elongation factor 1 alpha 1		Hs.422118	NM_001402	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
PA2G4	Proliferation-associated 2G4, 38kDa		Hs.374491	NM_006191	
GAPD	Glyceraldehyde-3-phosphate dehydrogenase		Hs.169476	NM_002046	
CHD4	Chromodomain helicase DNA bending protein 4		Hs.74441	NM_001273	
OID_7951	E2F-like protein (LOC51270)		Hs.142908	NM_016521	
DAB1	Disabled homolog 1 (Drosophila)		Hs.344127	NM_021080	
OID_3406	Hypothetical protein FLJ20356		Hs.61053	NM_018986	
OID_6986	462H9 EST		Hs.434526	AK093608	
OID_5962	EST 3' end		Hs.372917	AW452467	
OID_5152	EST 3' end		Hs.368921	AI392805	
S100A8	S100 calcium-binding protein A8 (calgranulin A)		Hs.416073	NM_002964	
HNRPU	HNRPU Heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)		Hs.103804	BM467823	
ERCC5	Excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))		Hs.48576	NM_000123	
RPS27	Ribosomal protein S27 (metallopanstimulin 1)		Hs.195453	NM_001030	
ACRC	acidic repeat containing (ACRC),		Hs.135167	NM_052957	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
PSMD11	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 11		Hs.90744	A1684022	
OID_1016	FLJ00048 protein		Hs.289034	AK024456	
OID_1309	AV706481 cDNA		None	AV706481	
OID-7582	Weakly similar to ZINC FINGER PROTEIN 142		Hs.16493	AK027866	
OID_4317	ta73c09.x1 3' end /clone=IMAGE:2049712 Ribosomal Protein S15		Hs.387179	AI318342	
OID_5889	3' end /clone=IMAGE:3083913		Hs.255698	AW297843	
UBL1	Ubiquitin-like 1 (sentrin)		Hs.81424	NM_003352	
OID_3687	EST		Zone	W03955	
OID_7371	EST 5'		Hs.290874	BE730505	
SH3BGR L3	SH3 domain binding glutamic acid-rich protein like 3		Hs.109051	NM_031286	
SEMA7A	Sema domain, immunoglobulin domain (Ig), and GPI membrane anchor, (semaphorin) 7A		Hs.24640	NM_003612	
OID_5708	EST 3' end		Hs.246494	AW081540	
OID_5992	EST 3' end		Hs.257709	AW467992	
IL21	Interleukin 21		Hs.302014	NM_021803	
HERC3	Hect domain and RLD 3 (HERC3)		Hs.35804	NM_014606	
OID_7799	AluJo/FLAM SINE/Alu			AW837717	
P11	26 serine protease		Hs.997	NM_006025	
OID_7766	EST 3' end		Hs.437931	AVJ294711	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
TTMM10	translocase of inner mitochondrial membrane 10 (yeast) homolog (TIMM10)		Hs.235750	NM_012456	
EGLN1	Egl nine homolog 1 (C. elegans)		Hs.6523	AJ310543	
TBCC	Tubulin-specific chaperone c		Hs.75064	NM_003192	
RNF3	Ring finger protein 3		Hs.8834	NM_006315	
OID_6451	170F9, hypothetical protein FLJ21439		Hs.288872	AL834168	
CCNDBP 1	cyclin D-type binding-protein 1 (CCNDBP1)		Hs.36794	NM_012142	
OID_8063	MUC18 gene exons 1&2		NA	X68264	
SW39H1	Suppressor of variegation 3-9 homolog 1 (Drosophila)		Hs.37936	NM_003173	
HSPC048	HSPC048 protein		Hs.278944	NM_014148	
OID_5625	EST 3' end from T cells		Hs.279121	AW063780	
WARS	Tryptophanyl-tRNA synthetase		Hs.82030	NM_004184	
OID_6823	107H8		Hs.169610	AL832642	
OID_7073	119F12		Has.13264	AL705961	
OID_5339	EST 3' end		Hs.436022	AI625119	
OID_4263	fetal retina 937202 cDNA clone IMAGE:565899		Hs.70877	AA136584	
MGC2676 6	Hypothetical protein MGC26766		Hs.288156	AK025472	
SERPINB 11	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 11		Hs.350958	NM_080475	
OID_6711	58G4, IMAGE:4359351 5'		none	BF968628	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
RNF10	Ring finger protein 10		Hs.5094	NM_014868	
MKRN1	Makorin, ring finger protein, 1		Hs.7838	NM_013446	
RPS16	ribosomal protein S16		Hs.397609	NM_001020	
BAZ1A	Bromodomain adjacent to zinc finger domain, 1A		Hs.8858	NM_013448	
OID_5998	EST 3' end		Hs.330268	AW468459	
ATP5L	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g		Hs.107476	NM_006476	
OID_6393	52B9		NA	52B9	
RoXaN	Ubiquitous tetrapeptide containing protein RoXaN		Hs.25347	BC004857	
NCBP2	Nuclear cap binding protein subunit 2, 20kDa		Hs.240770	NM_007362	
OID_6273	EST 3' end		Hs.158976	AW294774	
HZF12	zinc finger protein 12		Hs.164284	NM_033204	
CCL3	Chemokine (C-C motif) ligand 3		Hs.73817	D90144	
OID_4323	IMAGE:1283731 3'		Hs.370770	AA744774	
OID_5181	tg93h12.x1 NCI CGAP_CLL1 cDNA clone IMAGE:2116391 3' similar to contains TAR1.tl_ MER22		NA	A1400725	
PRDX4	Peroxiredoxin 4		Hs.83383	NM_006406	
BTk	Bruton agammaglobulinemia tyrosine kinase		Hs.159494	NM_000061	
OID_6298	Importin beta subunit mRNA		Hs.180446	A1948513	
PGK1	Phosphoglycerate kinase 1		Hs.78771	NM_000291	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
TNFRSF1 0A	Tumor necrosis factor receptor superfamily, member 10a		Hs.249190	NM_003844	
ADM	adrenomedullin		Hs.394	NM_001124	
OID_357	138G5		NA	138G5	
C20orf6	461A4 chromosome 20 open reading frame 6		Hs.88820	NM_016649	
OID_3226	DKFZP56400823 protein		Hs.105460	NM_015393	
ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase)	279	Hs.75811	NM_004315	611
ATFS	Activating transcription factor 5		Hs.9754	NM_012068	
OID_4887	hypothetical protein MGC14376		Hs.417157	NM_032895	
OID_4239	EST		Hs.177376	BQ022840	
MDM2	Mouse double minute 2, homolog of; p53-binding protein (MDM2), transcript variant MDM2,		Hs.170027	NM_002392	
XRN2	5'-3' exoribonuclease 2		Hs.268555	AF064257	
OID_6039	Endothelial differentiation, lysophosphatidic acid G-protein- coupled receptor, 4 (EDG4)		Hs.122575	BE502246	
OID_4210	IMAGE:4540096		Hs.374836	AI300700	
OID_7698	EST 3' end		Hs.118899	AA243283	
PRKRA	Protein kinase, interferon-inducible double stranded RNA dependent activator		Hs.18571	NM_003690	
OID_4288	IMAGE:2091815		Hs.309108	AI378046	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
OID_5620	EST 3' end from T cells		Hs.279116	AW063678	
OID_7384	EST 5'		Hs.445429	BF475239	
OID_1209	EST Weakly similar to hypothetical protein FLJ20378		Hs.439346	C14379	
CDKN1B	Cyclin-dependent kinase inhibitor 1B (p27, Kip1)		Hs.238990	NM_004064	
PLOD	Procollagen-lysine, 2- oxoglutarate 5- dioxygenase (lysine hydroxylase, Ehlers- Danlos syndrome type VI)		Hs.75093	NM_000302	
OID_5128	EST		Hs.283438	AK097845	
OID_5877	EST 3' end		Hs.438118	AW297664	
FZD4	Frizzled (Drosophila) homolog 4		Hs.19545	NM_012193	
HLA-B	Major histocompatibility complex, class I, B		Hs.77961	NM_005514	
OID_5624	EST 3' end from T cells		Hs.279120	AW063921	
FPR1	Formyl peptide receptor 1		Hs.753	NM_002029	
ODF2	Outer dense fiber of sperm tails 2		Hs.129055	NM_153437	
OID_5150	tg04g01.x1 cDNA, 3' end /clone=IMAGE:2107824	302	Hs.160981	AI392793	634
OID_5639	EST 3' end from T cells		Hs.279139	AW064243	
OID_6619	469A10		NA	469A10	
OID_6933	463C7, 4 EST hits. Aligned		Hs.86650	AI089520	
OID_7049	480E2		NA	480E2	
IL17C	Interleukin 17C		Hs.278911	NM_013278	
OID_5865	EST 3' end		Hs.255649	BM684739	
CD44	CD44		Hs.169610	AA916990	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
VPS45A	Vacuolar proteins sorting 45A (yeast)		Hs.6650	NM_007259	
OID_4932	aa92c03.r1 Stratagene fetal retina 937202 cDNA clone IMAGE:838756		NA	AA457757	
OID_7821	EST		NA	AA743221	
OID_4916	zr76a03.r1 Soares-NhHMPu_S1 cDNA clone IMAGE:669292		NA	AA252909	
OID_4891	Hypothetical protein LOC255488		Hs.294092	AL832329	
HADHB	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl- Coenzyme A thiolase/enoyl- Coenzyme A hydratase (trifunctional protein), beta subunit		Hs.146812	NM_000183	
FLJ22757	Hypothetical protein FLJ22757		Hs.236449	NM_024898	
RAC1	Ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)		Hs.173737	AK054993	
OID_6415	72D4, FLJ00290 protein	318	Hs.98531	CA407201	650
NMES1	Normal mucosa of esophagus specific 1		Hs.112242	NM_032413	
DMBT1	Deleted in malignant brain tumors 1, transcript variant 2		Hs.279611	NM_007329	
RPS23	ribosomal protein S23		Hs.3463	NM_001025	
ZF	HCF-binding transcription factor Zhangfei		Hs.29417	NM_021212	
NFE2L3	Nuclear factor (erythroid-derived 2)-like 3		Hs.22900	NM_004289	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
RAD9	RAD9 homolog (S. pombe)		Hs.240457	NM_004584	
OID_6295	EST 3' end		Hs.389327	A1880607	
DEFCAP	Death effector filament-forming Ced-4-like apoptosis protein, transcript variant B		Hs.104305	NM_014922	
RPL27A	Ribosomal protein L27a		Hs.76064	BF214146	
IL22	Interleukin 22 (IL22)		Hs.287369	NM_020525	
PSMA4	Proteasome (prosome, macropain) subunit, alpha type, 4, (PSMA4)		Hs.251531	NM_002789	
CCNI	cyclin I (CCNI)		Hs.79933	NM_006835	
THBD	Thrombomodulin		Hs.2030	NM_000361	
CGR19	Cell growth regulatory with ring finger domain		Hs.59106	NM_006568	
CLC	Charcot-Leyden crystal protein	NM_0018 28	779	4342	
TERF2IP	telomeric repeat binding factor 2, interacting protein	NM_0189 75	744	1775	
HLA-A	Major histocompatibility complex class I, A	NM_0021 16	735	125	1
OID_5891	EST 3'end	AW29794 9	730	7044.5	1
MSCP	mitochondrial solute carrier protein	NM_0185 79	730	3465.5	
DUSP5	dual specificity phosphatase 5	NM_0044 19	726	3122.5	

(continued)

Gene	Gene Name	SEQ ID		HS	ACC	SEQ ID RNA/cDN A
		50mer	5mer			
PRO1853	Hypothetical protein PRO1853	NM_0186 07		725	4153	
OID_6420	73A7, FLJ00290 protein	AK090404		725	7000.5	
CDSN	Comeodesmosin	NM_0012 64		722	2732	
OID_4269	EST	BM727677		715	5598.5	
RPS25	Ribosomal protein S25	NM_0010 28		710	164.5	
GAPD	Glyceraldehyde-3-phosphate dehydrogenase	NM_0020 46		707	215.5	
RPLP 1	Ribosomal protein, large, P1	NM_0010 03		703	157	
OID_5115	qz23b07.x1 CDNA, 3' tend /clone=IMAGE:2027701	AI364926		703	6629	1
SLC9A8	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 8	AB023156		702	2538.5	
OID_1512	IMAGE:3865861 5 clone 5'	BE618004		700	4008	1
POLR2D	Polymerase (RNA) II (DNA directed) polypeptide D	NM_0048 05		700	4190.5	
ARPC3	Actin related protein 2/3 complex, subunit 3, 21kDa	NM_0057 19		698	470.5	
OID_6282	EST 3' end	BC041913		697	4371.5	
PRO1073	PRO1073 protein	AF001542		697	6754	
OID_7222	EST, weakly similar to A43932 mucin 2 precursor, intestinal	BG260891		695	6759	

(continued)

Gene	Gene Name	SEQ ID		HS	ACC	SEQ ID RNA/cDN A
		50mer	5mer			
	FPRL1		NM_0014 62	692	4084.5	
	FKBPL		NM_0221 10	691	1780.5	
	PREB		NM_0133 88	690	3568	
	OID_1551		BE887646 LOC200227	689	6423	1
	OID_7595		NM_0156 53	689	3882.5	
	RNF19		NM_0154 35	689	7700.5	
	SMCY		NM_0046 53	687	6074.5	
	OID_4184		X17403	687	6810.5	
	OID_7504		NM_1523 12	686	6939	
	DNAJC3		NM_0062 60	686	3932.5	
	ARHU		NM_0212 05	686	7584	
	OID_7200		NM_0227 52	685	2804.5	
	SERPINB2		NM_0025 75	684	4690.5	
	ENO1		NM_0014 28	684	327	

(continued)

Gene	Gene Name		SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
	OID_7696	EST 3' end	AW29732 5	683	4875.5	
	OID_4173	CMV HCMVTRL2 (IRL2)	X17403	683	4010.5	
	CSF2RB	Upstream variant mRNA of colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage)	AL540399	683	3753	
	OID_7410	CM2-LT0042-281299-062.e 11 LT0042 cDNA, mRNA sequence	AW83771 7	682	7445	
	OID_4180	CMV HCMVUS28	X17403	681	4359	
	OID_5101	EST	BG461987	681	7272	
	MOP3	MOP-3	NM_0181 83	681	4085.5	1
	RPL18A	Ribosomal protein L18a	NM_0009 80	680	238	
	INPP5A	Inositol polyphosphate-5- phosphatase, 40kDa	NM_0055 39	680	4838.5	1
	hIAN7	Immune associated nucleotide	BG772661	680	4718	
	RPS29	Ribosomal protein S29	NM_0010 32	680	107.5	
	OID_6008	BST 3' end	AW59287 6	679	6560.5	
	OID_4186 OID_4186	CMV HCMVUL122	X17403	677	4788.5	
	VNN2	vanin 2	NM_0046 65	677	2620.5	

(continued)

Gene	Gene Name	SEQ ID		HS	ACC	SEQ ID RNA/cDN A
		50mer	5mer			
	OID_7703	KIAA0907 protein	NM_0149 49	676	6104.5	
	OID_7057	480F8	480F8	675	6862	
	OID_4291	EST	BC038439	674	5618.5	
	OID_1366	EST	AW85004 1	674	5590.5	1
	EEF1A1	Eukaryotic translation elongation factor 1 alpha 1	NM_0014 02	672	232	
	PA2G4	Proliferation-associated 2G4, 38kDa	NM_0061 91	672	4402	
	GAPD	Glyceraldehyde-3-phosphate dehydrogenase	NM_0020 46	671	194.5	
	CHD4	Chromodomain helicase DNA binding protein 4	NM_0012 73	671	2578.5	
	OID_7951	E2F-like protein (LOC51270)	NM_0165 21	671	4467	
	DAB1	Disabled homolog 1 (Drosophila)	NM_0210 80	670	6357.5	
	OID_3406	Hypothetical protein FLJ20356	NM_0189 86	669	2087	
	6986	462H9 EST	AK093608	669	4454	1
	OID_5962	EST 3' end	AW45246 7	668	5870.5	1
	OID_5152	EST 3' end	AI392805	668	6354.5	
	S100A8	S100 calcium binding protein A8 (calgranulin A)	NM_0029 64	668	134	

(continued)

Gene	Gene Name	SEQ ID		HS	ACC	SEQ ID RNA/cDN A
		50mer	5mer			
	HNRPU		BM467823	668	4108	
	ERCC5		NM_000123	668	6430.5	
	RPS27		NM_001030	668	160	
	ACRC		NM_052957	668	4871.5	1
	PSMD11		AI684022	668	4138	
	OID_1016		AK024456	667	5199	
	OID_1309		AV706481	667	7279.5	
	OID_7582		AK027866	667	5003.5	1
	OID_4317		AI318342	667	6499	
	OID_5889		AW297843	666	6837	1
	UBL1		NM_003352	666	1978.5	
	OID_3687		W03955	666	5519.5	

(continued)

Gene	Gene Name	SEQ ID 50mer		HS	ACC	SEQ ID RNA/cDN A
		EST 5'				
	OID_7371		BE730505	665	7751.5	
	SH3BGR1 3	SH3 domain binding glutamic acid-rich protein like 3	NM_0312 86	665	310	
	SEMA7A	Sema domain, immunoglobulin domain (Ig), and GPI membrane anchor, (semaphorin) 7A	NM_0036 12	665	3505.5	
	OID_5708	EST 3' end	AW08154 0	665	6224.5	
	OID_5992	EST 3' end	AW46799 2	665	5648	
	IL21	Interleukin 21	NM_0218 03	664	5036.5	
	HERC3	Hect domain and RLD 3 (HERC3)	NM_0146 06	664	3056.5	1
	OID_7799	AluJo/FLAM SINE/Alu	AW83771 7	664	3544	
	P11	26 serine protease	NM_0060 25	664	7173	
	OID_7766	EST 3' end	AW29471 1	663	7270.5	
	TIMM10	translocase of inner mitochondrial membrane 10 (yeast) homolog (TIMM10)	NM_0124 56	663	4779.5	
	EGLN1	Egl nine homolog 1 (C. elegans)	AJ310543	662	7172.5	
	TBCC	Tubulin-specific chaperone c	NM_0031 92	662	3384	

(continued)

Gene	Gene Name	SEQ ID		HS	ACC	SEQ ID RNA/cDN A
		50mer	5mer			
RNF3	Ring finger protein 3	NM_006315		661	4062	
OID_6451	170F9, hypothetical protein FLJ21439	AL834168		661	7126	1
CCNDBP1	cyclin D-type binding-protein 1 (CCNDBP1)	NM_012142		661	1919	
OID_8063	MUC18 gene exons 1&2	X68264		661	4692.5	
SUV39H1	Suppressor of variegation 3-9 homolog 1 (Drosophila)	NM_003173		661	5103	
HSPC048	HSPC048 protein	NM_004148		660	5981.5	
OID_5625	EST 3' end from T cells	AW063780		660	4437	
WARS	Tryptophanyl-tRNA synthetase	NM_004184		660	905.5	
OID_6823	107H8	AL832642		659	2619	
OID_7073	119F12	AL705961		659	6837.5	
OID_5339	EST 3' end	AI625119		658	4414.5	1
OID_4263	fetal retina 937202 cDNA clone IMAGE:565899	AA136584		658	5870	
MGC26766	Hypothetical protein MGC26766	AK025472		658	1892.5	
SERPINB1	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 11	NM_080475		658	7535.5	1
OID_6711	58G4, IMAGE:4359351 5'	BF968628		658	7264	

(continued)

Gene	Gene Name	SEQ ID		HS	ACC	SEQ ID RNA/cDN A
		50mer	5mer			
RNF10	Ring finger protein 10	NM_014868		658	3127.5	
MKRN1	Makorin, ring finger protein, 1	NM_013446		658	2228.5	
RPS16	ribosomal protein S16	NM_001020		657	165.5	
BAZ1A	Bromodomain adjacent to zinc finger domain, 1A	NM_013448		657	2533	
OID_5998	EST 3' end	AW468459		657	6339.5	
ATP5L	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g	NM_006476		657	1155	
OID_6393	52B9	52B9		657	7420.5	
RoXaN	Ubiquitous tetratricopeptide containing Protein RoXaN	BC004857		656	7378	
NCBP2	Nuclear cap binding protein subunit 2, 20kDa	NM_007362		656	4666.5	
OID_6273	EST 3' end	AW294774		656	5498.5	
HZF12	zinc finger protein 12	NM_033204		656	4715.5	
CCL3	Chemokine (C-C motif) ligand 3	D90144		656	4910	1
OID_4323	IMAGE:1283731 3'	AA744774		655	6406.5	1

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDNA
	OID_5181	tg93h12.x1 NCL CGAP_ CLL1 cDNA clone IMAGE: 2116391 3' similar to contains TAR1.tl MER22	655	4838	1
	PRDX4	Peroiredoxin 4	655	3397.5	
	BTK	Bruton agammaglobulinemia tyrosine kinase	655	2358	
	OID_6298	Importin beta subunit mRNA	655	2433.5	
	PGK1	Phosphoglycerate kinase 1	655	2059.5	
	TNFRSF1 0 A	Tumor necrosis factor receptor superfamily, member 10a	654	4897.5	1
	ADM	adrenomedullin	654	4235	
	OID_357	138G5	654	5427.5	1
	C20orf6	461A4 chromosome 20 open reading frame 6	654	6343	1
	OID_3226	DKFZP56400823 protein	653	6187.5	
279	ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	653	1003	
	ATF5	Activating transcription factor 5	653	4545.5	
	OID_4887	hypothetical protein MGC14376	653	2310	1

(continued)

Gene	Gene Name		SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
	OID_4239	EST	BQ022840	652	2774.5	
	MDM2	Mouse double minute 2, homolog of; p53-binding protein (MDM22), transcript variant MDM2,	NM_0023 92	652	4342	
	XRN2	5'-3' exoribonuclease 2	AF064257	652	6896.5	
	OID_6039	Endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 4 (EDG4)	BE502246	652	5147	
	OID_4210	IMAGE:4540096	AI300700	652	1330.5	
	OID_7698	EST 3' end	AA243283	652	7432.5	1
	PRKRA	Protein kinase, interferon-inducible double stranded RNA dependent activator	NM_0036 90	652	3512.5	
	OID_4288	IMAGE:2091815	AI378046	651	6401.5	
	OID_5620	EST 3' end from T cells	AW06367 8	651	6400	
	OID_7384	EST 3'	BF475239	651	6875	
	OID_1209	EST Weakly similar to hypothetical protein FLJ20378	C14379	651	1356.5	1
	CDKN1B	Cyclin-dependent kinase inhibitor 1B (p27, Kip1)	NM_0040 64	650	4272.5	
	PLOD	Procollagen-lysine, 2- oxoglutarate 5-dioxygenase (lysine hydroxylase, Ehlers- Danlos syndrome type VI)	NM_0003 02	650	3101	

(continued)

Gene	Gene Name		SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
	OID_5128	EST	AK097845	650	6476	
	OID_5877	EST 3' end	AW29766 4	650	6864.5	1
	FZD4	Frizzled (Drosophila) homolog 4	NM_0121 93	650	5816	
	HLA-B	Major histocompatibility complex, class I. B	NM_0055 14	650	229	
	OID_5624	EST 3' end from T cells	AW06392 1	649	7812.5	
	FPR1	Formyl peptide receptor 1	NM_0020 29	649	1156.5	
	ODF2	Outer dense fiber of sperm tails 2	NM_1534 37	649	4982.5	
302	OID_5150	tg04g01.x1 cDNA, 3' end /clone=IMAGE:2107824	AI392793	649	7638	
	OID_5639	EST 3' end from T cells	AW06424 3	648	6805	1
	OID_6619	469A10	469A10	647	7110	1
	OID_6933	463C7, 4 EST hits. Aligned	AI089520	647	6880.5	1
	OID_7049	480E2	480E2	647	7128.5	
	IL17C	Interleukin 17C	NM_0132 78	647	6411.5	
	OID_5866	EST 3' end	BM684739	647	6532	1
	CD44	CD44	AA916990	646	4758	
	VPS45A	Vacuolar protein sorting 45A (yeast)	NM_0072 59	646	3371	

(continued)

Gene	Gene Name	SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
OID_4932	aa92c03.r1 Stratagene fetal retina 937202 cDNA clone IMAGE:838756	AA457757	646	6057	1
OID_7821	EST	AA743221	645	7507	
OID_4916	zr76a03.r1 Soares_NhHMPu_S1 cDNA clone IMAGE:669292	AA252909	645	6962.5	1
OID_4891	Hypothetical protein LOC255488	AL832329	645	6148.5	
HADHB	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit	NM_000183	645	3212.5	
FLJ22757	Hypothetical protein FLJ22757	NM_024898	644	1965.5	1
RAC1	Ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	AK054993	644	1533	
318	OID_6415	CA407201	644	4881	
NMES1	Normal mucosa of esophagus specific 1	NM_032413	644	6217	1
DMBT1	Deleted in malignant brain tumors 1, transcript variant 2	NM_007329	644	7284	
RPS23	ribosomal protein S23	NM_001025	643	219.5	

(continued)

Gene	Gene Name		SEQ ID 50mer	HS	ACC	SEQ ID RNA/cDN A
	ZF	HCF-binding transcription factor Zhangfei	NM_021212	643	4069	
	NFE2L3	Nuclear factor (erythroid-derived 2)-like 3	NM_004289	643	3378	
	RAD9	RAD9 homolog (S. pombe)	NM_004584	643	6453	
	OID_6295	EST 3' end	AI880607	643	7493.5	
	DEFCAP	Death effector filament-forming Ced-4-like apoptosis protein, transcript variant B	NM_014922	643	3059	
	RPL27A	Ribosomal protein L27a	BF214146	642	6571	1
	IL22	Interleukin 22 (IL22)	NM_020525	642	3891	1
	PSMA4	Proteasome (prosome, macropain) subunit, alpha type, 4, (PSMA4)	NM_002789	641	1934.5	
	CCNI	cyclin I (CCNI)	NM_006835	641	980.5	
	THBD	Thrombomodulin	NM_000361	640	4732.5	
	CGR19	Cell growth regulatory with ring finger domain	NM_006568	640	5510	
			PCR Reverse Primer 1 SEQ ID	PCR Forward Primer 2 SEQ ID	PCR Reverse Primer 2 SEQ ID	PCR Probe 2 SEQ ID
	SEQ ID 50mer	SEQ ID RNA/cDNA	1031	1677	1925	2173
HBB	36	368	700	1362	1677	2173

(continued)

Gene	SEQ ID 50mer	SEQ ID RNA/cDNA	PCR Forward Primer 1 SEQ ID	PCR Reverse Primer 1 SEQ ID	PCR Probe 1 SEQ ID	PCR Forward Primer 2 SEQ ID	PCR Reverse Primer 2 SEQ ID	PCR Probe 2 SEQ ID
ITGB7	49	381	713	1044	1375			
PDCD1	52	384	716	1047	1378	1690	1938	2186
TNFSF6	77	409	741	1072	1403			
TCIRG1	78	410	742	1073	1404			
PRF1	79	411	743	1074	1405			
IL4	80	412	744	1075	1406			
OID_4460	85	417	749	1080	1411			
HBB	86	418	750	1081	1412			
BPGM	87	419	751	1082	1413			
MTHFD2	88	420	752	1083	1414			
OID_873	94	426	758	1089	1420			
KLRF1	101	433	765	1096	1427			
LYN	102	434	766	1097	1428			
CCL4	104	436	768	1099	1430			
OID_4281	107	439	771	1102	1433			
ASAH1	279	611	942	1273	1604	1850	2098	2346
OID_5150	302	634	965	1296	1627	1873	2121	2369
OID_6415	318	650	981	1312	1643	1889	2137	2385
Gene		Gene Name	SEQ ID 50mer	SEQ ID RNA/cDN A	n	Non-parametr ic Odds ratio	Fisher p-value	t-test p value
HBB		Hemoglobin, beta	86	418	55	8.33	0.00	0.00
OID_4365		Mitochondrial solute carrier			53	6.16	0.00	0.00
OID_873		KIAA1892 protein	94	426	55	5.09	0.01	0.01

(continued)

Gene	Gene Name	SEQ ID 50mer	SEQ ID RNA/cDNA	n	Non-parametric Odds ratio	Fisher p-value	t-test p value
IL4	Interleukin 4			46	4.90	0.02	0.01
OID_4281	EST	107	439	56	5.19	0.01	0.01
IGHM	Immunoglobulin heavy constant mu			52	2.89	0.09	0.01
BPGM	2,3-bisphosphoglycerate mutase	87	419	43	7.31	0.01	0.01
CTLA4	Cytotoxic T-lymphocyte-associated protein 4			52	1.84		0.02
SLC7A11	Solute carrier family 7, (cationic amino acid transporter, y+ system) member 11			48	2.50	0.15	0.03
IL13	Interleukin 13			29	4.95	0.07	0.04
OID_6207	EST			37	3.58	0.10	0.04
PRDM1	PR domain containing 1, with ZNF domain			57	1.44		0.07
LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	102	434	55	1.08		0.08
KPNA6	Karyopherin alpha 6 (importin alpha 7)	90	422	51	1.50		0.09
OID_7094	XDx EST 479G12			35	1.13		0.09
IL15	Interleukin 15			51	3.78	0.05	0.09
OID_4460	EST	85	417	47	2.73	0.14	0.10
OID_7016	EST			53	2.14	0.27	0.10
MTHFD2	Methylene tetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase			43	3.50	0.07	0.11

(continued)

Gene	Gene Name	SEQ ID 50mer	SEQ ID RNA/cDNA	n	Non-parametric Odds ratio	Fisher p-value	t-test p value
TCIRG1	T-cell, immune regulator 1, ATPase, H ⁺ transporting, lysosomal V0 protein a isoform 3	78	410	57	1.08		0.11
OID_5847	Hypothetical protein FLJ32919			45	1.08		0.12
CXCR4	Chemokine (C-X-C motif)			56	1.29		0.12
CXCR3	Chemokine (C-X-C motif)			54	2.10	0.27	0.12
GPI	Glucose phosphate isomer			57	1.44	0.60	0.12
KLRF1	Killer cell lectin-like receptor	101	433	50	1.68		0.13
CCL5	Chemokine (C-C motif)			34	1.96		0.13
CD47	CD47 antigen (Rh-related)			55	1.45		0.13
IL10	Interleukin 10			33	1.43		0.13
OID_253	EST			26	1.93		0.15
CXCL10	Chemokine (C-X-C motif)			53	1.75		0.16
IFNG	Interferon, gamma			41	1.33		0.16
PRF1	Perforin 1 (pore forming)	79	411	48	1.20		0.17
IL2	Interleukin 2			33	2.00		0.17
HLA-DRB1	Major histocompatibility			42	1.50		0.18
IL6	Interleukin 6			49	1.33		0.18
IL2RA	Interleukin 2 receptor, alpha			39	2.03	0.34	0.19
OID_573	KIAA1486 protein			8	3.00		0.19
CXCL9	Chemokine (C-X-C motif) ligand 9 (MIG)			46	1.71		0.20
OID_3	EST			49	2.19		0.20

(continued)

Gene	Gene Name	SEQ ID 50mer	SEQ ID RNA/cDNA	n	Non-parametric Odds ratio	Fisher p-value	t-test p value
CD8B1	CD8 antigen, beta polypeptide 1 (p37)			55	1.21		0.22
CD69	CD69 antigen (p60, early T-cell activation antigen)			30	1.71		0.23
OID_7605	EST			47	3.11	0.08	0.24
TNFSF6	Tumor necrosis factor (ligand) superfamily, member 6	77	409	54	1.36		0.25
CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, Alpha)			20	2.00		0.26
OID_5350	EST			49	2.08	0.26	0.28
CD8A	CD8 antigen, alpha polypeptide (p32)			57	1.39		0.28
CD4	CD4 antigen (p55)			55	1.64		0.28
PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)			46	2.05	0.37	0.29
GZMB	Granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)			40	1.81		0.33
CCL4	Chemokine (C-C motif) ligand 4, SCYA4			53	2.25		0.35
ChGn	Chondroitin beta 1,4 N-acetylgalactosaminyltransferase			31	2.57		0.36
TCRGC2	T cell receptor gamma constant 2			52	1.33		0.39
HLA-F	Major histocompatibility complex, class I, F			54	2.36	0.17	0.40

(continued)

Gene	Gene Name	SEQ ID 50mer	SEQ ID RNA/cDN A	n	Non-parametric Odds ratio	Fisher p-value	t-test p value
TAP1	Transporter 1, ATP-binding cassette, subfamily B (MDR1/TAP)			36	1.93		0.45
BY55	Natural killer cell receptor, immunoglobulin superfamily member			52	2.49	0.16	0.48
IL8	Interleukin 8			49	2.10	0.26	0.49
Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein		
ACTB	NM_001101			NP_001092			
GUSB	NM_000181			NP_000172			
B2M	NM_004048			NP_004039			
TSN	NM_004622			NP_004613			
CCR7	NM_001838			NP_001829			
IL1R2	NM_004633			NP_004624			
AIF-1	NM_004847			NP_004838			
IALAS2	NM_000032.1			NP_000023			
APELIN	NM_017413			NP_059109			

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDNA	RefSeq Peptide Accession #	SEQ ID Protein
CD80	NIM_00519 1		343	NP_005182	
EPB41	NIM_00443 7		344	NP_004428	
CBLB	NIM_00435 1		345	NP_733762	
CCR5	NIM_00057 9		346	NP_000570	
MME	NIM_00090 2		347	NP_000893	
KLRC1	NIM_00225 9		348	NP_002250	
FCGR3A	NIM_00056 9		349	NP_000560	
FCGR3B	NIM_00057 0		350	NP_000561	
LAG3	NIM_00228 6		351	NP_002277	
PECAM1	NIM_00044 2		352	NP_000433	
CD34	NIM_00177 3		353	NP_001764	
FCGR1A	NIM_00056 6		354	NP_000557	
TFRC	NIM_00323 4		355	NP_003225	
CMA1	NIM_00183 6		356	NP_001827	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
KIT	NIM_00022 2		357	NP_000213	
MPL	NIM_00537 3		358	NP_005364	
EphB6	NIM_00444 5		359	NP_004436	
EPO-R	NIM_00012 1.2		360	NP_000112	
Foxp3	NIM_01400 9		361	NP_054728	
GATA-1	NIM_00204 9		362	NP_002040	
ITGA2B	NIM_00041 9		363	NP_000410	
GNLY	NIM_00643 3		364	NP_006424	
GZMA	NIM_00614 4		365	NP_006135	
HBA	NIM_00055 8.3		366	NP_000549	
HBZ	NIM_00533 2.2		367	NP_005323	
HBD	NIM_00051 9.2		369	NP_000510	
HBB	NIM_00533 0		370	NP_005321	
HBG	NIM_00055 9.2		371	NP_000550	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
HBQ	NM_00533 1		372	NP_005322	
HLA-DP	NM_03355 4		373	NP_291032	
HLA-DQ	NM_00212 2		374	NP_002113	
ICOS	NM_01209 2		376	NP_036224	
IL18	NM_00156 2		377	NP_001553	
IL3	NM_00058 8		378	NP_000579	
ITGA4	NM_00088 5		379	NP_000876	
ITGAM	NM_00063 2		380	NP_000623	
ITGB7	NM_00088 9	49	381	NP_000880	2445
CEBPB	NM_00519 4		382	NP_005185	
NF-E2	NM_00616 3		383	NP_006154	
PDCD1	NM_00501 8	52	384	NP_005009	2448
PF4	NM_00261 9		385	NP_002610	
PRKCQ	NM_00625 7.1		386	NP_006248	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
PPARGC 1	NM_01326 1		387	NP_037393	
RAG1	NM_00044 8		388	NP_000439	
RAG2	NM_00053 6		389	NP_000527	
CXCL12	NM_00060 9		390	NP_000600	
TNFRSF4	NM_00332 7		391	NP_003318	
TNFSF4		NM_003326		NP_003317	
TPS1		NM_003293		NP_003284	
ADA		NM_000022		NP_000013	
CPM		NM_001874.1		NP_001865	
CSF2		NM_000758.2		NP_000749	
CSF3		NM_172219		NP_757373	
CRP		NM_000567.1		NP_000558	
FLT3		NM_004119		NP_004110	
GATA3		NM_002051.1		NP_002042	
IL7R		NM_002185.1		NP_002176	
KLF1		NM_006563.1		NP_006554	
LCK		NM_005356.2		NP_005347	
LEF1		NM_016269.2		NP_057353	
PLAUR		NM_002659.1		NP_002650	
TNFSF13 B		NM_006573.3		NP_006564	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
IL8	NM_000584			NP_000575	
GZMB	NM_004131			NP004122	
TNFSF6	NM_000639	77	409	NP_000630	2473
TCIRG1	NM_006019	78	410	NP_006010	2474
PRF1	NM_005041	79	411	NP_005032	2475
IL4	NM_000589	80	412	NP_000580	2476
IL13	NM_002188			NP_002179	
CTLA4	NM_005214			NP_005205	
CD8A	NM_001768			NP_001759	
BY55	NM_007053			NP_008984	
HBB	NM_000518	86	418	NP_000509	2481
BPGM	NM_001724	87	419	NP_001715	2482
MTHFD2	NM_006636	88	420	NP_006627	2483
TAP1	NM_000593			NP000584	
OID_873	AK000354	94	426NP_	056212	2485
CXCR4	NM_003467			NP_003458	2486
CD69	NM_001781			NP_001772	2487
CCL5	NM_002985			NP_002976	2488
IL6	NM_000600			NP_000591	2489
IL2	NM_000586			NP_000577	
KLRF1	NM_016523	101	433	NP_057607	2491
LYN	NM_002350	102	434	NP_002341	2492
IL2RA	NM_000417			NP_000408	
CCLA	NM_002984			NP_002975	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
ChGn	NM_018371			NP_060841	
CXCL9	NM_002416			NP_002407	
CXCL10	NM_001565			NP_001556	
IL17	NM_002190			NP_002181	
IL15	NM_000585			NP_000576	
IL10	NM_000572			NP_000563	
IFNG	NM_000619			NP_000610	
HLA-DRB 1	NM_002124			NP_002115	
CD8B1	NM_004931			NP_004922	
CD4	NM_000616			NP_000607	
CXCR3	NM_001504			NP_001495	
CXCL1	NM_001511			NP_001502	
GPI	NM_000175			NP_000166	
CD47	NM_001777			NP_001768	
HLA-F	NM_018950			NP_061823	
PTGS2	NM_000963			NP_000954	
OID_5847	NM_144588			NP_653189	
PRDMI	NM_001198			NP_001189	
CKB	NM_001823			NP_001814	
TNNT3	NM_000363			NP_000354	
TNNT2	NM_000364			NP_000355	
MB	NM_005368			NP_005359	
SLC7A11	NM_014331			NP_055146	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
TNFRSF5		NM_001250		NP_001241	
TNFRSF7		NM_001242		NP_001233	
CD86		NM_175862		NP_787058	
AIF1v2		NM_004847		NP_004838	
CMV IE1e3		NC_001347, compl		NP_040060	
CMV IE1e4		NC_001347, compl		NP_040060	
EV EBNA-1		NC_001345, compl		NP_039875	
EV BZLF-1		NC_001345, compl		NP_039871	
CMV IE1		NC_001347, compl		NP_040060	
CMV IE1		NC_001347, compl		NP_040060	
CLC		NM_001828		NP_001819	
TERF2IP		NM_018975		NP_061848	
HLA-A		NM_002116		NP_002107	
MSCP		NM_018579		NP_061049	
DUSP5		NM_004419		NP_004410	
PRO1853		NM_018607		NP_061077	
CDSN		NM_001264		NP_001255	
RPS25		NM_001028		NP_001019	
GAPD		NM_002046		NP_002037	
RPLP1		NM_001003		NP_000994	
POLR2D		NM_004805		NP_004796	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
ARPC3	NM_005719			NP_005710	
FPRL1	NM_001462			NP_001453	
FKBP1	NM_022110			NP_071393	
PREB	NM_013388			NP_037520	
OID_7595	NM_015653			NP_056468	
RNF19	NM_015435			NP_056250	
SMCY	NM_004653			NP_004644	
OID_7504	NM_152312			NP_689525	
DNAJC3	NM_006260			NP_006251	
ARHU	NM_021205			NP_067028	
OID_7200	NM_022752			NP_073589	
SERPINB 2	NM_002515			NP_002566	
ENO1	NM_001428			NP_001419	
MOP3	NM_018183			NP_060653	
RPL18A	NM_000980			NP_000971	
INPP5A	NM_005539			NP_005530	
RPS29	NM_001032			NP_001023	
VNN2	NM_004665			NP_004656	
OID7703	NM_014949			NP_055764	
EEF1A1	NM_001402			NP_001393	
PA2G4	NM_006191			NP_006182	
GAPD	NM_002046			NP_002037	
CHD4	NM_001273			NP_001264	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
OID_7951		NM_016521		NP_057605	
DAB1		NM_021080		NP_066566	
OID_3406		NM_018986		NP_061859	
S100A8		NM_002964		NP_002955	
ERCC5		NM_000123		NP_100114	
RPS27		NM_001030		NP_001021	
ACRC		NM_052957		NP_443189	
UBL1		NM_003352		NP_003343	
SH3BGR L3		NM_031286		NP_112576	
SEMA7A		NM_003612		NP_003603	
IL21		NM_021803		NP_068575	
HERC3		NM_014606		NP_055421	
P11		NM_006025		NP_006016	
TIMM10		NM_012456		NP_036588	
EGLN1		AJ310543		NP_071334	
TBCC		NM_003192		NP_003183	
RNF3		NM_006315		NP_006306	
CCNDBP 1		NM_012142		NP_036274	
SUV39H1		NM_003173		NP_003164	
HSPC048		NM_014148		NP_054867	
WARS		NM_004184		NP_004175	
SERPINB 11		NM_080475		NP_536723	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
RNF10	NM_014868			NP_055683	
MKRN1	NM_013446			NP_038474	
RPS16	NM_001020			NP_001011	
BAZ1A	NM_013448			NP_038476	
ATPSL	NM_006476			NP_006467	
NCBP2	NM_007362			NP_031388	
HZF12	NM_033204			NP_149981	
CCL3	D90144			NP_002974	
PRDX4	NM_006406			NP_006397	
BTX	NM_000061			NP_000052	
PGK1	NM_000291			NP_000282	
TNFRSF10A	NM_003844			NP_003835	
ADM	NM_001124			NP_001115	
C20orf6	NM_016649			NP_057733	
OID_3226	NM_015393			NP_056208	
ASAH1	NM_004315	279	611	NP_004306	2599
ATF5	NM_012068			NP_036200	
OID_4887	NM032895			NP_116284	
MDM2	NM_002392			NP_002383	
XRN2	AF064257			NP_036387	
PRKRA	NM_003690			NP_003681	
CDKN1B	NM_004064			NP_004055	
PLOD	NM_000302			NP_000293	

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
FZD4	NM_012193			NP_036325	
HLA-B	NM_005514			NP_005505	
FPR1	NM_002029			NP_002020	
ODF2	NM_153437			NP_702915	
IL17C	NM_013278			NP_037410	
VPS45A	NM_007259			NP_009190	
HADHB	NM_000183			NP_000174	
FLJ22757	NM_024898			NP_079174	
NMES1	NM_032413			NP_115789	
DMT1	NM_007329			NP_015568	
RPS23	NM_001025			NP_001016	
ZF	NM_021212			NP_067035	
NFE2L3	NM_004289			NP_004280	
RAD9	NM_004584			NP_004575	
DEFCAP	NM_014922o			NP_055737	
IL22	NM_020525			NP_065386	
PSMA4	NM_002789			NP_002780	
CCNI	NM_006835			NP_006826	
THBD	NM_000361			NP_000352	
CGR19	NM_006568			NP_006559	
HSRRN1 8S	X03205				
HBB	NG_000007	36	368		
HLA-DRB					

(continued)

Gene	ACC	SEQ ID 50mer		SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
OID_4460	AF150295	85		417		
KPNA6	AW021037	90		422		
OID_4365	AI114652					
IGHM	BC032249					
OID_573	AB040919					
OID_3	AW968823					
OID_6207	D20522					
OID_4281	AA053887	107		439		
OID_7094						
OID_7605	AA808018					
OID_253	AK091125					
OID_5350	AK055687					
TCRGC2	M17323					
OID_7016	BI018696					
EV EBV						
CMV p67	NC_001347					
CMV TRL7						
EV EBN						
EV EBNA-LP						
OID_5891	AW297949					
OID_6420	AK090404					
OID_4269	BM727677					

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
OID_5115	A1364926				
SLC9A8	AB023156				
OID_1512	BE618004				
OID_6282	BC041913				
PRO1073	AF001542				
OID_7222	BG260891				
OID_1551	BE887646				
OID_4184	X17403				
OID_7696	AW297325				
OID_4173	X17403				
CSF2RB	IAL540399				
OID_7410	AW837717				
OID_4180	X17403				
OID_5101	BG461987				
hIAN7	8G772661				
OID_6008	JAW592876				
OID_4186	X17403				
OID_7057	480F8				
OID_4291	BC038439				
OID_1366	AW850041				
OID_6986	AK093608				
OID_5962	AW452467				
OID_5152	AI392805				
HNRPU	BM467823				

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
PSMD11		AI684022			
OID_1016		Zap024456			
OID_1309		AV706481			
OID_7582		AK027866			
OID_4317		AI318342			
OID_5889		AW297843			
OID_3687		W03955			
OID_7371		BE730505			
OID_5708		AW081540			
OID_5992		AW467992			
OID_7799		AW837717			
OID_7766		AW294711			
OID_6451		AL834168			
OID_8063		X68264			
OID_5625		AW063780			
OID_6823		L832642			
OID_7073		AL705961			
OID_5339		AI625119			
OID_4263		AA136584			
MGC2676 6		AK025472			
OID_6711		BF968628			
OID_5998		AW468459			
OID_6393		52B9			

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
RoXaN		BC004857			
OID_6273		AW294774			
OID_4323		AA744774			
OID_5181		AI400725			
OID_6298		AI948513			
OID_357		138G5			
OID_4239		BQ022840			
OID_6039		BE502246			
OID_210		AI300700			
OID_7698		AA243283			
OID_4288		AI378046			
OID_5620		AW063678			
OID_7384		BF475239			
OID_1209		C14379			
OID_5128		AK097845			
OID_5877		AW297664			
OID_5624		AW063921			
OID_5150		AI392793	302	634	
OID_5639		AW064243			
OID_6619		469A10			
OID_6933		AI089520			
OID_7049		480E2			
OID_5866		BM684739			
CD44		AA916990			

(continued)

Gene	ACC	SEQ ID 50mer	SEQ ID RNA/cDN A	RefSeq Peptide Accession #	SEQ ID Protein
OID_4932	AA457757				
OID_7821	AA743221				
OID_4916	AA252909				
OID_4891	AL832329				
RAC1	AK054993				
OID_6415	CA407201	318	650		
OID_6295	A1880607				
RPL27A	BF214146				

Table 3: Viral genomes were used to design oligonucleotides for the microarrays. The accession numbers for the viral genomes used are given, along with the gene name and location of the region used for oligonucleotide design.

Virus	Gene Name	Genome Location
Adenovirus, type 2 Accession #J01917	E1a	1226..1542
	E1b_1	3270...3503
	E2a_2	complement(24089..25885)
	E3-1	27609..29792
	E4 (last exon at 3'-end)	complement(33193..32802)
	IX	3576..4034
	Iva2	complement(4081..5417)
Cytomegalovirus (CMV) Accession #X17403	DNA Polymerase	complement(5187..5418)
	HCMVTRL2 (IRL2)	1893..2240
	HCMVTRL7 (IRL7)	complement(6595..6843)
	HCMVUL21	complement(26497..27024)
	HCMVUL27	complement(32831..34657)
	HCMVUL33	43251..44423
	HCMVUL54	complement(76903..80631)
	HCMVUL75	complement(107901..110132)
	HCMVUL83	complement(119352..121037)
	HCMVUL106	complement(154947..155324)
	HCMVUL109	complement(157514..157810)
	HCMVUL113	161503..162800
	HCMVUL122	complement(169364..170599)
	HCMVUL123 (last exon at 3'-end)	complement(171006..172225)
	HCMVUS28	219200..220171
Epstein-Barr virus (EBV) Accession # NC_001345	Exon in EBNA-1 RNA	67477..67649
	Exon in EBNA-1 RNA	98364..98730
	BRLF1	complement(103366..105183)
	BZLF1 (first of 3 exons)	complement(102655..103155)
	BMLF1	complement(82743..84059)
	BALF2	complement(161384..164770)
Human Herpesvirus 6 (HHV6) Accession #NC_001664	U16/U17	complement(26259..27349)
	U89	complement(133091..135610)
	U90	complement(135664..135948)
	U86	complement(125989..128136)
	U83	123528..123821
	U22	complement(33739..34347)
	DR2 (DR2L)	791..2653
	DR7 (DR7L)	5629..6720
	U95	142941..146306
	U94	complement(141394..142866)
	U39	complement(59588..62080)
	U42	complement(69054..70598)
	U81	complement(121810..122577)
	U91	136485..136829

Table 4: Dependent variables for discovery of gene expression markers of cardiac allograft rejection.

Dependent Variable	Description	Number of Rejection Samples	Number of No-Rejection Samples
0 vs 1-4 Bx	Grade 0 vs. Grades 1-4, local biopsy reading	65	114
s0 vs 1B-4 HG	Stable Grade 0 vs Grades 1B-4, highest grade, Grade 1A not included	41	57
0-1A vs 1B-4 HG	Grades 0 and 1A vs Grades 1B-4, highest grade.	121	58
0 vs 3A HG	Grade 0 vs Grade 3A, highest grade. Grades 1A-2 and Grade 3B were not included.	56	29
0 vs 1B-4	Grade 0 vs Grades 1B-4, highest grade. Grade 1A was not included.	57	57
0 vs 1A-4	Grade 0 vs. Grades 1-4, highest grade	56	123

Table 5: Real-time PCR assay chemistries. Various combinations of reporter and quencher dyes are useful for real-time PCR assays.

Reporter	Quencher
FAM	TAMRA
	BHQ1
TET	TAMRA
	BHQ1
JOE	TAMRA
	BHQ1
HEX	TAMRA
	BHQ1
VIC	TAMRA
	BHQ1
ROX	BHQ2
TAMRA	BHQ2

Table 6: Real-time PCR results for rejection markers

Gene Array Probe SEQ ID	Phase 1				Phase 2				All Data			
	Fold	t-Test	NR	R	Fold	t-Test	NR	R	Fold	t-Test	NR	R
79	1.822	0.01146	6	7	0.63	0.04185	19	15	0.72	0.05632	35	26
3016	1.045	0.41017	12			10			1.001	0.49847	16	15
2766	0.956	0.43918	12	10	0.989	0.48275	19	14	0.978	0.45101	31	24

EP 2 194 145 B1

(continued)

Gene Array Probe SEQ ID	Phase 1				Phase 2				All Data			
	Fold	t-Test	NR	R	Fold	t-Test	NR	R	Fold	t-Test	NR	R
94	1.601	0.02418	11	10					1.831	0.00094	17	15
36	1.605	0.09781	12	8	2.618	0.01227	18	11	2.808	0.00015	38	23
80	5.395	0.00049	9	6	4.404	0.05464	10	10	2.33	0.02369	29	18
77	1.894	0.01602	10	10	0.537	0.01516	19	15	0.863	0.21987	35	29
2772	1.583	0.06276	10	6	0.714	0.13019	13	10	1.136	0.28841	28	17
102	1.245	0.05079	11	10	1.018	0.42702	17	15	1.117	0.08232	32	28
78	1.324	0.01985	12	9	0.967	0.33851	18	14	1.007	0.46884	38	24
88	1.282	0.14267	7	7	0.995	0.48504	17	14	1.008	0.47383	30	23
2781					0.492	0.01344	12	12	0.819	0.25555	17	15
101					0.652	0.04317	19	15	0.773	0.09274	29	22
87					4.947	0.02192	18	15	3.657	0.00389	30	23
104	2.292	0.0024	11	8	0.621	0.05152	19	15	0.913	0.34506	30	23
91	1.969	0.07789	11	8	4.047	0.00812	19	13	3.535	0.00033	37	23
85	0.95	0.43363	12	8	0.699	0.0787	13	13	0.633	0.01486	33	24
2783	0.945	0.46023	10	5	0.852	0.28701	17	10	0.986	0.48609	29	17
2784	2.12	0.00022	12	10	0.498	0.01324	18	13	0.935	0.37356	37	25
3018	1.523	0.1487	12	10	0.84	0.27108	18	13	1.101	0.33276	36	26
3019	1.268	0.21268	6	7	0.981	0.45897	16	10	1.012	0.46612	29	19
2790	0.881	0.17766	11	8	1.22	0.04253	18	10	0.966	0.33826	40	23
3020	1.271	0.10162	12	10	0.853	0.10567	19	13	0.965	0.36499	36	25
2794	1.936	0.00176	13	9	0.717	0.09799	19	14	0.877	0.22295	40	25

Table 7: Significance analysis for microarrays for identification of markers of acute rejection. In each case the highest grade from the 3 pathologists was taken for analysis. No rejection and rejection classes are defined. Samples are either used regardless of redundancy with respect to patients or a requirement is made that only one sample is used per patient or per patient per class. The number of samples used in the analysis is given and the lowest FDR achieved is noted.

No Rejection	Rejection	# Samples	Low FDR
All Samples			
Grade 0	Grade 3A-4	148	1
Grade 0	Grade 1B, 3A-4	158	1.5
Non-redundant within class			
Grade 0	Grade 3A-4	86	7
Grade 0	Grade 1 B, 3A-4	93	16

EP 2 194 145 B1

(continued)

Non-redundant (1 sample/patient)			
Grade 0	Grade 3A-4	73	11

5

10

15

20

25

30

35

40

45

50

55

Table 9

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster(Build 156)	Localization	Function
	IL15	Interleukin 15	NM_00585	NP_000576	Hs.168132	Secreted	T-cell activation and proliferation
79	PRF1	Perforin 1 (pore forming protein)	NM_003041	NP_005032	Hs.2200	Secreted	CD8, CTL effector, channel-forming protein capable of lysing non-specifically a variety of target cells; clearance of virally infected host cells and tumor cells;.
	IL17	Interleukin 17 (cytotoxic T-lymphocyte-associated serine esterase 8)	NM_002190	NP_002181	Hs.41724	Secreted	Induces stromal cells to produce proinflammatory and hematopoietic cytokines; enhances IL6, IL8 and ICAM-1 expression in fibroblasts; osteoclastic bone resorption in RA; expressed in only in activated CD4+T cells
	IL8	Interleukin 8	NM_000584	NP_000575	Hs.624	Secreted	Proinflammatory cytokine
	CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	NM_001511	NP_001502	Hs.789	Secreted	Neurogenesis, immune system development, signaling
	IFNG	Interferon, gamma	NM_000619	NP_000610	Hs.856	Secreted	Antiviral defense and immune activation
	IL2	Interleukin 2	NM_000586	NP_000577	Hs.89679	Secreted	Promotes growth of B and T cells
	B2M	beta 2 microglobulin	NM_004048	NP_004039	Hs.75415	Secreted	

(continued)

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster(Build 156)	Localization	Function
	CCL5	Chemokine (C-C motif) ligand (RANTES, SCYA5)	NM_002985	NP_002976	Hs.241392	Secreted	Chemoattractant for monocytes, memory T helper cells and eosinophils; causes release of histamine from basophils and activates eosinophils; One of the major HIV-suppressive factors produced by CD8+ cells
	IL10	Interleukin 10	NM_000572	NP_000563	Hs.193717	Secreted	Chemotactic factor for CD8+T cells; down-regulates expression of Th1 cytokines, MHC class II Ags, and costimulatory molecules on macrophages; enhances B cell survival, proliferation, and antibody production; blocks NF kappa B, JAK- STAT regulation;
80	IL4	Interleukin 4	NM_000589	NP_000580	Hs.73917	Secreted	TH2, cytokine, stimulates CTL
	IL7	Interleukin 7	NM_000880	NP_000871	Hs.72927	Secreted	Proliferation of lymphoid progenitors

(continued)

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster(Build 156)	Localization	Function
	CXCL10	Chemokine (C-X-C motif) ligand 10, SCYB10	NM_001565	NP_001556	Hs.2248	Secreted	Stimulation of monocytes; NK and T cell migration, modulation of adhesion molecule expression
	CCL17	Chemokine (C-C motif) ligand	NM_002987 17	NP_002978	Hs.66742	Secreted	T cell development, trafficking and activation
101	KLRF1	Killer cell lectin-like receptor subfamily F, member 1	NM_016523	NP_057607	Hs.183125	Secreted	Induction of IgE, IgG4, CD23, CD72, surface IgM, and class II MHC antigen in B cells
	IL6	Interleukin 6	NM_000600	NP_000591	Hs.93913	Secreted	B cell maturation
	CCL4	Chemokine (C-C motif) ligand	NM_002984 4	NP_002975	Hs.75703	Secreted	Inflammatory and chemokinetic properties; one of the major HIV-suppressive factors produced by CD8+ T cells
	GZMB	Granzyme B (granzyme 2, cytotoxic T-lymphocyte- associated serine esterase 1)	NM_004131	NP_004122	Hs.1051	Secreted	Apoptosis; CD8, CTL effector
	OID_4789	KIAA0963 protein	NM_014963	NP_055778	Hs.7724	Secreted	Proinflammatory; chemoattraction and activation of neutrophils
	XCL1	Chemokine (C motif) ligand (SCYC2)	NM_002995 1	NP_002986	Hs.3195	Secreted	Chemotactic factor for lymphocytes but not monocytes or neutrophils

(continued)

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster(Build 156)	Localization	Function
	PRDM1	PR domain containing 1, with ZNF domain	NM_001198	NP_001189	Hs.388346	Nuclear	Transcription factor; promotes B cell maturation, represses human beta-IFN gene expression
2781	TBX21	T-box 21	NM_013351	NP_037483	Hs.272409	Nuclear	TH1 differentiation, transcription factor
88	MTHFD2	Methylene tetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydr ofolate cyclohydrolase	NM_006636	NP_006627	Hs.154672	Mitochondrial	Folate metabolism
	IL2RA	Interleukin 2 receptor, alpha	NM_000417	NP_000408	Hs.1724	Membrane-bound and soluble forms	T cell mediated immune response
77	TNFSF6	Tumor necrosis factor (ligand) superfamily, member 6	NM_000639	NP_000630	Hs.2007	Membrane-bound and soluble forms	CD8, CTL effector; proapoptotic
	CD8B1	CD8 antigen, beta polypeptide (p37)	NM_004931.1	NP_004922	Hs.2299	Membrane-bound and soluble forms	CTL mediated killing
	PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	NM_000963	NP_000954	Hs.196384	Membrane-associated	Angiogenesis, cell migration, synthesis of inflammatory prostaglandins
	TAP1	Transporter 1, ATP-binding cassette, sub-family B (MDR1/TAP)	NM_000593 NP_000584		Hs.352018	ER membrane	Transports antigens into ER for association with MHC class I molecules
	IGHM	Immunoglobulin heavy constant mu	BC032249		Hs.300697	Cytoplasmic and secreted forms	Antibody subunit

(continued)

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster(Build 156)	Localization	Function
	GPI	Glucose phosphate isomerase	NM_000175	NP_000166	Hs.409162	Cytoplasmic and secreted forms	Glycolysis and gluconeogenesis (cytoplasmic); neurotrophic factor (secreted)
2783	GSN	Gelsolin (amyloidosis, Finnish type)	NM_000177	NP_000168	Hs.290070	Cytoplasmic and secreted forms	Controls actin filament assembly/disassembly
	STK39	Serine threonine kinase (STE20/SPS1 homolog, yeast)	NM_013233 39	NP_037365	Hs.199263	Cytoplasmic and nuclear	Mediator of stress activated signals; Serine/Thr Kinase, activated p38
	PSMB8	Proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional protease 7)	AK092738		Hs.180062	Cytoplasmic	Processing of MHC class 1 antigens
	LPXN	Leupaxin	NM_004811	NP_004802	Hs.49587	Cytoplasmic	Signal transduction
	ITK	IL2-inducible T-cell kinase	L10717		Hs.211576	Cytoplasmic	Intracellular kinase, T-cell proliferation and differentiation
90	KPNA6	Karyopherin alpha 6 (importin alpha 7)	AW021037		Hs.301553	Cytoplasmic	Nucleocytoplasmic transport
2794	SH2D2A	SH2 domain protein	NM_003975 2A	NP_003966 Hs.103527		Cytoplasmic	CD8 T activation, signal transduction
	TNFSF5	Tumor necrosis factor superfamily, member 5 (hyper-IgM syndrome)	NM_000074 (ligand)	NP_000065	Hs.652	Cellular membrane	B-cell proliferation, IgE production, immunoglobulin class switching; expressed on CD4+ and CD8+ T cells

(continued)

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster (Build 156)	Localization	Function
	CD69	CD69 antigen (p60, early activation antigen)	NM_001781 T-cell	NP_001772	Hs.82401	Cellular membrane	Activation of lymphocytes, monocytes, and platelets
	IL2RG	Interleukin 2 receptor, gamma (severe combined immunodeficiency)	NM_000206	NP_000197	Hs.84	Cellular membrane	Signalling component of many interleukin receptors (IL2, IL4, IL7, IL9, and IL15),
	CXCR4	Chemokine (C-X-C motif) receptor 4	NM_003467	NP_003458	Hs.89414	Cellular membrane	B-cell lymphopoiesis, leukocyte migration, angiogenesis; mediates intracellular calcium flux
2766	CD19	CD19 antigen	NM_001770	NP_001761	Hs.96023 membrane	Cellular	Signal transduction; B lymphocyte development, activation, and differentiation
		Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	NM_002211	NP_002202	Hs.287797	Cellular membrane	Cell-cell and cell-matrix interactions
	TRB	T cell receptor beta, constant region	K02885		Hs.300697	Cellular membrane	Antigen recognition
	CTLA4	Cytotoxic T-lymphocyte-associated protein 4	NM_005214	NP_005205	Hs.247824	Cellular membrane	Negative regulation of T cell activation, expressed by activated T cells
	CD8A	CD8 antigen, alpha polypeptide (p32)	NM_001768	NP_001759	Hs.85258	Cellular membrane	CD8 T-cell specific marker and class 1 MHC receptor
	HLA-DRB1	Major histocompatibility complex, class 11, DR beta 1	NM_002124	NP_002115	Hs.308026	Cellular membrane	Antigen presentation

(continued)

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster(Build 156)	Localization	Function
2772	CD3Z	CD3Z antigen, zeta polypeptide (TiT3 complex)	NM_000734	NP_000725	Hs.97087	Cellular membrane	T-cell marker; couples antigen recognition to several intracellular signal-transduction pathways
	ACTB	Actin, beta	NM_00101	NP_001092	Hs.288061	Cellular membrane	Cell adhesion and recognition
	ITGAL	Integrin, alpha L (antigen CD1 (p180), lymphocyte function-associated antigen 1; alpha Dolyptide)	1A NM_002209	NP_002200	Hs.174103	Cellular membrane	All leukocytes; cell-cell adhesion, signaling
78	TCIRG1	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 protein a isoform 3	NM_006019	NP_006010	Hs.46465	Cellular membrane	T cell activation
	CD72	CD72 antigen	NM_001782	NP_001773	Hs.1116481	Cellular membrane	B cell proliferation
	D12S2489E	DNA segment on chromosome (unique) 2489 expressed sequence	NM_007360	NP_031386	Hs.74085	Cellular membrane	NK cells marker
	MS4A1	Membrane-spanning 4-domains, subfamily A, member 1, CD20	NM_152866	NP_690605	Hs.89751	Cellular membrane	B-cell activation, plasma cell development
	TCRGC2	T cell receptor gamma constant 2	M17323		Hs.112259	Cellular membrane	
	CD4	CD4 antigen (p55)	NM_000616	NP_000607	Hs.17483	Cellular membrane	T cell activation, signal transduction, T-B cell adhesion

(continued)

Array Probe SEQ ID	Gene	Gene Name	mRNA Accession #	RefSeq Peptide Accession #	Current UniGene Cluster(Build 156)	Localization	Function
	CXCR3	Chemokine (C-X-C motif) receptor 3, GPR9	NM_001504	NP_001495	Hs.198252	Cellular membrane	Integrin activation, cytoskeletal changes and chemotactic migration of leukocytes
	CD33	CD33 antigen (gp67)	MM_001772	NP_001763	Hs.83731	Cellular membrane	Cell adhesion; receptor that inhibits the proliferation of normal and leukemic myeloid cells
	CD47	CD47 antigen (Rh-related antigen, integrin-associated signal transducer)	NM_001777	NP_001768	Hs.82685	Cellular membrane	Cell adhesion, membrane transport, signaling transduction, permeability
	BY55	Natural killer cell immunoglobulin superfamily member	NM_007053	NP_008984	Hs.81743	Cellular membrane	NK cells and CTLs, costim with MHC I
2784	KLRD1	Killer cell lectin-like receptor subfamily D, member 1	NM_002262	NP_002253	Hs.41682	Cellular membrane	NK cell regulation
	HLA-F	Major histocompatibility complex, class 1, F	NM_018950	NP_061823	Hs.377850	Cellular membrane	Antigen presentation
2752	PTPRCAP	Protein tyrosine phosphatase, receptor type, C-associated protein	NM_005608	NP_005599	Hs.155975	Cellular membrane	T cell activation

SEQUENCE LISTING

[0531]

5 <110> EXPRESSION DIAGNOSTICS, INC.
 wohlgemuth, Jay
 Fry, Kirk
 woodward, Robert
 Ly, Ngoc
 10 Prentice, James
 Morris, MacDonald
 Rosenberg, Steven

 <120> METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING TRANSPLANT REJECTION

15 <130> 506612000150

 <150> US 10/131,827
 <151> 2002-04-24

20 <150> US 10/325,899
 <151> 2002-12-20

 <160>

25 <170> PatentIn version 3.2

 <210> 36
 <211> 50
 <212> DNA
 <213> Homo sapiens

30 <400> 36
 aagtccaact actaaactgg gggatattat gaagggcctt gagcatctgg 50

35 <210> 49
 <211> 50
 <212> DNA
 <213> Homo sapiens

40 <400> 49
 gcaaccttgc atccatctgg gctacccac ccaagtatac aataaagtct 50

45 <210> 52
 <211> 50
 <212> DNA
 <213> Homo sapiens

50 <400> 52
 caccagcatc caggcggcca gcaggcacct gagggtctgg gacaagggat 50

 <210> 75

55 <210> 77
 <211> 50
 <212> DNA
 <213> Homo sapiens

EP 2 194 145 B1

	<400> 77	
	ccatcggtga aactaacaga taagcaagag agatgtttg gggactcatt	50
5	<210> 78	
	<211> 50	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 78	
	tgccagacct ccttcctgac ctctgaggca ggagaggaat aaagacggtc	50
	<210> 79	
	<211> 50	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 79	
	cctgtgatca ggctccaag tctggtccc atgaggtgag atgcaacctg	50
20	<210> 80	
	<211> 50	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 80	
	accagagtac gttggaaaac ttctggaaa ggctaaagac gatcatgaga	50
	<211> 50	
30	<210> 85	
	<211> 50	
	<212> DNA	
	<213> Homo sapiens	
35	<400> 85	
	gcaaaaagcc caagagcctg aatttagacc aatctatcat ctctctctc	50
	<210> 86	
40	<211> 50	
	<212> DNA	
	<213> Homo sapiens	
	<400> 86	
45	aagtccaact actaaactgg gggatattat gaaggcctt gagcatctgg	50
	<210> 87	
	<211> 50	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 87	
	ttctctttg gccacaagaa taagcagcaa ataaacaact atggctgttg	50
	<210> 88	
55	<211> 50	
	<212> DNA	
	<213> Homo sapiens	

EP 2 194 145 B1

<400> 88
 tgggcagctt gggtaagtac gcaacttact ttccaccaa agaactgtca 50

 <210> 90
 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 90
 10 acataggcga agaaaacatg gcattgagtg tgctgagtcc agacaaatgt 50

 <210> 91
 <211> 50
 <212> DNA
 15 <213> Homo sapiens

 <400> 91
 ccggcagctg tgttagccc ctccagatgg aagtttact tgaatgtaa 50

 <210> 94
 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 94
 25 ttgtactat tgctagacc tctctgtaa tgggtaatgc gtttgattgt 50

 <210> 99
 <211> 50
 <212> DNA
 30 <213> Homo sapiens

 <400> 99
 gcagtttgaa tatcctttgt ttcagagcca gatcatttct tggaaagtgt 50

 <210> 101
 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 101
 40 ttccaggctt ttgtactct tctcagct acaataaaca tcctgaatgt 50

 <210> 102
 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 102
 50 aaccggatat atacatagca tgacatttct ttgtgcttg gcttactgt 50

 <210> 104
 <211> 50
 <212> DNA
 55 <213> Homo sapiens

 <400> 104
 gtccactgtc actgtttctc tgctgtgca aatacatgga taacacattt 50

EP 2 194 145 B1

<210> 107
 <211> 50
 <212> DNA
 <213> Homo sapiens
 5
 <400> 107
 tgctgctaca gttgcaaac actggagcta gagaaaataa agtactgatc 50
 10
 <210> 189
 <211> 50
 <212> DNA
 <213> Homo sapiens
 15
 <400> 189
 gaacatcagg agaggagtcc agagcccacg tctactgagg aaaagtcagg 50
 20
 <210> 190
 <211> 50
 <212> DNA
 <213> Homo sapiens
 25
 <400> 190
 tcccacttca agttaagcac caaagcaatc actaattctg gagcacagga 50
 30
 <210> 191
 <211> 50
 <212> DNA
 <213> Homo sapiens
 35
 <400> 191
 ttcgtgggca ccaagtttcg caagaactac actgtctgct ggccgagttt 50
 40
 <210> 192
 <211> 50
 <212> DNA
 <213> Homo sapiens
 45
 <400> 192
 agcctggaat tctaagcagc agtttcacaa tctgtaattg cacgtttctg 50
 50
 <210> 193
 <211> 50
 <212> DNA
 <213> Homo sapiens
 55
 <400> 193
 aacagaaaca gctatggcaa cagcatcacc ctgagagcat caccaacttg 50
 60
 <210> 194
 <211> 50
 <212> DNA
 <213> Homo sapiens
 65
 <400> 194
 agcacaagcc acgcttcacc accaagaggc ccaacacctt cttctagggtg 50
 70
 <210> 195
 <211> 50

EP 2 194 145 B1

<212> DNA
 <213> Homo sapiens

 <400> 195
 5 ggaccattcc ggagcagccc cacatacctc actgtctcgt ctgtctatgt 50

 <210> 196
 <211> 50
 <212> DNA
 10 <213> Homo sapiens

 <400> 196
 cagagaacga aagtcaagtg cagcgagtg ggtggaagct gatagagcaa 50

 15 <210> 197
 <211> 50
 <212> DNA
 <213> Homo sapiens

 20 <400> 197
 agttggacta aatgctcttc cttcagagga ttatccgggg catctactca 50

 <210> 198
 <211> 50
 25 <212> DNA
 <213> Homo sapiens

 <400> 198
 ctggcacatc caggttttag agcaggcagc ctgagatttc aaaaatgagg 50
 30
 <210> 279
 <211> 50
 <212> DNA
 <213> Homo sapiens
 35
 <400> 279
 ataatacacag ttgtgttct gacactcaat aaacagtcac tggaagaggt 50

 <210> 302
 40 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 302
 45 ccacaagggt tagtttgggc cttaaaactg ccaaggagtt tccaaggatt 50

 <210> 318
 <211> 50
 <212> DNA
 50 <213> Homo sapiens

 <400> 318
 cagttcccag atgtgcgtgt tgtggtcccc aagtatcacc ttccaatttc 50

 55 <210> 368
 <211> 626
 <212> DNA
 <213> Homo sapiens

EP 2 194 145 B1

<400> 368

	acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcatc	60
5	tgactcctga ggagaagtct gccgttactg ccctgtggg caaggtgaac gtggatgaag	120
	ttggtggtga ggcctgggc aggctgctgg tggcttacct ttggaccag aggttctttg	180
	agtcctttgg ggatctgtcc actcctgatg ctgttatggg caaccctaag gtgaaggctc	240
	atggcaagaa agtgctcggg gccttttagt atggcctggc tcacctggac aacctcaagg	300
10	gcacctttgc cacactgagt gagctgact gtgacaagct gcacgtggat cctgagaact	360
	tcaggctcct gggcaacgtg ctggtctgtg tgctggcca tcactttggc aaagaattca	420
	ccccaccagt gcaggctgcc tatcagaaag tgggtggctgg tgtggctaata gccctggccc	480
	acaagtatca ctaagctcgc tttcttgctg tccaatttct attaaagggt cctttgttcc	540
15	ctaagtccaa ctactaaact gggggatatt atgaagggcc ttgagcatct ggattctgcc	600
	taataaaaaa catttatttt cattgc	626

<210> 381

20 <211> 2798

<212> DNA

<213> Homo sapiens

<400> 381

25

30

35

40

45

50

55

EP 2 194 145 B1

	cgttgctgtc gctctgcacg cacctatgtg gaaactaaag cccagagaga aagtctgact	60
	tgccccacag ccagtgagtg actgcagcag caccagaatc tggctctgtt cctgtttggc	120
5	tcttctacca ctacggcttg ggatctcggg catggtggct ttgccaatgg tccttgtttt	180
	gctgctggtc ctgagcagag gtgagagtga attggacgcc aagatcccat ccacagggga	240
	tgccacagaa tggcggaatc ctcacctgtc catgctgggg tcctgccagc cagccccctc	300
	ctgccagaag tgcatcctct cacaccccag ctgtgcatgg tgcaagcaac tgaacttcac	360
10	cgcgctggga gaggcggagg cgcggcgctg cggccgacga gaggagctgc tggctcgagg	420
	ctgcccgtg gaggagctgg aggagccccg cggccagcag gaggtgctgc aggaccagcc	480
	gctcagccag ggcgcccgcg gagaggggtg caccagctg gcgcccagc gggtcgggt	540
	cacgctgcgg cctggggagc cccagcagct ccaggtccgc ttccttcgtg ctgagggata	600
15	cccgggtggac ctgtactacc ttatggacct gagctactcc atgaaggacg acctggaacg	660
	cgtgcgccag ctcgggcacg ctctgctggt cggctgacg gaagtcaccc attctgtgcg	720
	cattggtttt ggttcctttg tggacaaaac ggtgctgccc tttgtgagca cagtaccctc	780
	caaactgctc caccctgcc ccacccggct ggagcgctgc cagtcaccat tcagctttca	840
20	ccatgtgctg tccctgacgg gggacgcaca agccttcgag cgggaggtgg ggcgccagag	900
	tgtgtccggc aatctggact cgcctgaagg tggcttcgat gccattctgc aggtgcact	960
	ctgccaggag cagattggct ggagaaatgt gtcccggctg ctggtgttca cttcagacga	1020
25	cacattccat acagctgggg acgggaagt tggcgccatt ttcattgcca gtgatgggca	1080
	ctgccacttg gacagcaatg gcctctacag tcgcagcaca gaggttgact acccttctgt	1140
	gggtcaggta gccaggccc tctctgcagc aaatatccag cccatctttg ctgtcaccag	1200
	tgccgcactg cctgtctacc aggagctgag taaactgatt cctaagtctg cagttgggga	1260
30	gctgagtgag gactccagca acgtggtaca gctcatcatg gatgcttata atagcctgtc	1320
	ttccaccgtg acccttgaac actcttact cctcctggg gtccacattt cttacgaatc	1380
	ccagtgtgag ggtcctgaga agagggaggg taaggctgag gatcgaggac agtgcaacca	1440
	cgtccgaatc aaccagacgg tgactttctg ggtttctctc caagccacc actgcctccc	1500
35	agagcccat ctcctgaggc tccgggccct tggcttctca gaggagctga ttgtggagt	1560
	gcacacgctg tgtgactgta attgcagtga caccagccc caggctcccc actgcagtga	1620
	tggccaggga cacctacaat gtggtgatg cagctgtgcc cctggccgcc taggtcggct	1680
	ctgtgagtg tctgtggcag agctgtcctc cccagacctg gaatctgggt gccgggctcc	1740
40	caatggcaca gggcccctgt gcagtggaaa gggtcactgt caatgtggac gctgcagctg	1800
	cagtggacag agctctgggc atctgtgcga gtgtgacgat gccagctgtg agcgacatga	1860
	gggcatcctc tgcggaggct ttggtcgtg ccaatgtgga gtatgtcact gtcattgcaa	1920
45	ccgcacgggc agagcatgag aatgcagtgg ggacatggac agttgcatca gtcccaggga	1980
	agggctctgc agtgggcatg gacgctgcaa atgcaaccgc tgccagtgct tggacggcta	2040
	ctatggtgct ctatgcgacc aatgcccagg ctgcaagaca ccatgcgaga gacaccggga	2100
	ctgtgcagag tgtggggcct tcaggactgg cccactggcc accaactgca gtacagctg	2160
50	tgcccatacc aatgtgacct tggccttggc ccctatcttg gatgatggct ggtgcaaaga	2220
	gcggaccctg gacaaccagc tgttcttctt cttggtggag gatgacgcca gaggcacggt	2280
	cgtgctcaga gtgagacccc aagaaaaggg agcagaccac acgcaggcca ttgtgctggg	2340
55	ctgcgtaggg ggcacgtgg cagtggggct ggggctggct ctggcttacc ggctctcgg	2400

EP 2 194 145 B1

	ggaaatctat gaccgccggg aatacagtcg ctttgagaag gagcagcaac aactcaactg	2460
	gaagcaggac agtaatcctc tctacaaaag tgccatcacg accaccatca atcctcgctt	2520
5	tcaagaggca gacagtccca ctctctgaag gaggaggga cacttaccca aggctcttct	2580
	ccttgaggga cagtgggaac tggagggtga gaggaagggt gggctctgtaa gaccttggtg	2640
	ggggactaat tctctggcga ggtgcggcca ccacctact tcattttcag agtgacaccc	2700
	aagagggtg cttcccatgc ctgcaacctt gcatccatct gggctacccc acccaagtat	2760
10	acaataaagt cttacctcag aaaaaaaaaa aaaaaaaaaa	2798

<210> 384

<211> 2106

<212> DNA

15 <213> Homo sapiens

<400> 384

20	agtttccctt ccgctcacct ccgcctgagc agtggagaag gcggcactct ggtggggctg	60
	ctccaggcat gcagatccca caggcgccct ggccagtcgt ctgggcgggtg ctacaactgg	120
	gctggcgggc aggatggttc ttagactccc cagacaggcc ctggaacccc cccaccttct	180
25	tcccagccct gctcgtggtg accgaagggg acaacgccac cttcacctgc agcttctcca	240
	acacatcgga gagcttcgtg ctaaactggt accgcatgag ccccagcaac cagacggaca	300
	agctggccgc cttccccgag gaccgcagcc agcccggcca ggaactgccgc ttccgtgtca	360
30	cacaactgcc caacgggctg gacttccaca tgagcgtggt cagggcccg gcaatgaca	420
	gcggcaccta cctctgtggg gccatctccc tggcccccaa ggcgcagatc aaagagagcc	480
35	tgcgggcaga gctcaggggtg acagagagaa gggcagaagt gcccacagcc cccccagcc	540
	cctcaccag gccagccggc cagttccaaa ccctggtggt tgggtgtcgtg ggcggcctgc	600

40

45

50

55

EP 2 194 145 B1

	tgggcagcct ggtgctgcta gtctgggtcc tggccgtcat ctgctcccgg gccgcacgag	660
	ggacaatagg agccaggcgc accggccagc ccctgaagga ggacccctca gccgtgcctg	720
5	tgttctctgt ggactatggg gagctggatt tccagtggcg agagaagacc ccggagcccc	780
	ccgtgccctg tgccctgag cagacggagt atgccaccat tgtctttcct agcggaatgg	840
	gcacctcatc ccccgcccgc aggggctcag ccgacggccc tcggagtgcc cagccactga	900
10	ggcctgagga tggacactgc tcttggtccc tctgaccggc ttccttggtc accagtgttc	960
	tgcagaccct ccaccatgag cccgggtcag cgcatttcct caggagaagc aggcagggtg	1020
	caggccattg caggccgtcc aggggctgag ctgcctgggg gcgaccgggg ctccagcctg	1080
15	cacctgcacc aggcacagcc ccaccacagg actcatgtct caatgcccac agtgagccca	1140
	ggcagcaggt gtcaccgtcc cctacaggga gggccagatg cagtactgc ttcaggctct	1200
20	gccagcacag agctgcctgc gtccagctcc ctgaatctct gctgctgctg ctgctgctgc	1260
	tgctgctgcc tggggcccgg ggctgaaggc gccgtggccc tgcctgacgc cccggagcct	1320
	cctgcctgaa cttgggggct ggttgagat ggccttgag cagccaaggc gccctggca	1380
25	gtggcatccc gaaacgccct ggacgcaggc cccaagactg ggcacaggag tgggaggtac	1440
	atggggctgg ggactcccca ggagttatct gtcacctgca ggcctagaga agtttcaggg	1500
30	aaggtcagaa gagctcctgg ctgtggtggg cagggcagga aaccctccc acctttacac	1560
	atgccaggc agcacctcag gccctttgtg gggcaggga gctgaggcag taagcgggca	1620
	ggcagagctg gaggcctttc aggcagcca gcactctggc ctctgcccgc cgcattccac	1680
35	cccagcccct cacaccactc gggagaggga catcctacgg tcccaaggc aggagggcag	1740
	ggctgggggt gactcaggcc cctcccagct gtggccacct ggggtgtggg agggcagaag	1800
	tgcaggcacc tagggccccc catgtgcccc ccctgggagc tctccttgga acccattcct	1860
40	gaaattattt aaaggggttg gccgggtccc caccagggcc tgggtgggaa ggtacaggcg	1920
	ttccccggg gcctagtacc cccgcgtggc ctatccactc ctcacatcca cacactgcac	1980
45	ccccactcct ggggcagggc caccagcatc caggcgcca gcaggcacct gagtggctgg	2040
	gacaagggat ccccttccc tgtggttcta ttatattata attataatta aatatgagag	2100
	catgct	2106
50	<210> 409	
	<211> 1909	
	<212> DNA	
	<213> Homo sapiens	
55	<400> 409	

EP 2 194 145 B1

	gaggtgtttc ccttagctat ggaaactcta taagagagat ccagcttgcc tcctcttgag	60
	cagtcagcaa caggggtccc tccttgacac ctcagcctct acaggactga gaagaagtaa	120
5	aaccgtttgc tggggctggc ctgactcacc agctgccatg cagcagccct tcaattaccc	180
	atatccccag atctactggg tggacagcag tgccagctct ccctgggccc ctccaggcac	240
	agttcttccc tgtccaacct ctgtgccag aaggcctggt caaaggaggc caccaccacc	300
	accgccaccg ccaccactac cacctccgcc gccgccgcca cactgcctc cactaccgct	360
10	gccaccctg aagaagagag ggaaccacag cacaggcctg tgtctccttg tgatgttttt	420
	catggttctg gttgccttgg taggattggg cctggggatg tttcagctct tccacctaca	480
	gaaggagctg gcagaactcc gagagtctac cagccagatg cacacagcat catctttgga	540
	gaagcaaata ggccacccca gtccaccccc tgaaaaaaag gagctgagga aagtggccca	600
15	tttaacaggc aagtccaact caagggtccat gcctctggaa tgggaagaca cctatggaat	660
	tgtcctgctt tctggagtga agtataagaa ggggtggcctt gtgatcaatg aaactgggct	720
	gtactttgta tattccaaag tatacttccg gggtaaatct tgcaacaacc tgcccctgag	780
	ccacaaggtc tacatgagga actctaagta tccccaggat ctggtgatga tggaggggaa	840
20	gatgatgagc tactgcacta ctgggcagat gtgggcccgc agcagctacc tgggggcagt	900
	gttcaatctt accagtgcg atcatattata tgtcaacgta tctgagctct ctctgggtcaa	960
	ttttgaggaa tctcagacgt ttttcggctt atataagctc taagagaagc actttgggat	1020
25	tctttccatt atgattcttt gttacaggca ccgagaatgt tgtattcagt gagggtcttc	1080
	ttacatgcat ttgaggtcaa gtaagaagac atgaaccaag tggaccttga gaccacaggg	1140
	ttcaaatgt ctgtagctcc tcaactcacc taatgtttat gagccagaca aatggaggaa	1200
	tatgacggaa gaacatagaa ctctgggctg ccatgtgaag agggagaagc atgaaaaagc	1260
30	agctaccagg tgttctacac tcatcttagt gcctgagagt atttaggcag attgaaaagg	1320
	acacctttta actcacctct caagggtgggc cttgctacct caagggggac tgtctttcag	1380
	atacatggtt gtgacctgag gatttaaggg atggaaaagg aagactagag gcttgcataa	1440
	taagctaaag aggctgaaag aggccaatgc cccactggca gcatcttcac ttctaaatgc	1500
35	atatcctgag ccacgggtga aactaacaga taagcaagag agatgttttg gggactcatt	1560
	tcattcctaa cacagcatgt gtatttccag tgcaattgta ggggtgtgtg tgtgtgtgtg	1620
	tgtgtgtgtg tgtgtatgac taaagagaga atgtagatat tgtgaagtac atattaggaa	1680
	aatatgggtt gcatttggtc aagattttga atgcttcctg acaatcaact ctaatagtgc	1740
40	ttaaaaatca ttgattgtca gctactaatg atgttttcct ataataataa aaatatttat	1800
	gtagatgtgc atttttgtga aatgaaaaca tgtaataaaa agtatatgtt aggatacaaa	1860
	aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	1909

45 <210> 410
 <211> 2700
 <212> DNA
 <213> Homo sapiens

50 <400> 410

	gcggcgcccta gtccggggct ggcgggagtg cagttctgag tcccggccgg cgtgcgcgga	60
	gcggggcagc cagcagcgga ggcgcggcgc gcagcacacc cggggaccat gggctccatg	120
55	ttccggagcg aggagggtggc cctgggtccag ctctttctgc ccacagcggc tgcctacacc	180

EP 2 194 145 B1

	tgctgagtc	ggctgggcga	gctgggcctc	gtggagttca	gagacctcaa	cgcctcgggtg	240
	agcgccttcc	agagacgctt	tgtggttgat	gttcggcgct	gtgaggagct	ggagaagacc	300
5	ttcaccttcc	tgcaggagga	ggtgcggcgg	gctgggctgg	tcctgcccc	gccaaagggg	360
	aggctgccgg	cacccccacc	ccgggacctg	ctgcgcatcc	aggaggagac	ggagcgcctg	420
	gcccaggagc	tgcgggatgt	gcggggcaac	cagcaggccc	tcggggccca	gctgcaccag	480
	ctgcagctcc	acgccgccgt	gctacgccag	ggccatgaac	ctcagctggc	agccgcccac	540
10	acagatgggg	cctcagagag	gacgcccctg	ctccaggccc	ccggggggcc	gcaccaggac	600
	ctgaggggtca	actttgtggc	aggtgccgtg	gagccccaca	aggcccctgc	cctagagcgc	660
	ctgctctgga	gggcctgccg	cggttctctc	attgccagct	tcaggagagct	ggagcagccg	720
	ctggagcacc	ccgtgacggg	cgagccagcc	acgtggatga	ccttctctcat	ctcctactgg	780
15	ggtgagcaga	tcggacagaa	gatccgcaag	atcacggact	gcttccactg	ccagctcttc	840
	ccgtttctgc	agcaggagga	ggcccgctc	ggggccctgc	agcagctgca	acagcagagc	900
	caggagctgc	aggaggtcct	cggggagaca	gagcggttcc	tgagccaggt	gctaggccgg	960
20	gtgctgcagc	tgtgccgcc	agggcagggtg	caggtccaca	agatgaaggc	cgtgtacctg	1020
	gccctgaacc	agtgcagcgt	gagcaccacg	cacaagtgcc	tcattgccga	ggcctggtgc	1080
	tctgtgcgag	acctgcccgc	cctgcaggag	gccctgctgg	acagctcgat	ggaggagggg	1140
	gtgagtgcg	tggctcaccg	catcccctgc	cgggacatgc	ccccacact	catccgcacc	1200
25	aaccgcttca	cggccagctt	ccagggcata	gtggatgcct	acggcgtagg	ccgctaccag	1260
	gagggtcaacc	ccgctcccta	caccatcatc	accttcccct	tcctgtttgc	tgtgatgttc	1320
	ggggatgtgg	gccacgggct	gctcatgttc	ctcttcgccc	tggccatggt	ccttgccggag	1380
	aaccgaccgg	ctgtgaaggc	cgcgcagaa	gagatctggc	agactttctt	caggggccgc	1440
30	tacctgtctc	tgttatggg	cctgttctcc	atctacaccg	gcttcatcta	caacgagtgc	1500
	ttcagtcgcy	ccaccagcat	cttcccctcg	ggctggagtg	tggccgccat	ggccaaccag	1560
	tctggctgga	gtgatgcatt	cctggcccag	cacacgatgc	ttaccctgga	tccaacgctc	1620
	accggtgtct	tcctgggacc	ctaccctttt	ggcatcgatc	ctatttgtag	cctggctgcc	1680
35	aaccacttga	gcttcctcaa	ctccttcaag	atgaagatgt	ccgtcatcct	ggcgctcgtg	1740
	cacatggcct	ttgggggtgg	cctcggagtc	ttcaaccacg	tgacttttgg	ccagaggcac	1800
	cggctgctgc	tggagacgct	gccggagctc	accttctctg	tgggactctt	cggttacctc	1860
40	gtgttcctag	tcatctacaa	gtggctgtgt	gtctgggctg	ccagggccgc	ctcggccccc	1920
	agcatcctca	tccacttcat	caacatgttc	ctcttctccc	acagccccag	caacaggctg	1980
	ctctaccccc	ggcaggaggt	ggtccaggcc	acgctgggtg	tcctggcctt	ggccatgggtg	2040
	cccatcctgc	tgcttggcac	accctgcac	ctgctgcacc	gccaccgccg	ccgcctgcgg	2100
45	aggaggcccg	ctgaccgaca	ggaggaaaac	aaggccgggt	tgctggacct	gcctgacgca	2160
	tctgtgaatg	gctggagctc	cgatgaggaa	aaggcagggg	gcctggatga	tgaagaggag	2220
	gccgagctcg	tccccctcga	ggtgctcatg	caccaggcca	tccacaccat	cgagttctgc	2280
	ctgggctgcy	tctccaacac	cgcctcttac	ctgcgcctgt	gggccctgag	cctggcccac	2340
50	gcccagctgt	ccgaggttct	gtgggccatg	gtgatgcgca	taggcctggg	cctgggccgg	2400
	gagggtggcg	tggcggtgtg	ggtgctggtc	cccattcttg	ccgcctttgc	cgtgatgacc	2460
	gtggctatcc	tgctgggtgat	ggagggactc	tcagccttcc	tgacgcccct	gcggctgcac	2520
	tgggtggaat	tccagaacaa	gttctactca	ggcacgggct	acaagctgag	tcccttcacc	2580
55	ttcgtgccca	cagatgacta	ggggccactg	caggtcctgc	cagacctcct	tcctgacctc	2640
	tgaggcagga	gaggaataaa	gacgggtccgc	cctggcagtg	aaaaaaaaaa	aaaaaaaaaa	2700

EP 2 194 145 B1

<210> 411
 <211> 1668
 <212> DNA
 <213> Homo sapiens

5

<400> 411

	atggcagccc gtctgtcctt cctgggcatc cttctcctgc tgctgccctt gcccgctcct	60
10	gccccgtgcc acacagccgc acgctcagag tgcaagcgca gccacaagtt cgtgcctggt	120
	gcatggctgg ccggggaggg tgtggacgtg accagcctcc gccgctcggg ctccttccca	180
	gtggacacac aaagggttctt gcggcccgac ggcacctgca ccctctgtga aaatgcccta	240
	caggagggca ccctccagcg cctgcctctg gcgctcacca actggcgggc ccagggtctt	300
15	ggctgccagc gccatgtaac caggggccaaa gtcagctcca ctgaagctgt ggccccggat	360
	gcggctcgta gcatccgcaa cgactggaag gtcgggctgg acgtgactcc taagcccacc	420
	agcaatgtgc atgtgtctgt ggccggctca cactcacagg cagccaactt tgagccccag	480
20	aagaccacc aggaccagta cagcttcagc actgacacgg tggagtgccg cttctacagt	540
	ttccatgtgg tacacactcc cccgctgcac cctgacttca agagggccct cggggacctg	600
	ccccaccact tcaacgcctc caccagccc gcctacctca ggcttatctc caactacggc	660
	accacttca tccgggctgt ggagctgggt ggccgcatac cggccctcac tgccctgcgc	720
25	acctgcgagc tggccctgga agggctcacg gacaacgagg tggaggactg cctgactgtc	780
	gaggccccagg tcaacatagg catccacggc agcatctctg ccgaagccaa ggctgtgag	840
	gagaagaaga agaagcaca gatgacggcc tccttcacc aaacctaccg ggagcgccac	900
	tcggaagtgg ttggcgcca tcacacctcc attaacgacc tgctgttcgg gatccaggcc	960
30	gggcccagc agtactcagc ctgggtaaac tccgtgcccg gcagccctgg cctggtggac	1020
	tacacctgg aaccctgca cgtgctgtg gacagccagg acccgggcg ggaggcactg	1080
	aggagggccc tgagtcahta cctgacggac agggctcgtt ggagggactg cagccggccg	1140
35	tgccaccag ggcggcagaa gagccccga gacctatgcc agtgtgtgtg ccatggctca	1200
	gcggtcacca cccaggactg ctgccctcgg cagaggggcc tggcccagct ggaggtgacc	1260
	ttcatccaag catggagcct gtggggggac tggttcactg ccacggatgc ctatgtgaag	1320
	ctcttctttg gtggccagga gctgaggacg agcaccgtgt gggacaataa caacccatc	1380
40	tggtcagtgc ggctggattt tggggatgtg ctcttgcca caggggggcc cctgaggttg	1440
	cagggtctgg atcaggactc tggcaggac gatgacctc ttggcacctg tgatcaggct	1500
	ccaagtctg gttcccatga ggtgagatgc aacctgaatc atggccacct aaaattccgc	1560
	tatcatgcca ggtgcttgcc ccacctggga ggaggcacct gcctggacta tgtccccaa	1620
45	atgcttctgg gggagcctcc aggaaccgg agtggggccg tgtggtga	1668

<210> 412
 <211> 921
 <212> DNA
 <213> Homo sapiens

50

<400> 412

55

EP 2 194 145 B1

	ttctatgcaa agcaaaaagc cagcagcagc cccaagctga taagattaat ctaaagagca	60
	aattatgggtg taatttccta tgctgaaact ttgtagttaa ttttttaaaa aggtttcatt	120
5	ttcctattgg tctgatttca caggaacatt ttacctgttt gtgaggcatt ttttctcctg	180
	gaagagaggt gctgattggc cccaagtgc tgacaatctg gtgtaacgaa aatttccaat	240
	gtaaactcat tttccctcgg tttcagcaat tttaaatcta tatatagaga tatctttgtc	300
	agcattgcat cgtagcttc tcctgataaa ctaattgcct cacattgtca ctgcaaatcg	360
10	acacctatta atgggtctca cctcccaact gcttccccct ctgttcttcc tgctagcatg	420
	tgccggcaac tttgtccacg gacacaagtg cgatatcacc ttacaggaga tcatcaaaac	480
	tttgaacagc ctcacagagc agaagactct gtgcaccgag ttgaccgtaa cagacatctt	540

15	tgctgcctcc aagaacacaa ctgagaagga aaccttctgc agggctgcga ctgtgctccg	600
	gcagttctac agccaccatg agaaggacac tcgctgcctg ggtgcgactg cacagcagtt	660
	ccacaggcac aagcagctga tccgattcct gaaacggctc gacaggaacc tctggggcct	720
	ggcgggcttg aattcctgtc ctgtgaagga agccaaccag agtacgttgg aaaacttctt	780
20	ggaaaggcta aagacgatca tgagagagaa atattcaaag tgttcgagct gaatatttta	840
	atttatgagt ttttgatagc tttatTTTTT aagtatttat atatttataa ctcatacataa	900
	aataaagtat atatagaatc t	921

25 <210> 417
 <211> 292
 <212> DNA
 <213> Homo sapiens

30 <400> 417

	tcttcaacaa ggggtaaatc agtcagtttc taaaactggt gggagggtctc cataaacctg	60
	ataacaagat cccaaactcc aaactgattg actgagttta ttcctgatca tttgggttga	120
35	acttaagagt tatacaagaa aatggtaggg gacgaggagg ttgtataaag gggaaaaaac	180
	aacaactgca aaaagcccaa gagcctgaat ttagaccaat ctatcatctt cctcctctta	240
	aaaagaaaac aattttaaag tttcaaaaaa aaaaaaaaaa aaaaaaaaaa aa	292

40 <210> 418
 <211> 626
 <212> DNA
 <213> Homo sapiens

45 <400> 418

50

55

EP 2 194 145 B1

	acatttgctt ctgacacaac tgtgttctact agcaacctca aacagacacc atggtgcatc	60
	tgactcctga ggagaagtct gccgttactg ccctgtggg caaggtgaac gtggatgaag	120
5	ttggtggtga ggccttggg aggctgctgg tggcttacct ttggaccag aggttctttg	180
	agtcctttgg ggatctgtcc actcctgatg ctgttatggg caaccctaag gtgaaggctc	240
	atggcaagaa agtgctcggg gccttttagtg atggcctggc tcacctggac aacctcaagg	300
	gcacctttgc cacactgagt gagctgcact gtgacaagct gcacgtggat cctgagaact	360
10	tcaggctcct gggcaacgtg ctggtctgtg tgctggccca tcactttggc aaagaattca	420
	ccccaccagt gcaggctgcc taccagaaag tgggtggctgg tgggctaata gccctggccc	480
	acaagtatca ctaagctcgc tttcttgctg tccaatttct attaaagggt cctttgttcc	540
	ctaagtccaa ctactaaact ggggatatt atgaagggcc ttgagcatct ggattctgcc	600
15	taataaaaaa cttttatctt cattgc	626

<210> 419

<211> 1764

<212> DNA

20 <213> Homo sapiens

<400> 419

25	cgtctggttc aggggctaga aaagagcgtc gatgccggcg gcagtgatga gtcctaggag	60
	gcgctggctc tttggcggct cggaggagcg gctgctgctg ctgctgctgc tgctgggtggc	120
	ccctttgcag atgtattgct gtccttgaat attagcccat ttgaaaacgc ctgggaagtt	180
	cagccatcag tatgtccaag taaaactta ttatgttaag acatggagag ggtgcttgga	240
30	ataaggagaa ccgtttttgt agctgggtgg atcagaaact caacagcgaa ggaatggagg	300
	aagctcggaa ctgtgggaag caactcaaag cgttaaaact tgagtttgat cttgtattca	360

35

40

45

50

55

EP 2 194 145 B1

	catctgtcct taatcgggtcc attcacacag cctggctgat cctggaagag ctaggccagg	420
	aatgggtgcc tgtggaaagc tcctggcgtc taaatgagcg tcaactatggg gccttgatcg	480
5	gtctcaacag ggagcagatg gctttgaatc atggtgaaga acaagtgagg ctctggagaa	540
	gaagctacaa tgtaaccccg cctcccattg aggagtctca tccttactac caagaaatct	600
	acaacgaccg gaggtataaa gtatgcatg tgcccttgga tcaactgcca cggtcggaaa	660
	gcttaaagga tgttctggag agactccttc cctattggaa tgaaaggatt gctcccgaa	720
10	tattacgtgg caaaaccatt ctgatatctg ctcatggaaa tagcagtagg gcactcctaa	780
	aacacctgga aggtatctca gatgaagaca tcatcaacat tactcttcct actggagtcc	840
	ccattcttct ggaattggat gaaaacctgc gtgctgttgg gcctcatcag ttcctgggtg	900
	accaagaggc gatccaagca gccattaaga aagtagaaga tcaaggaaaa gtgaacaag	960
15	ctaaaaata gtctttctca actgttggtc aagaagaaat gcaaaagaag tggcatagga	1020
	gtgtgttatg ggtgctgaac tctctctctt tttcccgat tttccagagc taggctgtgg	1080
	agtagagttt gtataggtaa ctaggtaact tattgtggcc cagataaggc tttaggatgc	1140
20	ctcagtgtt atgtcatagc cttatgagtt agctttcttg ctagccccct agtcggtcac	1200
	caactagta actagtgggg cttaatgaag gtcataagtt tctgagatgg gagagcaaca	1260
	agtagagatg aagttaaagg tatttatcat tcaagaaatc attattgagt caccattgac	1320
	aggcactatt ctaatcagta gttcacttta atatttaata agattttctg ggataacagt	1380
25	aagggatatt agataatata ccgtatgtat ttattactag tcttttcctc taggaaaagg	1440
	gatactttga taattaaggc cagaggcca ttagttgaga aagtcacaga tatatttctc	1500
	caagaaagcc aacaaccacc accacaatga cagaaatgac aacaaggccc ttttaactgt	1560
	cttctagttt agagacatcc ttcatttgac atttagtaga attcctcttt ggccacaaga	1620
30	ataagcagca aataaacaac tatggctgtt gaggttctca ttttggttg ttttaatttt	1680
	ttgaactttg ggtacctgta attagtttaa aaataaagtt cctgataata aagtgactga	1740
	aaatggcaaa aaaaaaaaaa aaaa	1764

35 <210> 420
 <211> 2154
 <212> DNA
 <213> Homo sapiens

40 <400> 420

45

50

55

EP 2 194 145 B1

	atataaccgc	gtggcccg	cgcgcgcttc	cctcccggcg	cagtcaccgg	cgcggtctat	60
	ggctgcgact	tctctaattgt	ctgctttggc	tgcccggctg	ctgcagcccg	cgcacagctg	120
5	ctcccttcgc	cttcgcccctt	tccacctcgc	ggcagttcga	aatgaagctg	ttgtcatttc	180
	tggaaggaaa	ctggcccagc	agatcaagca	ggaagtgcgg	caggaggtag	aagagtgggt	240
	ggcctcaggc	aacaaacggc	cacacctgag	tgtgatcctg	gttggcgaga	atcctgcaag	300
	tcactcctat	gtcctcaaca	aaaccagggc	agctgcagtt	gtgggaatca	acagtgcagac	360
10	aattatgaaa	ccagcttcaa	tttcagagga	agaattgttg	aatttaataca	ataaactgaa	420
	taatgatgat	aatgtagatg	gcctccttgt	tcagttgcct	cttcagagc	atattgatga	480
	gagaaggatc	tgcaatgctg	tttctccaga	caaggatgtt	gatggctttc	atgtaattaa	540
	tgtaggacga	atgtgtttgg	atcagtattc	catgttaccg	gctactccat	ggggtgtgtg	600
15	ggaaataatc	aagcgaactg	gcattccaac	cctaggggaa	aatgtgggtg	tggttggaag	660
	gtcaaaaaac	gttggaatgc	ccattgcaat	gttactgcac	acagatgggg	cgcagtgacg	720
	tcccgagggt	gatgccactg	ttacaatatc	tcacgatata	actcccaaag	agcagttgaa	780
	gaaacataca	attcttgtag	atattgtaat	atctgctgca	ggtattccaa	atctgatcac	840
20	agcagatatg	atcaagggaag	gagcagcagt	cattgatgtg	ggaataaata	gagttcacga	900
	tcctgtaact	gccaaacca	agttggttgg	agatgtggat	tttgaaggag	tcagacaaaa	960
	agctgggtat	atcactccag	ttcctggagg	tgttggtccc	atgacagtgg	caatgctaata	1020
25	gaagaatacc	attattgctg	caaaaaaggt	gctgaggctt	gaagagcgag	aagtgtctgaa	1080
	gtctaaagag	cttggggtag	ccactaatta	actactgtgt	cttctgtgtc	acaaacagca	1140
	ctccaggcca	gctcaagaag	caaagcaggc	caatagaaat	gcaatatttt	taattttattc	1200
30	tactgaaatg	gtttaaaatg	atgccttgta	tttattgaaa	gcttaaattg	gtgggtgttt	1260
	ctgcacatac	ctctgcagta	cctcaccagg	gagcattcca	gtatcatgca	gggtcctgtg	1320
	atctagccag	gagcagccat	taacctagt	attaatatgg	gagacattac	catatggagg	1380
	atggatgctt	cactttgtca	agcacctcag	ttacacattc	gccttttcta	ggattgcatt	1440
35	tccaagtgc	tattgcaata	acagttgata	ctcatttttag	gtaccagacc	ttttgagttc	1500
	aactgatcaa	accaaaggaa	aagtgttgct	agagaaaatt	ggggaaaagg	tgaaaaagaa	1560
	aaaaatggtag	taattgagca	gaaaaaaatt	aatttatata	tgtattgatt	ggcaaccaga	1620
40	tttatctaag	tagaactgaa	ttggctagga	aaaaagaaaa	actgcatgtt	aatcattttc	1680
	ctaagctgtc	cttttgaggc	ttagttagtt	tattgggaaa	atgttttagga	ttattccttg	1740
	ctatttagtac	tcattttatg	tatgttacc	ttcagtaagt	tctccccatt	ttagttttct	1800
	aggactgaaa	ggattccttt	ctacattata	catgtgtgtt	gtcatatttg	gcttttgcta	1860
45	tatactttta	cttcattgtt	aaatttttgt	attgtatagt	ttctttggtg	tatcttaaaa	1920
	cctatttttg	aaaaacaac	ttggcttgat	aatcattttg	gcagcttggg	taagtacgca	1980
	acttactttt	ccaccaaaaga	actgtcagca	gctgcctgct	tttctgtgat	gtatgtatcc	2040
	tgttgacttt	tccagaaatt	ttttaagagt	ttgagttact	attgaattta	atcagacttt	2100
50	ctgattaaag	ggttttcttt	cttttttaat	aaaacacatc	tgtctggtat	ggta	2154

<210> 422

<211> 456

<212> DNA

55 <213> Homo sapiens

<400> 422

EP 2 194 145 B1

	gcacgagtgg agttgggtgt cggctttttt agccagcttt tgtgggaatt gcctttgacc	60
	tattaaagaa ggaaagtggg taatggagtc ccagccactc aagagactgg atatcccccg	120
5	agaatggctt gggttaccag ctatggaccc ttggaagatg aatctaatac ttctcactgg	180
	tttttctttg caaatccatt tgctttttatt tttctaataa caataaactc tattttccat	240
	gttctcaggg cccctgggta gacagacaca gcttgatttc agagcagaca taggcgaaga	300
10	aaacatggca ttgagtgtgc tgagtccaga caaatgttat ttatatacac atccaaattt	360
	gaagagaaaa tgtattttctt taggtttcaa aactgtaat agatataaag caaaaataaa	420
15	aacctgttgc aaagttaaaa aaaaaaaaaa aaaaaa	456

	<210> 423
	<211> 691
	<212> DNA
20	<213> Homo sapiens

	<220>
	<221> misc_feature
	<222> (35)..(35)
25	<223> n is a, c, g, t or u

	<220>
	<221> misc_feature
	<222> (140)..(140)
30	<223> n is a, c, g, t or u

	<220>
	<221> misc_feature
	<222> (394)..(394)
35	<223> n is a, c, g, t or u

	<220>
	<221> misc_feature
	<222> (397)..(397)
40	<223> n is a, c, g, t or u

	<220>
	<221> misc_feature
	<222> (401)..(401)
45	<223> n is a, c, g, t or u

	<220>
	<221> misc_feature
	<222> (404)..(404)
50	<223> n is a, c, g, t or u

	<220>
	<221> misc_feature
	<222> (412)..(412)
55	<223> n is a, c, g, t or u

	<220>
	<221> misc_feature

<222> (536)..(536)
 <223> n is a, c, g, t or u

 <220>
 5 <221> misc_feature
 <222> (569)..(569)
 <223> n is a, c, g, t or u

 <220>
 10 <221> misc_feature
 <222> (581)..(581)
 <223> n is a, c, g, t or u

 <220>
 15 <221> misc_feature
 <222> (615)..(615)
 <223> n is a, c, g, t or u

 <220>
 20 <221> misc_feature
 <222> (619)..(619)
 <223> n is a, c, g, t or u

 <220>
 25 <221> misc_feature
 <222> (640)..(640)
 <223> n is a, c, g, t or u

 <220>
 30 <221> misc_feature
 <222> (651)..(651)
 <223> n is a, c, g, t or u

 <220>
 35 <221> misc_feature
 <222> (662)..(662)
 <223> n is a, c, g, t or u

 <220>
 40 <221> misc_feature
 <222> (677)..(677)
 <223> n is a, c, g, t or u

 <220>
 45 <221> misc_feature
 <222> (680)..(680)
 <223> n is a, c, g, t or u

 <220>
 50 <221> misc_feature
 <222> (687)..(687)
 <223> n is a, c, g, t or u

 <400> 423
 55

EP 2 194 145 B1

	tttttttttt tttttttttt tttttttttt ttttncaaaa tataaacttt attattttac	60
	attcaagtga aacttccatc tggaggggct aaacacagct gccggccaca ttcactgatt	120
5	tattactttg ttgccttttn cgttcacctg atggaagaat tcaaccctct taaaaacata	180
	acaacaacaa aaacagctgg agagtcccag ccgtaatact aggtgtagac acgcacaagc	240
	acacacacaa attcaaaaac ttctacatag aaaaataaag gataaacatt atccatctat	300
	tttgtactgt gtaatgcaac ttttatatac ataaattttt tttttttttt tttttttttt	360
10	ttttttttta ctgtttttcag tcactgcaaa ttttctnccc nccnctggga tntaaggatc	420
	cagggaggag gctgccacag tgaacaaaaa aagctacatt ctgcccaggg aggaaaaaaa	480
	aaagcaattt ctcgctcccc ttcccaagtc cttcctgtcc accaccacct cggatnttcc	540
	cgcacacagc cttccggtga gcgggctgnc cgtcccctcc nctctctaag gcattggggga	600
15	acaaaaggcc catangcanc ccctgccaaa aaaaaaatn atctaccttt naagaaaagg	660
	cnaggggctg ggatccngcn aaaaatnact t	691

<210> 426

<211> 3478

20 <212> DNA

<213> Homo sapiens

<400> 426

25

30

35

40

45

50

55

EP 2 194 145 B1

	attttccggg	ccgggcgac	taaggtg	ggccccggg	cccagtatat	gaccgcggt	60
	cctgctatcc	ttcgttccc	ccgccccatg	tggctgcggg	gccgcggcgg	cgctgccac	120
5	tatggccccg	aaagtagtta	gcaggaagcg	gaaagcgccc	gcctcgccgg	gagctgggag	180
	cgacgctcag	ggcccgag	tggctgggat	cactcgcttc	acaaaaggaa	aagacttcct	240
	cctgtgaaga	gatccttagt	atactacttg	aagaaccggg	aagtcaggct	acagaatgaa	300
	accagctact	ctcgagtgtt	gcatggttat	gcagcacagc	aacttcccag	tctcctgaag	360
10	gagagagagt	ttcaccttgg	gacccttaat	aaagtgtttg	catctcagtg	gttgaatcat	420
	aggcaagtgg	tgtgtggcac	aaaatgcaac	acgctatttg	tcgtagatgt	ccagacaagc	480
	cagatcacca	agatccccat	tctgaaagac	cgggagcctg	gaggtgtgac	ccagcagggc	540
	tgtggtatcc	atgccatcga	gctgcatcct	tctagaacac	tgctagccac	tggaggagac	600
15	aacccaaca	gtcttgccat	ctatcgacta	cctacgctgg	atcctgtgtg	tgtaggagat	660
	gatggacaca	aggactggat	cttttccatc	gcattggatca	gcgacactat	ggcagtgctt	720
	ggctcacgtg	atggttctat	gggactctgg	gaggtgacag	atgatgtttt	gaccaaaagt	780
	gatgagagac	acaatgtgtc	acgggtccct	gtgtatgcac	acatcactca	caaggcctta	840
20	aaggacatcc	ccaaagaaga	cacaaacctt	gacaactgca	aggttcgggc	tctggccttc	900
	aacaacaaga	acaaggaact	gggagcagtg	tctctggatg	gctactttca	tctctggaag	960
	gctgaaaata	cactatctaa	gctcctctcc	accaaactgc	catattgccg	tgagaatgtg	1020
25	tgcttggttt	atggtagtga	atggtcagtt	tatgcagtg	gctcccaagc	tcattgtctcc	1080
	ttcttggtatc	cacggcagcc	atcatacaac	gtcaagtctg	tctgttccag	ggagcgaggc	1140
	agtgaatcc	ggtcagttag	tttctacgag	cacatcatca	ctgtgggaac	agggcagggc	1200
	tccctgctgt	tctatgacat	ccgagctcag	agatttctgg	aagagaggct	ctcagcttgt	1260
30	tatgggtcca	agcccagact	agcaggggag	aatctgaaac	taaccactgg	caaaggctgg	1320
	ctgaatcatg	atgaaacctg	gaggaattac	ttttcagaca	ttgacttctt	ccccaatgct	1380
	gtttacaccc	actgtacga	ctcgtctgga	acgaaactct	ttgtggcagg	aggtcccctc	1440
	ccttcagggc	tccatggaaa	ctatgctggg	ctctggagtt	aatgacaact	ccccaaatgc	1500
35	agagatttac	actaacttcc	attctcagtt	tccttgtttc	ttttgatttt	ttttccta	1560
	tgtgtgaggc	tcttgtgttt	tagtgggaac	accaaagt	gcctatagtt	taggcactta	1620
	ataggaagaa	gctctgtaca	gaaatctgaa	agttgttttg	ctttttgttt	tcccccttg	1680
	taatcaaaat	tttactatct	tttattttt	ctggcttttc	aaccaaacat	tgttgcta	1740
40	ccctatTTTT	ccttaagtga	cacacattct	cctgtctctg	gcttcttcag	gctgaaatga	1800
	catagtcttt	ctcaccctta	cttcaactct	gagaggtagg	gctcctttat	aattacatgg	1860
	ttgtctcag	actttctgtg	aaagtttggg	agctgtgtgt	gtctgtgtgt	gtgtgagaga	1920
45	gagatcttgt	ctgcgtgtgt	gtgtgtgac	ttgtgtgcct	gtaggtactg	tgtgtcactg	1980
	aaattacctg	gagtggagat	tacttgta	taaaatattt	ataaaagaaa	caactttatt	2040
	cacagagtcc	agctttggga	ctagtctgta	tcttgttttt	taagtcta	aacactgata	2100
	ataggaagta	aaaacagaaa	ggaaaagaaa	ttaccactgg	gaaaatcttt	ttagttagat	2160
50	tgtaggcttc	ctggggcctc	ccatgccagg	actgcaaagt	gatccagccc	tacctgtctt	2220
	cccactgtg	tgtccccctg	gtgggaagt	ggtgtcactt	ccccttccca	ccctcacatc	2280
	tgcttagcca	gtagccacac	ccctaaaaca	tcagactcac	catccagggtg	cagctccaga	2340
55	ggctacaaaa	ggcttcatgg	gacttgaatc	cccatcctag	cttctctctc	cttccccctca	2400

EP 2 194 145 B1

	agacctgatac	tggttttaag	gggcctggag	ctgggagtct	caagtctgct	aagattcaca	2460
	tccatagccc	ccgtggcttt	gaggagaatc	ctctctgcca	ttcttccaat	ctccccagtg	2520
5	ggttttgcta	ttattttcta	aattgggtta	agtctaagaa	ggtgggggtg	agcagggggg	2580
	ttatctgtgt	gtagtgagt	cttcatgtgt	ggaatattca	ttttcttact	gcagtgggac	2640
	ttggggttga	agccacccct	cctactctgt	tggcttagcc	ctgagatggg	gacaggctgg	2700
	cctgcagtca	gcatcattgt	gcatgtgaca	gcatcaatgt	gattagtaat	ttgtctgttc	2760
10	ctcccttgaa	ctgtctgttt	agtctgaggt	ttttaaaact	gcaggcagct	gactgtgatg	2820
	tccacttggt	ccctgatttt	tacacatcat	gtcaaagata	acagctgttc	ccaccaccca	2880
	gttcctctaa	gcacatactc	tgcttttctg	tcaacatccc	attttgggga	aaggaaaagt	2940
	catatttatt	cctgcacccc	agttttttaa	cttgttctcc	cagtgtgtccc	cctcttctct	3000
15	gggtgtaaga	agggaaattg	gaaaaaaaaa	tatatatata	ttctcctttt	aatggtgggg	3060
	ggctactgga	gaggagagac	agcaagtcca	ccctaacttg	ttacacagca	cataccacag	3120
	gttcctggaat	tctcatcttc	gaacctagag	aaataggtgc	tataaacagg	gaattaagca	3180
	aaatgctgga	tgctatagat	cttttaattg	tcttaatttt	ttttctatta	ttaaactaca	3240
20	ggctgtagat	ttcttagttc	tcacagaact	tctatcattt	taaactgact	tgtatattta	3300
	aaaaaaaaat	cttcagtagg	atgttttgta	ctattgctag	accctcttct	gtaatgggta	3360
	atgcgtttga	ttgtttgaga	ctttctgttt	ttaaaaatgt	agcacttgac	tttttgccag	3420
25	gaaaaaaata	aaaattattc	cgtgcaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	3478

<210> 433
 <211> 1242
 <212> DNA
 <213> Homo sapiens

30
 <400> 433

	atttcatggt	atacttaata	aaacaaaaca	tacctgtata	cacacacatt	cactcacatt	60
35	gaagatgcaa	gatgaagaaa	gatacatgac	attgaatgta	cagtcaaaga	aaaggagttc	120
	tgcccaaaca	tctcaactta	catttaaaga	ttattcagtg	acgttgact	ggtataaaat	180
	cttactggga	atatctggaa	ccgtgaatgg	tattctcact	ttgactttga	tctccttgat	240
40	cctgttggtt	tctcaggagg	tattgctaaa	atgccaaaaa	ggaagttggt	caaatgccac	300

45

50

55

EP 2 194 145 B1

	tcagtatgag gacactggag atctaaaagt gaataatggc acaagaagaa atataagtaa	360
	taaggacctt tgtgcttcga gatctgcaga ccagacagta ctatgccaat cagaatggct	420
5	caaataccaa ggggaagtgtt attggttctc taatgagatg aaaagctgga gtgacagtta	480
	tgtgtattgt ttggaaagaa aatctcatct actaatcata catgaccaac ttgaaatggc	540
	ttttatacag aaaaacctaa gacaattaaa ctacgtatgg attgggctta actttacctc	600
	cttgaaaatg acatggactt ggggtggatgg ttctccaata gattcaaaga tattcttcat	660
10	aaagggacca gctaaagaaa acagctgtgc tgccattaag gaaagcaaaa ttttctctga	720
	aacctgcagc agtggtttca aatggatttg tcagtattag agtttgacaa aattcacagt	780
	gaaataatca atgatcacta tttttggcct attagtttct aatattaatc tccagggtga	840
	agatttttaa gtgcaattaa atgcaaaaat ctcttctccc ttctccctcc atcatcgaca	900
15	ctggtctagc ctcagagtaa cccctgttaa caaactaaaa tgtacacttc aaaattttta	960
	cgtgatagta taaaccaatg tgacttcatg tgatcatatc caggattttt attcgtcgct	1020
	tatttttatgc caaatgtgat caaattatgc ctgtttttct gtatcttgcg ttttaaattc	1080
20	ttaataaggt cctaaacaaa atttcttata tttctaattg ttgaattata atgtgggttt	1140
	atacattttt tacccttttg tcaaagagaa ttaactttgt ttccaggctt ttgctactct	1200
	tcactcagct acaataaaca tcctgaatgt tttcttaaaa aa	1242

25 <210> 434
 <211> 2298
 <212> DNA
 <213> Homo sapiens

30 <400> 434

35

40

45

50

55

EP 2 194 145 B1

	tcggccgagc ccagagacag ccagttcctc tcccgccgcg ccggggccgcg tgccgctcgc	60
	tccccggccg tggcgccctc gggccagacg cgtgcagcc tccagcccgc ggcaagcggg	120
5	cggggcggcc gcgccacccc cggccccgcg ccagcagccc ctgcgcgcgc gtccagcggt	180
	cccgccagc agcctcccca tacgcagtcc tgctggaccg ccccgctcgcg cccccactc	240
	tgaactcaag tcaccgtgga gctccgccgc ccgaaactt tcacgcgagc gggaaatatg	300
	ggatgtataa aatcaaaaagg gaaagacagc ttgagtgcg atggagtaga tttgaagact	360
10	caaccagtac gtaatactga aagaactatt tatgtgagag atccaacgtc caataaacag	420
	caaaggccag ttccagaatc tcagctttta cctggacaga ggtttcaaac taaagatcca	480
	gaggaacaag gagacattgt ggtagccttg taccctatg atggcatcca cccggacgac	540
	ttgtctttca agaaaggaga gaagatgaaa gtcctggagg agcatggaga atggtggaaa	600
15	gcaaagtccc ttttaacaaa aaaagaaggc ttcattccca gcaactatgt ggccaaactc	660
	aacaccttag aaacagaaga gtgggttttc aaggatataa ccaggaagga cgcagaaaagg	720
	cagcttttgg caccaggaag tagcgctgga gctttcctta ttagagaaaag tgaacatta	780
	aaaggaagct tctctctgtc tgtcagagac tttgaccctg tgcattggtga tgttattaag	840
20	cactacaaaa ttagaagtct ggataatggg ggctattaca tctctccacg aatcactttt	900
	ccctgtatca gcgacatgat taaacattac caaagcagg cagatggctt gtgcagaaga	960
	ttggagaagg cttgtattag tcccaagcca cagaagccat gggataaaga tgcctgggag	1020
25	atcccccggt agtccatcaa gttggtgaaa aggcttggcg ctgggcagtt tggggaagtc	1080
	tggatgggtt actataacaa cagtaccaag gtggctgtga aaaccctgaa gccaggaact	1140
	atgtctgtgc aagccttcct ggaagaagcc aacctcatga agaccctgca gcatgacaag	1200
	ctcgtgaggc tctacgctgt ggtcaccagg gaggagccca tttacatcat caccgagtac	1260
30	atggccaagg gcagtttctg ggatttcctg aagagcgtg aaggtggcaa agtgctgctt	1320
	ccaaagctca ttgacttttc tgctcagatt gcagagggaa tggcatacat cgagcggaag	1380
	aactacattc accgggacct gcgagcagct aatgttctgg tctccgagtc actaatgtgc	1440
35	aaaattgcag attttggcct tgctagagta attgaagata atgagtacac agcaagggaa	1500
	gggtgctaagt tccctattaa gtggacggct ccagaagcaa tcaacttttg atgtttcact	1560
	attaagtctg atgtgtggtc ctttggaatc ctctatagc aaattgtcac ctatgggaaa	1620
40	attccctacc cagggagaac taatgccgac gtgatgaccg ccctgtccca gggctacagg	1680
	atgccccgtg tggagaactg ccagatgag ctctatgaca ttatgaaaat gtgctggaaa	1740
	gaaaaggcag aagagagacc aacgtttgac tacttacaga gcgtcctgga tgatttctac	1800
	acagccacgg aagggaata ccagcagcag ccttagagca caggagacc cgtccatttg	1860
45	gcaggggtgg ctgcctcatt tagagaggaa aagtaaccat cactgggtgc acttatgatt	1920
	tcatgtgcgg ggatcatctg ccgtgcctgg atcctgaaat agaggctaaa ttactcagga	1980
	agaacaccct ctaaatggga agtattctg tactcttaga tggattctcc actcagttgc	2040
	aacttggaact tgcctcagc agctggtaat cttgctctgc ttgacaacat ctgagtgcag	2100
50	ccgtttgaga agaaaacatc tattctctcc aaaaatgcac ccaactagct ctatgtttac	2160
	aaatggacat aggactcaaa gtttcagaga ccattgcaat gaatcccaa taattgcaga	2220
	actaaactca ttataaagc taaaataacc ggatatatac atagcatgac atttctttgt	2280
55	gctttggctt acttgttt	2298

<210> 436
<211> 696
<212> DNA

EP 2 194 145 B1

<213> Homo sapiens

<400> 436

5	ttcccccccc ccccccccc ccccgccga gcacaggaca cagctgggtt ctgaagcttc	60
	tgagttctgc agcctcacct ctgagaaaac ctcttttcca ccaataccat gaagctctgc	120
	gtgactgtcc tgtctctcct catgctagta gctgccttct gctctccagc gctctcagca	180
10	ccaatgggct cagaccctcc caccgcctgc tgcttttctt acaccgcgag gaagcttcct	240
	cgcaactttg tggtagatta ctatgagacc agcagcctct gctcccagcc agctgtggta	300
	ttccaaacca aaagaagcaa gcaagtctgt gctgatccca gtgaatcctg ggtccaggag	360
	tacgtgtatg acctggaact gaactgagct gctcagagac aggaagtctt caggggaaggt	420
15	cacctgagcc cggatgcttc tccatgagac acatctcctc catactcagg actcctctcc	480
	gcagttcctg tcccttctct taatttaatc ttttttatgt gccgtgttat tgtattaggt	540
	gtcatttcca ttatttatat tagtttagcc aaaggataag tgcctatgg ggatgggtcca	600
	ctgtcactgt ttctctgctg ttgcaaatac atggataaca catttgattc tgtgtgtttt	660
20	ccataataaa actttaaaat aaaatgcaga cagtta	696

<210> 439

<211> 384

<212> DNA

25 <213> Homo sapiens

<220>

<221> misc_feature

<222> (144)..(145)

30 <223> n is a, c, g, t or u

<220>

<221> misc_feature

<222> (159)..(159)

35 <223> n is a, c, g, t or u

<220>

<221> misc_feature

<222> (165)..(165)

40 <223> n is a, c, g, t or u

<220>

<221> misc_feature

<222> (223)..(223)

45 <223> n is a, c, g, t or u

<220>

<221> misc_feature

<222> (309)..(309)

50 <223> n is a, c, g, t or u

<400> 439

55

EP 2 194 145 B1

	tttttttttt tttttttttt tcgaagatca gtactttatt ttctctagct ccagtgtttt	60
	gcaactgtag cagcatatca gaaacatccc cacacaaaaa cacacaattc tccccttctt	120
5	caaagagctg gcaacaattg aganncagaa acaatagtna ctacnggcat ttgagaaatt	180
	taagaaataa cacttgctca cccttgaaac atacattgtg cgncttgag gtcggaagca	240
	gcagtacatt tgtcattcaa agacacaatc atccttaaataa aaagttaaataa aaaaccttat	300
	tggcataana accgcgttgg agatgcagct ttatcgggga ctttgggagg aaggtgcttg	360
10	gaataagaca tgagcatttt aaaa	384

<210> 611
 <211> 2333
 <212> DNA
 <213> Homo sapiens
 <400> 611

20	ggcacgaggc tagagcgatg ccgggccgga gttgcgtcgc cttagtcctc ctggctgccg	60
	ccgtcagctg tgccgtcgcg cagcacgcgc cgccgtggac agaggactgc agaaaatcaa	120
	cctatcctcc ttcaggacca acgtacagag gtgcagttcc atggtagacc ataaatcttg	180
	acttaccacc ctacaaaaga tggcatgaat tgatgcttga caaggcacca atgctaaagg	240
25	ttatagtga ttctctgaag aatatgataa atacattcgt gccaaagtga aaagttagtc	300
	aggtggtgga tgaaaaattg cctggcctac ttggcaactt tcctggccct tttgaagagg	360
	aaatgaaggg tattgccgct gttactgata tacctttagg agagattatt tcattcaata	420
	ttttttatga attatttacc atttgactt caatagtagc agaagacaaa aaaggctatc	480
30	taatacatgg gagaaacatg gatatttgag tatttcttgg gtggaacata aataatgata	540
	cctgggtcat aactgagcaa ctaaaacctt taacagtga tttggatttc caaagaaaca	600
	acaaaactgt cttcaaggct tcaagctttg ctggctatgt gggcatgtta acaggattca	660
35	aaccaggact gttcagctt acactgaatg aacgtttcag tataaatggt ggttatctgg	720
	gtattctaga atggattctg ggaagaaaag atgccatgtg gatagggttc ctactagaa	780
	cagttctgga aaatagcaca agttatgaag aagccaagaa ttatttgacc aagaccaaga	840
	tattggcccc agcctacttt atcctgggag gcaaccagtc tggggaagggt tgtgtgatta	900
40	cacgagacag aaaggaatca ttggatgtat atgaactcga tgctaagcag ggtagatggt	960
	atgtggtaca aacaaattat gaccgttggg aacatccctt cttccttgat gatcgagaa	1020
	cgcttgcaaa gatgtgtctg aaccgcacca gccaaagagaa tatctcattt gaaaccatgt	1080
45	atgatgtcct gtcaacaaaa cctgtcctca acaagctgac cgtatacaca accttgatag	1140

EP 2 194 145 B1

	atgttaccaa aggtcaattc gaaacttacc tgcgggactg ccctgaccct tgtataggtt	1200
	ggtgagcaca cgtctggcct acagaatgcg gcctctgaga catgaagaca ccatctccat	1260
5	gtgaccgaac actgcagctg tctgaccttc caaagactaa gactcgcggc aggtttctctt	1320
	tgagtcaata gcttgtcttc gtccatctgt tgacaaatga cagatctttt tttttttccc	1380
	cctatcagtt gatttttctt atttacagat aacttcttta ggggaagtaa aacagtcac	1440
	tagaattcac tgagttttgt ttcactttga catttgggga tctggtgggc agtcgaacca	1500
10	tggtgaactc cacctccgtg gaataaatgg agattcagcg tgggtgttga atccagcacg	1560
	tctgtgtgag taacgggaca gtaaacactc cacattcttc agtttttcac ttctacctac	1620
	atatttgtat gtttttctgt ataacagcct tttccttctg gttctaactg ctgttaaaat	1680
	taatatatca ttatctttgc tgttattgac agcgatatta ttttattaca tatcattaga	1740
15	gggatgagac agacattcac ctgtatattt cttttaatgg gcacaaaatg ggcccttgcc	1800
	tctaaatagc actttttggg gttcaagaag taatcagtat gcaaagcaat cttttataca	1860
	ataattgaag tgttcccttt ttcataatta ctctacttcc cagtaaccct aaggaagttg	1920
	ctaacttaaa aaactgcatc ccacgttctg ttaatttagt aaataaacia gtcaaagact	1980
20	tgtggaaaat aggaagtga cccatatttt aaattctcat aagtagcatt gatgtaataa	2040
	acagggtttt agtttgttct tcagattgat agggagtttt aaagaaattt tagtagttac	2100
	taaaattatg ttactgtatt tttcagaaat caaactgctt atgaaaagta ctaatagaac	2160
25	ttgttaacct ttctaacctt cacgattaac tgtgaaatgt acgtcatttg tgcaagaccg	2220
	tttgtccact tcattttgta taatcacagt tgtgttcctg aactcaata aacagtcact	2280
	ggaaagagtg ccagtcagca gtcatgcacg ctgataaaaa aaaaaaaaaa aaa	2333

30 <210> 634
 <211> 359
 <212> DNA
 <213> Homo sapiens

35 <400> 634

	tttttttttt tttttttttt tttttttttt tttttttttt taaaacaaaa agggctttat	60
	taaaaccccc aaaaaaaact ttttaacaaaa ggggacccat accattcccc aaaaaagttt	120
	agctgaaaaa tggcaacaaa aaagggaag gcttttttta aacccccaaa aataaggttc	180
40	cacaaaaaag gacccgcaa aaccaaatta tagcggcaaa ttttttttgg ccataaatag	240
	ggatccccctt aaaaatcctt gaaactcct tggcagtttt aaggcccaaa ctaacccttg	300
	tgggccagtg gctcaccttc atcaaaaaaa ggaaccatt tggcaaaaaa attttggtt	359

45 <210> 650
 <211> 636
 <212> DNA
 <213> Homo sapiens

50 <400> 650

55

EP 2 194 145 B1

	ggcaacaccc tgtgataatt ccaggtgatt ctctacatct gcagcttgag gtgggaagtc	60
	tgaagctcag agagcctggg ccaatggtac aggtcacaca gcacatcagt ggctacatgt	120
5	gagctcagac ctgggtctgc tgctgtctgt cttcccaata tccatgacct tgactgatgc	180
	aggtgtctag ggatacgtcc atccccgtcc tgctggagcc cagagcacgg aagcctggcc	240
	ctccgaggag acagaaggga gtgtcggaca ccatgacgag agcttggcag aataaataac	300
	ttctttaaac aattttacgg catgaagaaa tctggaccag tttattaaat gggatttctg	360
10	ccacaaacct tggagaatc acatcatctt agcccaaggt gaaaactgtg ttgcgtaaca	420
	aagaacatga ctgcgctcca cacatacatc attgcccggc gaggcgggac acaagtcaac	480
	gacggaacac ttgagacagg cctacaactg tgcacggttc aaaagcagg ttaagccata	540
	cttctgtcag tgagactaca tttctgtcta aagaagatgt ccctgacttg atctgttttt	600
15	caactccagt tcccagatgt gcgtgtgtg gtcccc	636

	<210> 700	
	<211> 21	
	<212> DNA	
20	<213> Homo sapiens	
	<400> 700	
	aggcagaatc cagatgctca a	21
25	<210> 713	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
30	<400> 713	
	cctcgcttc aagaggcaga	20
	<210> 716	
	<211> 20	
35	<212> DNA	
	<213> Homo sapiens	
	<400> 716	
	gtgttgggag ggcagaagtg	20
40	<210> 741	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
45	<400> 741	
	ctacctaag ggggactgtc tt	23
	<210> 742	
50	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
	<400> 742	
55	gcacgggcta caagctgag	19
	<210> 743	
	<211> 21	

<212> DNA
 <213> Homo sapiens

 <400> 743
 5 agcaccgtgt gggacaataa c 21

 <210> 744
 <211> 20
 <212> DNA
 10 <213> Homo sapiens

 <400> 744
 gactgtgctc cggcagttct 20

 15 <210> 749
 <211> 22
 <212> DNA
 <213> Homo sapiens

 20 <400> 749
 ctggtgggag gtctcataa ac 22

 <210> 750
 <211> 20
 25 <212> DNA
 <213> Homo sapiens

 <400> 750
 ctggctcacc tggacaacct 20
 30
 <210> 751
 <211> 21
 <212> DNA
 <213> Homo sapiens
 35
 <400> 751
 ggccacaaga ataagcagca a 21

 40 <210> 752
 <211> 20
 <212> DNA
 <213> Homo sapiens

 <400> 752
 45 ttgggcagc ttggtaagt 20

 <210> 755
 <211> 20
 <212> DNA
 50 <213> Homo sapiens

 <400> 755
 tgtgtgtgct tgtgcgtgtc 20

 55 <210> 758
 <211> 21
 <212> DNA
 <213> Homo sapiens

	<400> 758	
	ggggctactg gagaggagag a	21
5	<210> 765	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 765	
	tgccaaaatc tcttccct tc	22
	<210> 766	
	<211> 20	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 766	
	acagggagac ccgtccattt	20
20	<210> 768	
	<211> 25	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 768	
	tgccgtgtta ttgtattagg tgtca	25
	<210> 771	
30	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 771	
35	tgctgtactc aggtggcact aact	24
	<210> 942	
40	<211> 27	
	<212> DNA	
	<213> Homo sapiens	
	<400> 942	
	actgttaac ctttctaacc ttcacga	27
45	<210> 965	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 965	
	gacccgcca aaccaaatta	20
	<210> 981	
55	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 981	
	gcacggttca aaagcaggtt	20

EP 2 194 145 B1

	<210> 1031	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
5		
	<400> 1031	
	ccctggccca caagtatcac	20
10		
	<210> 1044	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 1044	
	caccctccag ttcccactgt	20
20		
	<210> 1047	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
25		
	<400> 1047	
	cggccaacc ccttaaata	20
30		
	<210> 1072	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
35		
	<400> 1072	
	ttggcctctt tcagcctctt t	21
40		
	<210> 1073	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
45		
	<400> 1073	
	cctgcagtgg gccctagtc	19
50		
	<210> 1074	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
55		
	<400> 1074	
	gagcacatcc ccaaatcca	20
60		
	<210> 1075	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
65		
	<400> 1075	
	gatcagctgc ttgtgcctgt	20
70		
	<210> 1080	
	<211> 23	

<212> DNA
 <213> Homo sapiens

 <400> 1080
 5 agttcaaccc aaatgatcag gaa 23

 <210> 1081
 <211> 20
 <212> DNA
 10 <213> Homo sapiens

 <400> 1081
 gccaggagc ctgaagttct 20

 15 <210> 1082
 <211> 23
 <212> DNA
 <213> Homo sapiens

 20 <400> 1082
 accaaaatga gaacctcaac agc 23

 <210> 1083
 <211> 27
 25 <212> DNA
 <213> Homo sapiens

 <400> 1083
 aatttctgga aaagtcaaca ggataca 27
 30
 <210> 1086
 <211> 20
 <212> DNA
 <213> Homo sapiens
 35
 <400> 1086
 ccggccacat tcactgattt 20

 <210> 1089
 <211> 20
 40 <212> DNA
 <213> Homo sapiens

 <400> 1089
 45 gagaattccg gaacctgtgg 20

 <210> 1096
 <211> 25
 <212> DNA
 50 <213> Homo sapiens

 <400> 1096
 tgatcacatg aagtcacatt ggttt 25

 55 <210> 1097
 <211> 19
 <212> DNA
 <213> Homo sapiens

	<400> 1097		
	agatgatccc cgcacatga	19	
5	<210> 1099		
	<211> 23		
	<212> DNA		
	<213> Homo sapiens		
10	<400> 1099		
	ccatgtattt gcaacagcag aga	23	
	<210> 1102		
	<211> 23		
	<212> DNA		
15	<213> Homo sapiens		
	<400> 1102		
	cagtcatggg tgtcttgga gtg	23	
20	<210> 1273		
	<211> 23		
	<212> DNA		
	<213> Homo sapiens		
25	<400> 1273		
	tggcactctt tccagtgact gtt	23	
	<210> 1296		
	<211> 20		
30	<212> DNA		
	<213> Homo sapiens		
	<400> 1296		
35	gggccttaaa actgccaagg	20	
	<210> 1312		
	<211> 20		
	<212> DNA		
40	<213> Homo sapiens		
	<400> 1312		
	caacacgcac atctgggaac	20	
45	<210> 1362		
	<211> 34		
	<212> DNA		
	<213> Homo sapiens		
50	<400> 1362		
	ccaactacta aactggggga tattatgaag ggcc	34	
	<210> 1375		
	<211> 26		
	<212> DNA		
55	<213> Homo sapiens		
	<400> 1375		
	tgtccctccc tccttcagag agtggg	26	

<210> 1378
 <211> 20
 <212> DNA
 <213> Homo sapiens
 5
 <400> 1378 20
 tcccagggtg ggcacatggg 20
 <210> 1403
 <211> 29
 <212> DNA
 <213> Homo sapiens
 10
 <400> 1403
 ccatccctta aatcctcagg tcacaacca 29
 <210> 1404
 <211> 22
 <212> DNA
 <213> Homo sapiens
 20
 <400> 1404
 tcccttcacc ttcgctgcc ca 22
 <210> 1405
 <211> 22
 <212> DNA
 <213> Homo sapiens
 25
 <400> 1405
 aaccccatct ggtcagtgcg gc 22
 <210> 1406
 <211> 20
 <212> DNA
 <213> Homo sapiens
 35
 <400> 1406
 cgctgcctgg gtgcgactgc 20
 <210> 1411
 <211> 32
 <212> DNA
 <213> Homo sapiens
 40
 <400> 1411 32
 caagatcca aatccaaac tgattgactg ag 32
 <210> 1412
 <211> 24
 <212> DNA
 <213> Homo sapiens
 50
 <400> 1412
 tgcactgtga caagctgcac gtgg 24
 <210> 1413

<211> 28
 <212> DNA
 <213> Homo sapiens

5 <400> 1413
 tcctctttgg ccacaagaat aagcagca 28

<210> 1414
 <211> 26
 10 <212> DNA
 <213> Homo sapiens

<400> 1414
 15 ccaccaaaga actgtcagca gctgcc 26

<210> 1417
 <211> 24
 <212> DNA
 20 <213> Homo sapiens

<400> 1417
 aaaacagctg gagagtccca gccg 24

<210> 1420
 25 <211> 31
 <212> DNA
 <213> Homo sapiens

<400> 1420
 30 tgctgtgtaa caagttaggg tggacttgct g 31

<210> 1427
 <211> 27
 <212> DNA
 35 <213> Homo sapiens

<400> 1427
 ccctccatca tcgacactgg tctagcc 27

40 <210> 1428
 <211> 21
 <212> DNA
 <213> Homo sapiens

45 <400> 1428
 tggcaggggt ggctgcctca t 21

<210> 1430
 <211> 25
 50 <212> DNA
 <213> Homo sapiens

<400> 1430
 55 cccctatggg gatggtccac tgtca 25

<210> 1433
 <211> 27
 <212> DNA

<213> Homo sapiens
 <400> 1433
 catctgcagc cagttagtgc cacctga 27
 5
 <210> 1604
 <211> 23
 <212> DNA
 <213> Homo sapiens
 10
 <400> 1604
 ttcccaggct gcctctcctc acc 23
 <210> 1627
 <211> 24
 <212> DNA
 <213> Homo sapiens
 15
 <400> 1627
 aggctctgct cgttcctct cccc 24
 20
 <210> 1643
 <211> 22
 <212> DNA
 <213> Homo sapiens
 25
 <400> 1643
 cacagcctcg gtagcagcgg ga 22
 30
 <210> 1677
 <211> 22
 <212> DNA
 <213> Homo sapiens
 35
 <400> 1677
 gcccttcata atatcccca gt 22
 <210> 1690
 <211> 20
 <212> DNA
 <213> Homo sapiens
 40
 <400> 1690
 gggagaggga catcctacgg 20
 45
 <210> 1850
 <211> 27
 <212> DNA
 <213> Homo sapiens
 50
 <400> 1850
 acttgtaac ctttctaacc ttcacga 27
 <210> 1873
 <211> 21
 <212> DNA
 <213> Homo sapiens
 55

	<400> 1873	
	tggcagtttt aaggcccaaa c	21
5	<210> 1889	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 1889	
	ctgggccaat ggtacaggtc	20
	<210> 1925	
	<211> 18	
	<212> DNA	
15	<213> Homo sapiens	
	<400> 1925	
	ggtggctggt gtggctaa	18
20	<210> 1938	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 1938	
	cacttctgcc ctccaacac	20
	<210> 2098	
	<211> 28	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 2098	
35	tgtgattata caaatgaag tggacaaa	28
	<210> 2121	
	<211> 21	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 2121	
	tttgccaaa tgggttcctt t	21
	<210> 2137	
45	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 2137	
50	tgcatcagtc aaggtcatgg a	21
	<210> 2173	
	<211> 27	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 2173	
	tgccctggcc cacaagtatc actaagc	27

EP 2 194 145 B1

<210> 2186
 <211> 20
 <212> DNA
 <213> Homo sapiens
 5
 <400> 2186
 caaccccagc cctgccctcc 20
 <210> 2346
 10 <211> 30
 <212> DNA
 <213> Homo sapiens
 <400> 2346
 15 cgtcttgca caaatgacgt acattcaca 30
 <210> 2369
 <211> 21
 <212> DNA
 20 <213> Homo sapiens
 <400> 2369
 tgagccactg gccacaagg g 21
 25 <210> 2385
 <211> 25
 <212> DNA
 <213> Homo sapiens
 30 <400> 2385
 gggaagacag acagcagcag accca 25
 <210> 2445
 <211> 798
 35 <212> PRT
 <213> Homo sapiens
 <400> 2445
 40 Met Val Ala Leu Pro Met Val Leu Val Leu Leu Leu Val Leu Ser Arg
 1 5 10 15
 45
 50
 55

EP 2 194 145 B1

Gly Glu Ser Glu Leu Asp Ala Lys Ile Pro Ser Thr Gly Asp Ala Thr
 20 25 30
 Glu Trp Arg Asn Pro His Leu Ser Met Leu Gly Ser Cys Gln Pro Ala
 35 40 45
 Pro Ser Cys Gln Lys Cys Ile Leu Ser His Pro Ser Cys Ala Trp Cys
 50 55 60
 Lys Gln Leu Asn Phe Thr Ala Ser Gly Glu Ala Glu Ala Arg Arg Cys
 65 70 75 80
 Ala Arg Arg Glu Glu Leu Leu Ala Arg Gly Cys Pro Leu Glu Glu Leu
 85 90 95
 Glu Glu Pro Arg Gly Gln Gln Glu Val Leu Gln Asp Gln Pro Leu Ser
 100 105 110
 Gln Gly Ala Arg Gly Glu Gly Ala Thr Gln Leu Ala Pro Gln Arg Val
 115 120 125
 Arg Val Thr Leu Arg Pro Gly Glu Pro Gln Gln Leu Gln Val Arg Phe
 130 135 140
 Leu Arg Ala Glu Gly Tyr Pro Val Asp Leu Tyr Tyr Leu Met Asp Leu
 145 150 155 160
 Ser Tyr Ser Met Lys Asp Asp Leu Glu Arg Val Arg Gln Leu Gly His
 165 170 175
 Ala Leu Leu Val Arg Leu Gln Glu Val Thr His Ser Val Arg Ile Gly
 180 185 190
 Phe Gly Ser Phe Val Asp Lys Thr Val Leu Pro Phe Val Ser Thr Val
 195 200 205
 Pro Ser Lys Leu Arg His Pro Cys Pro Thr Arg Leu Glu Arg Cys Gln
 210 215 220
 Ser Pro Phe Ser Phe His His Val Leu Ser Leu Thr Gly Asp Ala Gln
 225 230 235 240
 Ala Phe Glu Arg Glu Val Gly Arg Gln Ser Val Ser Gly Asn Leu Asp
 245 250 255
 Ser Pro Glu Gly Gly Phe Asp Ala Ile Leu Gln Ala Ala Leu Cys Gln
 260 265 270
 Glu Gln Ile Gly Trp Arg Asn Val Ser Arg Leu Leu Val Phe Thr Ser
 275 280 285
 Asp Asp Thr Phe His Thr Ala Gly Asp Gly Lys Leu Gly Gly Ile Phe
 290 295 300
 Met Pro Ser Asp Gly His Cys His Leu Asp Ser Asn Gly Leu Tyr Ser
 305 310 315 320
 Arg Ser Thr Glu Phe Asp Tyr Pro Ser Val Gly Gln Val Ala Gln Ala
 325 330 335
 Leu Ser Ala Ala Asn Ile Gln Pro Ile Phe Ala Val Thr Ser Ala Ala
 340 345 350
 Leu Pro Val Tyr Gln Glu Leu Ser Lys Leu Ile Pro Lys Ser Ala Val
 355 360 365

EP 2 194 145 B1

	Gly	Glu	Leu	Ser	Glu	Asp	Ser	Ser	Asn	Val	Val	Gln	Leu	Ile	Met	Asp
		370					375					380				
5	Ala	Tyr	Asn	Ser	Leu	Ser	Ser	Thr	Val	Thr	Leu	Glu	His	Ser	Ser	Leu
	385					390					395					400
	Pro	Pro	Gly	Val	His	Ile	Ser	Tyr	Glu	Ser	Gln	Cys	Glu	Gly	Pro	Glu
					405					410					415	
10	Lys	Arg	Glu	Gly	Lys	Ala	Glu	Asp	Arg	Gly	Gln	Cys	Asn	His	Val	Arg
				420					425					430		
	Ile	Asn	Gln	Thr	Val	Thr	Phe	Trp	Val	Ser	Leu	Gln	Ala	Thr	His	Cys
			435					440					445			
15	Leu	Pro	Glu	Pro	His	Leu	Leu	Arg	Leu	Arg	Ala	Leu	Gly	Phe	Ser	Glu
		450					455					460				
	Glu	Leu	Ile	Val	Glu	Leu	His	Thr	Leu	Cys	Asp	Cys	Asn	Cys	Ser	Asp
	465					470					475					480
20	Thr	Gln	Pro	Gln	Ala	Pro	His	Cys	Ser	Asp	Gly	Gln	Gly	His	Leu	Gln
					485					490					495	
	Cys	Gly	Val	Cys	Ser	Cys	Ala	Pro	Gly	Arg	Leu	Gly	Arg	Leu	Cys	Glu
				500					505					510		
25	Cys	Ser	Val	Ala	Glu	Leu	Ser	Ser	Pro	Asp	Leu	Glu	Ser	Gly	Cys	Arg
			515					520					525			
	Ala	Pro	Asn	Gly	Thr	Gly	Pro	Leu	Cys	Ser	Gly	Lys	Gly	His	Cys	Gln
		530					535					540				
30	Cys	Gly	Arg	Cys	Ser	Cys	Ser	Gly	Gln	Ser	Ser	Gly	His	Leu	Cys	Glu
	545					550					555					560
	Cys	Asp	Asp	Ala	Ser	Cys	Glu	Arg	His	Glu	Gly	Ile	Leu	Cys	Gly	Gly
					565					570					575	
35	Phe	Gly	Arg	Cys	Gln	Cys	Gly	Val	Cys	His	Cys	His	Ala	Asn	Arg	Thr
				580					585					590		
	Gly	Arg	Ala	Cys	Glu	Cys	Ser	Gly	Asp	Met	Asp	Ser	Cys	Ile	Ser	Pro
			595					600					605			
40	Glu	Gly	Gly	Leu	Cys	Ser	Gly	His	Gly	Arg	Cys	Lys	Cys	Asn	Arg	Cys
		610					615					620				
	Gln	Cys	Leu	Asp	Gly	Tyr	Tyr	Gly	Ala	Leu	Cys	Asp	Gln	Cys	Pro	Gly
	625					630					635					640
45	Cys	Lys	Thr	Pro	Cys	Glu	Arg	His	Arg	Asp	Cys	Ala	Glu	Cys	Gly	Ala
					645					650					655	
50	Phe	Arg	Thr	Gly	Pro	Leu	Ala	Thr	Asn	Cys	Ser	Thr	Ala	Cys	Ala	His
				660					665					670		
	Thr	Asn	Val	Thr	Leu	Ala	Leu	Ala	Pro	Ile	Leu	Asp	Asp	Gly	Trp	Cys
			675					680					685			
55	Lys	Glu	Arg	Thr	Leu	Asp	Asn	Gln	Leu	Phe	Phe	Phe	Leu	Val	Glu	Asp
		690					695					700				

EP 2 194 145 B1

	Asp	Ala	Arg	Gly	Thr	Val	Val	Leu	Arg	Val	Arg	Pro	Gln	Glu	Lys	Gly
	705					710					715					720
5	Ala	Asp	His	Thr	Gln	Ala	Ile	Val	Leu	Gly	Cys	Val	Gly	Gly	Ile	Val
					725					730					735	
	Ala	Val	Gly	Leu	Gly	Leu	Val	Leu	Ala	Tyr	Arg	Leu	Ser	Val	Glu	Ile
				740					745					750		
10	Tyr	Asp	Arg	Arg	Glu	Tyr	Ser	Arg	Phe	Glu	Lys	Glu	Gln	Gln	Gln	Leu
			755					760					765			
	Asn	Trp	Lys	Gln	Asp	Ser	Asn	Pro	Leu	Tyr	Lys	Ser	Ala	Ile	Thr	Thr
	770						775					780				
15	Thr	Ile	Asn	Pro	Arg	Phe	Gln	Glu	Ala	Asp	Ser	Pro	Thr	Leu		
	785					790					795					
	<210> 2448															
	<211> 288															
20	<212> PRT															
	<213> Homo sapiens															
	<400> 2448															
25	Met	Gln	Ile	Pro	Gln	Ala	Pro	Trp	Pro	Val	Val	Trp	Ala	Val	Leu	Gln
	1				5					10					15	
	Leu	Gly	Trp	Arg	Pro	Gly	Trp	Phe	Leu	Asp	Ser	Pro	Asp	Arg	Pro	Trp
				20					25					30		
30	Asn	Pro	Pro	Thr	Phe	Phe	Pro	Ala	Leu	Leu	Val	Val	Thr	Glu	Gly	Asp
			35					40					45			
	Asn	Ala	Thr	Phe	Thr	Cys	Ser	Phe	Ser	Asn	Thr	Ser	Glu	Ser	Phe	Val
		50					55					60				
35	Leu	Asn	Trp	Tyr	Arg	Met	Ser	Pro	Ser	Asn	Gln	Thr	Asp	Lys	Leu	Ala
	65					70					75				80	
	Ala	Phe	Pro	Glu	Asp	Arg	Ser	Gln	Pro	Gly	Gln	Asp	Cys	Arg	Phe	Arg
					85					90					95	
40	Val	Thr	Gln	Leu	Pro	Asn	Gly	Arg	Asp	Phe	His	Met	Ser	Val	Val	Arg
				100					105					110		
	Ala	Arg	Arg	Asn	Asp	Ser	Gly	Thr	Tyr	Leu	Cys	Gly	Ala	Ile	Ser	Leu
45																
50																
55																

EP 2 194 145 B1

	115	120	125	
5	Ala Pro Lys Ala Gln Ile Lys Glu Ser Leu Arg Ala Glu Leu Arg Val 130 135 140			
	Thr Glu Arg Arg Ala Glu Val Pro Thr Ala His Pro Ser Pro Ser Pro 145 150 155 160			
10	Arg Pro Ala Gly Gln Phe Gln Thr Leu Val Val Gly Val Val Gly Gly 165 170 175			
	Leu Leu Gly Ser Leu Val Leu Leu Val Trp Val Leu Ala Val Ile Cys 180 185 190			
15	Ser Arg Ala Ala Arg Gly Thr Ile Gly Ala Arg Arg Thr Gly Gln Pro 195 200 205			
	Leu Lys Glu Asp Pro Ser Ala Val Pro Val Phe Ser Val Asp Tyr Gly 210 215 220			
20	Glu Leu Asp Phe Gln Trp Arg Glu Lys Thr Pro Glu Pro Pro Val Pro 225 230 235 240			
	Cys Val Pro Glu Gln Thr Glu Tyr Ala Thr Ile Val Phe Pro Ser Gly 245 250 255			
25	Met Gly Thr Ser Ser Pro Ala Arg Arg Gly Ser Ala Asp Gly Pro Arg 260 265 270			
	Ser Ala Gln Pro Leu Arg Pro Glu Asp Gly His Cys Ser Trp Pro Leu 275 280 285			
30	<210> 2473 <211> 281 <212> PRT <213> Homo sapiens			
35	<400> 2473			
	Met Gln Gln Pro Phe Asn Tyr Pro Tyr Pro Gln Ile Tyr Trp Val Asp 1 5 10 15			
40	Ser Ser Ala Ser Ser Pro Trp Ala Pro Pro Gly Thr Val Leu Pro Cys 20 25 30			
	Pro Thr Ser Val Pro Arg Arg Pro Gly Gln Arg Arg Pro Pro Pro Pro 35 40 45			
45	Pro Pro Pro Pro Pro Leu Pro Pro Pro Pro Pro Pro Pro Leu Pro 50 55 60			
	Pro Leu Pro Leu Pro Pro Leu Lys Lys Arg Gly Asn His Ser Thr Gly 65 70 75 80			
50	Leu Cys Leu Leu Val Met Phe Phe Met Val Leu Val Ala Leu Val Gly 85 90 95			
	Leu Gly Leu Gly Met Phe Gln Leu Phe His Leu Gln Lys Glu Leu Ala 100 105 110			
55	Glu Leu Arg Glu Ser Thr Ser Gln Met His Thr Ala Ser Ser Leu Glu 115 120 125			

EP 2 194 145 B1

	Lys	Gln	Ile	Gly	His	Pro	Ser	Pro	Pro	Pro	Glu	Lys	Lys	Glu	Leu	Arg
	130						135					140				
5	Lys	Val	Ala	His	Leu	Thr	Gly	Lys	Ser	Asn	Ser	Arg	Ser	Met	Pro	Leu
	145					150					155					160
	Glu	Trp	Glu	Asp	Thr	Tyr	Gly	Ile	Val	Leu	Leu	Ser	Gly	Val	Lys	Tyr
					165					170					175	
10	Lys	Lys	Gly	Gly	Leu	Val	Ile	Asn	Glu	Thr	Gly	Leu	Tyr	Phe	Val	Tyr
				180					185					190		
	Ser	Lys	Val	Tyr	Phe	Arg	Gly	Gln	Ser	Cys	Asn	Asn	Leu	Pro	Leu	Ser
			195					200					205			
15	His	Lys	Val	Tyr	Met	Arg	Asn	Ser	Lys	Tyr	Pro	Gln	Asp	Leu	Val	Met
		210					215					220				
	Met	Glu	Gly	Lys	Met	Met	Ser	Tyr	Cys	Thr	Thr	Gly	Gln	Met	Trp	Ala
	225					230					235					240
20	Arg	Ser	Ser	Tyr	Leu	Gly	Ala	Val	Phe	Asn	Leu	Thr	Ser	Ala	Asp	His
					245					250					255	
	Leu	Tyr	Val	Asn	Val	Ser	Glu	Leu	Ser	Leu	Val	Asn	Phe	Glu	Glu	Ser
				260					265					270		
25	Gln	Thr	Phe	Phe	Gly	Leu	Tyr	Lys	Leu							
			275					280								

<210> 2474
 <211> 830
 <212> PRT
 <213> Homo sapiens

 <400> 2474

35

40

45

50

55

EP 2 194 145 B1

Met Gly Ser Met Phe Arg Ser Glu Glu Val Ala Leu Val Gln Leu Phe
1 5 10 15
Leu Pro Thr Ala Ala Ala Tyr Thr Cys Val Ser Arg Leu Gly Glu Leu
20 25 30
Gly Leu Val Glu Phe Arg Asp Leu Asn Ala Ser Val Ser Ala Phe Gln
35 40 45
Arg Arg Phe Val Val Asp Val Arg Arg Cys Glu Glu Leu Glu Lys Thr
50 55 60
Phe Thr Phe Leu Gln Glu Glu Val Arg Arg Ala Gly Leu Val Leu Pro
65 70 75 80
Pro Pro Lys Gly Arg Leu Pro Ala Pro Pro Pro Arg Asp Leu Leu Arg
85 90 95
Ile Gln Glu Glu Thr Glu Arg Leu Ala Gln Glu Leu Arg Asp Val Arg
100 105 110
Gly Asn Gln Gln Ala Leu Arg Ala Gln Leu His Gln Leu Gln Leu His
115 120 125
Ala Ala Val Leu Arg Gln Gly His Glu Pro Gln Leu Ala Ala Ala His
130 135 140
Thr Asp Gly Ala Ser Glu Arg Thr Pro Leu Leu Gln Ala Pro Gly Gly
145 150 155 160

EP 2 194 145 B1

Pro His Gln Asp Leu Arg Val Asn Phe Val Ala Gly Ala Val Glu Pro
 165 170 175
 His Lys Ala Pro Ala Leu Glu Arg Leu Leu Trp Arg Ala Cys Arg Gly
 180 185 190
 Phe Leu Ile Ala Ser Phe Arg Glu Leu Glu Gln Pro Leu Glu His Pro
 195 200 205
 Val Thr Gly Glu Pro Ala Thr Trp Met Thr Phe Leu Ile Ser Tyr Trp
 210 215 220
 Gly Glu Gln Ile Gly Gln Lys Ile Arg Lys Ile Thr Asp Cys Phe His
 225 230 235 240
 Cys His Val Phe Pro Phe Leu Gln Gln Glu Glu Ala Arg Leu Gly Ala
 245 250 255
 Leu Gln Gln Leu Gln Gln Gln Ser Gln Glu Leu Gln Glu Val Leu Gly
 260 265 270
 Glu Thr Glu Arg Phe Leu Ser Gln Val Leu Gly Arg Val Leu Gln Leu
 275 280 285
 Leu Pro Pro Gly Gln Val Gln Val His Lys Met Lys Ala Val Tyr Leu
 290 295 300
 Ala Leu Asn Gln Cys Ser Val Ser Thr Thr His Lys Cys Leu Ile Ala
 305 310 315 320
 Glu Ala Trp Cys Ser Val Arg Asp Leu Pro Ala Leu Gln Glu Ala Leu
 325 330 335
 Arg Asp Ser Ser Met Glu Glu Gly Val Ser Ala Val Ala His Arg Ile
 340 345 350
 Pro Cys Arg Asp Met Pro Pro Thr Leu Ile Arg Thr Asn Arg Phe Thr
 355 360 365
 Ala Ser Phe Gln Gly Ile Val Asp Ala Tyr Gly Val Gly Arg Tyr Gln
 370 375 380
 Glu Val Asn Pro Ala Pro Tyr Thr Ile Ile Thr Phe Pro Phe Leu Phe
 385 390 395 400
 Ala Val Met Phe Gly Asp Val Gly His Gly Leu Leu Met Phe Leu Phe
 405 410 415
 Ala Leu Ala Met Val Leu Ala Glu Asn Arg Pro Ala Val Lys Ala Ala
 420 425 430
 Gln Asn Glu Ile Trp Gln Thr Phe Phe Arg Gly Arg Tyr Leu Leu Leu
 435 440 445
 Leu Met Gly Leu Phe Ser Ile Tyr Thr Gly Phe Ile Tyr Asn Glu Cys
 450 455 460
 Phe Ser Arg Ala Thr Ser Ile Phe Pro Ser Gly Trp Ser Val Ala Ala
 465 470 475 480
 Met Ala Asn Gln Ser Gly Trp Ser Asp Ala Phe Leu Ala Gln His Thr
 485 490 495
 Met Leu Thr Leu Asp Pro Asn Val Thr Gly Val Phe Leu Gly Pro Tyr

EP 2 194 145 B1

	500	505	510
5	Pro Phe Gly Ile Asp Pro Ile Trp Ser Leu Ala Ala Asn His Leu Ser 515 520 525		
	Phe Leu Asn Ser Phe Lys Met Lys Met Ser Val Ile Leu Gly Val Val 530 535 540		
10	His Met Ala Phe Gly Val Val Leu Gly Val Phe Asn His Val His Phe 545 550 555 560		
	Gly Gln Arg His Arg Leu Leu Leu Glu Thr Leu Pro Glu Leu Thr Phe 565 570 575		
15	Leu Leu Gly Leu Phe Gly Tyr Leu Val Phe Leu Val Ile Tyr Lys Trp 580 585 590		
	Leu Cys Val Trp Ala Ala Arg Ala Ala Ser Ala Pro Ser Ile Leu Ile 595 600 605		
20	His Phe Ile Asn Met Phe Leu Phe Ser His Ser Pro Ser Asn Arg Leu 610 615 620		
	Leu Tyr Pro Arg Gln Glu Val Val Gln Ala Thr Leu Val Val Leu Ala 625 630 635 640		
25	Leu Ala Met Val Pro Ile Leu Leu Leu Gly Thr Pro Leu His Leu Leu 645 650 655		
	His Arg His Arg Arg Arg Leu Arg Arg Arg Pro Ala Asp Arg Gln Glu 660 665 670		
30	Glu Asn Lys Ala Gly Leu Leu Asp Leu Pro Asp Ala Ser Val Asn Gly 675 680 685		
	Trp Ser Ser Asp Glu Glu Lys Ala Gly Gly Leu Asp Asp Glu Glu Glu 690 695 700		
35	Ala Glu Leu Val Pro Ser Glu Val Leu Met His Gln Ala Ile His Thr 705 710 715 720		
	Ile Glu Phe Cys Leu Gly Cys Val Ser Asn Thr Ala Ser Tyr Leu Arg 725 730 735		
40	Leu Trp Ala Leu Ser Leu Ala His Ala Gln Leu Ser Glu Val Leu Trp 740 745 750		
	Ala Met Val Met Arg Ile Gly Leu Gly Leu Gly Arg Glu Val Gly Val 755 760 765		
45	Ala Ala Val Val Leu Val Pro Ile Phe Ala Ala Phe Ala Val Met Thr 770 775 780		
	Val Ala Ile Leu Leu Val Met Glu Gly Leu Ser Ala Phe Leu His Ala 785 790 795 800		
50	Leu Arg Leu His Trp Val Glu Phe Gln Asn Lys Phe Tyr Ser Gly Thr 805 810 815		
55	Gly Tyr Lys Leu Ser Pro Phe Thr Phe Ala Ala Thr Asp Asp 820 825 830		

<210> 2475

<211> 555

<212> PRT
<213> Homo sapiens

<400> 2475

5

10

15

20

25

30

35

40

45

50

55

EP 2 194 145 B1

	Met	Ala	Ala	Arg	Leu	Leu	Leu	Leu	Gly	Ile	Leu	Leu	Leu	Leu	Pro	
	1				5					10				15		
5	Leu	Pro	Val	Pro	Ala	Pro	Cys	His	Thr	Ala	Ala	Arg	Ser	Glu	Cys	Lys
				20					25					30		
	Arg	Ser	His	Lys	Phe	Val	Pro	Gly	Ala	Trp	Leu	Ala	Gly	Glu	Gly	Val
			35					40					45			
10	Asp	Val	Thr	Ser	Leu	Arg	Arg	Ser	Gly	Ser	Phe	Pro	Val	Asp	Thr	Gln
		50					55					60				
	Arg	Phe	Leu	Arg	Pro	Asp	Gly	Thr	Cys	Thr	Leu	Cys	Glu	Asn	Ala	Leu
	65					70					75					80
15	Gln	Glu	Gly	Thr	Leu	Gln	Arg	Leu	Pro	Leu	Ala	Leu	Thr	Asn	Trp	Arg
					85					90					95	
	Ala	Gln	Gly	Ser	Gly	Cys	Gln	Arg	His	Val	Thr	Arg	Ala	Lys	Val	Ser
				100					105					110		
20	Ser	Thr	Glu	Ala	Val	Ala	Arg	Asp	Ala	Ala	Arg	Ser	Ile	Arg	Asn	Asp
			115					120					125			
	Trp	Lys	Val	Gly	Leu	Asp	Val	Thr	Pro	Lys	Pro	Thr	Ser	Asn	Val	His
		130					135					140				
25	Val	Ser	Val	Ala	Gly	Ser	His	Ser	Gln	Ala	Ala	Asn	Phe	Ala	Ala	Gln
	145					150				155						160
	Lys	Thr	His	Gln	Asp	Gln	Tyr	Ser	Phe	Ser	Thr	Asp	Thr	Val	Glu	Cys
				165						170					175	
30	Arg	Phe	Tyr	Ser	Phe	His	Val	Val	His	Thr	Pro	Pro	Leu	His	Pro	Asp
				180					185					190		
	Phe	Lys	Arg	Ala	Leu	Gly	Asp	Leu	Pro	His	His	Phe	Asn	Ala	Ser	Thr
			195					200					205			
35	Gln	Pro	Ala	Tyr	Leu	Arg	Leu	Ile	Ser	Asn	Tyr	Gly	Thr	His	Phe	Ile
		210					215					220				
	Arg	Ala	Val	Glu	Leu	Gly	Gly	Arg	Ile	Ser	Ala	Leu	Thr	Ala	Leu	Arg
	225					230					235					240
40	Thr	Cys	Glu	Leu	Ala	Leu	Glu	Gly	Leu	Thr	Asp	Asn	Glu	Val	Glu	Asp
				245					250						255	
	Cys	Leu	Thr	Val	Glu	Ala	Gln	Val	Asn	Ile	Gly	Ile	His	Gly	Ser	Ile
				260					265					270		
45	Ser	Ala	Glu	Ala	Lys	Ala	Cys	Glu	Glu	Lys	Lys	Lys	Lys	His	Lys	Met
			275					280					285			
50	Thr	Ala	Ser	Phe	His	Gln	Thr	Tyr	Arg	Glu	Arg	His	Ser	Glu	Val	Val
		290					295					300				
	Gly	Gly	His	His	Thr	Ser	Ile	Asn	Asp	Leu	Leu	Phe	Gly	Ile	Gln	Ala
	305					310					315					320
55	Gly	Pro	Glu	Gln	Tyr	Ser	Ala	Trp	Val	Asn	Ser	Val	Pro	Gly	Ser	Pro
					325					330					335	

EP 2 194 145 B1

Gly Leu Val Asp Tyr Thr Leu Glu Pro Leu His Val Leu Leu Asp Ser
 340 345 350
 5 Gln Asp Pro Arg Arg Glu Ala Leu Arg Arg Ala Leu Ser Gln Tyr Leu
 355 360 365
 Thr Asp Arg Ala Arg Trp Arg Asp Cys Ser Arg Pro Cys Pro Pro Gly
 370 375 380
 10 Arg Gln Lys Ser Pro Arg Asp Pro Cys Gln Cys Val Cys His Gly Ser
 385 390 395 400
 Ala Val Thr Thr Gln Asp Cys Cys Pro Arg Gln Arg Gly Leu Ala Gln
 405 410 415
 15 Leu Glu Val Thr Phe Ile Gln Ala Trp Ser Leu Trp Gly Asp Trp Phe
 420 425 430
 Thr Ala Thr Asp Ala Tyr Val Lys Leu Phe Phe Gly Gly Gln Glu Leu
 435 440 445
 20 Arg Thr Ser Thr Val Trp Asp Asn Asn Asn Pro Ile Trp Ser Val Arg
 450 455 460
 Leu Asp Phe Gly Asp Val Leu Leu Ala Thr Gly Gly Pro Leu Arg Leu
 465 470 475 480
 25 Gln Val Trp Asp Gln Asp Ser Gly Arg Asp Asp Asp Leu Leu Gly Thr
 485 490 495
 Cys Asp Gln Ala Pro Lys Ser Gly Ser His Glu Val Arg Cys Asn Leu
 500 505 510
 30 Asn His Gly His Leu Lys Phe Arg Tyr His Ala Arg Cys Leu Pro His
 515 520 525
 Leu Gly Gly Gly Thr Cys Leu Asp Tyr Val Pro Gln Met Leu Leu Gly
 530 535 540
 35 Glu Pro Pro Gly Asn Arg Ser Gly Ala Val Trp
 545 550 555

<210> 2476
 <211> 153
 <212> PRT
 <213> Homo sapiens

<400> 2476

EP 2 194 145 B1

	Met	Gly	Leu	Thr	Ser	Gln	Leu	Leu	Pro	Pro	Leu	Phe	Phe	Leu	Leu	Ala
	1				5					10					15	
5	Cys	Ala	Gly	Asn	Phe	Val	His	Gly	His	Lys	Cys	Asp	Ile	Thr	Leu	Gln
				20					25					30		
	Glu	Ile	Ile	Lys	Thr	Leu	Asn	Ser	Leu	Thr	Glu	Gln	Lys	Thr	Leu	Cys
			35					40					45			
10	Thr	Glu	Leu	Thr	Val	Thr	Asp	Ile	Phe	Ala	Ala	Ser	Lys	Asn	Thr	Thr
		50					55					60				
	Glu	Lys	Glu	Thr	Phe	Cys	Arg	Ala	Ala	Thr	Val	Leu	Arg	Gln	Phe	Tyr
	65					70					75					80
15	Ser	His	His	Glu	Lys	Asp	Thr	Arg	Cys	Leu	Gly	Ala	Thr	Ala	Gln	Gln
							85				90				95	
20	Phe	His	Arg	His	Lys	Gln	Leu	Ile	Arg	Phe	Leu	Lys	Arg	Leu	Asp	Arg
				100					105					110		
	Asn	Leu	Trp	Gly	Leu	Ala	Gly	Leu	Asn	Ser	Cys	Pro	Val	Lys	Glu	Ala
			115					120					125			
25	Asn	Gln	Ser	Thr	Leu	Glu	Asn	Phe	Leu	Glu	Arg	Leu	Lys	Thr	Ile	Met
		130					135					140				
	Arg	Glu	Lys	Tyr	Ser	Lys	Cys	Ser	Ser							
	145					150										

30 <210> 2481
 <211> 147
 <212> PRT
 <213> Homo sapiens

35 <400> 2481

40

45

50

55

EP 2 194 145 B1

Met Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp
1 5 10 15
Gly Lys Val Asn Val Asp Glu Val Gly Gly Glu Ala Leu Gly Arg Leu
5 20 25 30
Leu Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp
35 40 45
Leu Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His
10 50 55 60
Gly Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp
65 70 75 80
Asn Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys
15 85 90 95
Leu His Val Asp Pro Glu Asn Phe Arg Leu Leu Gly Asn Val Leu Val
100 105 110
Cys Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln
20 115 120 125
Ala Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His
25 130 135 140
Lys Tyr His
145

<210> 2482
<211> 259
<212> PRT
<213> Homo sapiens

<400> 2482

EP 2 194 145 B1

Met Ser Lys Tyr Lys Leu Ile Met Leu Arg His Gly Glu Gly Ala Trp
 1 5 10 15
 Asn Lys Glu Asn Arg Phe Cys Ser Trp Val Asp Gln Lys Leu Asn Ser
 20 25 30
 Glu Gly Met Glu Glu Ala Arg Asn Cys Gly Lys Gln Leu Lys Ala Leu
 35 40 45
 Asn Phe Glu Phe Asp Leu Val Phe Thr Ser Val Leu Asn Arg Ser Ile
 50 55 60
 His Thr Ala Trp Leu Ile Leu Glu Glu Leu Gly Gln Glu Trp Val Pro
 65 70 75 80
 Val Glu Ser Ser Trp Arg Leu Asn Glu Arg His Tyr Gly Ala Leu Ile
 85 90 95
 Gly Leu Asn Arg Glu Gln Met Ala Leu Asn His Gly Glu Glu Gln Val
 100 105 110
 Arg Leu Trp Arg Arg Ser Tyr Asn Val Thr Pro Pro Pro Ile Glu Glu
 115 120 125
 Ser His Pro Tyr Tyr Gln Glu Ile Tyr Asn Asp Arg Arg Tyr Lys Val
 130 135 140
 Cys Asp Val Pro Leu Asp Gln Leu Pro Arg Ser Glu Ser Leu Lys Asp
 145 150 155 160
 Val Leu Glu Arg Leu Leu Pro Tyr Trp Asn Glu Arg Ile Ala Pro Glu
 165 170 175
 Val Leu Arg Gly Lys Thr Ile Leu Ile Ser Ala His Gly Asn Ser Ser
 180 185 190
 Arg Ala Leu Leu Lys His Leu Glu Gly Ile Ser Asp Glu Asp Ile Ile
 195 200 205
 Asn Ile Thr Leu Pro Thr Gly Val Pro Ile Leu Leu Glu Leu Asp Glu
 210 215 220
 Asn Leu Arg Ala Val Gly Pro His Gln Phe Leu Gly Asp Gln Glu Ala
 225 230 235 240
 Ile Gln Ala Ala Ile Lys Lys Val Glu Asp Gln Gly Lys Val Lys Gln
 245 250 255
 Ala Lys Lys

<210> 2483
 <211> 344
 <212> PRT
 <213> Homo sapiens
 <400> 2483

EP 2 194 145 B1

Met Ser Ala Leu Ala Ala Arg Leu Leu Gln Pro Ala His Ser Cys Ser
 1 5 10 15
 Leu Arg Leu Arg Pro Phe His Leu Ala Ala Val Arg Asn Glu Ala Val
 5 20 25 30
 Val Ile Ser Gly Arg Lys Leu Ala Gln Gln Ile Lys Gln Glu Val Arg
 35 40 45
 Gln Glu Val Glu Glu Trp Val Ala Ser Gly Asn Lys Arg Pro His Leu
 10 50 55 60
 Ser Val Ile Leu Val Gly Glu Asn Pro Ala Ser His Ser Tyr Val Leu
 65 70 75 80
 Asn Lys Thr Arg Ala Ala Ala Val Val Gly Ile Asn Ser Glu Thr Ile
 15 85 90 95
 Met Lys Pro Ala Ser Ile Ser Glu Glu Glu Leu Leu Asn Leu Ile Asn
 100 105 110
 Lys Leu Asn Asn Asp Asp Asn Val Asp Gly Leu Leu Val Gln Leu Pro
 115 120 125
 Leu Pro Glu His Ile Asp Glu Arg Arg Ile Cys Asn Ala Val Ser Pro
 130 135 140
 Asp Lys Asp Val Asp Gly Phe His Val Ile Asn Val Gly Arg Met Cys
 145 150 155 160
 Leu Asp Gln Tyr Ser Met Leu Pro Ala Thr Pro Trp Gly Val Trp Glu
 165 170 175
 Ile Ile Lys Arg Thr Gly Ile Pro Thr Leu Gly Lys Asn Val Val Val
 180 185 190
 Ala Gly Arg Ser Lys Asn Val Gly Met Pro Ile Ala Met Leu Leu His
 195 200 205

EP 2 194 145 B1

	Thr	Asp	Gly	Ala	His	Glu	Arg	Pro	Gly	Gly	Asp	Ala	Thr	Val	Thr	Ile
	210						215					220				
5	Ser	His	Arg	Tyr	Thr	Pro	Lys	Glu	Gln	Leu	Lys	Lys	His	Thr	Ile	Leu
	225					230					235					240
	Ala	Asp	Ile	Val	Ile	Ser	Ala	Ala	Gly	Ile	Pro	Asn	Leu	Ile	Thr	Ala
					245					250					255	
10	Asp	Met	Ile	Lys	Glu	Gly	Ala	Ala	Val	Ile	Asp	Val	Gly	Ile	Asn	Arg
				260					265					270		
	Val	His	Asp	Pro	Val	Thr	Ala	Lys	Pro	Lys	Leu	Val	Gly	Asp	Val	Asp
			275					280					285			
15	Phe	Glu	Gly	Val	Arg	Gln	Lys	Ala	Gly	Tyr	Ile	Thr	Pro	Val	Pro	Gly
		290					295					300				
	Gly	Val	Gly	Pro	Met	Thr	Val	Ala	Met	Leu	Met	Lys	Asn	Thr	Ile	Ile
	305					310					315					320
20	Ala	Ala	Lys	Lys	Val	Leu	Arg	Leu	Glu	Glu	Arg	Glu	Val	Leu	Lys	Ser
					325					330					335	
	Lys	Glu	Leu	Gly	Val	Ala	Thr	Asn								
				340												
25	<210> 2485															
	<211> 453															
	<212> PRT															
	<213> Homo sapiens															
30	<400> 2485															
	Met	Ala	Arg	Lys	Val	Val	Ser	Arg	Lys	Arg	Lys	Ala	Pro	Ala	Ser	Pro
	1				5					10					15	

EP 2 194 145 B1

	Gly	Ala	Gly	Ser ₂₀	Asp	Ala	Gln	Gly	Pro ₂₅	Gln	Phe	Gly	Trp	Asp ₃₀	His	Ser
5	Leu	His	Lys ₃₅	Arg	Lys	Arg	Leu	Pro ₄₀	Pro	Val	Lys	Arg	Ser ₄₅	Leu	Val	Tyr
	Tyr	Leu ₅₀	Lys	Asn	Arg	Glu	Val ₅₅	Arg	Leu	Gln	Asn	Glu ₆₀	Thr	Ser	Tyr	Ser
10	Arg	Val	Leu	His	Gly	Tyr ₇₀	Ala	Ala	Gln	Gln	Leu ₇₅	Pro	Ser	Leu	Leu	Lys ₈₀
	Glu	Arg	Glu	Phe ₈₅	His	Leu	Gly	Thr	Leu	Asn ₉₀	Lys	Val	Phe	Ala	Ser ₉₅	Gln
15	Trp	Leu	Asn ₁₀₀	His	Arg	Gln	Val	Val	Cys ₁₀₅	Gly	Thr	Lys	Cys	Asn ₁₁₀	Thr	Leu
	Phe	Val	Val ₁₁₅	Asp	Val	Gln	Thr	Ser ₁₂₀	Gln	Ile	Thr	Lys	Ile ₁₂₅	Pro	Ile	Leu
20	Lys	Asp ₁₃₀	Arg	Glu	Pro	Gly	Gly ₁₃₅	Val	Thr	Gln	Gln	Gly ₁₄₀	Cys	Gly	Ile	His
	Ala	Ile	Glu	Leu	Asn	Pro ₁₅₀	Ser	Arg	Thr	Leu	Leu ₁₅₅	Ala	Thr	Gly	Gly	Asp ₁₆₀
25	Asn	Pro	Asn	Ser	Leu ₁₆₅	Ala	Ile	Tyr	Arg	Leu ₁₇₀	Pro	Thr	Leu	Asp	Pro ₁₇₅	Val
	Cys	Val	Gly	Asp ₁₈₀	Asp	Gly	His	Lys	Asp ₁₈₅	Trp	Ile	Phe	Ser	Ile ₁₉₀	Ala	Trp
30	Ile	Ser	Asp ₁₉₅	Thr	Met	Ala	Val	Ser ₂₀₀	Gly	Ser	Arg	Asp	Gly ₂₀₅	Ser	Met	Gly
	Leu	Trp ₂₁₀	Glu	Val	Thr	Asp	Asp ₂₁₅	Val	Leu	Thr	Lys	Ser ₂₂₀	Asp	Ala	Arg	His
35	Asn	Val	Ser	Arg	Val	Pro ₂₃₀	Val	Tyr	Ala	His	Ile ₂₃₅	Thr	His	Lys	Ala	Leu ₂₄₀
	Lys	Asp	Ile	Pro	Lys ₂₄₅	Glu	Asp	Thr	Asn	Pro ₂₅₀	Asp	Asn	Cys	Lys	Val ₂₅₅	Arg
40	Ala	Leu	Ala	Phe ₂₆₀	Asn	Asn	Lys	Asn	Lys ₂₆₅	Glu	Leu	Gly	Ala	Val ₂₇₀	Ser	Leu
	Asp	Gly	Tyr ₂₇₅	Phe	His	Leu	Trp	Lys ₂₈₀	Ala	Glu	Asn	Thr	Leu ₂₈₅	Ser	Lys	Leu
45	Leu	Ser ₂₉₀	Thr	Lys	Leu	Pro	Tyr ₂₉₅	Cys	Arg	Glu	Asn	Val ₃₀₀	Cys	Leu	Ala	Tyr
	Gly	Ser	Glu	Trp	Ser	Val ₃₁₀	Tyr	Ala	Val	Gly	Ser ₃₁₅	Gln	Ala	His	Val	Ser ₃₂₀
50	Phe	Leu	Asp	Pro	Arg ₃₂₅	Gln	Pro	Ser	Tyr	Asn ₃₃₀	Val	Lys	Ser	Val	Cys ₃₃₅	Ser
	Arg	Glu	Arg	Gly ₃₄₀	Ser	Gly	Ile	Arg	Ser ₃₄₅	Val	Ser	Phe	Tyr	Glu ₃₅₀	His	Ile
55																

EP 2 194 145 B1

	Ile	Thr	Val	Gly	Thr	Gly	Gln	Gly	Ser	Leu	Leu	Phe	Tyr	Asp	Ile	Arg
			355					360					365			
5	Ala	Gln	Arg	Phe	Leu	Glu	Glu	Arg	Leu	Ser	Ala	Cys	Tyr	Gly	Ser	Lys
		370					375					380				
	Pro	Arg	Leu	Ala	Gly	Glu	Asn	Leu	Lys	Leu	Thr	Thr	Gly	Lys	Gly	Trp
	385					390					395					400
10	Leu	Asn	His	Asp	Glu	Thr	Trp	Arg	Asn	Tyr	Phe	Ser	Asp	Ile	Asp	Phe
					405					410					415	
	Phe	Pro	Asn	Ala	Val	Tyr	Thr	His	Cys	Tyr	Asp	Ser	Ser	Gly	Thr	Lys
				420					425					430		
15	Leu	Phe	Val	Ala	Gly	Gly	Pro	Leu	Pro	Ser	Gly	Leu	His	Gly	Asn	Tyr
			435					440					445			
	Ala	Gly	Leu	Trp	Ser											
		450														

20 <210> 2486
 <211> 352
 <212> PRT
 <213> Homo sapiens

25 <400> 2486

30

35

40

45

50

55

EP 2 194 145 B1

Met Glu Gly Ile Ser Ile Tyr Thr Ser Asp Asn Tyr Thr Glu Glu Met
1 5 10 15

5 Gly Ser Gly Asp Tyr Asp Ser Met Lys Glu Pro Cys Phe Arg Glu Glu
20 25 30

10 Asn Ala Asn Phe Asn Lys Ile Phe Leu Pro Thr Ile Tyr Ser Ile Ile
35 40 45

Phe Leu Thr Gly Ile Val Gly Asn Gly Leu Val Ile Leu Val Met Gly
50 55 60

15 Tyr Gln Lys Lys Leu Arg Ser Met Thr Asp Lys Tyr Arg Leu His Leu
65 70 75 80

20 Ser Val Ala Asp Leu Leu Phe Val Ile Thr Leu Pro Phe Trp Ala Val
85 90 95

25 Asp Ala Val Ala Asn Trp Tyr Phe Gly Asn Phe Leu Cys Lys Ala Val
100 105 110

His Val Ile Tyr Thr Val Asn Leu Tyr Ser Ser Val Leu Ile Leu Ala
115 120 125

30 Phe Ile Ser Leu Asp Arg Tyr Leu Ala Ile Val His Ala Thr Asn Ser
130 135 140

35 Gln Arg Pro Arg Lys Leu Leu Ala Glu Lys Val Val Tyr Val Gly Val
145 150 155 160

40

45

50

55

EP 2 194 145 B1

	Trp	Ile	Pro	Ala	Leu	Leu	Leu	Thr	Ile	Pro	Asp	Phe	Ile	Phe	Ala	Asn	
					165					170					175		
5	Val	Ser	Glu	Ala	Asp	Asp	Arg	Tyr	Ile	Cys	Asp	Arg	Phe	Tyr	Pro	Asn	
				180					185					190			
10	Asp	Leu	Trp	Val	Val	Val	Phe	Gln	Phe	Gln	His	Ile	Met	Val	Gly	Leu	
			195					200					205				
15	Ile	Leu	Pro	Gly	Ile	Val	Ile	Leu	Ser	Cys	Tyr	Cys	Ile	Ile	Ile	Ser	
		210					215					220					
20	Lys	Leu	Ser	His	Ser	Lys	Gly	His	Gln	Lys	Arg	Lys	Ala	Leu	Lys	Thr	
	225					230					235					240	
25	Thr	Val	Ile	Leu	Ile	Leu	Ala	Phe	Phe	Ala	Cys	Trp	Leu	Pro	Tyr	Tyr	
				245						250					255		
30	Ile	Gly	Ile	Ser	Ile	Asp	Ser	Phe	Ile	Leu	Leu	Glu	Ile	Ile	Lys	Gln	
				260					265					270			
35	Gly	Cys	Glu	Phe	Glu	Asn	Thr	Val	His	Lys	Trp	Ile	Ser	Ile	Thr	Glu	
			275					280					285				
40	Ala	Leu	Ala	Phe	Phe	His	Cys	Cys	Leu	Asn	Pro	Ile	Leu	Tyr	Ala	Phe	
		290					295					300					
45	Leu	Gly	Ala	Lys	Phe	Lys	Thr	Ser	Ala	Gln	His	Ala	Leu	Thr	Ser	Val	
	305					310					315					320	
50	Ser	Arg	Gly	Ser	Ser	Leu	Lys	Ile	Leu	Ser	Lys	Gly	Lys	Arg	Gly	Gly	
				325						330					335		
55	His	Ser	Ser	Val	Ser	Thr	Glu	Ser	Glu	Ser	Ser	Ser	Phe	His	Ser	Ser	
				340					345					350			
60	<210>	2487															
	<211>	199															
	<212>	PRT															
	<213>	Homo sapiens															
65	<400>	2487															
70	Met	Ser	Ser	Glu	Asn	Cys	Phe	Val	Ala	Glu	Asn	Ser	Ser	Leu	His	Pro	
	1				5					10					15		

EP 2 194 145 B1

	Glu Ser Gly Gln Glu Asn Asp Ala Thr Ser Pro His Phe Ser Thr Arg	
	20 25 30	
5	His Glu Gly Ser Phe Gln Val Pro Val Leu Cys Ala Val Met Asn Val	
	35 40 45	
10	Val Phe Ile Thr Ile Leu Ile Ile Ala Leu Ile Ala Leu Ser Val Gly	
	50 55 60	
15	Gln Tyr Asn Cys Pro Gly Gln Tyr Thr Phe Ser Met Pro Ser Asp Ser	
	65 70 75 80	
20	His Val Ser Ser Cys Ser Glu Asp Trp Val Gly Tyr Gln Arg Lys Cys	
	85 90 95	
25	Tyr Phe Ile Ser Thr Val Lys Arg Ser Trp Thr Ser Ala Gln Asn Ala	
	100 105 110	
30	Cys Ser Glu His Gly Ala Thr Leu Ala Val Ile Asp Ser Glu Lys Asp	
	115 120 125	
35	Met Asn Phe Leu Lys Arg Tyr Ala Gly Arg Glu Glu His Trp Val Gly	
	130 135 140	
40	Leu Lys Lys Glu Pro Gly His Pro Trp Lys Trp Ser Asn Gly Lys Glu	
	145 150 155 160	
45	Phe Asn Asn Trp Phe Asn Val Thr Gly Ser Asp Lys Cys Val Phe Leu	
	165 170 175	
50	Lys Asn Thr Glu Val Ser Ser Met Glu Cys Glu Lys Asn Leu Tyr Trp	
	180 185 190	
55	Ile Cys Asn Lys Pro Tyr Lys	
	195	
55	<210> 2488	
	<211> 91	
	<212> PRT	
	<213> Homo sapiens	
55	<400> 2488	
	Met Lys Val Ser Ala Ala Ala Leu Ala Val Ile Leu Ile Ala Thr Ala	
	1 5 10 15	
55	Leu Cys Ala Pro Ala Ser Ala Ser Pro Tyr Ser Ser Asp Thr Thr Pro	
	20 25 30	

EP 2 194 145 B1

Cys Cys Phe Ala Tyr Ile Ala Arg Pro Leu Pro Arg Ala His Ile Lys
35 40 45

5 Glu Tyr Phe Tyr Thr Ser Gly Lys Cys Ser Asn Pro Ala Val Val Phe
50 55 60

10 Val Thr Arg Lys Asn Arg Gln Val Cys Ala Asn Pro Glu Lys Lys Trp
65 70 75 80

Val Arg Glu Tyr Ile Asn Ser Leu Glu Met Ser
85 90

15
<210> 2489
<211> 212
<212> PRT
<213> Homo sapiens
20
<400> 2489

Met Asn Ser Phe Ser Thr Ser Ala Phe Gly Pro Val Ala Phe Ser Leu
1 5 10 15

Gly Leu Leu Leu Val Leu Pro Ala Ala Phe Pro Ala Pro Val Pro Pro
20 25 30

30 Gly Glu Asp Ser Lys Asp Val Ala Ala Pro His Arg Gln Pro Leu Thr
35 40 45

35 Ser Ser Glu Arg Ile Asp Lys Gln Ile Arg Tyr Ile Leu Asp Gly Ile
50 55 60

Ser Ala Leu Arg Lys Glu Thr Cys Asn Lys Ser Asn Met Cys Glu Ser
65 70 75 80

Ser Lys Glu Ala Leu Ala Glu Asn Asn Leu Asn Leu Pro Lys Met Ala
85 90 95

45 Glu Lys Asp Gly Cys Phe Gln Ser Gly Phe Asn Glu Glu Thr Cys Leu
100 105 110

50 Val Lys Ile Ile Thr Gly Leu Leu Glu Phe Glu Val Tyr Leu Glu Tyr
115 120 125

Leu Gln Asn Arg Phe Glu Ser Ser Glu Glu Gln Ala Arg Ala Val Gln
130 135 140

EP 2 194 145 B1

	Met	Ser	Thr	Lys	Val	Leu	Ile	Gln	Phe	Leu	Gln	Lys	Lys	Ala	Lys	Asn	
	145					150					155					160	
5	Leu	Asp	Ala	Ile	Thr	Thr	Pro	Asp	Pro	Thr	Thr	Asn	Ala	Ser	Leu	Leu	
					165					170					175		
10	Thr	Lys	Leu	Gln	Ala	Gln	Asn	Gln	Trp	Leu	Gln	Asp	Met	Thr	Thr	His	
				180					185						190		
15	Leu	Ile	Leu	Arg	Ser	Phe	Lys	Glu	Phe	Leu	Gln	Ser	Ser	Leu	Arg	Ala	
			195					200					205				
20	Leu	Arg	Gln	Met													
	210																
	<210>	2490															
	<211>	153															
	<212>	PRT															
	<213>	Homo sapiens															
25	<400>	2490															
	Met	Tyr	Arg	Met	Gln	Leu	Leu	Ser	Cys	Ile	Ala	Leu	Ser	Leu	Ala	Leu	
	1				5					10					15		
30	Val	Thr	Asn	Ser	Ala	Pro	Thr	Ser	Ser	Ser	Thr	Lys	Lys	Thr	Gln	Leu	
				20					25						30		
35	Gln	Leu	Glu	His	Leu	Leu	Leu	Asp	Leu	Gln	Met	Ile	Leu	Asn	Gly	Ile	
			35					40					45				
40	Asn	Asn	Tyr	Lys	Asn	Pro	Lys	Leu	Thr	Arg	Met	Leu	Thr	Phe	Lys	Phe	
	50						55					60					
45	Tyr	Met	Pro	Lys	Lys	Ala	Thr	Glu	Leu	Lys	His	Leu	Gln	Cys	Leu	Glu	
	65					70					75					80	
50	Glu	Glu	Leu	Lys	Pro	Leu	Glu	Glu	Val	Leu	Asn	Leu	Ala	Gln	Ser	Lys	
					85				90						95		
55	Asn	Phe	His	Leu	Arg	Pro	Arg	Asp	Leu	Ile	Ser	Asn	Ile	Asn	Val	Ile	
				100					105					110			
60	Val	Leu	Glu	Leu	Lys	Gly	Ser	Glu	Thr	Thr	Phe	Met	Cys	Glu	Tyr	Ala	
			115					120					125				
65	Asp	Glu	Thr	Ala	Thr	Ile	Val	Glu	Phe	Leu	Asn	Arg	Trp	Ile	Thr	Phe	
	130						135					140					

EP 2 194 145 B1

Cys Gln Ser Ile Ile Ser Thr Leu Thr
145 150

<210> 2491
<211> 231
<212> PRT
<213> Homo sapiens

<400> 2491

Met Gln Asp Glu Glu Arg Tyr Met Thr Leu Asn Val Gln Ser Lys Lys
1 5 10 15

Arg Ser Ser Ala Gln Thr Ser Gln Leu Thr Phe Lys Asp Tyr Ser Val
20 25 30

Thr Leu His Trp Tyr Lys Ile Leu Leu Gly Ile Ser Gly Thr Val Asn
35 40 45

Gly Ile Leu Thr Leu Thr Leu Ile Ser Leu Ile Leu Leu Val Ser Gln
50 55 60

Gly Val Leu Leu Lys Cys Gln Lys Gly Ser Cys Ser Asn Ala Thr Gln
65 70 75 80

Tyr Glu Asp Thr Gly Asp Leu Lys Val Asn Asn Gly Thr Arg Arg Asn
85 90 95

Ile Ser Asn Lys Asp Leu Cys Ala Ser Arg Ser Ala Asp Gln Thr Val
100 105 110

Leu Cys Gln Ser Glu Trp Leu Lys Tyr Gln Gly Lys Cys Tyr Trp Phe
115 120 125

Ser Asn Glu Met Lys Ser Trp Ser Asp Ser Tyr Val Tyr Cys Leu Glu
130 135 140

Arg Lys Ser His Leu Leu Ile Ile His Asp Gln Leu Glu Met Ala Phe
145 150 155 160

Ile Gln Lys Asn Leu Arg Gln Leu Asn Tyr Val Trp Ile Gly Leu Asn
165 170 175

Phe Thr Ser Leu Lys Met Thr Trp Thr Trp Val Asp Gly Ser Pro Ile
180 185 190

Asp Ser Lys Ile Phe Phe Ile Lys Gly Pro Ala Lys Glu Asn Ser Cys
195 200 205

Ala Ala Ile Lys Glu Ser Lys Ile Phe Ser Glu Thr Cys Ser Ser Val
210 215 220

Phe Lys Trp Ile Cys Gln Tyr
225 230

<210> 2492
<211> 512
<212> PRT
<213> Homo sapiens

<400> 2492

EP 2 194 145 B1

Met Gly Cys Ile Lys Ser Lys Gly Lys Asp Ser Leu Ser Asp Asp Gly
1 5 10 15

Val Asp Leu Lys Thr Gln Pro Val Arg Asn Thr Glu Arg Thr Ile Tyr
5 20 25 30

Val Arg Asp Pro Thr Ser Asn Lys Gln Gln Arg Pro Val Pro Glu Ser
35 40 45

Gln Leu Leu Pro Gly Gln Arg Phe Gln Thr Lys Asp Pro Glu Glu Gln
10 50 55 60

Gly Asp Ile Val Val Ala Leu Tyr Pro Tyr Asp Gly Ile His Pro Asp
65 70 75 80

Asp Leu Ser Phe Lys Lys Gly Glu Lys Met Lys Val Leu Glu Glu His
15 85 90 95

Gly Glu Trp Trp Lys Ala Lys Ser Leu Leu Thr Lys Lys Glu Gly Phe
100 105 110

Ile Pro Ser Asn Tyr Val Ala Lys Leu Asn Thr Leu Glu Thr Glu Glu
20 115 120 125

Trp Phe Phe Lys Asp Ile Thr Arg Lys Asp Ala Glu Arg Gln Leu Leu

25

30

35

40

45

50

55

EP 2 194 145 B1

	130					135						140					
5	Ala 145	Pro	Gly	Asn	Ser	Ala 150	Gly	Ala	Phe	Leu	Ile 155	Arg	Glu	Ser	Glu	Thr 160	
	Leu	Lys	Gly	Ser	Phe 165	Ser	Leu	Ser	Val	Arg 170	Asp	Phe	Asp	Pro	Val 175	His	
10	Gly	Asp	Val	Ile 180	Lys	His	Tyr	Lys	Ile 185	Arg	Ser	Leu	Asp	Asn 190	Gly	Gly	
	Tyr	Tyr	Ile 195	Ser	Pro	Arg	Ile	Thr 200	Phe	Pro	Cys	Ile	Ser 205	Asp	Met	Ile	
15	Lys	His 210	Tyr	Gln	Lys	Gln	Ala 215	Asp	Gly	Leu	Cys	Arg 220	Arg	Leu	Glu	Lys	
	Ala 225	Cys	Ile	Ser	Pro	Lys 230	Pro	Gln	Lys	Pro	Trp 235	Asp	Lys	Asp	Ala	Trp 240	
20	Glu	Ile	Pro	Arg	Glu 245	Ser	Ile	Lys	Leu	Val 250	Lys	Arg	Leu	Gly	Ala 255	Gly	
	Gln	Phe	Gly	Glu 260	Val	Trp	Met	Gly	Tyr 265	Tyr	Asn	Asn	Ser	Thr 270	Lys	Val	
25	Ala	Val	Lys 275	Thr	Leu	Lys	Pro	Gly 280	Thr	Met	Ser	Val	Gln 285	Ala	Phe	Leu	
	Glu	Glu 290	Ala	Asn	Leu	Met	Lys 295	Thr	Leu	Gln	His	Asp 300	Lys	Leu	Val	Arg	
30	Leu 305	Tyr	Ala	Val	Val	Thr 310	Arg	Glu	Glu	Pro	Ile 315	Tyr	Ile	Ile	Thr	Glu 320	
	Tyr	Met	Ala	Lys	Gly 325	Ser	Leu	Leu	Asp	Phe 330	Leu	Lys	Ser	Asp	Glu 335	Gly	
35	Gly	Lys	Val	Leu 340	Leu	Pro	Lys	Leu	Ile 345	Asp	Phe	Ser	Ala	Gln 350	Ile	Ala	
	Glu	Gly	Met 355	Ala	Tyr	Ile	Glu	Arg 360	Lys	Asn	Tyr	Ile	His 365	Arg	Asp	Leu	
40	Arg	Ala 370	Ala	Asn	Val	Leu 375	Val	Ser	Glu	Ser	Leu	Met 380	Cys	Lys	Ile	Ala	
	Asp 385	Phe	Gly	Leu	Ala	Arg 390	Val	Ile	Glu	Asp	Asn 395	Glu	Tyr	Thr	Ala	Arg 400	
45	Glu	Gly	Ala	Lys	Phe 405	Pro	Ile	Lys	Trp	Thr 410	Ala	Pro	Glu	Ala	Ile 415	Asn	
	Phe	Gly	Cys	Phe 420	Thr	Ile	Lys	Ser	Asp 425	Val	Trp	Ser	Phe	Gly 430	Ile	Leu	
50	Leu	Tyr	Glu 435	Ile	Val	Thr	Tyr	Gly 440	Lys	Ile	Pro	Tyr	Pro 445	Gly	Arg	Thr	
	Asn	Ala 450	Asp	Val	Met	Thr	Ala 455	Leu	Ser	Gln	Gly	Tyr 460	Arg	Met	Pro	Arg	
55	Val 465	Glu	Asn	Cys	Pro	Asp 470	Glu	Leu	Tyr	Asp	Ile 475	Met	Lys	Met	Cys	Trp 480	

EP 2 194 145 B1

Lys Glu Lys Ala Glu Glu Arg Pro Thr Phe Asp Tyr Leu Gln Ser Val
485 490 495

Leu Asp Asp Phe Tyr Thr Ala Thr Glu Gly Gln Tyr Gln Gln Gln Pro
500 505 510

5

<210> 2599

<211> 395

<212> PRT

10

<213> Homo sapiens

<400> 2599

15

20

25

30

35

40

45

50

55

EP 2 194 145 B1

Met Pro Gly Arg Ser Cys Val Ala Leu Val Leu Leu Ala Ala Ala Val
 1 5 10 15
 Ser Cys Ala Val Ala Gln His Ala Pro Pro Trp Thr Glu Asp Cys Arg
 20 25 30
 Lys Ser Thr Tyr Pro Pro Ser Gly Pro Thr Tyr Arg Gly Ala Val Pro
 35 40 45
 Trp Tyr Thr Ile Asn Leu Asp Leu Pro Pro Tyr Lys Arg Trp His Glu
 50 55 60
 Leu Met Leu Asp Lys Ala Pro Met Leu Lys Val Ile Val Asn Ser Leu
 65 70 75 80
 Lys Asn Met Ile Asn Thr Phe Val Pro Ser Gly Lys Val Met Gln Val
 85 90 95
 Val Asp Glu Lys Leu Pro Gly Leu Leu Gly Asn Phe Pro Gly Pro Phe
 100 105 110
 Glu Glu Glu Met Lys Gly Ile Ala Ala Val Thr Asp Ile Pro Leu Gly
 115 120 125
 Glu Ile Ile Ser Phe Asn Ile Phe Tyr Glu Leu Phe Thr Ile Cys Thr
 130 135 140
 Ser Ile Val Ala Glu Asp Lys Lys Gly His Leu Ile His Gly Arg Asn
 145 150 155 160
 Met Asp Phe Gly Val Phe Leu Gly Trp Asn Ile Asn Asn Asp Thr Trp
 165 170 175
 Val Ile Thr Glu Gln Leu Lys Pro Leu Thr Val Asn Leu Asp Phe Gln
 180 185 190
 Arg Asn Asn Lys Thr Val Phe Lys Ala Ser Ser Phe Ala Gly Tyr Val
 195 200 205
 Gly Met Leu Thr Gly Phe Lys Pro Gly Leu Phe Ser Leu Thr Leu Asn
 210 215 220
 Glu Arg Phe Ser Ile Asn Gly Gly Tyr Leu Gly Ile Leu Glu Trp Ile
 225 230 235 240
 Leu Gly Lys Lys Asp Ala Met Trp Ile Gly Phe Leu Thr Arg Thr Val
 245 250 255
 Leu Glu Asn Ser Thr Ser Tyr Glu Glu Ala Lys Asn Leu Leu Thr Lys
 260 265 270
 Thr Lys Ile Leu Ala Pro Ala Tyr Phe Ile Leu Gly Gly Asn Gln Ser
 275 280 285
 Gly Glu Gly Cys Val Ile Thr Arg Asp Arg Lys Glu Ser Leu Asp Val
 290 295 300

EP 2 194 145 B1

Tyr Glu Leu Asp Ala Lys Gln Gly Arg Trp Tyr Val Val Gln Thr Asn
 305 310 315 320
 5 Tyr Asp Arg Trp Lys His Pro Phe Phe Leu Asp Asp Arg Arg Thr Pro
 325 330 335
 Ala Lys Met Cys Leu Asn Arg Thr Ser Gln Glu Asn Ile Ser Phe Glu
 340 345 350
 10 Thr Met Tyr Asp Val Leu Ser Thr Lys Pro Val Leu Asn Lys Leu Thr
 355 360 365
 Val Tyr Thr Thr Leu Ile Asp Val Thr Lys Gly Gln Phe Glu Thr Tyr
 370 375 380
 15 Leu Arg Asp Cys Pro Asp Pro Cys Ile Gly Trp
 385 390 395

<210> 2736
 <211> 50
 20 <212> DNA
 <213> Homo sapiens

<400> 2736
 25 tggccactg tcactgttc tctgctgtg caaatacatg gataacacat 50

<210> 2751
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 2751
 30 ctccaccacc tgaccagagt gtctcttca gaggactggc tctttccca

<210> 2752
 35 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 2752
 40 ccccaaccac aggcacagg caaccattg aaataaaact ccttcagcct

<210> 2760
 <211> 50
 <212> DNA
 45 <213> Homo sapiens

<400> 2760
 aggagctatg attagacttc tgtagactt cctcactcta tcaccacat 50

50 <210> 2766
 <211> 50
 <212> DNA
 <213> Homo sapiens

55 <400> 2766
 ggccagcctg gaccaatca tgaggaagat gcagactctt atgagaacat 50

<210> 2772

<211> 50
 <212> DNA
 <213> Homo sapiens

5 <400> 2772
 tgctattgcc ttctatttt gcataataaa tgcttcagtg aaaatgcagc 50

<210> 2781
 <211> 50
 10 <212> DNA
 <213> Homo sapiens

<400> 2781
 actgagagtg gtgtctggat atattccttt tgtcttcac ctttctgaa 50

15
 <210> 2783
 <211> 50
 <212> DNA
 <213> Homo sapiens

20
 <400> 2783
 agccctgcaa aaattcagag tccttgcaaa attgtctaaa atgtcagtg 50

25
 <210> 2784
 <211> 50
 <212> DNA
 <213> Homo sapiens

30
 <400> 2784
 gggcagagaa ggtggagagt aaagacccaa cattactaac aatgatacag 50

35
 <210> 2790
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 2790
 acattttcct ccgcataagc ctgcgtcaga ttaaactact gaactgacaa 50

40
 <210> 2794
 <211> 50
 <212> DNA
 <213> Homo sapiens

45
 <400> 2794
 gattctgtc ttgctaataa atcatcacca actgccttct cctacaggga 50

50
 <210> 3016
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 3016
 gggaggaaca ctgcactctt aagcttccgc cgtctcaacc cctcacagga 50

55
 <210> 3018
 <211> 50
 <212> DNA

EP 2 194 145 B1

<213> Homo sapiens
 <400> 3018
 cacctgattc cccctcttgc ccacaggact ctgctgttgt ttccattctg 50
 5
 <210> 3019
 <211> 50
 <212> DNA
 <213> Homo sapiens
 10
 <400> 3019
 attatatttg tccctatcag aatcctcgaa tccctagcag ccagtccttg 50
 15
 <210> 3020
 <211> 50
 <212> DNA
 <213> Homo sapiens
 20
 <400> 3020
 gtcccttagg ggaggagag ttgtcctctt tgcccacagt ctaccctcag 50
 25
 <210> 3021
 <211> 50
 <212> DNA
 <213> Homo sapiens
 30
 <400> 3021
 cttgggccag actgtcaggg ttcaaggagg gcatcaggag cagacggaga 50
 35
 <210> 3022
 <211> 50
 <212> DNA
 <213> Homo sapiens
 40
 <400> 3022
 ctctcaagg ggtctacatg gcaactgtga ggaggggaga ttcagtgtgg 50
 45
 <210> 3023
 <211> 50
 <212> DNA
 <213> Homo sapiens
 50
 <400> 3023
 taagcataaa acctgacacg ttaaaatccc tgcccttgg tgagcccact 50
 55
 <210> 3024
 <211> 50
 <212> DNA
 <213> Homo sapiens
 60
 <400> 3024
 tgctggtatt ctactgcca catthttgga aacctgtatt acacctaaa 50
 65
 <210> 3025
 <211> 50
 <212> DNA
 <213> Homo sapiens

EP 2 194 145 B1

<400> 3025
 cagtcactgg gtctatatta aacagcaacc agagcaacaa atggcaaaca 50

 5 <210> 3026
 <211> 50
 <212> DNA
 <213> Homo sapiens

 10 <400> 3026
 tctagcccag cattgatcta gaagcagagg aatcccagcg ccttttaaaa 50

 <210> 3027
 <211> 50
 <212> DNA
 15 <213> Homo sapiens

 <400> 3027
 tgggcaagac atgattaatg aatcagaatc ctgttcatt ggtgacttgg 50

 20 <210> 3028
 <211> 50
 <212> DNA
 <213> Homo sapiens

 25 <400> 3028
 tgcagattcc tagtagcatg ccttacctac agcactatgt gcatttgctg 50

 <210> 3029
 <211> 50
 30 <212> DNA
 <213> Homo sapiens

 <400> 3029
 ggtctgagag tctgtgaaga tggcccagtc ttctatcccc cacctaaaaa 50
 35
 <210> 3030
 <211> 50
 <212> DNA
 <213> Homo sapiens
 40
 <400> 3030
 cttgaccaaa cccacagcct gtctctctc ttgttagtt acttacggca 50

 45 <210> 3031
 <211> 50
 <212> DNA
 <213> Homo sapiens

 50 <400> 3031
 ccactgtcac tgtttctctg ctgttgcaaa tacatggata acacattga 50

 <210> 3032
 <211> 50
 <212> DNA
 55 <213> Homo sapiens

 <400> 3032
 tgcagccact attgttagtc tcttgattca taatgactta agcacacttg 50

EP 2 194 145 B1

<210> 3033
 <211> 50
 <212> DNA
 <213> Homo sapiens
 5
 <400> 3033
 cctgatggag agaagaaggc atatgtcga ctggctcctg attacgatgc 50
 10
 <210> 3034
 <211> 50
 <212> DNA
 <213> Homo sapiens
 15
 <400> 3034
 gccgaattgt cttggtgct ttcactgt gttttaaata aaggatttt 50
 20
 <210> 3035
 <211> 50
 <212> DNA
 <213> Homo sapiens
 25
 <400> 3035
 ctggaggacc tgtgatgct agttcagagt atcaccaaga gctggagagg 50
 30
 <210> 3036
 <211> 50
 <212> DNA
 <213> Homo sapiens
 35
 <400> 3036
 accatccaat cggacaagct ttcagaacct tattgaagga ttgaagcac 50
 40
 <210> 3037
 <211> 50
 <212> DNA
 <213> Homo sapiens
 45
 <400> 3037
 agaaatgttc agtaatgaaa aaatatatcc aatcagagcc atcccgaata 50
 50
 <210> 3038
 <211> 50
 <212> DNA
 <213> Homo sapiens
 55
 <400> 3038
 agcggactca ggctccagct gtggctacaa catagggttt ttatacaga 50
 60
 <210> 3039
 <211> 50
 <212> DNA
 <213> Homo sapiens
 65
 <400> 3039
 accagactga caaatgtgta tcggatgctt ttgttcaggc ctgtgatcgg 50
 70
 <210> 3040
 <211> 50

EP 2 194 145 B1

<212> DNA
 <213> Homo sapiens

 <400> 3040
 5 ccactgtcac tgttctctg ctgttgcaaa tacatggata acacattga 50

 <210> 3041
 <211> 50
 <212> DNA
 10 <213> Homo sapiens

 <400> 3041
 aagtgacaa aataagcaac taaatgagac ctaataattg gccttcgatt 50

 15 <210> 3042
 <211> 50
 <212> DNA
 <213> Homo sapiens

 20 <400> 3042
 tctgttgata gctggagaac ttagtttca agtactacat tgtgaaagca 50

 <210> 3043
 <211> 50
 25 <212> DNA
 <213> Homo sapiens

 <400> 3043
 gacctcatct ccaagatgga gaagatctga cctccacgga gccgctgtcc 50
 30
 <210> 3044
 <211> 50
 <212> DNA
 <213> Homo sapiens
 35
 <400> 3044
 ctgaccgccca ctctcacatt tgggctcttc gctggccttg gtggagctgg 50

 <210> 3045
 40 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 3045
 45 atacagcccc ggcagaaaac gcctaaagtc agatgagaga ccagtacata 50

 <210> 3046
 <211> 50
 <212> DNA
 50 <213> Homo sapiens

 <400> 3046
 aacagaagtc aagagaacat agaccaactt gctgcatgag taaggtagct 50

 55 <210> 3047
 <211> 50
 <212> DNA
 <213> Homo sapiens

EP 2 194 145 B1

<400> 3047
 agaggactat agtggagtg aaagcattct gtgtttactc ttgcattaa 50

5
 <210> 3048
 <211> 20
 <212> DNA
 <213> Homo sapiens

10
 <400> 3048
 caggtaacc cccaccggac ctacaaccg cagtcccaca tcattcagg 50

15
 <210> 3049
 <211> 50
 <212> DNA
 <213> Homo sapiens

20
 <400> 3049
 cctccagaa gctacgaaa agggagctgt taaatttaa taaatctctg 50

25
 <210> 3050
 <211> 50
 <212> DNA
 <213> Homo sapiens

30
 <400> 3050
 ccaatggata ttctgtatt actagggagg catttacagt cctctaattg 50

35
 <210> 3051
 <211> 50
 <212> DNA
 <213> Homo sapiens

40
 <400> 3051
 atctgacatt attgtaacta ccgtgtgatc agtaagattc ctgtaagaaa 50

45
 <210> 3052
 <211> 50
 <212> DNA
 <213> Homo sapiens

50
 <400> 3052
 tccaatgcag tccattctt tatggcctat agtctcactc ccaactacc 50

55
 <210> 3053
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 3053
 tcctaggtag ggttaaatcc ccagtaaaat tgccatattg cacatgtctt 50

<210> 3055
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 3055
 aaagggtttt atccactgtc attcaattg gataacattt tgtcaagttt 50

EP 2 194 145 B1

<210> 3056
 <211> 50
 <212> DNA
 <213> Homo sapiens
 5
 <400> 3056
 aattcctgaa ccgttgatc acctctgtc agtccatcat ctccaccctg 50
 10
 <210> 3057
 <211> 50
 <212> DNA
 <213> Homo sapiens
 15
 <400> 3057
 ggcaggagg tctcactgtt gtgaagggtg tagacgtgtg gtaatgtgtt 50
 20
 <210> 3058
 <211> 50
 <212> DNA
 <213> Homo sapiens
 25
 <400> 3058
 gtgggtaagg ggctcaagct gtgatgctgc tggtttatc tctagtgaag 50
 30
 <210> 3059
 <211> 50
 <212> DNA
 <213> Homo sapiens
 35
 <400> 3059
 agacaaagag agcataaata tagctctact catgggtacc ataccagtgt 50
 40
 <210> 3060
 <211> 50
 <212> DNA
 <213> Homo sapiens
 45
 <400> 3060
 cttcaggccc aagttcaacg ggttaaagag gtccgctccc aaattattct 50
 50
 <210> 3061
 <211> 50
 <212> DNA
 <213> Homo sapiens
 55
 <400> 3061
 aacaagcat gttgcccta gtccaggatt gcctcactg agacttgcta 50
 60
 <210> 3062
 <211> 50
 <212> DNA
 <213> Homo sapiens
 65
 <400> 3062
 tcttctggg aatgtgatgt gttttcact ggttctaatt ctgtcttct 50
 70
 <210> 3063
 <211> 50

EP 2 194 145 B1

<212> DNA
 <213> Homo sapiens

 5 <400> 3063 tgatctgact ggaaaacaat cctgtatccc ctcccaaaga atcàtgggct 50

 <210> 3064
 <211> 50
 <212> DNA
 <213> Homo sapiens
 10

 <400> 3064
 aacaagccat gtttgccta gtccaggatt gcctcactg agacttgcta 50

 <210> 3065
 15 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 3065
 20 acagcatgag aaactgtag tacgcatacc tcagttcaaa ctttaggga 50

 <210> 3067
 <211> 50
 <212> DNA
 25 <213> Homo sapiens

 <400> 3067
 gagagcttcc tccccgctt cagtttctga tggatctagc catgttgaaa 50
 30
 <210> 3068
 <211> 50
 <212> DNA
 <213> Homo sapiens
 35
 <400> 3068
 ggaggaatgg ctgtgcccggt cccctccact taagcgacct gagtctccag 50

 <210> 3069
 <211> 50
 40 <212> DNA
 <213> Homo sapiens

 <400> 3069
 45 ggaaatgttg ctgtggggga ttcattgtaa ctctcctgt gaactgctca 50

 <210> 3070
 <211> 50
 <212> DNA
 <213> Homo sapiens
 50

 <400> 3070
 aggtgggctg gacttctacc tgccctcaag ggtgtgtata ttgtataggg 50

 <210> 3071
 55 <211> 50
 <212> DNA
 <213> Homo sapiens

EP 2 194 145 B1

<400> 3071
 cagacgctcc agtgctgccg aggttagtgt gttattaga cctgaaatga 50

 5 <210> 3072
 <211> 50
 <212> DNA
 <213> Homo sapiens

 10 <400> 3072
 cctgggctg agtttgctg tctgaagat tacagtttg gtagagaga 50

 <210> 3073
 <211> 50
 <212> DNA
 15 <213> Homo sapiens

 <400> 3073
 ttcggttta ccttttgct gttgtggtc ttgttctg ctggttgct 50

 20 <210> 3074
 <211> 50
 <212> DNA
 <213> Homo sapiens

 25 <400> 3074
 tgtgctaagc ctgatgaaat gtgctcctc aatctccatg aaaccatcgt 50

 <210> 3075
 <211> 50
 30 <212> DNA
 <213> Homo sapiens

 <400> 3075
 ttctgtctc catgttggtg tcaagattgc cattgcttc ctgagttca 50
 35
 <210> 3076
 <211> 50
 <212> DNA
 <213> Homo sapiens
 40
 <400> 3076
 ataacagact ccagctcctg gtccacccgg catgtcagtc agcactctgg 50

 <210> 3077
 <211> 50
 45 <212> DNA
 <213> Homo sapiens

 <400> 3077
 50 gggagccatc cctctctacc aagggtgcaa tgatggaggg aactgcatg 50

 <210> 3078
 <211> 50
 <212> DNA
 55 <213> Homo sapiens

 <400> 3078
 ttgacctccc atttttacta ttgccaata ccttttcta ggaatgtgct 50

EP 2 194 145 B1

<210> 3079
 <211> 50
 <212> DNA
 <213> Homo sapiens
 5
 <400> 3079
 ggaggcagcc agggcttacc tgtactga cttgagacca gttgaataaa 50
 10
 <210> 3080
 <211> 50
 <212> DNA
 <213> Homo sapiens
 15
 <400> 3080
 tagttagagt ccaagacatg gttcctccc cttgtctgt acatcctggc 50
 20
 <210> 3081
 <211> 50
 <212> DNA
 <213> Homo sapiens
 25
 <400> 3081
 ttgcatccc gagtttgta ttccaagaaa atcaaagggg gccaatgtt 50
 30
 <210> 3083
 <211> 463
 <212> DNA
 <213> Homo sapiens
 35
 <400> 3083
 ttggcttgac tcaggattta aaaactggaa cgggtgaaggt gacagcagtc ggttggacga 60
 gcatcccca aagttcaca tgtggccgag gactttgatt gcacattgtt gttttttaat 120
 agtcattcca aatatgagat gcattgttac aggaagtccc ttgccatcct aaaagcaccc 180
 cacttctctc taaggagaat ggcccagtc tctcccaagt ccacacaggg gagggatagc 240
 attgctttcg tgtaaattat gtaatgcaa attttttta tcttcgcctt aatctttttt 300
 attttgtttt attttgatg atgagccttc gtgcccccc ttccccctt tttcccccaa 360
 cttgagatgt atgaaggctt ttggtctccc tgggagtggg tggaggcagc cgggcttacc 420
 40
 tgtactga cttgagacca gttgaataaa agtgcacacc tta 463
 45
 <210> 3084
 <211> 491
 <212> DNA
 <213> Homo sapiens
 50
 <400> 3084
 55

EP 2 194 145 B1

	gaagagtacc agaaaagtct gctagagcag taccatctgg gtctggatca aaaacgcaga	60
	aaatatgtgg ttggagagct ctttggaat ttgccgatt tcatgactga acagtcaccg	120
5	acgagagtgc tggggaataa aaaggggatc ttcactcggc agagacaacc aaaaagtgc	180
	gcgttccttt tgcgagagag atactggaag attgccaatg aaaccaggta tccccactca	240
	gtagccaagt cacaatgttt ggaaaacagc ccgtttactt gagcaagact gataccacct	300
	gcgtgtccct tcctccccga gtcagggcga cttccacagc agcagaacaa gtgcctcctg	360
10	gactgttcac ggcagaccag aacgtttctg gcctggggtt tgtgggtcatc tattctagca	420
	gggaacacta aagggtggaaa taaaagattt tctattatgg aaataaagag ttggcatgaa	480
	agtcgctact g	491

15
 <210> 3085
 <211> 20
 <212> DNA
 <213> Homo sapiens

20
 <400> 3085
 cacaatgtgg ccgaggactt 20

25
 <210> 3086
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 3086
 tgtggccgag gactttgatt 20

30
 <210> 3087
 <211> 20
 <212> DNA
 <213> Homo sapiens

35
 <400> 3087
 tggcttttag gatggcaagg 20

40
 <210> 3088
 <211> 20
 <212> DNA
 <213> Homo sapiens

45
 <400> 3088
 gggggcttag ttgcttcct 20

50
 <210> 3089
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 3089
 aagtcagcg ttcctttgc 20

55
 <210> 3090
 <211> 20
 <212> DNA
 <213> Homo sapiens

EP 2 194 145 B1

	<400> 3090		
	agcgttcctt ttgcgagaga	20	
5	<210> 3091		
	<211> 20		
	<212> DNA		
	<213> Homo sapiens		
10	<400> 3091		
	cgggctgttt tccaaacatt	20	
	<210> 3092		
	<211> 20		
15	<212> DNA		
	<213> Homo sapiens		
	<400> 3092		
	gaagggacac gcaggtgga	20	
20	<210> 3093		
	<211> 20		
	<212> DNA		
	<213> Homo sapiens		
25	<400> 3093		
	taccacctgc gtgtcccttc	20	
	<210> 3064		
	<211> 21		
30	<212> DNA		
	<213> Homo sapiens		
	<400> 3094		
35	gaggcacttg ttctgtctct g	21	
	<210> 3096		
	<211> 327		
	<212> DNA		
40	<213> Homo sapiens		
	<400> 3096		
	ggggactctg gaggccctct tgtgtgtaac aaggtggccc agggcattgt ctcctatgga	60	
45	cgaaaacaatg gcatgcctcc acgagcctgc accaaagtct caagctttgt aacttgata	120	
	aagaaaacca tgaaacgcta ctaactacag gaagcaaact aagccccgc tgtaatgaaa	180	
	caccttctct ggagccaagt ccagatttac actgggagag gtgccagcaa ctgaataaat	240	
	acctctcca gtgtaaatct ggagccaagt ccagatttac actgggagag gtgccagcaa	300	
50	ctgaataaat acctcttagc tgagtgg	327	
	<210> 3097		
	<211> 20		
	<212> DNA		
55	<213> Homo sapiens		
	<400> 3097		
	acgagcctgc accaaagtct	20	

EP 2 194 145 B1

	<210> 3098	
	<211> 20	
	<212> DNA	
5	<213> Homo sapiens	
	<400> 3098	
	aaacaatggc atgcctccac	20
10	<210> 3099	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
15	<400> 3099	
	tcattacagc gggggcttag	20
20	<210> 3100	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 3100	
	gggggcttag ttgcttct	20
25	<210> 3101	
	<211> 5252	
	<212> DNA	
	<213> Homo sapiens	
30	<400> 3101	
35		
40		
45		
50		
55		

EP 2 194 145 B1

	ctctctccca gaacgtgtct ctgctgcaag gcaccgggccc ctttcgctct gcagaactgc	60
	acttgcaaga ccattatcaa ctccaatcc cagctcagaa agggagcctc tgcgactcat	120
5	tcacgcctt ccaggactga ctgcattgca cagatgatgg atatttacgt atgtttgaaa	180
	cgaccatcct ggatggtgga caataaaaga atgaggactg cttcaaattt ccagtggctg	240
	ttatcaacat ttattcttct atatctaata aatcaagtaa atagccagaa aaagggggct	300
	cctcatgatt tgaagtgtgt aactaacaat ttgcaagtgt ggaactgttc ttggaaagca	360
10	ccctctggaa caggccgtgg tactgattat gaagtttgca ttgaaaacag gtcccgttct	420
	tgttatcagt tggagaaaac cagtattaaa attccagctc tttcacatgg tgattatgaa	480
	ataacaataa attctctaca tgattttgga agttctacaa gtaaattcac actaaatgaa	540
	caaaacgttt ccttaattcc agatactcca gagatcttga atttgtctgc tgattttctca	600
15	acctctacat tatacctaaa gtggaacgac aggggttcag tttttccaca ccgctcaaat	660
	gttatctggg aaattaaagt tctacgtaaa gagagtatgg agctcgtaaa attagtgacc	720
	cacaacacaa ctctgaatgg caaagataca cttcatcact ggagttgggc ctcagatatg	780
	cccttggaat gtgccattca ttttgtggaa attagatgct acattgacaa tcttcatttt	840
20	tctggtctcg aagagtggag tgactggagc cctgtgaaga acatttcttg gatacctgat	900
	tctcagacta aggtttttcc tcaagataaa gtgatacttg taggctcaga cataacattt	960
	tgttgtgtga gtcaagaaaa agtgttatca gcactgattg gccatacaaa ctgccccttg	1020
25	atccatcttg atggggaaaa tgttgcaatc aagattcgta atatttctgt ttctgcaagt	1080
	agtggaaaca atgtagtttt tacaaccgaa gataacatat ttggaaccgt tatttttgc	1140
	ggatatccac cagatactcc tcaacaactg aattgtgaga cacatgattt aaaagaaatt	1200
	atatgtagtt ggaatccagg aagggtgaca gcgttggtgg gccacgtgc tacaagctac	1260
30	actttagttg aaagtttttc aggaaaatat gttagactta aaagagctga agcacctaca	1320
	aacgaaagct atcaattatt atttcaaata cttccaaatc aagaaatata taattttact	1380

35

40

45

50

55

EP 2 194 145 B1

	ttgaatgctc	acaatccgct	gggtcgatca	caatcaacaa	ttttagttaa	tataactgaa	1440
	aaagtttatc	cccatactcc	tacttcattc	aaagtgaagg	atattaattc	aacagctggt	1500
	aaactttctt	ggcatttacc	aggcaacttt	gcaaagatta	attttttatg	tgaaattgaa	1560
5	attaagaaat	ctaattcagt	acaagagcag	cggaatgtca	caatcaaagg	agtagaaaat	1620
	tcaagttatc	ttgttgctct	ggacaagtta	aatccataca	ctctatatac	ttttcggatt	1680
	cgttgttcta	ctgaaacttt	ctggaaatgg	agcaaattgga	gcaataaaaa	acaacattta	1740
10	acaacagaag	ccagtccttc	aaaggggcct	gatacttgga	gagagtggag	ttctgatgga	1800
	aaaaatttaa	taatctattg	gaagccttta	cccattaatg	aagctaattg	aaaaatactt	1860
	tcctacaatg	tatcgtgttc	atcagatgag	gaaacacagt	ccctttctga	aatccctgat	1920
	cctcagcaca	aagcagagat	acgacttgat	aagaatgact	acatcatcag	cgtagtggct	1980
15	aaaaattctg	tgggctcatc	accaccttcc	aaaatagcga	gtatggaaat	tccaaatgat	2040
	gatctcaaaa	tagaacaagt	tggtgggatg	ggaaagggga	ttctcctcac	ctggcattac	2100
	gacccaaca	tgacttgca	ctacgtcatt	aagtgggtga	actcgtctcg	gtcggaaacca	2160
	tgctttatgg	actggagaaa	agttccctca	aacagcactg	aaactgtaat	agaatctgat	2220
20	gagtttcgac	cagggtataag	atataatttt	ttctgtatg	gatgcagaaa	tcaaggatat	2280
	caattattac	gctccatgat	tggatatata	gaagaattgg	ctcccattgt	tgacacaaat	2340
	tttactgttg	aggatacttc	tgagattcgc	atattagtaa	aatgggaaga	cattcctgtg	2400
	gaagaactta	gaggcttttt	aagaggatat	ttgttttact	ttggaaaagg	agaaagagac	2460
25	acatctaaga	tgagggtttt	agaatcaggt	cgtctcgaca	taaaagttaa	gaatattact	2520
	gacatatccc	agaagacact	gagaattgct	gatcttcaag	gtaaaacaag	ttaccacctg	2580
	gtcttgcgag	cctatacaga	tggtggagtg	ggcccgagga	agagtatgta	tgtggtgaca	2640
	aaggaaaatt	ctgtgggatt	aattattgcc	attctcatcc	cagtggcagt	ggctgtcatt	2700
30	gttggagtgg	tgacaagtat	cctttgctat	cggaaacgag	aatggattaa	agaaaccttc	2760
	taccttgata	ttccaaatcc	agaaaactgt	aaagcattac	agtttcaaaa	gagtgtctgt	2820
	gaggggaagca	gtgctcttaa	aacattggaa	atgaatcctt	gtaccccaaa	taatgttgag	2880
	gttctggaaa	ctcgatcagc	atttcctaaa	atagaagata	cagaaataat	ttccccagta	2940
35	gctgagcgtc	ctgaagatcg	ctctgatgca	gagcctgaaa	accatgtggg	tgtgtcctat	3000
	tgtccaccca	tcattgagga	agaaatacca	aaccagccg	cagatgaagc	tgaggggact	3060
	gcacagggtta	tttacattga	tgttcagtcg	atgtatcagc	ctcaagcaaa	accagaagaa	3120
	gaacaagaaa	atgacctgt	aggaggggca	ggctataagc	cacagatgca	cctccccatt	3180
40	aattctactg	tggaagatat	agctgcagaa	gaggacttag	ataaaactgc	gggttacaga	3240
	cctcaggcca	atgtaaatat	atggaattta	gtgtctccag	actctcctag	atccatagac	3300
	agcaacagtg	agattgtctc	atgtggaagt	ccatgtctca	ttaattcccg	acaatttttg	3360
45	attcctccta	aagatgaaga	ctctcctaaa	tctaattggag	gaggggtggc	ctttacaaac	3420
	ttttttcaga	acaaacaaaa	cgattaacag	tgtcacctg	tcacttcagt	cagccatctc	3480
	aataagctct	tactgctagt	gttgctacat	cagcactggg	cattcttgga	gggatcctgt	3540
	gaagtattgt	taggaggtga	acttcactac	atgttaagtt	acactgaaag	ttcatgtgct	3600
50	tttaatgtag	tctaaaagcc	aaagtatagt	gactcagaat	cctcaatcca	caaaactcaa	3660
	gattgggagc	tctttgtgat	caagccaaaag	aattctcatg	tactctacct	tcaagaagca	3720
	tttcaaggct	aatacctact	tgtacgtaca	tgtaaaacaa	atcccgccgc	aactgttttc	3780
	tgttctgttg	tttgtggttt	tctcatatgt	atacttggtg	gaattgtaag	tggaattgca	3840
55	ggccaggggag	aaaatgtcca	agtaacaggt	gaagtttatt	tgcttgacgt	ttactccttt	3900
	ctagatgaaa	accaagcaca	gattttaaaa	cttctaagat	tattctcctc	tatccacagc	3960

EP 2 194 145 B1

	attcacaaaa	attaatataa	tttttaatgt	agtgacagcg	atttagtggt	ttgtttgata	4020
	aagtatgctt	atttctgtgc	ctactgtata	atgggttatca	aacagttgtc	tcaggggtac	4080
5	aaactttgaa	aacaagtgtg	acactgacca	gcccaaatca	taatcatggt	ttcttgctgt	4140
	gataggtttt	gcttgccttt	tcattatttt	ttagctttta	tgcttgcttc	cattatttca	4200
	gttggttgcc	ctaataattt	aaatttacac	ttctaagact	agagaccac	attttttaa	4260
	aatcatttta	ttttgtgata	cagtgcagc	tttatatgag	caaattcaat	attattcata	4320
10	agcatgtaat	tccagtgcct	tactatgtga	gatgactact	aagcaatata	tagcagcgtt	4380
	agttccatat	agttctgatt	ggatttcgtt	cctcctgagg	agaccatgcc	gttgagcttg	4440
	gctaccagg	cagtgggtgat	ctttgcacac	ttctggtgga	tgctcctccc	actcatgagt	4500
	cttttcatca	tgccacatta	tctgatccag	tcctcacatt	tttaaatata	aaactaaaga	4560
15	gagaatgctt	cttacaggaa	cagttacca	agggtgtttt	cttagtaact	gtcataaact	4620
	gatctggatc	catgggcata	cctgtgttcg	agggtgcagca	attgcttggt	gagctgtgca	4680
	gaattgattg	ccttcagcac	agcatcctct	gcccaccctt	gtttctcata	agcgtgtctt	4740
	ggagtgattg	tggttcttgg	aaaagcagaa	ggaaaaacta	aaaagtgtat	cttgatattt	4800
20	ccctgccctc	aggttgccta	tgtattttac	cttttcatat	ttaaggcaaa	agtacttgaa	4860
	aattttaagt	gtccgaataa	gatagtcttt	ttttgtttgt	tttttttgtt	tggttggttg	4920
	ttttttatca	tctgagattc	tgtaatgtat	ttgcaaataa	tggaatcaat	aatttttttt	4980
25	gaagctcata	ttgtatcttt	ttaaaaacca	gtttgtggaa	aaaagccaga	gtgacaagtg	5040
	acaaaatcta	tttaggaact	ctgtgtatga	atcctgattt	taactgctag	gattcagcta	5100
	aatttctgag	ctttatgata	tgtggaaatt	tggaatgaaa	tcgaattcat	ttgtacata	5160
	catagtatat	taaaactata	taatagttca	tagaaatgtt	cagtaatgaa	aaaatatatc	5220
30	caatcagagc	catcccgaag	aaaaaaaaaa	aa			5252

<210> 3012

<211> 5252

<212> DNA

35 <213> Homo sapiens

<220>

<221> misc_feature

<222> (3967)..(3988)

40 <223> n is a, c, g, t or u

<400> 3102

45

50

55

EP 2 194 145 B1

	ctctctccca gaacgtgtct ctgctgcaag gcaccgggcc ctttcgctct gcagaactgc	60
	acttgcaaga ccattatcaa ctccaatcc cagctcagaa agggagcctc tgcgactcat	120
	tcatcgccct ccaggactga ctgcattgca cagatgatgg atatttacgt atgtttgaaa	180
5	cgaccatcct ggatggtgga caataaaaga atgaggactg cttcaaattt ccagtggctg	240
	ttatcaacat ttattcttct atatctaataa aatcaagtaa atagccagaa aaagggggct	300
	cctcatgatt tgaagtgtgt aactaacaat ttgcaagtgt ggaactgttc ttggaaagca	360
10	ccctctggaa caggccgtgg tactgattat gaagtttgca ttgaaaacag gtcccgttct	420
	tgttatcagt tggagaaaac cagtattaaa attccagctc ttccacatgg tgattatgaa	480
	ataacaataa attctctaca tgattttgga agttctacaa gtaaattcac actaaatgaa	540
	caaaacgttt ccttaattcc agatactcca gagatcctga atttgtctgc tgatttctca	600
15	acctctacat tatacctaaa gtggaacgac aggggttcag tttttccaca ccgctcaaat	660
	gttatctggg aaattaaagt tctacgtaaa gagagtatgg agctcgtaaa attagtgacc	720
	cacaacacaa ctctgaatgg caaagataca cttcatcact ggagttgggc ctcagatatg	780
	cccttggaaat gtgccattca ttttgtggaa attagatgct acattgacaa tcttcatttt	840
20	tctggtctcg aagagtggag tgactggagc cctgtgaaga acatttcttg gatacctgat	900

25

30

35

40

45

50

55

EP 2 194 145 B1

	tctcagacta aggtttttcc tcaagataaa gtgatacttg taggctcaga cataacattt	960
	tggtgtgtga gtcaagaaaa agtggttatca gacttgattg gccatacaaa ctgccccttg	1020
	atccatcttg atggggaaaa tggtgcaatc aagattcgta atatttctgt ttctgcaagt	1080
5	agtggaacaa atgtagtttt tacaaccgaa gataacatat ttggaaccgt tatttttgct	1140
	ggatatccac cagatactcc tcaacaactg aattgtgaga cacatgattt aaaagaaatt	1200
	atatgtagtt ggaatccagg aagggtgaca gcgttggtgg gccacgtgc tacaagctac	1260
10	actttagttg aaagtttttc aggaaaatat gttagactta aaagagctga agcacctaca	1320
	aacgaaagct atcaattatt atttcaaatg cttccaaatc aagaaatata taattttact	1380
	ttgaatgctc acaatccgct gggtcgatca caatcaacaa ttttagttaa tataactgaa	1440
	aaagtttatc cccatactcc tacttcattc aaagtgaagg atattaattc aacagctggt	1500
15	aaactttctt ggcatttacc aggcaacttt gcaagatta attttttatg tgaaattgaa	1560
	attaagaaat ctaattcagt acaagagcag cggaatgtca caatcaaagg agtagaaaat	1620
	tcaagttatc ttgttgctct ggacaagtta aatccataca ctctatatac ttttcggatt	1680
	cggtgttcta ctgaaacttt ctggaaatgg agcaaattgga gcaataaaaa acaacattta	1740
20	acaacagaag ccagtccttc aaaggggcct gatacttgga gagagtggag ttctgatgga	1800
	aaaaatttaa taatctattg gaagccttta cccattaatg aagctaattg aaaaatactt	1860
	tcctacaatg tatcgtgttc atcagatgag gaaacacagt ccctttctga aatccctgat	1920
	cctcagcaca aagcagagat acgacttgat aagaatgact acatcatcag cgtagtggtc	1980
25	aaaaattctg tgggtctatc accaccttcc aaaatagcga gtatggaaat tccaaatgat	2040
	gatctcaaaa tagaacaagt tggtgggatg ggaaagggga ttctcctcac ctggcattac	2100
	gacccaaca tgacttgca ctacgtcatt aagtgggtga actcgtctcg gtcggaacca	2160
	tgctttatgg actggagaaa agttccctca aacagcactg aaactgtaat agaactgat	2220
30	gagtttcgac caggataag atataatttt ttctgtatg gatgcagaaa tcaaggatat	2280
	caattattac gctccatgat tggatatata gaagaattgg ctcccattgt tgcaccaa	2340
	tttactgttg aggatacttc tgcagattcg atattagtaa aatgggaaga cattcctgtg	2400
	gaagaactta gaggtctttt aagaggatat ttgttttact ttggaaaagg agaaagagac	2460
35	acatctaaga tgagggtttt agaatcaggt cgtcttgaca taaaagttaa gaatattact	2520
	gacatatccc agaagacact gagaattgct gatcttcaag gtaaaacaag ttaccacctg	2580
	gtcttgagag cctatacaga tgggtggagt ggcccgagaga agagtatgta tgtggtgaca	2640
	aaggaaaatt ctgtgggatt aattattgcc attctcatcc cagtggcagt ggctgtcatt	2700
40	gttgagtggtg tgacaagtat cctttgctat cggaaacgag aatggattaa agaaaccttc	2760
	taccctgata ttccaaatcc agaaaactgt aaagcattac agtttcaaaa gagtgtctgt	2820
	gaggggaagca gtgctcttaa aacattggaa atgaatcctt gtaccccaaa taatgttgag	2880
45	gttctggaaa ctcgatcagc atttcctaaa atagaagata cagaaataat ttccccagta	2940
	gctgagcgtc ctgaagatcg ctctgatgca gagcctgaaa accatgtggt tgtgtcctat	3000
	tgtccacca tcattgagga agaaatacca aaccagccg cagatgaagc tggagggact	3060
	gcacagggtta ttacattga tgttcagtcg atgtatcagc ctcaagcaaa accagaagaa	3120
50	gaacaagaaa atgacctgt aggaggggca ggctataagc cacagatgca cctccccatt	3180
	aattctactg tggaagatat agctgcagaa gaggacttag ataaaactgc gggttacaga	3240
	cctcaggcca atgtaataac atggaattta gtgtctccag actctcctag atccatagac	3300
	agcaacagtg agattgtctc atttggaggt ccatgctcca ttaattcccg acaatttttg	3360
55	attcctccta aagatgaaga ctctcctaaa tctaattggag gaggggtggtc ctttacaac	3420
	ttttttcaga acaaaccaaa cgattaacag tgtcaccgtg tcacttcagt cagccatctc	3480

EP 2 194 145 B1

	aataagctct tactgctagt gttgctacat cagcactggg cattcttgga gggatcctgt	3540
	gaagtattgt taggaggtga acttcactac atgttaagtt aactgaaaag ttcattgtgt	3600
5	tttaattgtag tctaaaagcc aaagtatagt gactcagaat cctcaatcca caaaactcaa	3660
	gattgggagc tctttgtgat caagccaaag aattctcatg tactctacct tcaagaagca	3720
	tttcaaggct aatacctact tgtacgtaca tgtaaaacaa atcccgccgc aactgttttc	3780
	tgttctgttg tttgtggttt tctcatatgt atacttggtg gaattgtaag tggatttgca	3840
10	ggccagggag aaaatgtcca agtaacaggt gaagtttatt tgcctgacgt ttactccttt	3900
	ctagatgaaa accaagcaca gattttaaaa ctcttaagat tattctcctc tatccacagc	3960
	attcacnnnn nnnnnnnnnn nnnnnnnngt agtgacagcg atttagtggt ttgtttgata	4020
	aagtatgctt atttctgtgc ctactgtata atgggttatca aacagttgtc tcaggggtac	4080
15	aaactttgaa aacaagtgtg aactgacca gcccaaatca taatcatgtt ttcttgctgt	4140
	gataggtttt gcttgccctt tcattatttt ttagctttta tgcttgcttc cattatttca	4200
	gttggttgcc ctaatattta aaatttacac ttctaagact agagaccac attttttaa	4260
20	aatcatttta ttttgtgata cagtgcagc tttatatgag caaattcaat attattcata	4320
	agcatgtaat tccagtgact tactatgtga gatgactact aagcaatatc tagcagcgtt	4380
	agttccatat agttctgatt ggatttcgtt cctcctgagg agaccatgcc gttgagcttg	4440
	gctaccagc cagtgggtgat ctttgacacc ttctgggtgga tgctcctccc actcatgagt	4500
25	cttttcatca tgccacatta tctgatccag tcctcacatt tttaaatata aaactaaaga	4560
	gagaatgctt cttacaggaa cagttacca agggctgttt cttagtaact gtcataaact	4620
	gatctggatc catgggcata cctgtgttcg aggtgcagca attgcttggg gagctgtgca	4680
	gaattgattg ccttcagcac agcatcctc gccaccctt gtttctcata agcgatgtct	4740
30	ggagtgattg tggttcttgg aaaagcagaa ggaataaact aaaagtgtat cttgtatttt	4800
	ccctgccctc aggttgccca tgtattttac cttttcatat ttaaggcaaa agtacttgaa	4860
	aattttaagt gtccgaataa gatatgtctt ttttgtttgt tttttttggt tgggtgtttg	4920
	ttttttatca tctgagattc tgtaattgtat ttgcaaataa tggatcaatt aatttttttt	4980
35	gaagctcata ttgtatcttt ttaaaaacca tgttgtggaa aaaagccaga gtgacaagtg	5040
	acaaaatcta tttaggaact ctgtgtatga atcctgattt taactgctag gattcagcta	5100
	aatttctgag ctttatgatc tgtggaaatt tggaatgaaa tcgaattcat tttgtacata	5160
40	catagtatat taaaactata taatagttca tagaaatgtt cagtaatgaa aaaatatatc	5220
	caatcagagc catcccgaaa aaaaaaaaaa aa	5252

<210> 3103

<211> 841

45 <212> DNA

<213> Homo sapiens

<400> 3103

50

55

EP 2 194 145 B1

	tttttttttt ttttcttaaa tagcatttat tttctctcaa aaagcctatt atgtactaac	60
	aagtgttcct ctaaattaga aaggcatcac tactaaaatt ttatacatat tttttatata	120
	agagaaggaa tattgggtta caatctgaat ttctctttat gatttctctt aaagtataga	180
5	acagctatta aaatgactaa tattgctaaa atgaaggcta ctaaatttcc ccaagaattt	240
	cggtggaatg cccaaaaatg gtgttaagat atgcagaagg gcccathttca agcaaagcaa	300
	tctctccacc ccttcataaa agatttaagc taaaaaaaaa aaaaaaagaa gaaaatccaa	360
10	cagctgaaga cattgggcta tttataaatc ttctcccagt ccccagaca gcctcacatg	420
	ggggctgtaa acagctaact aaaatatctt tgagactctt atgtccacac cactgacac	480
	aaggagagct gtaaccacag tgaaactaga ctttgctttc ctttagcaag tatgtgccta	540
15	tgatagtaaa ctggagtaaa tgtaacagta ataaaacaaa ttttttttaa aaataaaaat	600
	tatacctttt tctccaacaa acggtaaaga ccacgtgaag acatccataa aattaggcaa	660
	ccagtaaaga tgtggagaac cagtaaactg tcgaaattca tcacattatt ttcatacttt	720
	aatacagcag ctttaattat tggagaacat caaagtaatt aggtgccgaa aaacattggt	780
20	attaatgaag ggaacccctg acgtttgacc ttttctgtac catctatagc cctggacttg	840
	a	841

	<210> 3104
	<211> 841
25	<212> DNA
	<213> Homo sapiens
	<220>
	<221> misc_feature
30	<222> (94)..(121)
	<223> n is a, c, g, t or u
	<220>
	<221> misc_feature
35	<222> (569)..(604)
	<223> n is a, c, g, t or u
	<400> 3104

40

45

50

55

EP 2 194 145 B1

	tttttttttt ttttcttaaa tagcatttat tttctctcaa aaagcctatt atgtactaac	60
	aagtgttctt ctaaattaga aaggcatcac tacnnnnnnn nnnnnnnnnn nnnnnnnnnn	120
5	ngagaaggaa tattgggtta caatctgaat tttcttttat gatttctctt aaagtataga	180
	acagctatta aaatgactaa tattgctaaa atgaaggcta ctaaatttcc ccaagaattt	240
	cggtggaatg cccaaaaatg gtgttaagat atgcagaagg gcccatattca agcaaagcaa	300
	tctctccacc ccttcataaa agatttaagc taaaaaaaaa aaaaaaagaa gaaaatccaa	360
10	cagctgaaga cattgggcta ttataaatc ttctcccagt ccccagaca gcctcacatg	420
	ggggctgtaa acagctaact aaaatatctt tgagactctt atgtccacac ccactgacac	480
	aaggagagct gtaaccacag tgaaactaga ctttgctttc ctttagcaag tatgtgccta	540
	tgatagtaaa ctggagtaaa tgtaacagnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	600
15	nnnncctttt tctccaacaa acggtaaaaga ccacgtgaag acatccataa aattaggcaa	660
	ccagtaaaga tgtggagaac cagtaaaactg tcgaaattca tcacattatt ttcatacttt	720
	aatacagcag ctttaattat tggagaacat caaagtaatt aggtgccgaa aaacattggt	780
	attaatgaag ggaacccctg acgtttgacc ttttctgtac catctatagc cctggacttg	840
20	a	841

<210> 3105
 <211> 63
 <212> DNA
 <213> Homo sapiens

<400> 3105
 ggccagtga ttgtaatacg actcactata gggaggcggg tttttttt tttttttt 60

ttt 63

<210> 3106
 <211> 609
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (303)..(304)
 <223> n is a, c, g, t or u

<400> 3106
 acatcagtgg ctacatgtga gctcagacct gggctgctg ctgtctgtct tccaatatc 60

	catgaccttg actgatgcag gtgtctaggg atacagggtca cacagccgct catccccgtc	120
	ctgtcggagc ccagagcacg gaagcctggc cctccgagga gacagaaggg agtgctggac	180
	accatgacga gagcttgga gaataaataa cttcttttaa caattttacg gcatgaagaa	240
50	atctggacca gtttattaaa tgggatttct gccacaaacc ttggaagaat cacatcatct	300
	tanncccaag tgaaaactgt gttgcgtaac aaagaacatg actgcgctcc acacatacat	360
	cattgcccgg cgaggcggga cacaagtcaa cgacggaaca cttgagacag gcctacaact	420
	gtgcacgggt cagaagcaag ttttaagccat acttgctgca gtgagactac atttctgtct	480
55	atagaagata cctgacttga tctgtttttc agctccagtt cccagatgtg cgtgtgtgtg	540
	tccccaagta tcaccttcca atttctggga gcagtgtctt ggccggatcc ttgccgcgcg	600
	gataaaaac	609

Claims

1. Use of a nucleic acid for assessing cardiac allograft rejection in a patient, wherein the expression level of said nucleic acid in said patient is detected in a blood sample from said patient to assess cardiac allograft rejection versus non-rejection, wherein said nucleic acid comprises the nucleotide sequence SEQ ID NO: 52 and wherein said expression level is detected by measuring the RNA level expressed by said nucleic acid.
2. The use according to claim 1, wherein the expression level is detected in peripheral blood leukocytes.
3. The use according to any one of claims 1-2, further including isolating RNA from said blood sample from said patient prior to detecting said RNA level expressed by said nucleic acid.
4. The use according to claim 3, wherein said RNA level is detected by PCR.
5. The use according to claim 3, wherein said RNA level is detected by hybridization.
6. The use according to claim 3, wherein said RNA level is detected by hybridization to an oligonucleotide.
7. The use according to claim 6, wherein said oligonucleotide comprises DNA, RNA, cDNA, PNA, genomic DNA, or synthetic oligonucleotides.

Patentansprüche

1. Verwendung einer Nucleinsäure zum Beurteilen von Abstoßung eines allogenen Herztransplantats in einem Patienten, wobei die Expressionsstärke der Nucleinsäure im Patienten in einer Blutprobe vom Patienten nachgewiesen wird, um Abstoßung gegenüber Nichtabstoßung des allogenen Herztransplantats zu beurteilen, wobei die Nucleinsäure die Nucleotidsequenz SEQ ID NO: 52 umfasst, und wobei die Expressionsstärke durch Messen der RNA-Konzentration nachgewiesen wird, die durch die Nucleinsäure exprimiert wird.
2. Verwendung nach Anspruch 1, wobei die Expressionsstärke in Leukozyten des peripheren Blutes nachgewiesen wird.
3. Verwendung nach Anspruch 1 bis 2, ferner umfassend ein Isolieren von RNA aus der Blutprobe vom Patienten vor dem Nachweisen der RNA-Konzentration, die durch die Nucleinsäure exprimiert wird.
4. Verwendung nach Anspruch 3, wobei die RNA-Konzentration durch PCR nachgewiesen wird.
5. Verwendung nach Anspruch 3, wobei die RNA-Konzentration durch Hybridisierung nachgewiesen wird.
6. Verwendung nach Anspruch 3, wobei die RNA-Konzentration durch Hybridisierung mit einem Oligonucleotid nachgewiesen wird.
7. Verwendung nach Anspruch 6, wobei das Oligonucleotid DNA, RNA, cDNA, PNA, genomische DNA oder synthetische Oligonucleotide umfasst.

Revendications

1. Utilisation d'un acide nucléique pour évaluer un rejet d'allogreffe cardiaque chez un patient, dans laquelle le niveau d'expression dudit acide nucléique chez ledit patient étant détecté dans un échantillon sanguin provenant dudit patient pour évaluer un rejet d'allogreffe cardiaque versus un non-rejet, dans laquelle ledit acide nucléique comprend la séquence nucléotide SEQ ID NO: 52 et dans laquelle ledit niveau d'expression étant détecté par mesure du niveau d'ARN exprimé par ledit acide nucléique.
2. Utilisation selon la revendication 1, dans laquelle le niveau d'expression est détecté dans des leucocytes du sang périphérique.

EP 2 194 145 B1

3. Utilisation selon l'une quelconque des revendications 1 et 2, comprenant en outre l'isolement d'ARN dudit échantillon sanguin provenant dudit patient avant de détecter ledit niveau d'ARN exprimé par ledit acide nucléique.
4. Utilisation selon la revendication 3, dans laquelle ledit niveau d'ARN est détecté par PCR.
5. Utilisation selon la revendication 3, dans laquelle ledit niveau d'ARN est détecté par hybridation.
6. Utilisation selon la revendication 3, dans laquelle ledit niveau d'ARN est détecté par hybridation à un oligonucléotide.
7. Utilisation selon la revendication 6, dans laquelle ledit oligonucléotide comprend l'ADN, l'ARN, l'ADNc, l'ANP, l'ADN génomique ou des oligonucléotides synthétiques.

Figure 1: Novel Gene Sequence Analysis

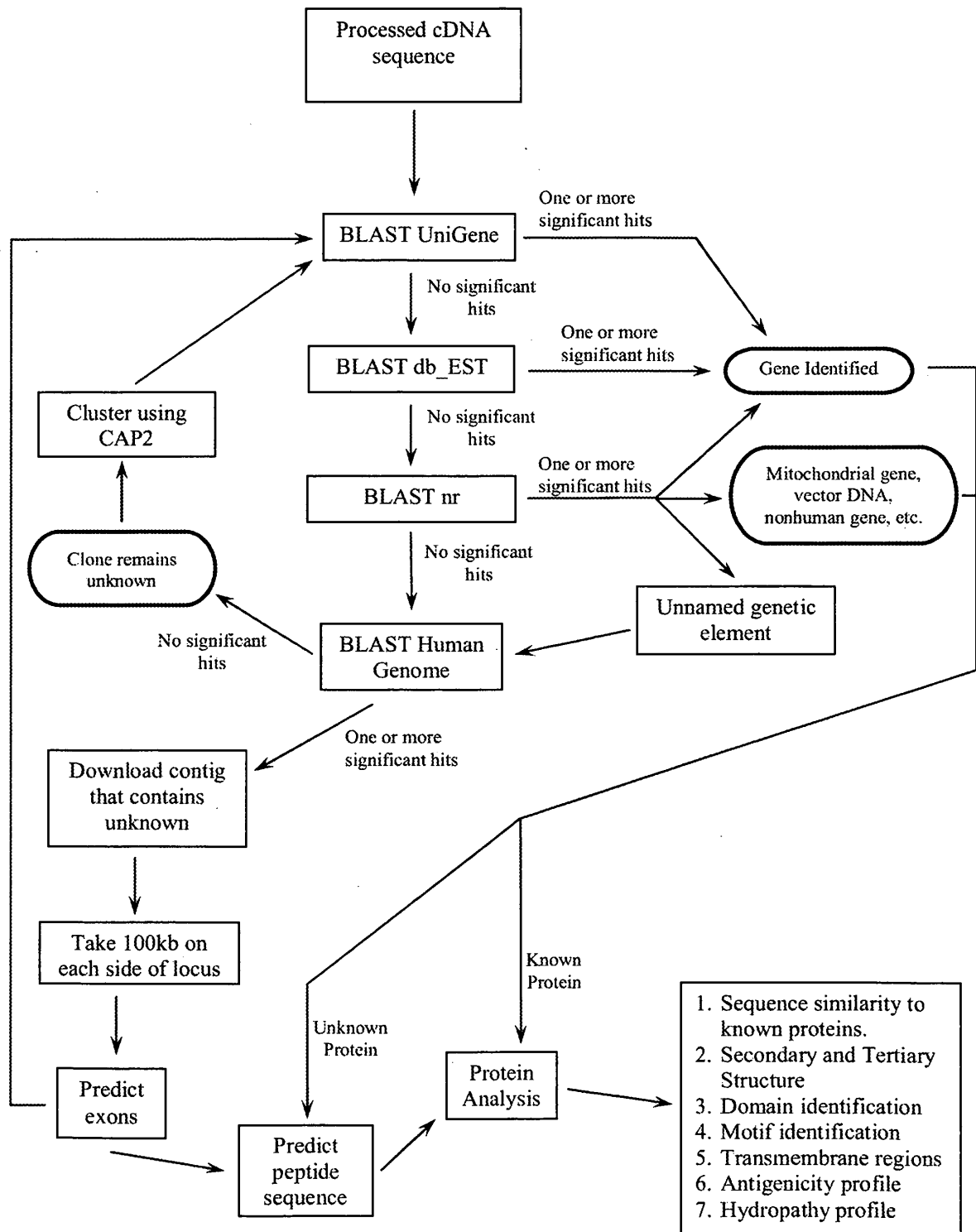


Figure 2. Automated Mononuclear Cell RNA Isolation Device

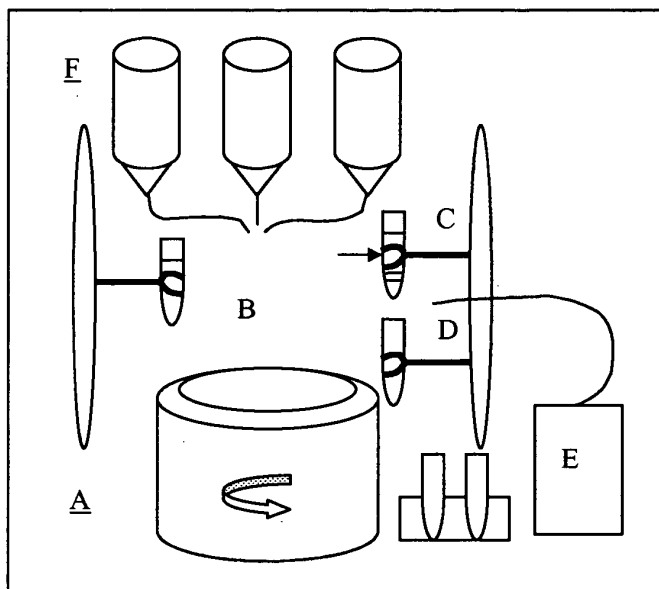


Figure 3

**Comparison of Guanine-Silica (GS) to
Acid-Phenol (AP) RNA Purification**

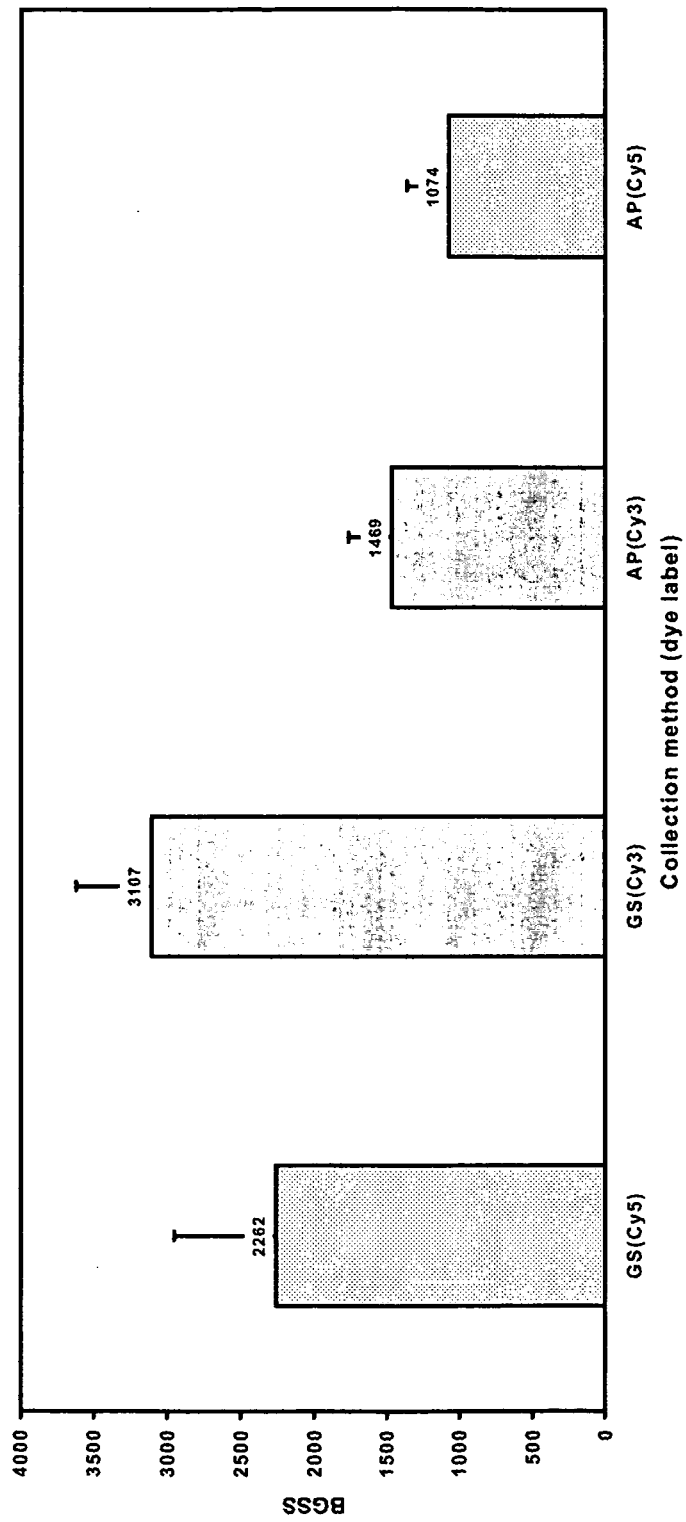


Figure 4
Expression of Leukocyte Specific Genes

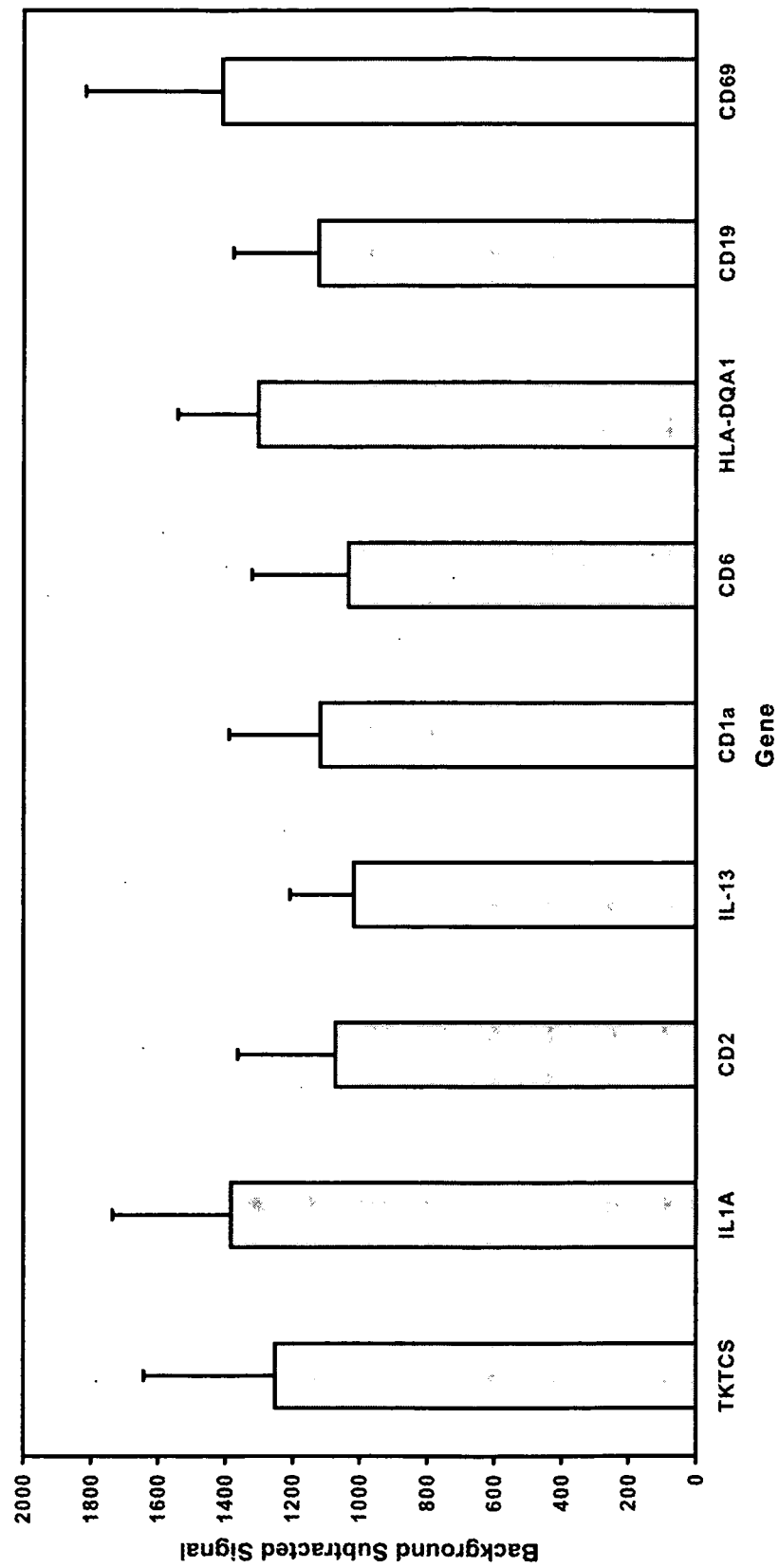


Figure 5
Expression of Leukocyte-Specific Genes

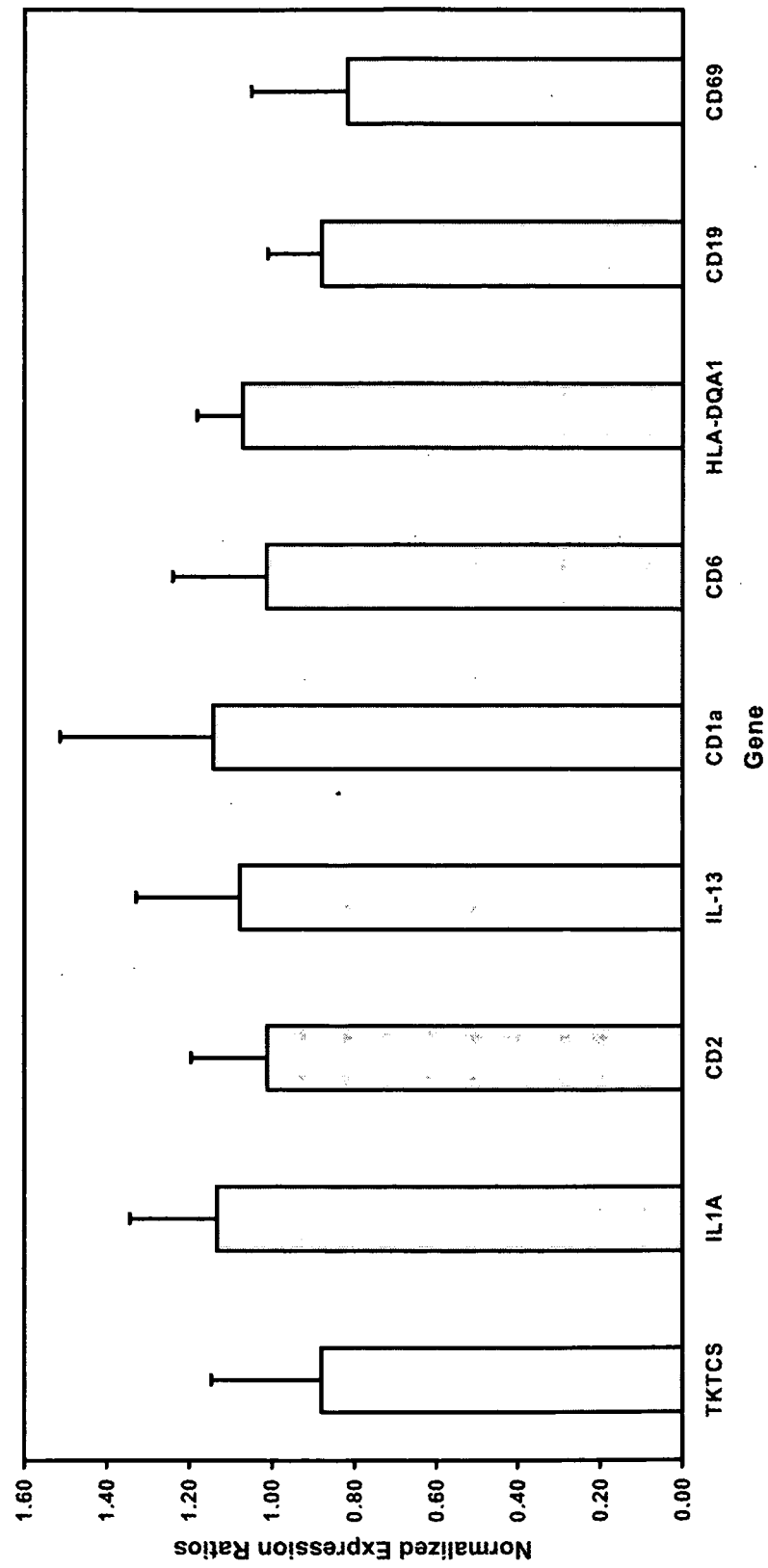


FIGURE 6

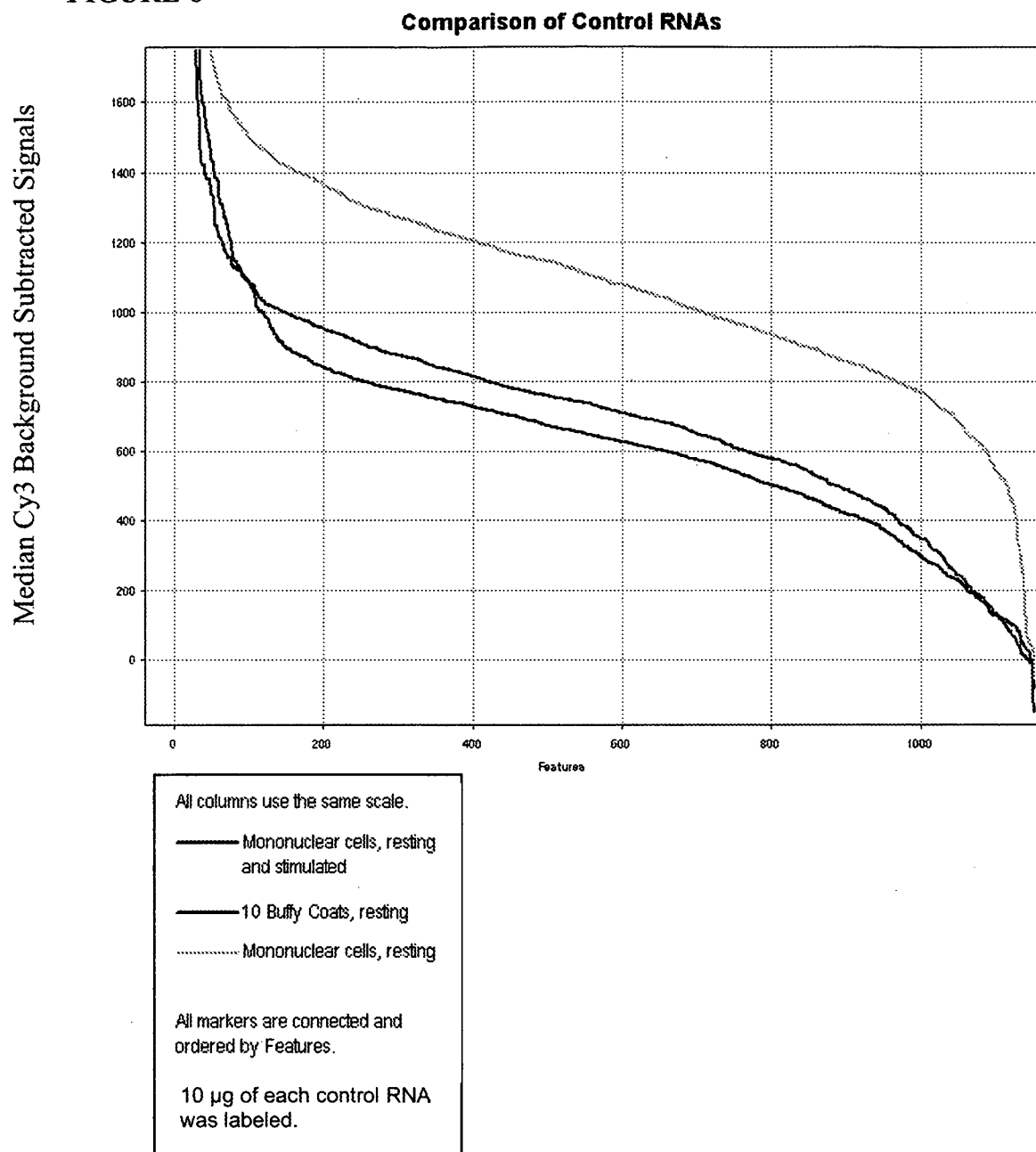
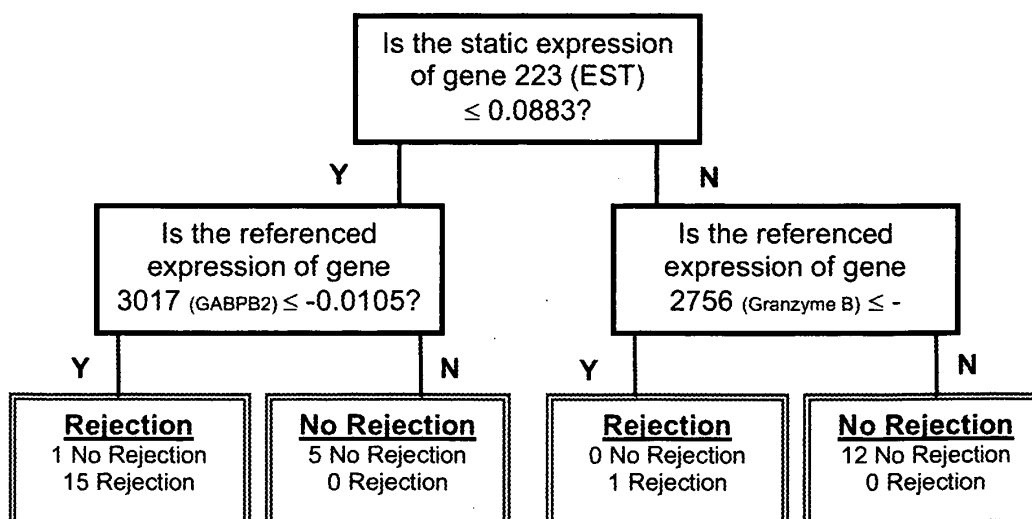


Figure 7: Cardiac Allograft rejection diagnostic genes.

A.

Sample	Grade	Marker Gene Expression Ratios				
		3020	3019	2760	3018	85
12-0025-02	0	3.90	3.69	5.49	3.24	3.34
12-0024-04	0	3.66	4.05	5.89	3.75	3.03
15-0024-01	0	3.55	4.01	5.61	2.90	3.23
12-0029-03	0	3.44	3.12	4.25	3.55	3.07
12-0024-03	0	2.88	2.54	2.56	2.20	2.38
14-0021-05	0	1.31	1.03	1.07	0.91	0.99
14-0005-06	3A	0.42	0.27	0.51	0.22	0.26
14-0012-07	3A	0.60	0.62	0.70	0.42	0.61
14-0001-06	3A	0.93	0.71	0.58	0.37	0.44
14-0009-01	3A	0.71	0.63	0.68	0.61	0.66
12-0012-02	3A	0.86	0.85	0.73	0.41	0.72
12-0001-01	3A	1.08	0.97	1.01	0.40	1.06
Average Grade 0:		3.13	3.07	4.14	2.76	2.67
Average Grade 3A:		0.77	0.68	0.70	0.40	0.62
Fold Difference:		4.08	4.55	5.91	6.82	4.28

B. CART classification model.



C. Surrogates for the CART classification model.

Primary Splitter	static 223	ref 3017	ref 4
Surrogate 1	ref 167	ref 102	ref 2761
Surrogate 2	ref 3016	static 36	ref 2762
Surrogate 3	ref 1760	ref 2764	ref 3016
Surrogate 4	ref 85	ref 2759	ref 2757
Surrogate 5	ref 2763	ref 2761	ref 2758

Figure 8A: Validation of differential expression of Granzyme B in CMV patients using Real-time PCR

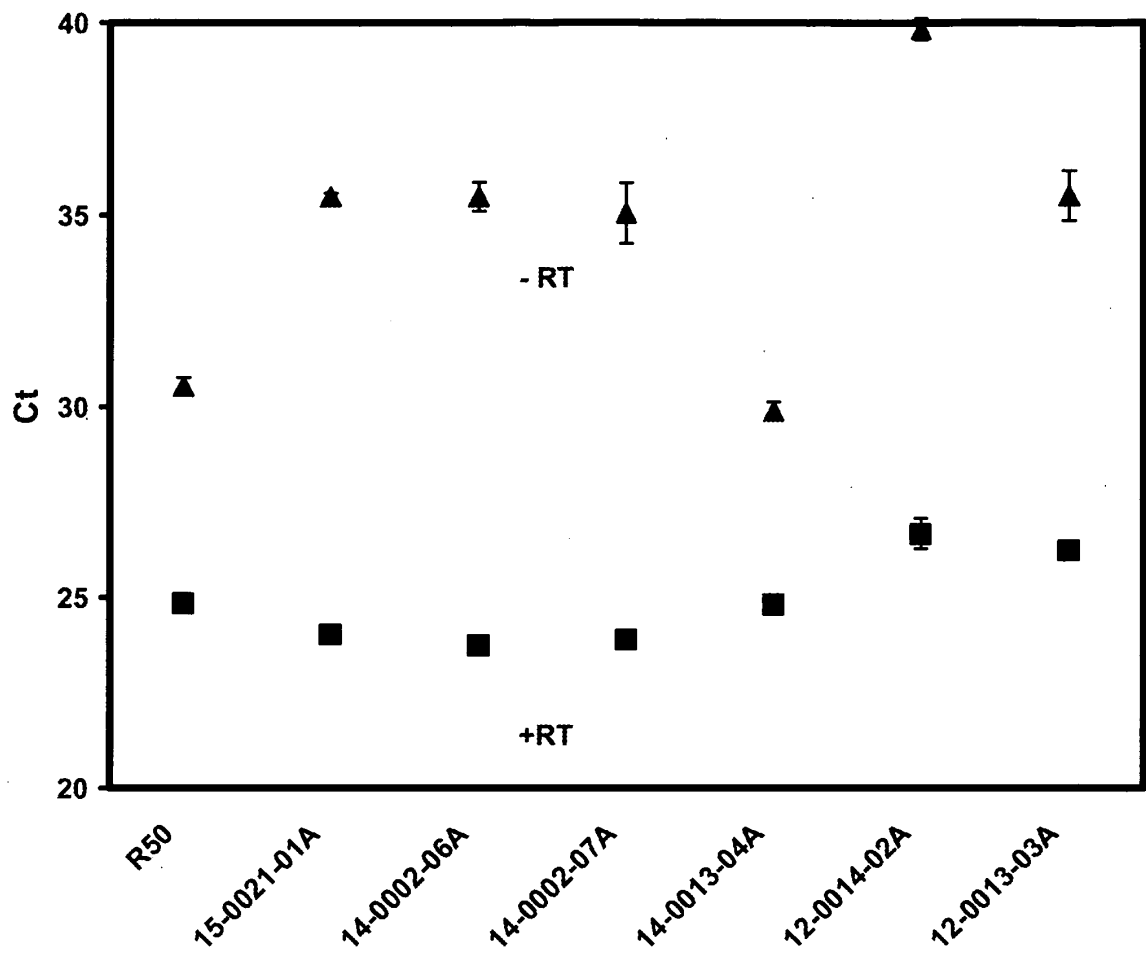


Figure 8B.

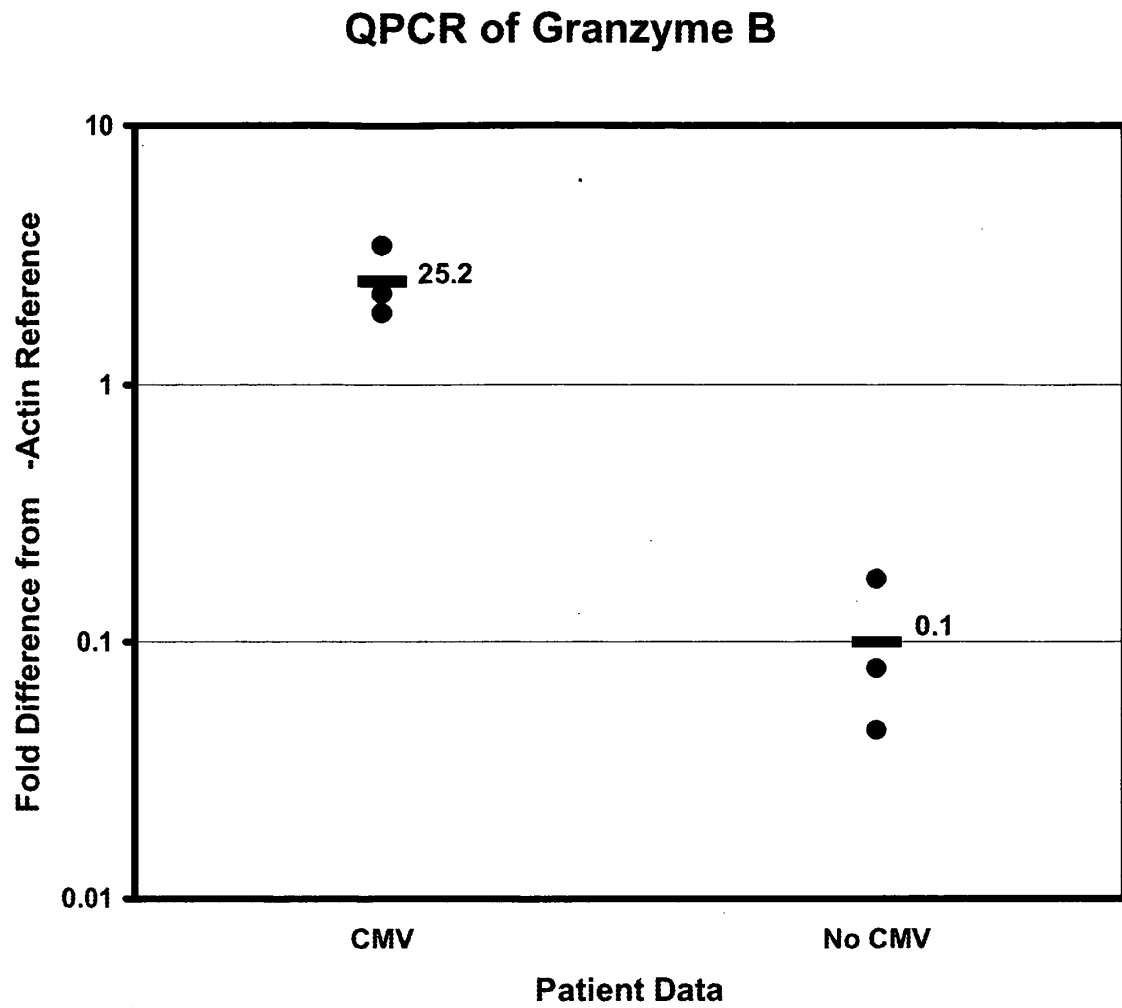
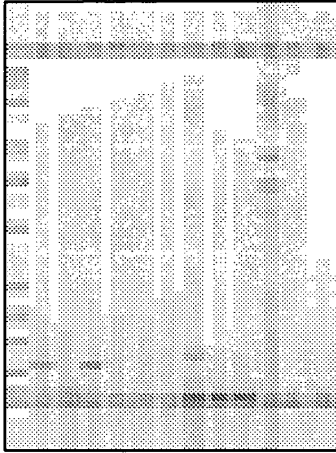


Figure 9

A.



B.

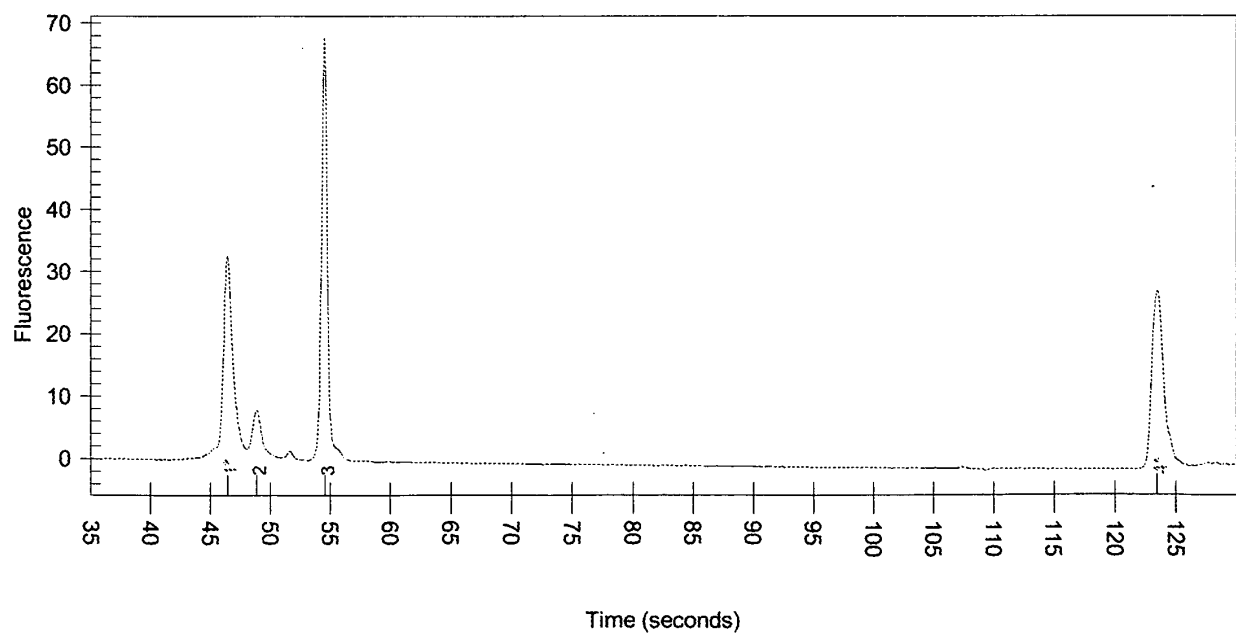
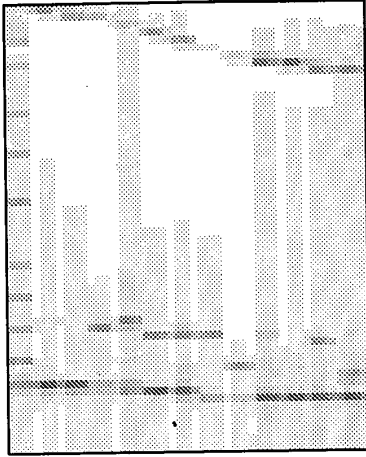


Figure 9

C.



D.

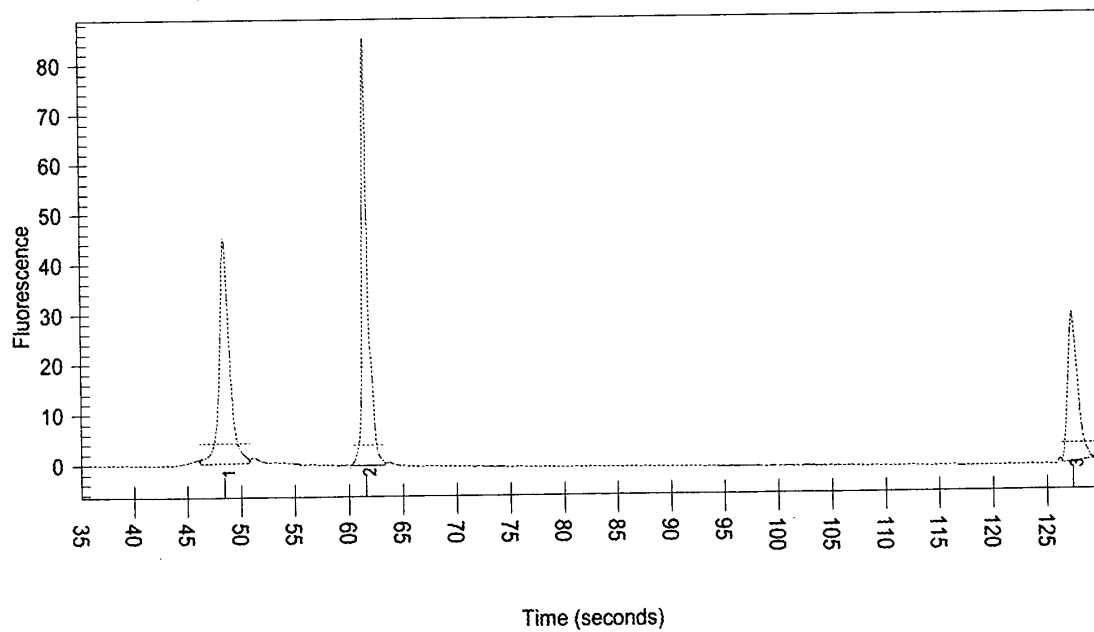


Figure 10

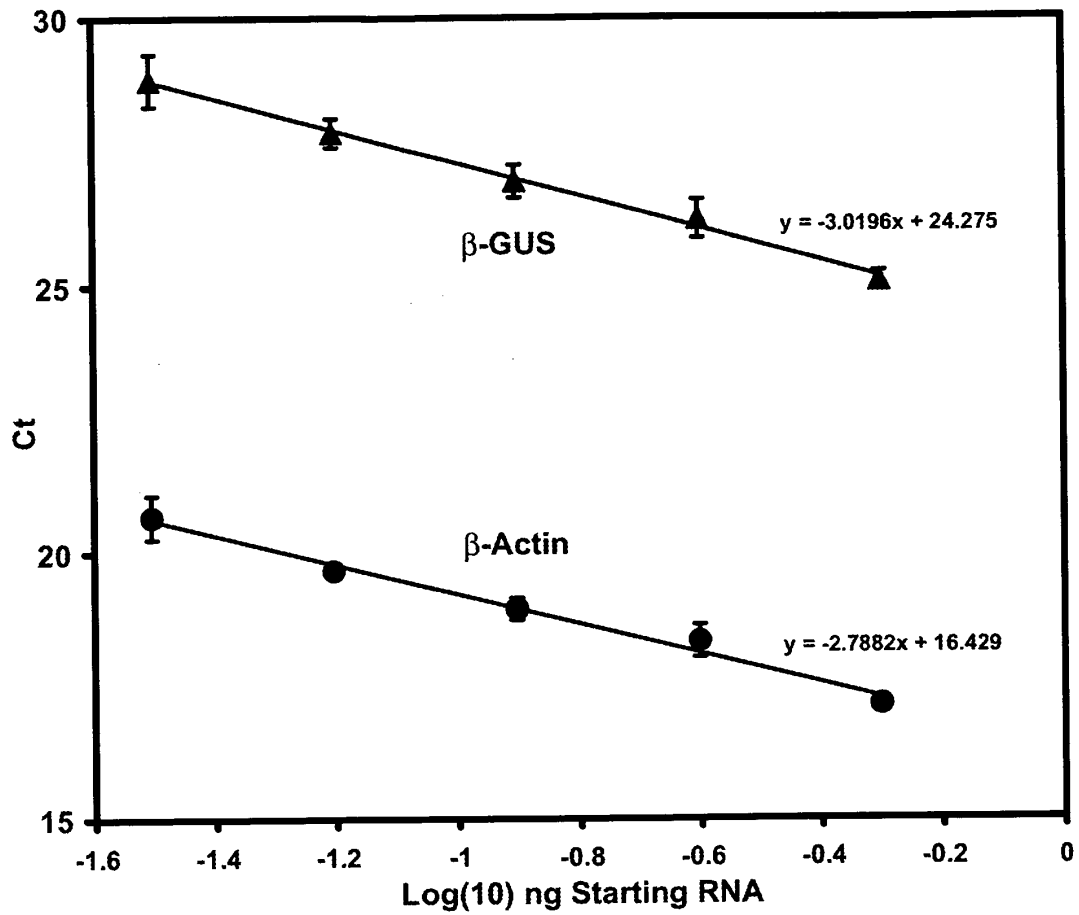
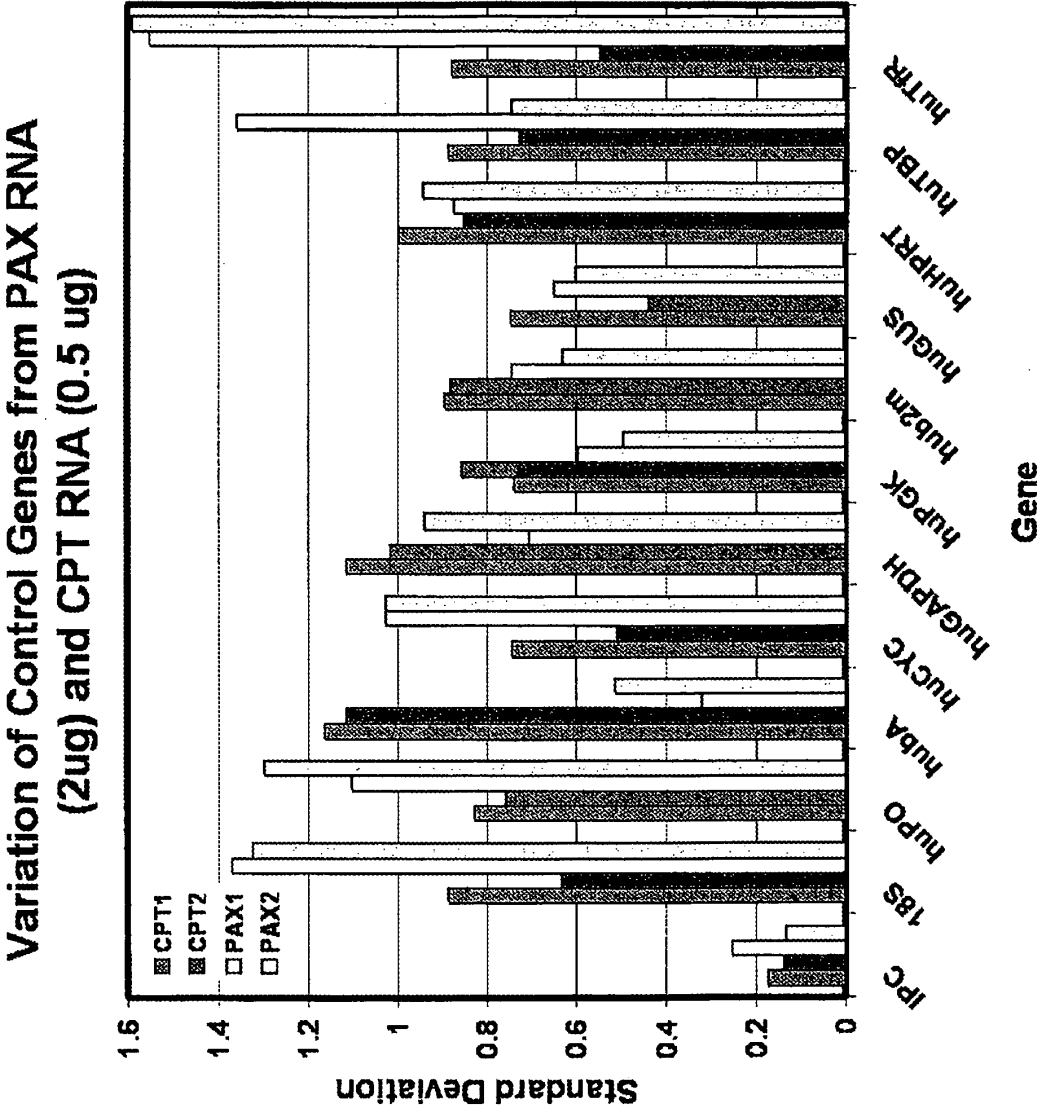


Figure 11



**Intensity of Control Genes from PAX RNA
(2ug) and CPT RNA (0.5 ug)**

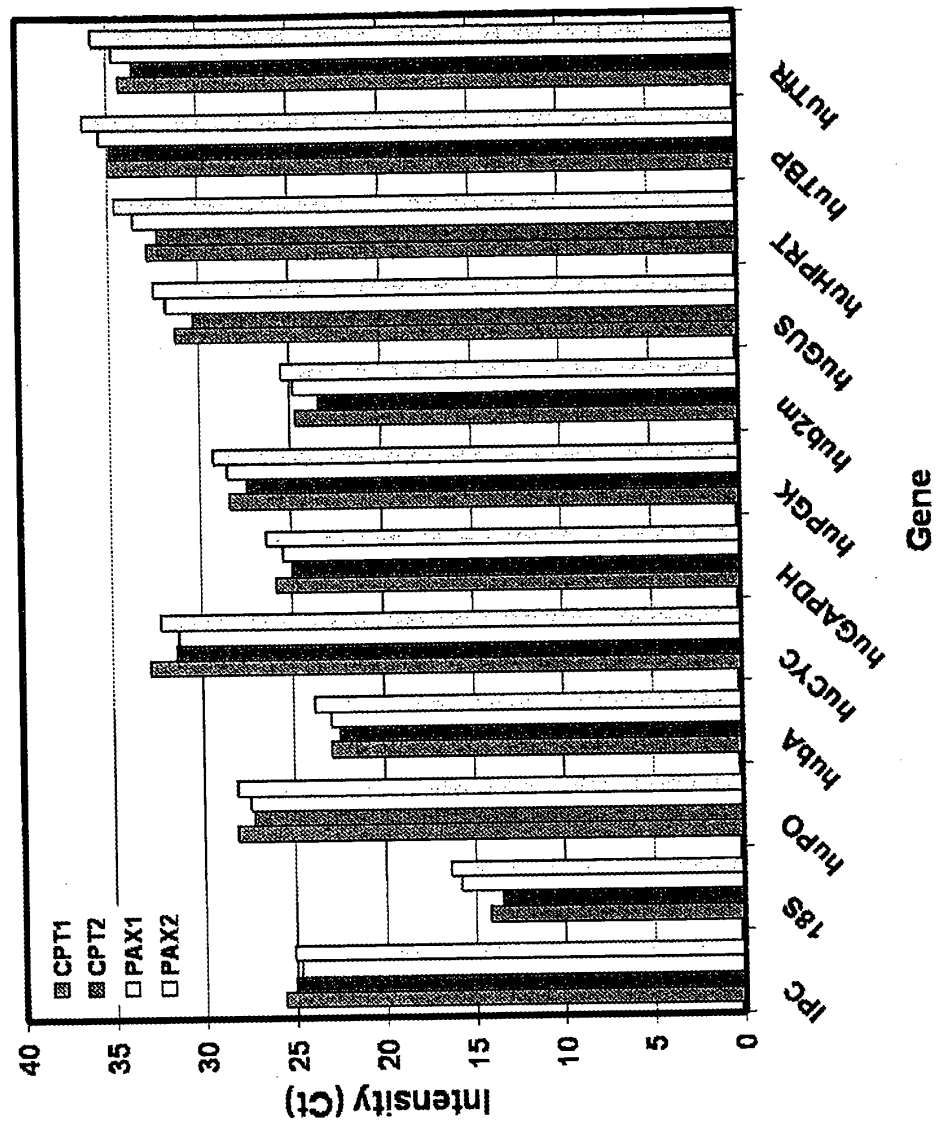


Figure 12

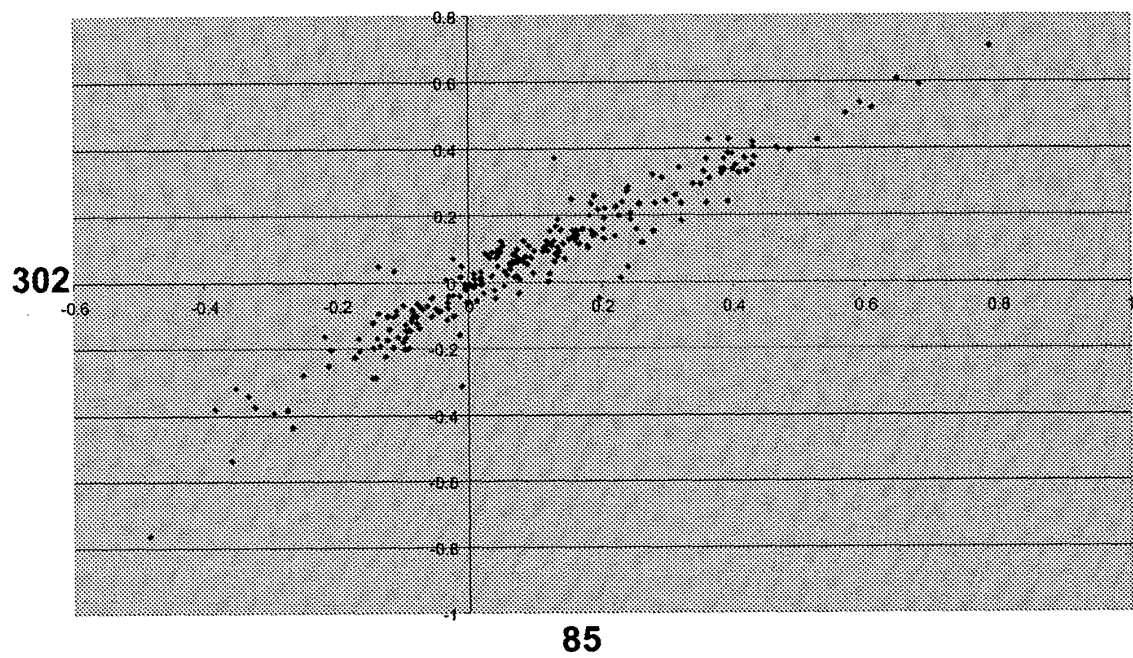
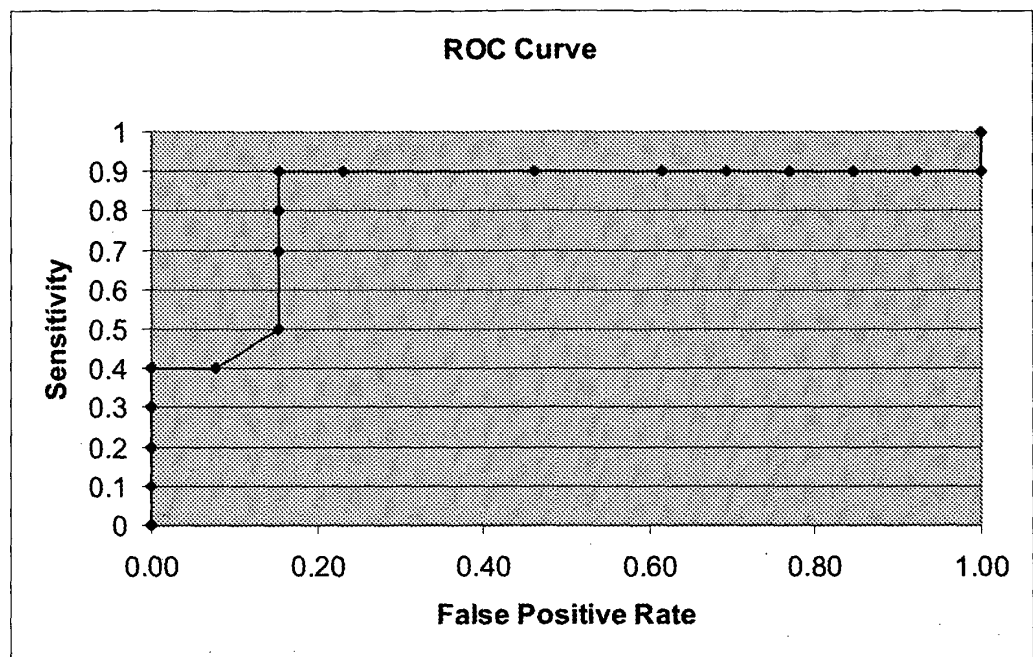


Figure 13



REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 6087477 A [0012]
- US 6099823 A [0012]
- US 6124433 A [0012]
- WO 9730065 A [0012]
- US 6033860 A [0012]
- WO 02057414 A [0014]
- US 4683202 A [0077]
- US 5426039 A [0077]
- US 5728822 A [0093]
- US 4843155 A [0093] [0237]
- US 6132997 A [0097] [0119]
- US 5143854 A [0116] [0117]
- US 5837832 A [0116]
- US 6087112 A [0116]
- US 5215882 A [0116]
- US 5707807 A [0116]
- US 5807522 A [0116]
- US 5958342 A [0116]
- US 5994076 A [0116]
- US 6004755 A [0116]
- US 6048695 A [0116]
- US 6060240 A [0116]
- US 6090556 A [0116]
- US 6040138 A [0116]
- US 5346994 A [0237]
- US 5973137 A [0238]
- US 6020186 A [0238]
- US 4215051 A [0254]
- US 773567 A [0258]
- US 4376110 A [0265]
- US 4946778 A [0267]
- US 10131827 B [0531]
- US 10325899 B [0531]

专利名称(译)	基于标志物表达水平诊断和监测心脏移植排斥		
公开(公告)号	EP2194145B1	公开(公告)日	2013-06-05
申请号	EP2008016970	申请日	2003-04-24
申请(专利权)人(译)	XDX INC.		
当前申请(专利权)人(译)	XDX INC.		
[标]发明人	WOHLGEMUTH JAY FRY KIRK WOODWARD ROBERT LY NGOC PRENTICE JAMES MORRIS MACDONALD ROSENBERG STEVEN		
发明人	WOHLGEMUTH, JAY FRY, KIRK WOODWARD, ROBERT LY, NGOC PRENTICE, JAMES MORRIS, MACDONALD ROSENBERG, STEVEN		
IPC分类号	C12Q1/68 G01N33/564 C12N15/09 C07H21/00 C12N15/12 C12P19/34 C12Q C12Q1/00 C12Q1/04 G01N G01N1/00 G01N24/08 G01N33/53 G01N33/567		
CPC分类号	C12Q1/6883 C12Q1/6881 C12Q1/6888 C12Q2600/158 G01N33/564 G01N33/6863 G01N2800/245		
优先权	10/131831 2002-04-24 US 10/325899 2002-12-20 US		
其他公开文献	EP2194145A1		
外部链接	Espacenet		

摘要(译)

描述了通过检测患者中一种或多种基因的表达水平来诊断或监测患者中的移植排斥，特别是心脏移植排斥的方法。还描述了用于诊断或监测移植排斥，特别是心脏移植排斥的诊断寡核苷酸和包含其的试剂盒或系统。

$$P_0 = (14)^L * 20$$