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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 March 2006 (16.03.2006)

PCT

(10) International Publication Number
WO 2006/028867 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2005/031038

(22) International Filing Date: 31 August 2005 (31.08.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/931,780 1 September 2004 (01.09.2004) US
11/110,209 20 April 2005 (20.04.2005) US

(71) Applicant *(for all designated States except US)*: **THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS** [US/US]; 2404 North University Avenue, Little Rock, AR 72207-3608 (US).

(72) Inventors: **SHAUGHNESSY, John**; 4317 Old Oak Drive, Little Rock, AR 72205 (US). **BARLOGIE, Bart**; 3118 Edgerstone Drive, Little Rock, AR 72205 (US). **FENGHUANG, Zhan**; 2002 Cherrybend Dr., Little Rock, AR 72211 (US).

(74) Agent: **ADLER, Benjamin, A.**; Adler & Associates, 8011 Candle Lane, Houston, TX 77071 (US).

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

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

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

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

WO 2006/028867 A2

(54) Title: DIAGNOSIS, PROGNOSIS AND IDENTIFICATION OF POTENTIAL THERAPEUTIC TARGETS OF MULTIPLE MYELOMA BASED ON GENE EXPRESSION PROFILING

(57) Abstract: Gene expression profiling is a powerful tool that has varied utility. It enables classification of multiple myeloma into subtypes and identifies genes directly involved in disease pathogensis and clinical manifestation. The present invention reported a 15-gene model that classifies myeloma into 7 groups. The present invention also used gene expression profiling in large uniformly treated population of patients with myeloma to identify genes associated with poor prognosis. It also demonstrated that over-expression of CKS 1 B gene, mainly due to gene amplification could impart poor prognosis in multiple myeloma. It is further contemplated that therapeutic strategies that directly target CKS 1 B or related pathways may represent novel and more specific means of treating high-risk myeloma and prevent its secondary evolution.

DIAGNOSIS, PROGNOSIS AND IDENTIFICATION OF POTENTIAL THERAPEUTIC TARGETS OF MULTIPLE MYELOMA BASED ON GENE EXPRESSION PROFILING

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to the field of cancer research. More specifically, the present invention relates to gene expression profiling of cancer patient and the use of these genes in the diagnosis and prognosis of the disease.

Description of the Related Art

Multiple myeloma (MM) is a uniformly fatal tumor of terminally differentiated plasma cells (PCs) that home to and expand in the bone marrow. Although initial transformation events leading to the development of multiple myeloma are thought to occur at a post-germinal center stage of development as suggested by the presence of somatic hypermutation of IGV genes, progress in understanding the biology and genetics of multiple myeloma has been slow.

Multiple myeloma cells are endowed with a multiplicity of anti-apoptotic signaling mechanisms that account for their resistance to current chemotherapy and thus the ultimately fatal outcome for most patients. While aneuploidy by interphase fluorescence *in situ* hybridization (FISH) and DNA flow cytometry are observed in >90% of cases, cytogenetic abnormalities in this typically hypoproliferative tumor are informative in only about 30% of cases and are typically complex, involving on average 7 different chromosomes. Given this "genetic chaos" it has been difficult to establish correlations between genetic abnormalities and clinical outcomes. Only recently has chromosome 13 deletion been identified as a distinct clinical entity with a grave prognosis. However, even with the most comprehensive analysis of laboratory parameters, such as β2-microglobulin (β2M), C-reactive protein (CRP), plasma cell labeling index (PCLI), metaphase karyotyping, and FISH, the clinical course of patients afflicted with multiple myeloma can only be approximated because no more than 20% of clinical heterogeneity can be accounted

for. Thus, there are distinct clinical subgroups of multiple myeloma, and modern molecular tests may provide help in identifying these entities.

Monoclonal gammopathy of undetermined significance (MGUS) and multiple myeloma are the most frequent forms of monoclonal gammopathies. MGUS is the most common plasma cell dyscrasia with an incidence of up to 10% of population over age 75. The molecular basis of MGUS and multiple myeloma are not very well understood and it is not easy to differentiate these two disorders. Diagnosis of multiple myeloma or MGUS is identical in 2/3 of cases using classification systems that are based on a combination of clinical criteria such as the amount of bone marrow plasmocytosis, the concentration of monoclonal immunoglobulin in urine or serum, and the presence of bone lesions. Especially in early phases of multiple myeloma, differential diagnosis is associated with a certain degree of uncertainty.

Furthermore, in the diagnosis of multiple myeloma, clinician must exclude other disorders in which a plasma cell reaction may occur. These other disorders include rheumatoid arthritis, connective tissue disorders, and metastatic carcinoma where the patient may have osteolytic lesions associated with bone metastases. Therefore, given that multiple myeloma is thought to have an extended latency and clinical features are recognized many years after development of the malignancy, new molecular diagnostic techniques are needed for differential diagnosis of multiple myeloma, e.g., MGUS versus multiple myeloma, or recognition of various subtypes of multiple myeloma.

Additionally, although this malignancy of B-cell origin initially resides in the bone marrow, it can transform into an aggressive disease with an abnormal karyotype, increased proliferation, elevated LDH and extra-medullary manifestations (Barlogie, B et al., 2001). Specific molecular genetic lesions and tumor cell-stroma interaction influence the clinical course and response to therapy (Kuehl, W.M. et al., 2002; Shaughnessy, J et al., 2003). Although complete responses can be obtained in more than 40% of patients with high-dose therapy, survival varies widely from few months to more than 15 years (Attal, M. et al., 2003; Barlogie, B. et al., 2004). High-risk disease is best captured by abnormal metaphase cytogentics, present in one-third of newly diagnosed patients and reflecting high proliferative capacity of the malignant disease (Shaughnessy, J. et al., 2003).

Global gene expression profiling has emerged as a powerful tool for classifying disease subtypes and developing robust prognostic models in leukemia and lyphoma. In myeloma, this technology helped identify genes directly involved in disease pathogenesis and clinical manifestation (Zhan, F et al., 2002).

Thus, the prior art is deficient in methods of differential diagnosing and identifying distinct and prognostically relevant clinical subgroups of multiple myeloma. Additionally, the prior art is also deficient in methods for identifying genes associated with poor prognosis in patients with myeloma. The present invention fulfills this long-standing need and desire in the art.

SUMMARY OF THE INVENTION

Bone marrow plasma cells from 74 patients with newly diagnosed multiple myeloma, 5 patients with MGUS, and 31 normal volunteers (normal plasma cells) were purified by $CD138^+$ selection. Gene expression of purified plasma cells and 7 multiple myeloma cell lines were profiled using high-density oligonucleotide microarrays interrogating ~6,800 genes. Using hierarchical clustering analysis, normal and multiple myeloma plasma cells were differentiated and four distinct subgroups of multiple myeloma (MM1, MM2, MM3 and MM4) were identified. The gene expression patterns of MM1 was similar to that of normal plasma cells and monoclonal gammopathy of undetermined significance, whereas MM4 was similar to multiple myeloma cell lines. Clinical parameters linked to poor prognosis such as abnormal karyotype (p =0.0003) and high serum β2-microglobulin levels (p = 0.0004) were most prevalent in MM4. Genes involved in DNA metabolism and cell cycle control were overexpressed in MM4 as compared to MM1.

Using chi square and Wilcoxon rank sum tests, 120 novel candidate disease genes that discriminated between normal and malignant plasma cells (p < .0001) were identified. Many of these candidate genes are involved in adhesion, apoptosis, cell cycle, drug resistance, growth arrest, oncogenesis, signaling and transcription. In addition, a total of 156 genes, including *FGFR3* and *CCND1*, exhibited highly elevated ("spiked") expression in at least 4 of the 74 multiple myeloma cases (range of spikes: 4 to 25). Elevated expression of *FGFR3* and *CCND1* were caused by translocation t(4;14)(p16;q32) or t(11;14)(q13;q32).

Additionally, multivariate stepwise discriminant analysis was used to identify a minimum subset of genes whose expression was intimately associated with malignant features of multiple myeloma. Fourteen genes were defined as predictors that are able to differentiate plasma cells of multiple myeloma patients from normal plasma cells with a high degree of accuracy, and 24 genes were identified as predictors that are able to differentiate distinct subgroups of multiple myeloma (MM1, MM2, MM3 and MM4) described herein. Furthermore, data disclosed herein indicate that multiple myeloma can be placed into a developmental schema parallel to that of normal plasma cell differentiation. Based on gene expression profiling, the MM4, MM3 and MM2 subgroups described above were found to have similarity with tonsil B cells, tonsil plasma cells and bone marrow plasma cells respectively. These data suggest that the enigmatic multiple myeloma is amendable to a gene expression/development stage-based classification system.

Thus, gene expression profiling using DNA microarray and hierarchical clustering analysis could be used to classify subgroups of multiple myeloma, identify genes with differential expression in subsets of multiple myeloma patients, and identify potential therapeutic targets for multiple myeloma. For example, there are provided methods of diagnosis for multiple myeloma or subgroups of multiple myeloma based on the expression of a group of 14 genes or a group of 24 genes respectively. There is also provided a method of multiple myeloma diagnosis based on the expression levels of 15 genes that classify patients into 7 subgroups of myeloma.

In another aspect of the present invention multiple myeloma can be treated using methods that involve inhibiting or enhancing expression of genes that are found to be over-expressed or down-regulated respectively in multiple myeloma patients as disclosed herein. Additionally, multiple myeloma can be classified based on developmental stage by comparing gene expression profiling between multiple myeloma cells and normal B or plasma cells.

The present invention provides a method of diagnosis for multiple myeloma. Such a method comprises first isolating biological samples form an individual. This is followed by examining the expression levels of a group of genes comprising CCND1, MAF, MAFB, FGFR3 and MMSET or determining chromosomal translocation represented by over-expression of a group of genes comprising CCND1, MAF, MafB, FGFR2 and MMSET and examining the

expression levels of a group of genes comprising CST6, RAB7L1, MAP4K3, HRASLS2, TRAIL, IG, FGL2, GNG11, MCM2 and FLJ10709, where results from these steps would classify the individual into one of 7 groups of myleoma (multiple myeloma groups 1-7). The present invention also provides a method of identifying genes associated with a phenotype of interest in a population. Such a method comprises isolating cells from individuals within the population. Subsequently, oligonucleotide microarrays are used to profile gene expression in the cells. Further, the gene expression profiles in the cells are correlated with the phenotype of interest, thereby identifying genes associated with the phenotype of interest in the population.

The present invention provides a method of diagnosing an individual with myeloma. This method comprises (a) obtaining bone marrow sample from the individual and (b) determining an amplification of a gene on chromosome 1q21 either alone or in combination with a deletion of a gene on chromosome 13q14 in the sample, thereby diagnosing the individual with myeloma. The present invention further provides a method of identifying an individual having an aggressive form of a cancer. This method comprises (a) obtaining a bone marrow sample from the individual, and (b) determining an amplification of a gene on chromosome 1q21, where the amplification of the gene is associated with poor survival , thereby identifying the individual having the aggressive form of the cancer.

The present invention also provides a method of screening for a drug useful in treating cancer. Such a method comprises contacting a sample comprising CKS1B gene or CKS1B gene product with a test compound and determining the effect of the compound on amplification, over-expression or activity of the CKS1B gene or the CKS1B gene product, where an inhibitory effect of the test compound indicates that the test compound is the drug useful in treating the cancer. The present invention still further provides a method of treating an individual having cancer. Such a method comprises administering to the individual a pharmaceutically effective amount of a compound that inhibits amplification, over-expression or activity of CKS1B gene or CKS1B gene product. The present invention also provides a method of treating an individual having high-risk myeloma. This method comprises administering to the individual pharmaceutically effective amounts of a compound that inhibits amplification, over-expression or activity of CKS1B gene or CKS1B gene product and a vector comprising DNA sequence encoding *RFP2* gene. The

present invention further provides a kit. This kit compirses (a) probe specific for CKS1B gene, (b) probe specific for RFP2 gene or combinations of these.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention. These embodiments are given for the purpose of disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A shows cluster-ordered data table. The clustering is presented graphically as a colored image. Along the vertical axis, the analyzed genes are arranged as ordered by the clustering algorithm. The genes with the most similar patterns of expression are placed adjacent to each other. Likewise, along the horizontal axis, experimental samples are arranged; those with the most similar patterns of expression across all genes are placed adjacent to each other. Both sample and gene groupings can be further described by following the solid lines (branches) that connect the individual components with the larger groups. The color of each cell in the tabular image represents the expression level of each gene, with red representing an expression greater than the mean, green representing an expression less than the mean, and the deeper color intensity representing a greater magnitude of deviation from the mean. **Figure 1B** shows amplified gene cluster showing genes downregulated in MM. Most of the characterized and sequence-verified cDNA-encoded genes are known to be immunoglobulins. **Figure 1C** shows cluster enriched with genes whose expression level was correlated with tumorigenesis, cell cycle, and proliferation rate. Many of these genes were also statistically significantly upregulated in multiple myeloma (χ^2 and WRS test) (see Table 5). **Figure 1D** shows dendrogram of hierarchical cluster. 74 cases of newly diagnosed untreated multiple myeloma, 5 monoclonal gammopathy of undetermined significance, 8 multiple myeloma cell lines, and 31 normal bone marrow plasma cell samples clustered based on the correlation of 5,483 genes (probe sets). Different-colored branches represent normal plasma cell (green), MGUS (blue arrow), multiple myeloma (tan) and multiple myeloma cell lines (brown arrow). **Figure 1E** shows dendrogram of a hierarchical cluster analysis of 74 cases of newly diagnosed untreated multiple myeloma alone. Two major branches contained two distinct cluster groups. The

subgroups under the right branch, designated MM1 (light blue) and MM2 (blue) were more related to the MGUS cases in Figure 1D. The two subgroups under the left branch, designated MM3 (violet) and MM4 (red) represent samples that were more related to the multiple myeloma cell lines in Figure 1D.

Figure 2 shows the spike profile distributions of FGFR3, CST6, IFI27, and CCND1 gene expression. The normalized average difference (AD) value of fluorescence intensity of streptavidin-phycoerythrin stained biotinylated cRNA as hybridized to probes sets is on the vertical axis and samples are on the horizontal axis. The samples are ordered from left to right: normal plasma cells (NPCs) (green), MM1 (light blue), MM2 (dark blue), MM3 (violet), and MM4 (red). Note relatively low expression in 31 plasma cells and spiked expression in subsets of multiple myeloma samples. The *p* values of the test for significant nonrandom spike distributions are noted.

Figure 3A shows GeneChip HuGeneFL analysis of MS4A2 (CD20) gene expression. The normalized average difference (AD) value of fluorescence intensities of streptavidin-phycoerythrin stained biotinylated cRNA as hybridized to two independent probes sets (accession numbers M27394 (blue) and X12530 (red) located in different regions of the MS4A2 gene is on the vertical axis and samples are on the horizontal axis. Note relatively low expression in 31 normal plasma cells (NPCs) and spiked expression in 5 of 74 multiple myeloma samples (multiple myeloma plasma cells). Also note similarity in expression levels detected by the two different probe sets. **Figure 3B** shows immunohistochemistry for CD20 expression on clonal multiple myeloma plasma cells: (1) bone marrow biopsy section showing asynchronous type multiple myeloma cells (H&E, x500); (2) $CD20^+$ multiple myeloma cells (x100; inset x500); (3) biopsy from a patient with mixed asynchronous and Marschalko-type multiple myeloma cells (H&E, x500); and (4) $CD20^+$ single lymphocyte and $CD20^-$ multiple myeloma cells (x200). CD20 immunohistochemistry was examined without knowledge of clinical history or gene expression findings.

Figure 4 shows the gene expression correlates with protein expression. Gene and protein expression of CD markers known to be differentially expressed during B-cell differentiation were compared between the multiple myeloma cell line CAG (left panel) and the Epstein–Barr virus (EBV) transformed B-lymphoblastoid line ARH-77 (right panel). In both panels, the 8 CD markers are listed in the left column of each panel. Flow cytometric analysis of protein expression is presented in

the second column; the average difference (AD) and absolute call (AC) values of gene expression are presented in the third and fourth columns. Note the strong expression of both the gene and protein for CD138 and CD38 in the CAG cells but the low expression in the ARH-77 cells. The opposite correlation is observed for the remaining markers.

Figure 5 shows an overview of the strategy used to classify and validate myeloma subgroups described in Example 12.

Figure 6 shows hierarchic cluster analysis of myeloma samples. Left panel shows thumbnail overview of two-way hierarchic clustering of 177 myeloma specimens (columns) and 2,856 variably expressed genes (rows). Mean-centered gene expression is depicted by a normalized-signal pseudocolor scale. Red and green indicate the genes are overexpressed and underexpressed, respectively. Right panel shows an enlarged view of the sample dendrogram, the color bars stand for the seven different myeloma subtypes. Subpanels I–VII show selected genes whose locations are indicated by vertical colored bars. Because of space limitations, only selected genes are indicated. The dendrogram of the 177 cases at the top of the matrix, in which the pattern and length of branches reflect the relatedness of the samples, indicates that the samples are clustered into seven branches on the basis of similar gene expression profiles.

Figures 7A-B show correlation of spike genes and unique genes with seven subgroups of myeloma. Figures **7A-B** show spike profile distributions of *FGFR3*, *MMSET*, *MAF*, *MAFB*, *CCND1*, and *CCND3* gene expression among the gene expression–defined myeloma subgroups. Signals from 177 myeloma samples in the training set (Figure 7A) and from 174 myeloma samples in the testing set (Figure 7B) are on the vertical axis, and samples are on the horizontal axis. Clearly, patient specimens with *CCND1*, *MAF/MAFB*, and *FGFR3/MMSET* spikes were assigned to group II, group V, and group VII, respectively. **Figures 7C-D** show supervised hierarchic clustering (specimens shown in columns) of 177 training-set specimens (Figure 7C) and 174 testing-set specimens (Figure 7D) versus 350 genes (rows). The genes used in this analysis were the top 50 up-regulated genes chosen by SAM that were most highly correlated with the seven specific class distinctions. The normalized expression value for each gene is indicated by a color, with red representing high expression and green representing low expression. Normalized expression scale is shown at the bottom.

Figures 8A-F show Kaplan-Meier estimates of event-free survival (EFS) and overall survival (OS) in the cluster-defined subgroup samples. Significant differences in event-free survival were present between the subgroups in the training set (Figure 8A) ($p = 0.0002$), testing set (Figure 8B) ($p = 0.021$), and predicting set by the 15-gene model (Figure 8C) ($p = 0.0014$). Significant differences in overall survival were present between the subgroups in the training set (Figure 8D) ($p = 0.0033$), testing set (Figure 8E) ($p = 0.0015$), and predicting set (Figure 8F) by the 15-gene model ($p = 0.0003$).

Figure 9 shows a 15-gene model representing seven subgroups of myeloma. The data show supervised hierarchic clustering of 177 training-set specimens (left panel) and 174 testing-set specimens (right panel) versus 15 genes (specimens shown in rows, genes shown in columns). The genes used in this analysis were the five genes related to recurrent translocation (*FGFR3/MMSET*, *MAF/MAFB*, and *CCND1*) and the 10 genes chosen by multivariate stepwise discriminant analysis that were most highly correlated with the seven specific class distinctions.

Figure 10 shows a decision matrix for the 7-group classification using the 15-gene model.

Figure 11 shows two-dimensional hierarchical cluster analysis of experimental expression profiles and gene behavior of 30 EDG-MM. B cells, tonsil and bone marrow plasma cells, and multiple myeloma (MM) samples were analyzed using a cluster-ordered data table. The tonsil B cell, tonsil plasma cell, bone marrow plasma cell samples are indicated by red, blue, and golden bars respectively. The nomenclature for the 74 MM samples is as indicated in Zhan et al. (2002a). Along the vertical axis, the analyzed genes are arranged as ordered by the clustering algorithm. The genes with the most similar patterns of expression are placed adjacent to each other. Both sample and gene groupings can be further described by following the solid lines (branches) that connect the individual components with the larger groups. The tonsil B cell cluster is identified by the horizontal red bar. The color of each cell in the tabular image represents the expression level of each gene, with red representing an expression greater than the mean, green representing an expression less than the mean, and the deeper color intensity representing a greater magnitude of deviation from the mean.

Figure 12 shows two-dimensional hierarchical cluster analysis of experimental expression profiles and gene behavior of 50 LDG-MM1 genes. Genes

are plotted along the vertical axis (right side), and experimental samples are plotted along the top horizontal axis by their similarity. The tonsil plasma cell cluster is identified by a horizontal blue bar. Tonsil B cell, tonsil plasma cell, and bone marrow plasma cell samples are indicated as in Figure 18.

Figure 13 shows two-dimensional hierarchical cluster analysis of experimental expression profiles and gene behavior of 50 LDG-MM2 genes. Genes are plotted along the vertical axis (right side), and experimental samples are plotted along the top horizontal axis by their similarity. The bone marrow plasma cell cluster is identified by a horizontal golden bar. Tonsil B cell, tonsil plasma cell, and bone marrow plasma cell samples are indicated as in Figure 18.

Figure 14 shows variation in expression of proliferation genes reveals similarities between tonsil B cells and MM4. The data are shown as boxplot of Kruskal-Wallis test values. The seven groups analyzed (tonsil B cells, tonsil plasma cells, bone marrow plasma cells, and gene expression defined subgroups MM1, MM2, MM3, and MM4) are distributed along the x-axis and the natural log transformed average difference is plotted on the y axis. *EZH2*; $p = 7.61\times10^{-11}$; *KNSL1*, $p = 3.21\times10^{-8}$; *PRKDC*, $p = 2.86\times10^{-11}$; *SNRPC*, $p = 5.44\times10^{-12}$; *CCNB1*, $p = 2.54\times10^{-8}$; *CKS2*, $p = 9.49\times10^{-11}$; *CKS1*, $p = 5.86\times10^{-9}$; *PRIM1*, $p = 4.25\times10^{-5}$.

Figure 15A-B show that *CKS1B* expression by myeloma plasma cells is variable and high levels define a high-risk myeloma entity. **Figure 15A** shows box plots of log base 2-transformed Affymetric signal plotted on the y-axis with respect to 351 cases according to the quartile expression levels (x-axis). **Figure 15B** shows Kaplan-Meier plots of overall survival that revealed inferior outcome among the 88 patients with 4th quartile expression levels of *CKS1B* compared to the remaining 263 patients with quartile 1-3 expression levels.

Figure 16A-D show that increased *CKS1B* expression is related to *CKS1B* DNA copy number and degree of DNA amplification is linked to poor survival. **Figure 16A** shows metaphase fluorescence in situ hybridization analysis of *CKS1B* at 1q21 (red signal) and *ASPM* at 1q31 (green signal) performed on plasma cells from a patient with myeloma. The tandem duplications of *CKS1B* (arrows) and the greater degree of amplification of 1q21 were compared to 1q31. **Figure 16B** shows box plots of log base 2-transformed Affymetrix signal (y-axis) by *CKS1B* amplification (N=197). In box plots, the top, bottom and middle lines of each box

correspond to the 75th percentile (top quartile), 25th percentile (bottom quartile) and 50th percentile (median) respectively and. the whiskers extend to the nearest point not beyond 1.5 times the inter-quartile range, with observations beyond these indicated by individual lines. A Wilcoxon rank sum test was used to compare Signal across copy number categories. **Figure 16C** shows Kaplan-Meier plot of overall survival in the validation cohort that depicts inferior outcomes among the 89 patients with *CKS1B* amplification compared to the remaining 135, as determined by interphase fluorescence in situ hybridization. **Figure 16D** shows the Kaplan-Meier plot as shown in 28C for the combined sample of 421 patients.

Figure 17 shows that *CKS1B* expression increases in relapsed myeloma. This figure shows CKS1B signal for 32-paired diagnosis and relapse arrays. The quartile 4 reference line is taken from the complete (N=351) sample of arrays at diagnosis. A majority of samples showed increased expression at relapse and the most dramatic changes were in those with quartile1 to quartile3 expression levels at diagnosis. For a Welch-modified paired t-test was used to compare log-scale Signal at diagnosis and relapse.

Figures 18A-E show that CKS1B mRNA is correlated with nuclear protein levels and inversely correlated with p27^{Kip1}. **Figure 18A** shows *CKS1B* and **Figure 18B** shows *CDKN1B* (p27^{Kip1}) gene expression signal in 1000 unit increments plotted on the y-axis. Primary myelomas with *CKS1B* expression in quartile 1 (n=13) and quartile 4 (n=14) and myeloma cell lines (n=7) were grouped and plotted from left to right along the X-axis. Each bar represents a sample and the height indicates the level of gene expression in the sample. **Figures 18C-E** show Western Blot analysis of nuclear protein extracts for CKS1B (**Figure 18C**), phos-thr-187-p27^{Kip1} (**Figure 18D**) and Histone 1A (**Figure 18E**) (loading control) respectively from aliquots of same plasma cells used in **Figure 18A** and **Figure 18B**. Samples are ordered from left to right in the exact same order in all panels. There is a high degree of correlation between *CKS1B* gene expression and protein expression. *CDKN1B (p27^{Kip1})* gene expression is not correlated with *CKS1B* gene expression, protein levels or p27^{Kip1} protein levels. However, there is a strong inverse correlation between CKS1B protein levels and p27^{Kip1} protein levels. Uniform histone 1A protein levels indicate equal protein loading across all samples.

DETAILED DESCRIPTION OF THE INVENTION

There is now strong evidence that global gene expression profiling can reveal molecular heterogeneity of similar or related hematopoietic malignancies that have been difficult to distinguish. Genes exhibiting significant differential expression between normal and malignant cells can be used in the development of clinically relevant diagnostics as well as provide clues into the basic mechanisms of cellular transformation. In fact, these profiles might even be used to identify malignant cells even in the absence of any clinical manifestations. In addition, biochemical pathways in which the products of these genes act may be targeted by novel therapeutics.

The present invention demonstrates both normal and malignant plasma cells can be purified to homogeneity from bone marrow aspirates using anti-CD138-based immunomagnetic bead-positive selection. Using these cells, the present invention provides the first comprehensive global gene expression profiling of newly diagnosed multiple myeloma patients and contrasted these expression patterns with those of normal plasma cells. Novel candidate multiple myeloma disease genes were identified and this profiling leads to development of a gene-based classification system for multiple myeloma.

Results from hierarchical cluster analysis on multiple myeloma plasma cells and normal plasma cells, as well as the benign plasma cell dyscrasia MGUS and end-stage-like multiple myeloma cell lines revealed normal plasma cells are unique and that primary multiple myeloma is either like MGUS or multiple myeloma cell lines. In addition, multiple myeloma cell line gene expression was homogeneous as evidenced by tight clustering in the hierarchical analysis. Similarity of multiple myeloma cell line expression patterns to primary newly diagnosed forms of multiple myeloma support the validity of using multiple myeloma cell lines as models for multiple myeloma.

Four distinct clinical multiple myeloma subgroups (MM1 to MM4) were distinguished upon hierarchical clustering of multiple myeloma. The MM1 subgroup contained samples that were more like monoclonal gammopathy of undetermined significance, whereas the MM4 subgroup contained samples more like multiple myeloma cell lines. The most significant gene expression patterns differentiating MM1 and MM4 were cell cycle control and DNA metabolism genes, and the MM4 subgroup was more likely to have abnormal cytogenetics, elevated

serum β2M, elevated creatinine, and deletions of chromosome 13. These are important variables that historically have been linked to poor prognosis.

Gene Expression Changes in Multiple Myeloma

Data disclosed herein indicated that the MM4 subgroup likely represents the most high-risk clinical entity. Thus, knowledge of the molecular genetics of this particular subgroup should provide insight into its biology and possibly provide a rationale for appropriate subtype-specific therapeutic interventions. The most significant gene expression changes differentiating the MM1 and MM4 subgroups code for activities that clearly implicate MM4 as having a more proliferative and autonomous phenotype. The most significantly altered gene in the comparison, *TYMS* (thymidylate synthase), which functions in the prymidine biosynthetic pathway, has been linked to resistance to fluoropyrimidine chemotherapy and also poor prognosis in colorectal carcinomas. Other notable genes upregulated in MM4 were the CAAX farnesyltransferase gene, *FTNA*. Farnesyltransferase prenylates RAS, a post translational modification required to allow RAS to attach to the plasma membrane. These data suggest that farnesyltransferase inhibitors may be effective in treating patients with high levels of *FTNA* expression.

Two other genes coding for components of the proteasome pathway, *POH1* (26S proteasome-associated pad1 homolog) and *UBL1* (ubiquitin-like protein 1) were also overexpressed in MM4. Overexpression of *POH1* confers P-glycoprotein-independent, pleotropic drug resistance to mammalian cells. *UBL1*, also known as sentrin, is involved in many processes including associating with RAD51, RAD52, and p53 proteins in the double-strand repair pathway; conjugating with RANGAP1 involved in nuclear protein import; and importantly for multiple myeloma, protecting against both Fas/Apo-1 (*TNFRSF6*) or *TNFR1*-induced apoptosis. In contrast to normal plasma cells, more than 75% of multiple myeloma plasma cells express abundant mRNA for the multidrug resistance gene, lung-resistance-related protein (*MVP*). These data are consistent with previous reports showing expression of *MVP* in multiple myeloma is a poor prognostic factor. Given the uniform development of chemotherapy resistance in multiple myeloma, the combined overexpression of *POH1* and *MVP* may have profound influences on this phenotype. The deregulated expression of many genes whose products function in the proteasome pathway may be used in pharmacogenomic analysis in determining

the efficacy of proteasome inhibitors like PS-341 (Millennium Pharmaceuticals, Cambridge, MA).

Another significantly upregulated gene in MM4 was the single stranded DNA-dependent ATP-dependent helicase (*G22P1*), which is also known as Ku70 autoantigen. The DNA helicase II complex, made up of p70 and p80, binds preferentially to fork-like ends of double-stranded DNA in a cell cycle-dependent manner. Binding to DNA is thought to be mediated by p70 and dimerization with p80 forms the ATP-dependent DNA-unwinding enzyme (helicase II) and acts as the regulatory component of a DNA-dependent protein kinase (*DNPK*) which was also significantly upregulated in MM4. The involvement of helicase II complex in DNA double-strand break repair, V(D)J recombination, and notably chromosomal translocations has been proposed. Another upregulated gene was the DNA fragmentation factor (*DFFA*). Caspase-3 cleaves DFFA-encoded 45 kD subunit at two sites to generate an active factor that produces DNA fragmentation during apoptosis signaling. In light of the many blocks to apoptosis in multiple myeloma, DFFA activation could result in DNA fragmentation, which in turn would activate the helicase II complex that may facilitate chromosomal translocations. It is of note that abnormal karyotypes, and thus chromosomal translocations, are associated with the MM4 subgroup which tended to overexpress these two genes.Hence, it was demonstrated that direct comparison of gene expression patterns in multiple myeloma and normal plasma cells can identified novel genes that could represent fundamental changes associated with malignant transformation of plasma cells.

Progression of multiple myeloma as a hypoproliferative tumor is thought to be linked to a defect in programmed cell death rather than rapid cell replication. Two genes, prohibitin (*PHB*) and quiescin Q6 (*QSCN6*), overexpressed in multiple myeloma are involved in growth arrest. Overexpression of these genes may be responsible for the typically low proliferation indices seen in multiple myeloma. It is hence conceivable that therapeutic downregulation of these genes that results in enhanced proliferation could render multiple myeloma cells more susceptible to cell cycle-active chemotherapeutic agents.

The gene coding for CD27, *TNFRSF7*, the second most significantly underexpressed gene in multiple myeloma, is a member of the tumor necrosis factor receptor (TNFR) superfamily that provides co-stimulatory signals for T and B cell proliferation, B cell immunoglobulin production and apoptosis. Anti-CD27

significantly inhibits induction of Blimp-1 and J-chain transcripts which are turned on in cells committed to plasma cell differentiation, suggesting that ligation of CD27 on B cells may prevent terminal differentiation. CD27 ligand (CD70) prevents IL-10-mediated apoptosis and directs differentiation of $CD27^+$ memory B cells toward plasma cells in cooperation with IL-10. Thus, it is possible that downregulation of CD27 gene expression in multiple myeloma may block an apoptotic program.

Overexpression of CD47 in multiple myeloma may be related to escape of multiple myeloma cells from immune surveillance. Studies have shown that cells lacking CD47 are rapidly cleared from the bloodstream by splenic red pulp macrophages and CD47 on normal red blood cells prevents this elimination.

The gene product of DNA methyltransferase 1, *DNMT1*, overexpressed in multiple myeloma is responsible for cytosine methylation in mammals and has an important role in epigenetic gene silencing. In fact, aberrant hypermethylation of tumor suppressor genes plays an important role in the development of many tumors. De novo methylation of *p16/INK4a* is a frequent finding in primary multiple myeloma. Also, recent studies have shown that upregulated expression of *DNMT*s may contribute to the pathogenesis of leukemia by inducing aberrant regional hypermethylation. DNA methylation represses genes partly by recruitment of the methyl-CpG-binding protein MeCP2, which in turn recruits a histone deacetylase activity. It has been shown that the process of DNA methylation mediated by *Dnmt1* may depend on or generate an altered chromatin state via histone deacetylase activity. It is significant that multiple myeloma cases also demonstrate significant overexpression of metastasis-associated 1 (*MTA1*) gene. *MTA1* was originally identified as being highly expressed in metastatic cells. *MTA1* has more recently been discovered to be one subunit of the NURD (NUcleosome Remodeling and histone Deacetylation) complex which contains not only ATP-dependent nucleosome disruption activity, but also histone deacetylase activity. Thus, over expression of *DNMT1* and *MTA1* may have dramatic effects on repressing gene expression in multiple myeloma.

Oncogenes activated in multiple myeloma included *ABL* and *MYC*. Although it is not clear whether ABL tyrosine kinase activity is present in multiple myeloma, it is important to note that overexpression of *abl* and *c-myc* results in accelerated development of mouse plasmacytomas. Thus, it may be more than a coincidence that multiple myeloma cells significantly overexpresses *MYC* and *ABL*.

Chromosomal translocations involving the *MYC* oncogene and *IGH* and *IGL* genes that result in dysregulated *MYC* expression are hallmarks of Burkitt's lymphoma and experimentally induced mouse plasmacytomas; however, *MYC/IGH*-associated translocations are rare in multiple myeloma. Although high *MYC* expression was a common feature in our panel of multiple myeloma, it was quite variable, ranging from little or no expression to highly elevated expression. It is also of note that the *MAZ* gene whose product is known to bind to and activate *MYC* expression was significantly upregulated in the MM4 subgroup. Given the important role of *MYC* in B cell neoplasia, overexpression of *MYC*, and possibly *ABL*, in multiple myeloma may have biological and possibly prognostic significance.

EXT1 and *EXT2*, which are tumor suppressor genes involved in hereditary multiple exostoses, heterodimerize and are critical in the synthesis and display of cell surface heparan sulfate glycosaminoglycans (GAGs). *EXT1* is expressed in both multiple myeloma and normal plasma cells. *EXT2L* was overexpressed in multiple myeloma, suggesting that a functional glycosyltransferase could be created in multiple myeloma. It is of note that syndecan-1 (*CD138/SDC1*), a transmembrane heparan sulfate proteoglycan, is abundantly expressed on multiple myeloma cells and, when shed into the serum, is a negative prognostic factor. Thus, abnormal GAG-modified SDC1 may be important in multiple myeloma biology. The link of SDC1 to multiple myeloma biology is further confirmed by the recent association of SDC1 in the signaling cascade induced by WNT proto-oncogene products. It has been showed that syndecan-1 (*SDC1*) is required for Wnt-1-induced mammary tumorigenesis. Data disclosed herein indicated a significant downregulation of *WNT10B* in primary multiple myeloma cases. It is also of note that the *WNT5A* gene and the *FRZB* gene, which codes for a decoy WNT receptor, were also marginally upregulated in newly diagnosed multiple myeloma. Given that WNTs represent a novel class of B cell regulators, deregulating the expression of these growth factors (*WNT5A*, *WNT10B*) and their receptors (e.g., *FRZB*) and genes products that modulate receptor signaling (e.g., *SDC1*) may be important in the genesis of multiple myeloma.

Genes identified by the present invention that shows significantly up-regulated or down-regulated expression in multiple myeloma are potential therapeutic targets for multiple myeloma. Over-expressed genes may be targets for small molecules or inhibitors that decrease their expression. Methods and materials that can

be used to inhibit gene expression, e.g. small drug molecules, anti-sense oligo, or antibody would be readily apparent to a person having ordinary skill in this art. On the other hand, under-expressed genes can be replaced by gene therapy or induced by drugs.

Gene Profiles Defining Disease Subgroups

A multivariate stepwise discriminant analysis was used to identify a minimum subset of genes whose expression is intimately associated with malignant features of multiple myeloma. By applying linear regression analysis to the top 50 differentially expressed genes, 14 genes were defined as predictors that are able to differentiate multiple myeloma from normal plasma cells with a high degree of accuracy. When the model was applied to a validation group consisting of 118 multiple myeloma, 6 normal plasma cells and 7 cases of MGUS (MGUS), an accuracy of classification of more than 99% was achieved. Importantly, 6 of the 7 MGUS cases were classified as multiple myeloma, indicating that MGUS has gene expression features of malignancy. Thus the altered expression of 14 genes out of over 6,000 genes interrogated are capable of defining multiple myeloma. Similar multivariate discriminant analysis also identified a set of 24 genes that can distinguish between the four multiple myeloma subgroups described above.

In addition to identifying genes that were statistically different between normal plasma cells and multiple myeloma plasma cells, the present invention also identified genes, like *FGFR3* and *CCND1*, that demonstrate highly elevated "spiked" expression in subsets of multiple myelomas. Patients with elevated expression of these genes can have significant distribution differences among the four gene expression cluster subgroups. For example, *FGFR3* spikes are found in MM1 and MM2 whereas spikes of *IFI27* are more likely to be found in MM3 and MM4. Highly elevated expression of the interferon-induced gene *IFI27* may be indicative of viral infection, either systemic or specifically within the plasma cells from these patients. Correlation analysis has shown that *IFI27* spikes are significantly linked (Pearson correlation coefficient values of .77 to .60) to elevated expression of 14 interferon-induced genes, including *MX1*, *MX2*, *OAS1*, *OAS2*, *IFIT1*, *IFIT4*, *PLSCR1*, and *STAT1*. More recent analysis of a large population of multiple myeloma patients (N = 280) indicated that nearly 25% of all patients had spikes of the *IFI27* gene. It is of interest to determine whether or not the *IFI27* spike patients who cluster in the

MM4 subgroup are more likely to have a poor clinical course and to identify the suspected viral infection causing upregulation of this class of genes. In conclusion, spiked gene expression may also be used in the development of clinically relevant prognostic groups.

Finally, the 100% coincidence of spiked *FGFR3* or *CCND1* gene expression with the presence of t(4;14)(p14;q32) or t(11;14)(q13;q32) translocations, as well as the strong correlations between protein expression and gene expression represent important validations of the accuracy of gene expression profiling and suggests gene expression profiling may eventually supplant the labor intensive and expensive clinical laboratory procedures, such as cell surface marker immunophenotyping and molecular and cellular cytogenetics.

In another embodiment, the present example also uses feature-subset selection to extract genes relevant to specific myeloma subtypes. In this regard, multivariate stepwise discriminant analysis was applied and identified 15 genes that could correctly separate tissue samples into seven subtypes. Examining the expression of the 15 genes identified herein not only would provide important new insights into the diagnosis and pathogenesis of these myeloma subtypes but also may pinpoint useful targets against which novel therapeutic agents could be developed.

Comparison of Multiple Myeloma with Normal Plasma Cell Development

Further, it was also shown that multiple myeloma could be placed into a developmental schema parallel to that of normal plasma cell differentiation. Global gene expression profiling reveals distinct changes in transcription associated with human plasma cell differentiation. Hierarchical clustering analyses with 4866 genes segregated tonsil B cells, tonsil plasma cells, and bone marrow plasma cells. Combining χ^2 and Wilcoxon rank sum tests, 359 previously defined and novel genes significantly (p <0.0005) discriminated tonsil B cells from tonsil plasma cells, and 500 genes significantly discriminated tonsil plasma cells from bone marrow plasma cells. Genes that were differentially expressed in the tonsil B cell to tonsil plasma cell transition were referred as "early differentiation genes" (EDGs) and those differentially expressed in the tonsil plasma cell to bone marrow plasma cell transition were referred as "late differentiation genes" (LDGs). One-way ANOVA was then applied to EDGs and LDGs to identify statistically significant expression differences

between multiple myeloma (MM) and tonsil B cells (EDG-MM), tonsil plasma cells (LDG-MM1), or bone marrow plasma cells (LDG-MM2).

Hierarchical cluster analysis revealed that 13/18 (p=.00005) MM4 cases (a putative poor-prognosis subtype) clustered tightly with tonsil B cells. The other groups (MM1, 2 and 3) failed to show such associations. In contrast, there was tight clustering between tonsil plasma cells and 14/15 (p=.00001) MM3, and significant similarities were found between bone marrow plasma cells and 14/20 (p=.00009) MM2 cases. MM1 showed no significant linkage with the normal cell types studied. In addition, *XBP1*, a transcription factor essential for plasma cell differentiation, exhibited a significant, progressive reduction in expression from MM1 to MM4, consistent with developmental-stage relationships. Therefore, global gene expression patterns linked to late-stage B cell differentiation confirmed and extended a global gene expression-defined classification system of multiple myeloma, suggesting that multiple myeloma represents a constellation of distinct subtypes of disease with unique origins.

Identification of Genes That Could Be Useful As Diagnostic, Prognostic And Potential Targets In Myeloma.

The present invention also identified genes that could be useful as diagnostic, prognostic and potential targets in myeloma. 70 genes whose expression was significantly correlated with disease-related survival was identified by performing gene expression analysis of highly purified plasma cells from newly diagnosed myeloma patients; 30% of these genes mapped to chromosome 1. Increased expression of 1q genes and reduced expression of 1p genes were consistent with cytogenetic data of frequent 1q gains and 1p loses in myeloma karyotypes (Nilsson et al., 2003; Gutierrez et al., 2004). Tandem duplications and jumping translocations involving 1q21, caused by decondensation of pericentromeric heterochromatin are features of end stage disease (Sawyer et al., 1998).

Over expression and amplification of *CKS1B*, mapping to 1q21 was linked to poor prognosis in newly diagnosed myeloma. Its role in controlling SCF^{Skp2}-mediated ubiquitinylation and proteasomal degradation of the cyclin-dependent kinase inhibitor $p27^{Kip1}$ made it an attractive candidate gene. CKS1B protein levels were correlated with gene expression and both inversely correlated with $p27^{Kip1}$ protein levels. Investigations in *S.cerevisiae* demonstrated an essential role of

cks1 in promoting mitosis by modulating the transcriptional activation of *CDC20* (Morris et al., 2003). A strong correlation between *CKS1B* and *CDC20* expression ($r=0.78$; $p<0.0001$) was observed in the present invention, consistent with *CKS1B* promoting mitosis by regulating *CDC20* expression in human cells. Therefore, the results obtained in the present invention demonstrate that gene dosage-related increase in *CKS1B* expression led to enhanced degradation of $p27^{Kip1}$ and possibly activation of *CDC20*.

In context of the recently recognized prognostically relevant genetic subgroups, CKS1B hyper-activation was less frequent in cases with hyperdiploid and normal karyotypes; one-third of those with CCND1-translocations had high CKS1B levels and upto two-thirds of high-risk entities, MAF, MAFB and hyperdiploidy displayed CKS1B hyperactivation (Table 15A). In addition to conferring a poor prognosis among newly diagnosed patients, CKS1B over-expression and amplification were common at relapse in patients lacking these features at diagnosis. Hence, it will be important to determine whether CKS1B amplification emerges in all subgroups and when present, portends rapid disease progression and death.

CKS1B gene amplification along with abnormal metaphase cytogenetics and chromosome 13 deletion accounted for almost 40% of the observed survival variability, underscored that myeloma risk was best assessed by studying molecular and cellular genetics. It is therefore recommended to routinely apply such studies, performed on a single bone marrow sample for appropriate patient stratification in therapeutic trial design. Additionally, it was also noteworthy that loss of expression of a single gene mapping to chromosome 13q14, RFP2/LEUS was significantly linked to poor survival in the present invention. The RFP2/LEUS gene contains a zinc finger domain of the RING type and shares significant homology with the RET finger protein (RFP) and the tumor suppressor gene BRCA1 and was previously identified as a candidate tumor suppressor gene on 13q14 in B-CLL. As demonstrated in the present invention and in previous studies, loss of chromosome 13q14 was linked to poor prognosis in myeloma. Furthermore, since loss of RFP2 was common to these highly related B-cell malignancies, an in-depth investigation of RFP2 function in late stage B-cell biology was defensible.

The impact of new agents such as bortezomib and thalidomide and its derivatives will be profound if their clinical efficacy is extended to genetically defined high-risk myeloma. Since *CKS1B* directly or indirectly interacts with

ubiquitin ligases and/or the proteasome to regulate multiple cell cycle checkpoints (Pagano and Benmaamar, 2003), it is contemplated that new therapeutic strategies that directly target CKS1B or related pathways might represent novel and more specific means of treating de novo high-risk myeloma and might prevent its secondary evolution. Similarly, due to the significance of RFP2 gene deletion as discussed above, it is also contemplated that new therapeutic strategies that directly target RFP2 or related pathways might also represent novel and more specific means of treating de novo high-risk myeloma.

Further, fluorescence in-situ hybridization (FISH) tests for 1q21 amplification and 13 deletion represented a better test for risk assessment that existed and could be used as a powerful staging system in myeloma. Since the expression of CKS1B highly correlated with the gene copy number and the risk of death increased with the increase in gene copy number, this test could also be used to stage patients at diagnosis, detect residual diseases, predict recurrence and progression of the disease.

In summary, the present invention is drawn to a method of gene expression profiling for multiple myeloma applying nucleic acid samples of isolated plasma cells derived from individuals with or without multiple myeloma to a DNA microarray, and performing hierarchical clustering analysis on data obtained from the microarray to classify the individuals into distinct subgroups such as the MM1, MM2, MM3 and MM4 subgroups disclosed herein. Gene expression profiling will also identify genes with elevated expression in subsets of multiple myeloma patients or genes with significantly different levels of expression in multiple myeloma patients as compared to normal individuals. These genes are potential therapeutic targets for multiple myeloma.

In another embodiment, there is provided a method of identifying a group of genes that can distinguish between normal plasma cells and plasma cells of multiple myeloma. Nucleic acid samples of isolated plasma cells derived from individuals with or without multiple myeloma were applied to a DNA microarray, and hierarchical clustering analysis was performed on data obtained from the microarray. Genes with statistically significant differential expression patterns were identified, and linear regression analysis was used to identify a group of genes that is capable of accurate discrimination between normal plasma cells and plasma cells of multiple myeloma. This analysis can also identify a group of genes that is capable of accurate discrimination between subgroups of multiple myeloma.

The expression levels of a group of 14 genes or a group of 24 genes could be used for diagnosis of multiple myeloma. Significant differential expression of these genes would indicate that such individual has multiple myeloma or a subgroup of multiple myeloma. Gene expression levels can be examined at nucleic acid level or protein level according to methods well known to one of skill in the art.

In another embodiment, there is provided a method of diagnosis for multiple myeloma based on examining the expression levels of a group of genes comprising *CST6, RAB7L1, MAP4K3, HRASLS2, TRAIL, IG, FGL2, GNG11, MCM2, FLJ10709, CCND1, MAF, MAFB, FGFR3,* and *MMSET. CCND1, MAF, MAFB, FGFR3,* and *MMSET* expression can be determined by and are correlated with chromosomal translocation such as t(4;14)(p21;q32), t(14;16)(q32;q23), t(14;20)(q32;q13), and t(11;14)(q13;q32). Gene expression levels of this group of genes would classify an individual into one of 7 groups of myeloma (groups 1-7). Individual in myeloma groups 1, 2, 3 and 6 would have poor prognosis compared to individual in myeloma groups 4, 5 and 7. Group 1 is defined by downregulation of *CST6, RAB7L1, MAP4K3, HRASLS2, GNG11, MCM2* and *FLJ10709*, and overexpression of *TRAIL, IG* and *FGFL2*. Group 2 is defined by downregulation of *CST6, RAB7L1, MAP4K3, HRASLS2, IG, FGFL2, MCM2* and *FLJ10709*, and overexpression of *TRAIL* and *GNG11*. Group 3 is defined by overexpression of *CCND1*, or translocation t(11;14)(q13;q32). Group 4 is defined by downregulation of *CST6, RAB7L1, MAP4K3, HRASLS2, IG, FGFL2,* and *GNG11*, and overexpression of *TRAIL, MCM2* and *FGFL2*. Group 5 is defined by overexpression of *MAF* or *MAFB*, or translocation t(14;16)(32;q23) or t(14;20)(q32;q13). Group 6 is defined by overexpression of *CST6, RAB7L1, MAP4K3,* and *HRASLS2,* and downregulation of *TRAIL*. Group 7 is defined by overexpression of *FGFR3* or *MMSET*, or translocation t(4;14)(p21;q32).

In yet another embodiment of the present invention, there are provided methods of treatment for multiple myeloma. Such methods involve inhibiting expression of one of the genes, or increasing expression of one of the genes. Methods of inhibiting or increasing gene expression such as those using anti-sense oligonucleotide or antibody are well known to one of skill in the art. Inhibiting gene expression can be achieved through RNA interference using so called siRNA. Gene expression enhancement might be through gene therapy.

The present invention is also drawn to a method of developmental stage-based classification for multiple myeloma. Nucleic acid samples of isolated B cells and plasma cells derived from individuals with or without multiple myeloma were applied to a DNA microarray, and hierarchical clustering analysis performed on data obtained from the microarray will classify the multiple myeloma cells according to the developmental stages of normal B cells and plasma cells. In general, normal B cells and plasma cells are isolated from tonsil, bone marrow, mucoal tissue, lymph node or peripheral blood.

The present invention is drawn to a method of identifying genes associated with a phenotype of interest in a population, comprising: isolating cells from individuals within the population, profiling gene expression in the cells using oligonucleotide microarrays, and correlating the gene expression profiles in the cells with the phenotype of interest, thereby identifying genes associated with the phenotype of interest in the population. Generally, the phenotype of interest is disease-related survival and response to a drug.

The present invention is also drawn to a method of diagnosing an individual with myeloma, comprising the steps of: (a) obtaining a bone marrow sample from the individual, and (b) determining an amplification of a gene on chromosome 1q21 either alone or in combination with a deletion of a gene on chromosome 13q14 in the sample, thereby diagnosing the individual with myeloma. This method comprises predicting relapse or progression of the disease in the individual based on the extent of amplification of the gene on chromosome 1q21, where an increase in the amplification of the gene indicates relapse or progression of the disease in the individual. Specifically, the amplified gene is CKS1B and the deleted gene is RFP2.

The present invention is further drawn to a method of identifying an individual having an aggressive form of a cancer, comprising the steps of: (a) obtaining a bone marrow sample from the individual; and (b) determining an amplification of a gene on chromosome 1q21, where the amplification of the gene is associated with poor survival, thereby identifying the individual having the aggressive form of cancer. Specifically, the amplified gene is CKS1B. Examples of cancer are not limited to but include breast cancer, colon cancer, prostate cancer or myeloma.

The present invention is still further drawn to a method of screening for a drug useful in treating cancer, comprising: contacting a sample comprising

CKS1B gene or CKS1B gene product with a test compound and determining the inhibitory effect of the compound on the amplification, overexpression or activity of CKS1B gene or CKS1B gene product, where an inhibitory effect of the test compound indicates that the test compound is the drug useful for treating the cancer. Examples of the cancer that the utility of the drug is assessed for are not limited to but include aggressive forms of breast cancer, colon cancer, prostate cancer or myeloma. As is known to one skilled in the art that such a drug will target components that are either upstream or downstream of this gene or protein in a cell cycle-related pathway.

The present invention is further drawn to a method of treating an individual having a cancer, comprising: administering to the individual a pharmacologically effective amount of a compound that inhibits amplification, over-expression or activity of CKS1B gene or CKS1B gene product. Examples of compounds with the inhibitory effect on the CKS1B gene although not limited to include compounds such as a peptide nucleic acid (PNA), RNA-mediated interference. Examples of compounds with the inhibitory effect on the CKS1B gene product although not limited to include compounds such as an antibody, a CKS1B antisense RNA or a small molecule inhibitor. Examples of cancer that can be treated using such compounds are as described supra.

The present invention is still further drawn to a method of treating an individual having a high-risk myeloma, comprising: administering to the individual pharamcologically effective amounts of a compound that inhibits amplification, over-expression or activity of CKS1B gene or CKS1B gene product and a vector comprising DNA sequence encoding RFP2 gene. Examples of the compounds that could be used in this method is the same as discussed supra. Additionally, the vector that is used in this method could be replication deficient adenoviral vector, retroviral vector or any of the vectors that are known to one skilled in the art.

The present invention is aslo drawn to a kit, comprising: (a) probe specific for CKS1B gene; (b) probe specific for RFP2 gene; or combinations of these.

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. One skilled in the art will appreciate readily that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those objects, ends and advantages inherent herein. Changes therein and

other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

EXAMPLE 1

Cell Isolation And Analysis

Samples for the following studies included plasma cells from 74 newly diagnosed cases of multiple myeloma, 5 subjects with monoclonal gammopathy of undetermined significance, 7 samples of tonsil B lymphocytes (tonsil BCs), 11 samples of tonsil plasma cells (tonsil PCs), and 31 bone marrow PCs derived from normal healthy donor. Multiple myeloma cell lines (U266, ARP1, RPMI-8226, UUN, ANBL-6, CAG, and H929 (courtesy of P.L. Bergsagel) and an Epstein–Barr virus (EBV)-transformed B-lymphoblastoid cell line (ARH-77) were grown as recommended (ATCC, Chantilly, VA).

Tonsils were obtained from patients undergoing tonsillectomy for chronic tonsillitis. Tonsil tissues were minced, softly teased and filtered. The mononuclear cell fraction from tonsil preparations and bone marrow aspirates were separated by a standard Ficoll-Hypaque gradient (Pharmacia Biotech, Piscataway, NJ). The cells in the light density fraction (S.G. ≤1.077) were resuspended in cell culture media and 10% fetal bovine serum, RBC lysed, and several PBS wash steps were performed. Plasma cell isolation was performed with anti-CD138 immunomagnetic bead selection as previously described (Zhan et al., 2002a). B lymphocyte isolation was performed using directly conjugated monoclonal mouse anti-human CD19 antibodies and the AutoMacs automated cell sorter (Miltenyi-Biotec, Auburn, CA).

For cytology, approximately 40,000 purified tonsil BC and PC mononuclear cells were cytocentrifuged at 1000 × g for 5 min at room temperature. For morphological studies, the cells were immediately processed by fixing and staining with DiffQuick fixative and stain (Dade Diagnostics, Aguada, PR).

For immunofluorescence, slides were treated essentially as described (Shaughnessy et al., 2000). Briefly, slides were air-dried overnight, then fixed in 100% ethanol for 5 min at room temperature and baked in a dry 37°C incubator for 6 hr. The slides were then stained with 100 µl of a 1:20 dilution of goat anti-human-kappa immunoglobulin light chain conjugated with 7-amino-4-methylcourmarin-3-acitic acid (AMCA) (Vector Laboratories, Burlingame, CA) for 30 min in a

humidified chamber. After incubation, the slides were washed two times in 1 × PBS + 0.1% NP-40 (PBD). To enhance the AMCA signal, the slides were incubated with 100 µl of a 1:40 dilution of AMCA-labeled rabbit–anti-goat IgG antibody and incubated for 30 min at room temperature in a humidified chamber. Slides were washed 2 times in 1 × PBD. The slides were then stained with 100 µl of a 1:100 dilution of goat anti-human-lambda immunoglobulin light chain conjugated with FITC (Vector Laboratories, Burlingame, CA) for 30 min in a humidified chamber; the slides were washed two times in 1 × PBD. Then the slides were stained with propidium iodide at 0.1 µg/ml in 1 × PBS for 5 min, washed in 1 × PBD, and 10 µl anti-fade (Molecular Probes, Eugene, OR) was added. Cytoplasmic immunoglobulin light chain-positive PCs were visualized using an Olympus BX60 epi-fluorescence microscope equipped with appropriate filters. The images were captured using a Quips XL genetic workstation (Vysis, Downers Grove, IL).

Both unpurified mononuclear cells and purified fractions from tonsil BCs, tonsil PCs, and bone marrow PCs were subjected to flow cytometric analysis of CD marker expression using a panel of antibodies directly conjugated to FITC or PE. Markers used in the analysis included FITC-labeled CD20, PE-labeled CD38, FITC-labeled or ECD-labeled CD45, PE- or PC5-labled CD138 (Beckman Coulter, Miami, FL). For detection of CD138 on PCs after CD138 selection, an indirect detection strategy was used using a FITC-labeled rabbit anti-mouse IgG antibody (Beckman Coulter) to detect the mouse monoclonal anti-CD138 antibody BB4 used in the immunomagnetic selection technique. Cells were taken after Ficoll Hypaque gradient or after cell purification, washed in PBS, and stained at 4°C with CD antibodies or isotype-matched control G1 antibodies (Beckman Coulter). After staining, cells were resuspended in 1 × PBS and analyzed using a Epics XL-MCL flow cytometry system.

EXAMPLE 2

Preparation Of Labeled cRNA And Hybridization To High-Density Microarray

Total RNA was isolated with RNeasy Mini Kit (Qiagen, Valencia, CA). Double-stranded cDNA and biotinylated cRNA were synthesized from total RNA and hybridized to HuGeneFL GeneChip microarrays (Affymetrix, Santa Clara, CA), which were washed and scanned according to procedures developed by the manufacturer. The arrays were scanned using Hewlett Packard confocal laser scanner

and visualized using Affymetrix 3.3 software (Affymetrix). Arrays were scaled to an average intensity of 1,500 and analyzed independently.

EXAMPLE 3

Genechip Data Analysis

To efficiently manage and mine high-density oligonucleotide DNA microarray data, a new data-handling tool was developed. GeneChip-derived expression data was stored on an MS SQL Server. This database was linked, via an MS Access interface called Clinical Gene-Organizer to multiple clinical parameter databases for multiple myeloma patients. This Data Mart concept allows gene expression profiles to be directly correlated with clinical parameters and clinical outcomes using standard statistical software. All data used in the present analysis were derived from Affymetrix 3.3 software. GeneChip 3.3 output files are given (1) as an average difference (AD) that represents the difference between the intensities of the sequence-specific perfect match probe set and the mismatch probe set, or (2) as an absolute call (AC) of present or absent as determined by the GeneChip 3.3 algorithm. Average difference calls were transformed by the natural log after substituting any sample with an average difference of <60 with the value 60 (2.5 times the average Raw Q). Statistical analysis of the data was performed with software packages SPSS 10.0 (SPSS, Chicago, IL), S-Plus 2000 (Insightful Corp., Seattle, WA), and Gene Cluster/Treeview (Eisen et al., 1998).

To differentiate four distinct subgroups of multiple myeloma (MM1, MM2, MM3 and MM4), hierarchical clustering of average linkage clustering with the centered correlation metric was employed. The clustering was done on the average difference data of 5,483 genes. Either Chi square (χ^2) or Fisher's exact test was used to find significant differences between cluster groups with the AC data. To compare the expression levels, the non-parametric Wilcoxon rank sum (WRS) test was used. This test uses a null hypothesis that is based on ranks rather than on normally distributed data. Before the above tests were performed, genes that were absent (AC) across all samples were removed; 5,483 genes were used in the analyses. Genes that were significant ($p < .0001$) for both the χ^2 test and the WRS test were considered to be significantly differentially expressed.

Clinical parameters were tested across multiple myeloma cluster groups. ANOVA test was used to test the continuous variables, and χ^2 test of

independence or Fisher's exact test was applied to test discrete variables. The natural log of the average difference data was used to find genes with a "spiked profile" of expression in multiple myeloma. Genes were identified that had low to undetectable expression in the majority of patients and normal samples (no more than 4 present absolute calls [P-AC]). A total of 2,030 genes fit the criteria of this analysis. The median expression value of each of the genes across all patient samples was determined. For the i^{th} gene, this value was called medgene (i). The i^{th} gene was a "spiked" gene if it had at least 4 patient expression values >2.5 + medgene (i). The constant 2.5 was based on the log of the average difference data. These genes that were "spiked" were further divided into subsets according to whether or not the largest spike had an average difference expression value greater than 10,000.

To determine transcriptional changes associated with human plasma cell differentiation, a total of 4866 genes were scanned across 7 cases each of tonsil B cells, tonsil plasma cells, and bone marrow plasma cells. The 4866 genes were derived from 6800 by filtering out all control genes, and genes not fulfilling the test of Max-Min <1.5 (1.5 being the natural log of the average difference). The χ^2 test was used to eliminate genes with absent absolute call (AAC). For example, in the tonsil plasma cell to bone marrow plasma cell comparison, genes with χ^2 values greater than 3.84 (p <0.05) or having "present" AC (PAC) in more than half of the samples in each group were retained. In the tonsil B cell to tonsil plasma cell and tonsil plasma cell to bone marrow plasma cell comparisons, 2662 and 2549 genes were retained as discriminating between the two groups, respectively. To compare gene expression levels, the non-parametric Wilcoxon Rank Sum (WRS) test was used to compare two groups using natural log transformed AD. The cutoff p value depended on the sample size, the heterogeneity of the two comparative populations (tonsil B cells, tonsil plasma cells, and bone marrow plasma cells showed a higher degree of stability in AD), and the degree of significance. In this analysis, 496 and 646 genes were found to be significantly (p <0.0005) differentially expressed in the tonsil B cell versus tonsil plasma cell and tonsil plasma cell versus bone marrow plasma cell comparisons, respectively. To define the direction of significance (expression changes being up or down in one group compared with the other), the non-parametric Spearman correlation test of the AD was employed.

Genes that were significantly differentially expressed in the tonsil B cell to tonsil plasma cell transition were referred as "early differentiation genes"

(EDGs) and those differentially expressed in the tonsil plasma cell to bone marrow plasma cell transition were referred as "late differentiation genes" (LDGs). Previously defined and novel genes were identified that significantly discriminated tonsil B cells from tonsil plasma cells (359 genes) and tonsil plasma cells from bone marrow plasma cells (500 genes).

To classify multiple myeloma with respect to EDG and LDG, 74 newly diagnosed cases of multiple myeloma and 7 tonsil B cell, 7 tonsil plasma cell, and 7 bone marrow plasma cell samples were tested for variance across the 359 EDGs and 500 LDGs. The top 50 EDGs that showed the most significant variance across all samples were defined as early differentiation genes for myeloma (EDG-MM). Likewise, the top 50 LDGs showing the most significant variance across all samples were identified as late differentiation genes for myeloma-1 (LDG-MM1). Subtracting the LDG-MM1 from the 500 LDG and then applying one-way ANOVA test for variance to the remaining genes identified the top 50 genes showing similarities between bone marrow plasma cells and multiple myeloma. These genes were defined as LDG-MM2.

Hierarchical clustering was applied to all samples using 30 of the 50 EDG-MM. A total of 20 genes were filtered out with Max-Min < 2.5. This filtering was performed on this group because many of the top 50 EDG-MM showed no variability across multiple myeloma and thus could not be used to distinguish multiple myeloma subgroups. A similar clustering strategy was employed to cluster the samples using the 50 LDG-MM1 and 50 LDG-MM2; however, in these cases all 50 significant genes were used in the cluster analysis.

EXAMPLE 4

RT-PCR And Immunohistochemistry

RT-PCR for the *FGFR3* MMSET was performed on the same cDNAs used in the microarray analysis. Briefly, cDNA was mixed with the IGJH2 (5'-CAATGGTCACCGTCTCTTCA-3', SEQ ID No. 1) primer and the MMSET primer (5'-CCTCAATTTCCTGAAATTGGTT-3', SEQ ID No. 2). PCR reactions consisted of 30 cycles with a 58° C annealing temperature and 1-minute extension time at 72° C using a Perkin–Elmer GeneAmp 2400 thermocycler (Wellesley, MA). PCR products were visualized by ethidium bromide staining after agarose gel electrophoresis.

Immunohistochemical staining was performed on a Ventana ES (Ventana Medical Systems, Tucson, AZ) using Zenker-fixed paraffin-embedded bone marrow sections, an avidin-biotin peroxidase complex technique (Ventana Medical Systems), and the antibody L26 (CD20, Ventana Medical Systems). Heat-induced epitope retrieval was performed by microwaving the sections for 28 minutes in a 1.0-mmol/L concentration of citrate buffer at pH 6.0.

EXAMPLE 5

Interphase FISH

For interphase detection of the t(11;14)(q13;q32) translocation fusion signal, a LSI IGH/CCND1 dual-color, dual-fusion translocation probe was used (Vysis, Inc, Downers Grove, IL). The TRI-FISH procedure used to analyze the samples has been previously described. Briefly, at least 100 clonotypic plasma cells identified by cIg staining were counted for the presence or absence of the translocation fusion signal in all samples except one, which yielded only 35 plasma cells. An multiple myeloma sample was defined as having the translocation when >25% of the cells contained the fusion.

EXAMPLE 6

Hierarchical Clustering of Plasma Cell Gene Expression Demonstrates Class Distinction

As a result of 656,000 measurements of gene expression in 118 plasma cell samples, altered gene expression in the multiple myeloma samples was identified. Two-dimensional hierarchical clustering differentiated cell types by gene expression when performed on 5,483 genes that were expressed in at least one of the 118 samples (Figure 1A). The sample dendrogram derived two major branches (Figure 1A and 1D). One branch contained all 31 normal samples and a single MGUS case whereas the second branch contained all 74 multiple myeloma and 4 MGUS cases and the 8 cell lines. The multiple myeloma-containing branch was further divided into two sub-branches, one containing the 4 MGUS and the other the 8 multiple myeloma cell lines. The cell lines were all clustered next to one another, thus showing a high degree of similarity in gene expression among the cell lines. This suggested that multiple myeloma could be differentiated from normal plasma cells and that at least

two different classes of multiple myeloma could be identified, one more similar to MGUS and the other similar to multiple myeloma cell lines.

Hierarchical clustering analysis with all 118 samples together with duplicate samples from 12 patients (plasma cells taken 24 hr or 48 hr after initial sample) were repeated to show reproducibility of the technique and analysis. All samples from the 12 patients studied longitudinally were found to cluster adjacent to one another. This indicated that gene expression in samples from the same patient were more similar to each other than they were to all other samples.

In addition to the demonstration of reproducibility of clustering noted above, three microarray analyses were also performed on a single source of RNA from one patient. When included in the cluster analysis, the three samples clustered adjacent to one another. Consistent with the manufacturer's specification, an analysis of the fold changes seen in the samples showed that <2% of all genes had a >2-fold difference. Hence, these data indicated reproducibility for same samples.

The clustergram (Figure 1A) showed that genes of unrelated sequence but similar function clustered tightly together along the vertical axis. For example, a particular cluster of 22 genes, primarily those encoding immunoglobulin molecules and major histocompatibility genes, had relatively low expression in multiple myeloma plasma cells and high expression in normal plasma cells (Figure 1B). This was anticipated, given that the plasma cells isolated from multiple myeloma are clonal and hence only express single immunoglobulin light-chain and heavy-chain variable and constant region genes, whereas plasma cells from normal donors are polyclonal and express many different genes of these two classes. Another cluster of 195 genes was highly enriched for numerous oncogenes/growth-related genes (e.g., *MYC, ABL1, PHB*, and *EXT2*), cell cycle-related genes (e.g., *CDC37*, *CDK4*, and *CKS2*), and translation machinery genes (*EIF2, EIF3, HTF4A*, and *TFIIA*) (Figure 1C). These genes were all highly expressed in MM, especially in multiple myeloma cell lines, but had low expression levels in normal plasma cells.

EXAMPLE 7

Hierarchical Clustering of Newly Diagnosed Multiple Myeloma Identifies Four Distinct Subgroups

Two-dimensional cluster analysis was performed on the 74 multiple myeloma cases alone. The sample dendrogram identified two major branches with

two distinct subgroups within each branch (Figure 1E). The four subgroups were designated MM1, MM2, MM3, and MM4 containing 20, 21, 15, and 18 patients respectively. The MM1 subgroup represented the patients whose plasma cells were most closely related to the MGUS plasma cells and MM4 were most like the multiple myeloma cell lines (see Figure 1D). These data suggested that the four gene expression subgroups were authentic and might represent four distinct clinical entities.

Differences in gene expression across the four subgroups were then examined using the χ^2 and WRS tests (Table 1). As expected the largest difference was between MM1 and MM4 (205 genes) and the smallest difference was between MM1 and MM2 (24 genes). Next, the top 30 genes turned on or upregulated in MM4 as compared with MM1 were examined (Table 2). The data demonstrated that 13 of 30 most significant genes (10 of the top 15 genes) were involved in DNA replication/repair or cell cycle control. Thymidylate synthase (*TYMS*), which was present in all 18 samples comprising the MM4 subgroup, was only present in 3 of the 20 MM1 samples and represented the most significant gene in the χ^2 test. The DNA mismatch repair gene, mutS (*E. coli*) homolog 2 (*MSH2*) with a WRS *p* value of 2.8 x 10^{-6} was the most significant gene in the WRS test. Other notable genes in the list included the CAAX farnesyltransferase (*FNTA*), the transcription factors enhancer of zeste homolog 2 (*EZH2*) and *MYC*-associated zinc finger protein (*MAZ*), eukaryotic translation initiation factors (*EIF2S1* and *EIF2B1*), as well as the mitochondrial translation initiation factor 2 (*MTIF2*), the chaperone (*CCT4*), the UDP-glucose pyrophosphorylase 2 (*IUGP2*), and the 26S proteasome-associated pad1 homolog (*POH1*).

To assess the validity of the clusters with respect to clinical features, correlations of various clinical parameters across the 4 subgroups were analyzed (Table 3). Of 17 clinical variables tested, the presence of an abnormal karyotype (p = .0003) and serum β2M levels (p = .0005) were significantly different among the four subgroups and increased creatinine (p = .06) and cytogenetic deletion of chromosome 13 (p = .09) were marginally significant. The trend was to have higher β2M and creatinine as well as an abnormal karyotype and chromosome 13 deletion in the MM4 subgroup as compared with the other 3 subgroups.

TABLE 1. Differences In Gene Expression Among Multiple Myeloma Subgroups

Comparison	Range of WRS* *p* Values	Number of Genes
MM1 vs MM4	.00097 to 9.58×10^{-7}	205
MM2 vs MM4	.00095 to 1.0410^{-6}	162
MM3 vs MM4	.00098 to 3.7510^{-6}	119
MM1 vs MM3	.00091 to 6.2710^{-6}	68
MM2 vs MM3	.00097 to 1.9810^{-5}	44
MM1 vs MM2	.00083 to 2.9310^{-5}	24

*Wilcoxon rank sum test. Comparisons are ordered based on the number of significant genes. Comparisons have a WRS *p* value < 0.001.

TABLE 2. The 30 Most Differentially Expressed Genes In A Comparison Of MM1 And MM4 Subgroups

Accession	Function	Gene Symbol	MM1 (N=20)	MM4 (N=18)	Chi Square	WRS‡ *p* Value
D00596	DNA replication	*TYMS*	3	18	24.35	1.26×10^{-4}
U35835	DNA repair	*PRKDC*	2	17	23.75	4.55×10^{-6}
U77949	DNA replication	*CDC6*	1	13	15.62	5.14×10^{-6}
U91985	DNA fragmentation	*DFFA*	1	12	13.38	6.26×10^{-5}
U61145	transcription	*EZH2*	4	15	12.77	1.67×10^{-4}
U20979	DNA replication	*CHAF1A*	2	12	10.75	1.10×10^{-4}
U03911	DNA repair	*MSH2*	0	9	10.48	2.88×10^{-6}
X74330	DNA replication	*PRIM1*	0	9	10.48	9.36×10^{-6}
X12517	SnRNP	*SNRPC*	0	9	10.48	5.26×10^{-6}
D85131	transcription	*MAZ*	0	9	10.48	1.08×10^{-5}
L00634	farnesyltransferase	*FNTA*	10	18	9.77	7.28×10^{-5}
U21090	DNA replication	*POLD2*	11	18	8.27	8.05×10^{-5}
X54941	cell cycle	*CKS1*	10	17	7.07	1.26×10^{-4}
U62136	cell cycle	*UBE2V2*	13	18	5.57	4.96×10^{-6}
D38076	cell cycle	*RANBP1*	13	18	5.57	7.34×10^{-6}
X95592	unknown	C1D†	13	18	5.57	1.10×10^{-4}
X66899	cell cycle	*EWSR1*	14	18	4.35	1.89×10^{-4}
L34600	translation initiation	*MTIF2*	14	18	4.35	3.09×10^{-5}
U27460	Metabolism	*IUGP2*	15	18	3.22	1.65×10^{-4}
U15009	SnRNP	*SNRPD3*	15	18	3.22	1.47×10^{-5}
J02645	translation initiation	*EIF2S1*	16	18	2.18	7.29×10^{-5}

Accession	Function	Gene Symbol	MM1 (N=20)	MM4 (N=18)	Chi Square	WRS‡ *p* Value
X95648	translation initiation	*EIF2B1*	16	18	2.18	$1.45x10^{-4}$
M34539	calcium signaling	*FKBP1A*	18	18	0.42	$1.71x10^{-5}$
J04611	DNA repair	*G22P1*	18	18	0.42	$7.29x10^{-5}$
U67122	anti-apoptosis	*UBL1*	20	18	0.00	$7.29x10^{-5}$
U38846	chaperon	*CCT4*	20	18	0.00	$1.26x10^{-4}$
U80040	metabolism	*ACO2*	20	18	0.00	$8.38x10^{-5}$
U86782	proteasome	*POH*†	20	18	0.00	$5.90x10^{-5}$
X57152	signaling	*CSNK2B*	20	18	0.00	$7.29x10^{-5}$
D87446	unknown	*KIAA0257*	20	18	0.00	$1.26x10^{-5}$

Accession numbers listed are GeneBank numbers. † Gene symbol not HUGO approved.

‡ Wilcoxon rank sum test.

TABLE 3. Clinical Parameters Linked To Multiple Myeloma Subgroups

	Multiple Myeloma Subgroups				
Clinical Parameter	1	2	3	4	*p* value
Abnormal cytogenetics	40.0%	5.3%	53.3%	72.2%	.00028
Average β2-microglobulin (mg/L)	2.81	2.73	4.62	8.81	.00047

ANOVA, Chi square, and Fisher's exact tests were used to determine significance.

EXAMPLE 8

Gene Expression "Spikes" in Subsets of Multiple Myeloma

A total of 156 genes not identified as differently expressed in the statistical analysis of multiple myeloma versus normal plasma cells, yet highly overexpressed in subsets of multiple myeloma, were also identified. A total of 25 genes with an AD spike greater than 10,000 in at least one sample are shown (Table 4). With 27 spikes, the adhesion associated gene *FBLN2* was the most frequently spiked. The gene for the interferon induced protein 27, *IFI27,* with 25 spikes was the second most frequently spiked gene and contained the highest number of spikes over 10,000 (N = 14). The *FGFR3* gene was spiked in 9 of the 74 cases (Figure 2A). It was the only gene for which all spikes were greater than 10,000 AD. In fact, the lowest AD value was 18,961and the highest 62,515, which represented the highest of all spikes. The finding of *FGFR3* spikes suggested that these spikes were induced by the multiple myeloma-specific, *FGFR3*-activating t(4;14)(p21;q32) translocation. To

test the above hypothesis, RT-PCR for a t(4;14)(p21;q32) translocation-specific fusion transcript between the *IGH* locus and the gene *MMSET* was performed (data not shown). The translocation-specific transcript was present in all 9 *FGFR3* spike samples but was absent in 5 samples that did not express *FGFR3*. These data suggested that the spike was caused by the t(4;14)(p21;q32) translocation.

The *CCND1* gene was spiked with AD values greater than 10,000 in 13 cases. TRI-FISH analysis for the t(11;14)(q13;q32) translocation was performed (Table 5). All 11 evaluable samples were positive for the t(11;14)(q13;q32) translocation by TRI-FISH; 2 samples were not analyzable due to loss of cell integrity during storage. Thus, all *FGFR3* and *CCND1* spikes could be accounted for by the presence of either the t(4;14)(p21;q32) translocation or the t(11;14)(q13;q32) translocation respectively. Next, the distribution of the *FGFR3, CST6, IFI27*, and *CCND1* spikes within the gene expression-defined multiple myeloma subgroups was determined (Figure 2). *FGFR3* and *CST6* spikes were more likely to be found in MM1 or MM2 ($p < .005$) whereas the spikes for *IFI27* were associated with an MM3 and MM4 distribution ($p < .005$). *CCND1* spikes were not associated with any specific subgroup ($p > .1$). Both *CST6* and *CCND1* map to 11q13 and had no overlap in spikes. The five spikes for *MS4A2* (CD20) were found in either the MM1 (3 spikes) or MM2 (2 spikes) subgroups.

The gene *MS4A2* which codes for the CD20 molecule was also found as a spiked gene in four cases. To investigate whether spiked gene expression correlated with protein expression, immunohistochemistry for CD20 was performed on biopsies from multiple myeloma samples (Figure 3B). All four cases that had spiked *MS4A2* gene expression were positive for CD20 protein expression, whereas 11 that had no *MS4A2* gene expression were negative for CD20. To validate to the gene expression profiling, a comparison of CD marker protein and gene expression in the multiple myeloma cell line CAG and the EBV-transformed lymphoblastoid cell line ARH-77 were performed. The expression of CD138 and CD38 protein and gene expression was high in CAG but absent in ARH-77 cells. Expression of CD19, CD20, CD21, CD22, CD45, and CDw52 was in ARH-77 and absent in CAG cells. The nearly 100% coincidence of *FGFR3* or *CCND1* spiked gene expression with the presence of the t(4;14)(p14;q32) or t(11;14)(q13;q32) translocation; the correlation of CD20 and *MS42A* gene expression in primary multiple myeloma; and correlation of

CD marker protein and gene expression in B cells and plasma cells represent important validations of the accuracy of the gene expression profiling herein.

TABLE 4. Genes with "Spiked" Expression in Plasma Cells from Newly Diagnosed Multiple Myeloma Patients

Accession No.	Function	Gene Symbol	# of Spikes	Spikes >10K	Max Spike
M64347	signaling	*FGFR3*	9	9	62,515
U89922	immunity	*LTB*	4	2	49,261
X67325	interferon signaling	*IFI27*	25	14	47,072
X59798	cell cycle	*CCND1*	6	13	42,814
U62800	cysteine protease inhibitor	*CST6*	17	6	36,081
U35340	eye lens protein	*CRYBB1*	4	1	35,713
X12530	B-cell signaling	*MS4A2*	5	5	34,487
X59766	unknown	*AZGP1*	18	4	28,523
U58096	unknown	*TSPY*	4	1	23,325
U52513	interferon signaling	*IFIT4*	5	2	21,078
X76223	vesicular trafficking	*MAL*	19	5	20,432
X92689	O-linked glycosylation	*GALNT3*	4	1	18,344
D17427	adhesion	*DSC3*	8	7	17,616
L11329	signaling	*DUSP2*	14	1	15,962
L13210	adhesion, macrophage lectin	*LGALS3BP*	8	2	14,876
U10991	unknown	*G2*†	7	1	14,815
L10373	integral membrane protein	*TM4SF2*	4	2	14,506
U60873	unknown	*137308*	12	1	12,751
M65292	complement regulation	*HFL1*	9	1	12,718
HT4215	phospholipid transport	*PLTP*	23	1	12,031
D13168	growth factor receptor	*ENDRB*	18	1	11,707
AC002077	signaling	*GNAT1*	21	1	11,469
M92934	growth factor	*CTGF*	4	1	11,201
X82494	adhesion	*FBLN2*	27	7	10,648
M30703	*growth factor*	AR	*5*	*1*	*10,163*

Accession numbers listed are GeneBank numbers, except those beginning with "HT", which are provided by the Institute of Genomic Research (TIGR). † Gene symbol not HUGO approved.

TABLE 5. Correlation of *CCND1* Spikes with FISH-Defined t(11;14)(q13;q32)

GC PT*	*CCND1* Spike (AD value)†	FISH t(11;14)	Percent PCs with Translocation	Cells Counted
P168	42,813	Yes	59%	113
P251	33,042	Yes	80%	124
P91	31,030	Not done	—	—
P99	29,862	Yes	65%	111
P85	26,737	Yes	92%	124
P241	25,611	Yes	96%	114
P56	23,654	Yes	100%	106
P63	22,358	Yes	98%	104
P199	18,761	Yes	60%	35
P107	15,205	Yes	100%	147
P75	14,642	Yes	100%	105
P187	14,295	Yes	25%	133
P124	10,594	Not done	—	—

*GC PT = patient identifier; † AD = average difference call.

EXAMPLE 9

A 15-Gene Model Representing Seven Subgroups of Multiple Myeloma

In this example, feature-subset selection was used to extract genes relevant to specific myeloma subtypes. In this regard, multivariate stepwise discriminant analysis (MSDA) was applied and identified 15 genes that could correctly separate tissue samples into seven subtypes. This small, subtle set of genes gave a final prediction accuracy of more than 92% and 95% in a training and an independent testing sets, respectively. Examining the expression of the 15 genes identified here not only would provide important new insights into the diagnosis and pathogenesis of these myeloma subtypes but also may pinpoint useful targets against which therapeutic agents could be developed.

Identification of Myeloma Subtypes

Samples from patients with newly diagnosed multiple myeloma (n = 351) were studied by microarray analysis. Plasma cell isolation, RNA extraction, and hybridization to GeneChip® U133 Plus 2.0 (Affymetrix) Arrays with 35,000 unique genes were performed as described previously (Zhan et al., 2002a, 2003). All data

used in the following analyses were derived with the Affymetrix Microarray Suite (MAS) 5.0 software. Affymetrix signal call was transformed by log base 2 and normalized to give a mean of 0 and variance of 1 for each sample.

The software package SPSS 11.0 (SPSS Inc.) and the significance analysis of microarrays (SAM) method were used (Tusher et al., 2001). Genes were selected for analysis on the basis of detection and false discovery rate. In each comparison, genes having "present" detection calls in more than one-third of the samples in the overexpressed gene group or a chi-square value >3.84 ($p < 0.5$) were retained for statistical analysis. For two-class and multiclass supervised analyses, the SAM method was used with sample-label permutations to evaluate statistical significance. Student's t test and the chi-square test were performed to identify survival-related genes. Hierarchic clustering or average linkage clustering with the centered correlation metric was employed (Eisen et al., 1998). A total of 2,856 probe sets were scanned across 177 myeloma patients in the training set. The 2,856 probe sets across all samples were derived from 57,000 sets by filtering out all genes with "absent" detection calls and genes not fulfilling the test of standard deviation >0.6 (0.6 being the normalized signal). For class-prediction analysis, the software tool called Prediction Analysis for Microarrays (PAM) in R version 1.7.1 (Tibshirani et al., 2002) was used. The method of the nearest shrunken centroid identified a subgroup of genes that best characterized a predefined class. The prediction error was calculated by means of 10-fold cross-validation within the training set followed by the use of a second validation set (almost one-half of the patients). Multivariate analysis based on stepwise multiple linear discriminant analysis (MSDA) with the Wilks lambda criterion was used to select the best combination of genes that differentiated myeloma subgroups. Figure 5 shows the overview of the strategy used to classify and validate myeloma subgroups based on gene expression profiles. Figure 6 shows the results of hierarchic cluster analysis of myeloma samples. The dendrogram of the 177 cases at the top of the matrix, in which the pattern and length of branches reflect the relatedness of the samples, indicates that the samples are clustered into seven branches on the basis of similar gene expression profiles.

Distinct Subtypes Associated With Recurrent Translocations

The distribution of spikes of *FGFR3/MMSET*, *MAF*, *MAFB*, *CCND1*, and *CCND3* within the myeloma subgroups described above was determined next.

These genes are known as the most translocation-related genes in the 177-sample training set (Figure 7). Myeloma samples exhibiting one of the five spikes accounted for 40% of the total. To determine chromosome 14q32 translocation, triple-color interphase FISH (TRI-FISH) protocol described by Shaughnessy et al. (2003) and Santra et al. (2003) was followed, and an LSI® IGH dual-color, break-apart rearrangement probe (Vysis Inc.) was used. This probe is a mixture of two probes that hybridize to opposite sides of the J segment through constant regions of the IgH locus. Approximately 900-kb SpectrumGreen™-labeled LSI® IGH probe covers essentially the entire IgH variable region. The hybridization target of approximately 250-kb SpectrumOrange™-labeled LSI® IGH 3′ flanking probe lies completely 3′ to the IgH locus. As a result of this probe design, any translocation with a breakpoint at the J segment or within switch sequences produced separate orange and green signals. For the purpose of this analysis, 14q required split IgH signal in at least 20% of approximately 100 Ig light chain–restricted clonal plasma cells.

With the use of TRI-FISH and four different probe combinations of *VH/CH,* disassociation of the *VH* from the *CH* probes was observed in 100% (13/13) of *MMSET*, 100% (3/3) of *MAF/MAFB*, 100% (15/15) of *CCND1*, and 100% (5/5) of *CCND3* spike samples. In view of the exact correlation of spikes with the translocations, all 24 *MMSET,* 5 *MAF,* 5 *MAFB,* 34 *CCND1,*and 4 *CCND3* spikes could be accounted for by the presence, respectively, of the t(4;14)(p21;q32), t(14;16)(q32;q23), t(14;20)(q32;q13), t(11;14)(q13;q32), and t(6;14)(p21;q32) translocations. With the exception of the *CCND3* spike, which was randomly distributed among subgroups, the other spikes were clearly clustered together.

CCND1-IgH. Thirty (88%) of 34 myeloma samples with a *CCND1* spike, which was caused by t(11;14)(q13;q32), were grouped in cluster III (Figure ~~10~~ 7A). Five specimens in this cluster did not show a spike. SAM and chi-square analysis revealed that *CCND1* was the most discriminative gene for this cluster (14-fold increase). Interestingly, two early-B-cell factor target genes, *VPREB3* and *CD20*, were highly correlated with this cluster (rank numbers 2 and 7 by mean ratio, respectively), perhaps because of increased expression of *PAX5* with *CCND1* activation.

MMSET-IgH or FGFR3-IgH. In group VII, 92% (22/24) of cases involved myeloma with t(4;14)(p21;q32) (Figure ~~10~~ 7A). Two specimens in this cluster showed neither a *FGFR3* or an *MMSET* spike, but TRI-FISH for Ig heavy

chain revealed a disassociation of *VH* from the *CH* probes. SAM and chi-square analysis showed that genes for Krüppel-like factor 4 (*KLF4*) and desmoglein 2 (*DSG2*) were extremely up-regulated in this class, whereas the *Wnt*-signaling antagonist *FRZB* and transcription factor *ISL2* were the most down-regulated genes in this group.

MAF-IgH and MAFB-IgH. All five *MAFB* spikes representing t(14;20)(q32;q13) and five of six *MAF* spikes representing t(14;16)(q32;q23) were included in group V. SAM and chi-square analysis demonstrated a strong up-regulation of Toll-like-receptor 4 (*TLR4*), chemokine (C-X3-C motif) receptor 1(*CX3CR1*), osteopontin (*SPP1*), and *CCND2* genes. The activation of *CX3CR1* in mast cells was identified via *TLR4* signaling; the high correlation probably suggests that cell-cycle-control gene *CCND2* and tumor adhesion- and metastasis-associated gene *SPP1* were activated through *TLR4* signaling by *MAF* or *MAFB* activation.

Distinct Subtypes Associated with Cytogenetic Abnormality

Cytogenetic abnormality was observed in 33% (51/153) of patients; 67% of patients with multiple myeloma had normal karyotype. Of the 51 patients with a cytogenetic abnormality, 29 (57%) had hyperdiploidy and 22 (43%) hypodiploidy. Group IV contained most of the cases of myeloma with cytogenetic abnormality (76%). SAM and chi-square analysis revealed that genes controlling cell cycle, such as *CCNB2, CCNB1, MCM2, CDCA2, BUB1,CDC2*, and *TYMS*, and cancer–testis antigens, including *MAGEA6, MAGEA3, GAGE1*, and *GAGE4*, were the most up-regulated genes. Samples from patients in group I exhibited the lowest frequency for cytogenetic abnormality; only 14% (4/28) of specimens demonstrated an abnormal karyotype. Up-regulation of genes in this group, including Ig genes, neutrophil cell marker *DEFA*, and macrophage marker *CD163*, indicated that the samples had a strong immune response and were probably contaminated by some normal plasma cells and other types of cells.

Consistent with other reports, patients with IgH translocations were significantly more likely to have hypodiploid myeloma than those without IgH translocations as detected by cytogenetics. Patients with hypodiploid myeloma were detected in most translocation groups (57%; $p = 0.002$), including group III with a *CCND1* spike, group V with a *MAF/MAFB* spike, and group VII with a *MMSET/FGFR3* spike, whereas patients with hyperdiploid myeloma were less likely

to have IgH translocations and their disease fell mostly into group II (40%). The up-regulated genes in group II distinguished the specimens contained in it: guanine nucleotide binding protein (*GNG11*), wingless-type MMTV integration site family member 4 (*Wnt4*), oncogene *c-Kit*, and estrogen receptor 2 (*ESR2*). The *Wnt* signaling antagonists *DKK1* and *FRZB* were significantly up-regulated in this cluster.

The top uniquely and spike-like expressed gene in group VI was cystatin M (*CST6*), which occurred in 10% (17/177) of myeloma specimens. Most specimens in this group had a normal karyotype, which predicts a favorable outcome. Up-regulated genes included tumor transformation-related gene *SCAPIN1*, interleukin-6 receptor gene *IL6R*, and pre-B-cell colony-enhancing factor gene *PBEF*.

Validation And Prediction Studies

Genes discriminating the seven subgroups were selected by SAM and chi-square analysis with a 5,000-permutation adjustment. The number of discriminant probe sets per myeloma subtype, determined at a statistical significance level that had a false discovery rate of less than 1%, were as follows: group I, 404 probe sets; group II, 2,368 probe sets; group III, 1,879 probe sets; group IV, 749 probe sets; group V, 398 probe sets; group VI, 238 probe sets; and group VII, 525 probe sets. The marked differences in the number of discriminating genes for the various myeloma subtypes suggest that significant differences exist in the global gene expression profiles of multiple myeloma that are likely to depend on specific defining genetic lesions.

To evaluate whether those uniquely expressed genes in the subgroups were sufficiently robust for subclasses to be correctly assigned, PAM (prediction analysis for microarrays) based on nearest shrunken centroids to devise a cross-validated gene expression predictor with the top 50 ranked probe sets from each of the seven subgroups was then used. With 350 features (probe sets), PAM predicted correct diagnosis in 91% of training specimens. Figure 7C illustrates seven distinct groups of either overexpressed or underexpressed genes in a two-dimensional clustering analysis.

PAM was then used to predict the classes of 174 independently tested cases. Each test-set sample was individually assigned to a subgroup by determining whether its expression signature across the 350 predictive genes was more highly correlated with the average signature in the seven subtype training sets. With this

procedure, the same seven subgroups that were highly correlated with the training group were defined in the testing set, which yielded seven distinct hot spots (Figure 7D). Notably, when the same procedure was applied to subgroups of samples with recurrent translocation and ploidy change, it also identified classes associated with recurrent translocations and hyperdiploidy or hypodiploidy. A strong correspondence was observed between samples represented in subtype groups II, V, and VII and those demonstrating *CCND1*, *MAF/MAFB*, and *MMSET* spikes, respectively (Figure 7D). In addition, group III contained myeloma subtypes that were mostly hyperdiploid and rarely hypodiploid (p = 0.012; Table 7), whereas groups associated with recurrent translocations contained more myeloma subtypes with hypodiploidy (Table 7).

Prognosis And Clinical Parameters In Both Training And Testing Groups

To investigate whether there are significant differences in outcome in patient cohorts from the seven groups, Kaplan-Meier analysis was performed for both event-free survival and overall survival. The probability of event-free survival and overall survival were significantly or marginally significantly different among the seven subtypes in both the training (p = 0.053) and the testing (p = 0.047) sets (data not shown). Because the event percentage was small (12%), the seven subgroups were further combined into two risk groups based on visual appearance of the survival curves. Groups I, II, III, and VI were combined as group A, and groups IV, V, and VII as group B. Notably, it was clear that group A was highly correlated with a good outcome and group B with a poor outcome with respect to event-free survival and overall survival in both training (Figures 8A and D) and testing (Figures 8B and E) sets. Group A is thus a low risk group, and group B is a high risk group.

To further assess the validity of the clusters with respect to clinical features, correlations of various clinical parameters were analyzed across the seven subgroups in both training and testing sets (Tables 6 and 7). Of 15 clinical variables tested, increases in IgA isotype ($p < 0.001$) and serum β_2-microglobulin ($p < 0.05$) were significantly different in groups IV, V, and VII. These clinical features, together with the gene expression profile results, were consistent with a poor outcome for patients in groups IV, V, and VII and with a favorable clinical course for patients in groups I, II, III, and VI.

Importantly, the incidence of lytic bone lesion disease was significantly different among the seven subgroups (Tables 10 and 11). In group VI,

the frequency of a lytic bone lesion in training and testing sets was, respectively, 30% (average 70%) and 50% (average 80%) for patients with at least one lesion on magnetic resonance imaging (MRI) and only 23% (average 60%) and 25% (average, 65%) in patients with at least three lesions on MRI. Expression levels of *FRZB* and *DKK1*, the *Wnt*-signaling antagonist, and the association of lytic bone lesions with myeloma were significantly lower in group VI than in other groups.

Subgroup-Representative Genes, A 15-Gene Model

A search for smaller numbers of genes that could be used to distinguish the classes on the basis of expression signature of individual samples was performed. The number of genes required for an optimal assignment of class varied among the classes. Because the spike genes associated with 14q translocation were easy to identify, a signal gene such as *FGFR3* or *MMSET*, *MAF* or *MAFB*, and *CCND1*, was sufficient to give an almost 100% accuracy for translocation-related groups. Thus, 40% of patients were assigned to one of the three translocation subgroups. Next, using the signal intensity of the top 50 significant genes from each of the other four subgroups, we applied multivariate stepwise discriminant analysis (MSDA) and identified all the possible representative genes that could correctly distinguish tissue samples. A set of 120 samples from nontranslocation groups after defining the translocation subgroups was used. Among 200 genes, MSDA identified 10 genes that had an accuracy of 84% in predicting the sample types in the four nontranslocation subgroups and an accuracy of more than 92% in predicting the sample types in all seven training subgroups.

Next, the three translocation-related groups were defined by searching for spikes of *CCND1*, *MAF/MAFB*, and *MMSET* in an independent set of 174 samples (validation set). Sixty-three cases (36%) were assigned to one of the three translocation subgroups. The remaining 111 patients, corresponding to 31 samples in group I, 49 samples in group II, 16 samples in group IV, and 15 samples in group VI, were assigned to one of the four nontranslocation groups with the 10 genes identified by MSDA in the training set. MSDA was again performed on these 111 testing samples and achieved a 90% accuracy from these four groups. A final prediction accuracy of more than 95% was achieved for the entire 174-sample validation set. Figure 9 illustrates the 15-gene model representing the seven subgroups in both training and testing sets.

In addition to the recurrent translocation-related genes, we discovered several genes implicated in diverse biological processes and pathways. Genes that distinguished group I from the other groups included secreted Ig light chain gene and fibrinogen-like 2 (*FGL2*). Genes that discriminated group II from the remaining groups included tumor necrosis factor–related apoptosis-inducing ligand (*TRAIL*), signal transduction G-protein–coupled receptor protein signaling pathway (*GNG11*), and insulin gene enhancer protein (*ISL2*). Genes distinguishing group IV from other groups included those implicated in the onset of DNA replication and cell division (*MCM2*) and an unknown gene. Genes for two *RAS*-related signaling molecules, *RAB7L1* and *HRASLS,* and for mitogen-activated protein kinase kinase kinase kinase 3 (*MAP4K3*) were shown to distinguish group VI from the other groups.

To further test the validity of this model, we analyzed clinical parameters and event-free and overall survival (n=176). Seven new subgroups were generated with the 15-gene-model, and Kaplan-Meier analysis was performed to determine event-free and overall survival and clinical parameters. Patients with samples in groups IV, V, and VII (group B) had a poorer outcome than those with samples in groups I, II, III, and VI (group A) (Figures 8C and F; Table 8). Table 9 shows a comparison of patient classification using 350 genes and the 15-gene model.

Summary

Gene expression profiles (GEP) of ~33,000 genes were determined in CD-138 enriched plasma cells from 351 newly diagnosed multiple myeloma cases. A total of 2,800 genes exhibited a high degree of expression variation (S.D. >0.6) across a training set of 177 cases. Unsupervised hierarchical cluster analysis identified seven distinct subgroups with strong correlations to known translocations and cytogenetic abnormalities. The sample dendrogram created two major branches. Cases of multiple myeloma that tended to have normal karyotypes (Group I, n=38) and hyperdiploid karyotypes (Group II, n=36) were located on one branch. These two are cytogenetically-defined low risk entities. The other main branch consisted of 5 minor branches and each branch had a unique constellation of genes and cytogenetic features. *CCND1* spikes, a feature indicative of the presence of a t(11;14)(q13q34) translocation, defined Group III (n=30). Group IV (n=16) was defined by a proliferation-signature either in the context of a hyper or non-hyperdiploid karyotype. *MAF* and *MAFB* spikes, indicative of the presence of either the t(14;16)(q34;p16) or

t(14;20)(q34) translocations, clustered together and defined Group V (n=10). Group VI (n=21), which was not defined by any of the above parameters, was characterized by elevated expression of *CST6*. Cases with *MMSET* spikes resulting from a t(4;14)(p16;q32) translocation defined Group VII (n=26).

These groups also differed with respect to multiple clinical characteristics. Multiple myeloma cases with an IgA isotype were predominant in Groups V, VI, and VII ($p = 0.009$). β2M levels were significantly elevated in Group IV (mean 7.66 mg/L) and Group V (mean 9.09 mg/L) ($p = 0.003$). Creatinine (≥2.0 mg/dL) ($p = 0.005$) and LDH (UI/L) ($p = 0.003$) were significantly elevated in Group IV. Cytogenetic abnormalities predominated in Groups IV and VII ($p = 0.0002$). When abnormal karyotypes were present, they tended to be hyperdiploid in Groups II and IV ($p = 0.001$) and hypodiploid in Groups IV and VII ($p = 0.002$). Group VI had a low incidence of MRI-defined focal lesions < 3 MRI lesions ($p = 0.001$). Finally, as expected, FISH detectable 14q32 translocations predominated in Groups III, V, and VII ($p < 0.0001$).

SAM (significance analysis of microarrays) and Chi-square analysis were used to identify 50 genes whose expression uniquely defined each of these seven subgroups. Combining the 350 derived genes (7 x 50), a PAM analysis (prediction analysis for microarrays) was performed on the 177 training set to determine the degree of correlation with the original unsupervised cluster analysis. This analysis revealed 91% accuracy between the two tests. The 350 gene PAM model was then applied to a test group of 174 cases. The resulting cluster designations were highly correlated with a similar distribution of cytogenetic and clinical parameters seen in training and test sets. Kaplan-Meier test of event-free survival (EFS) and overall survival (OS) on the seven different groups resulted in a marginal separation. Based on the trends in these data and previously published data, the groups were separated into low-risk (Groups I, II, III, VI) and high-risk (Groups IV, V, VII) and Kaplan-Meier tests were repeated. The data showed significant differences in both the training set (EFS, $p < 0.0002$; OS, $p = 0.0033$) and the test set (EFS, $p = 0.021$; OS, $p = 0.0015$).

Multivariate Discriminant Analysis was then used to derive a model based on the expression levels of 15 genes that could accurately discriminate the seven groups in training and test sets. When all 351 cases were classified by the 15 gene model and characterized as low and high risk as defined above, Kapalan Meier

analysis revealed significant differences in both event-free survival and overall survival ($p < 0.05$). Taken together these data suggests that multiple myeloma is broad descriptor of seven discrete molecular entities each with unique mechanisms of transformation that give rise to group-specific deregulated gene expression and differing outcomes. A small scale quantitative RT-PCR-based assay based on the 15-gene model could allow rapid application of this new 15 gene classification schema as a novel risk stratification model.

TABLE 6. Characteristic of 177 Multiple Myeloma Patients In The Training Set In Relation To The Seven Subgroups Identified By Hierarchical Clustering

Group	I	II	III	IV	V	VI	VII	*P*-value
Age>=65	12/38 (32%	6/36 (17	5/30 (17%)	2/16 (13%)	3/10 (30%)	5/21 (24%)	4/26 (15%)	0.563*
IgA Isotype	6/32 (19%)	6/33 (18%)	4/24 (17%)	2/15 (13%)	**5/10 (50%)**	**7/20 (35%)**	**14/26 (54%)**	0.009 *
Albumin<3.5 g/dL	4/38 (11%)	5/36 (14%)	5/30 (17%)	4/16 (25%)	0/10 (0%)	3/21 (14%)	4/26 (15%)	0.741 *
B2M (mg/L)	2.73	3.49	5.68	**7.66**	**9.09**	5.59	4.26	0.003#
CRP>=4 mg/L	19/38 (50%)	19/34 (56%)	17/30 (57%)	10/16 (63%)	4/10 (40%)	10/21 (48%)	13/25 (52%)	0.927
Cytogenetics abnormality	7/38 (18%)	15/36 (41%)	7/30 (23%)	**12/16 (75%)**	2/10 (20%)	4/21 (19%)	**14/26 (53%)**	0.000 2
Hyperdiploidy	4/38 (10%)	**13/36 (36%)**	0/30 (0%)	**6/16 (37%)**	1/10 (10%)	2/21 (9%)	5/26 (19%)	0.001 1
Hypodiploidy	2/38 (5%)	3/36 (8%)	5/30 (16%)	**6/16 (37%)**	1/10 (10%)	2/21 (9%)	**10/26 (38%)**	0.002 2
MRI>1	28/36 (77%)	26/36 (72%)	28/30 (93%)	14/15 (93%)	7/10 (70%)	**10/20 (50%)**	18/26 (69%)	0.014 4
MRI>3	24/36 (66%)	19/36 (52%)	22/30 (73%)	14/15 (93%)	4/10 (40%)	**5/20 (25%)**	16/26 (61%)	0.001 0
Chr14>20	8/16 (50%)	7/21 (33%)	**15/15 (100%)**	8/12 (66%)	**3/3 (100%)**	4/10 (40%)	**13/13 (100%)**	3.82x 10^{-5}
LDH(UI/L)	183.11	172.69	179.27	**246.25**	187.10	160.52	155.04	0.003#
Creatinine>=2.0 mg/dL	0/38 (0%)	1/36 (3%)	5/30 (17%)	**4/16 (25%)**	2/10 (20%)	4/21 (19%)	2/26 (8%)	0.005 *
HGB<10 g/dL	8/38 (21%)	6/36 (17%)	6/30 (20%)	5/16 (31%)	2/10 (20%)	8/21 (38%)	7/26 (27%)	0.618 *

* Fisher's exact test was used. # ANOVA test was used. The top significant group was highlighted with bold number.

TABLE 7. Characteristic of 174 Multiple Myeloma Patients In The Testing Set In Relation To The Seven Subgroups Identified By PAM Prediction

Group	I	II	III	IV	V	VI	VII	*P*-value
Age>=65	9/31 (29%)	6/49 (12%)	7/29 (24%)	3/16 (19%)	2/10 (20%)	1/15 (7%)	4/24 (17%)	0.468*
IgA Isotype	4/29 (14%)	4/45 (9%)	5/25 (20%)	3/14 (21%)	**5/9 (56%)**	**8/12 (67%)**	**10/22 (45%)**	<.001*
Albumin<3.5 g/dL	2/31 (6%)	6/49 (12%)	6/29 (21%)	**7/16 (44%)**	1/10 (10%)	1/15 (7%)	4/24 (17%)	0.058*
B2M(mg/L)	3.3	3.63	4.73	**7.81**	**8.04**	3.57	4.52	0.029#
CRP>=4 mg/L	18/31 (58%)	26/49 (53%)	17/29 (59%)	10/16 (63%)	5/9 (56%)	8/15 (53%)	9/24 (38%)	0.743*
Cytogenetics abnormality	5/31 (16%)	21/49 (42%)	7/29 (24%)	**9/16 (56%)**	**5/10 (50%)**	2/15 (13%)	**14/24 (58%)**	0.0025
Hyperdiploidy	4/31 (12%)	18/49 (36%)	1/29 (3%)	4/16 (25%)	1/10 (10%)	2/15 (13%)	4/24 (16%)	0.012
Hypodiploidy	1/31 (3%)	3/49 (6%)	2/29 (6%)	4/16 (25%)	**5/10 (50%)**	0/15 (0%)	**10/24 (41%)**	3.82×10^{-6}
MRI>1	29/30 (96%)	36/46 (78%)	21/27 (77%)	15/16 (93%)	6/9 (66%)	**4/13 (30%)**	15/24 (62%)	0.0001
MRI>3	21/30 (70%)	26/46 (56%)	18/27 (66%)	11/16 (68%)	3/9 (33%)	**3/13 (23%)**	12/24 (50%)	0.049
Chr14>20	7/17 (41%)	8/29 (27%)	**9/11 (81%)**	4/9 (44%)	**3/3 (100%)**	3/7 (42%)	**15/15 (100%)**	6.94×10^{-5}
LDH>=190 UI/L	184.45	159.22	172.97	193.94	207.67	171.90	195.78	0.314#
Creatinine>=2.0 mg/dL	3/31 (10%)	2/49 (4%)	6/29 (21%)	3/16 (19%)	1/10 (10%)	0/15 (0%)	4/24 (17%)	0.129*
HGB<10 g/dL	6/31 (19%)	10/49 (20%)	11/29 (38%)	**9/16 (56%)**	3/10 (30%)	6/15 (40%)	6/24 (25%)	0.091*

* Fisher's exact test was used. # ANOVA test was used. The top significant group was highlighted with bold number.

TABLE 8. Characteristic of 176 Multiple Myeloma Patients In The Predicting Set In Relation To The Seven Subgroups Identified By The 15-Gene Model

Group	I	II	III	IV	V	VI	VII	*P*-value
Age>=65	9/36 (25%)	5/52 (10%)	3/25 (12%)	2/12 (17%)	0/4 (0%)	1/18 (6%)	5/29 (17%)	0.438*
IgA Isotype	5/35 (14%)	6/51 (12%)	3/24 (13%)	2/12 (17%)	1/4 (25%)	6/17 (35%)	**11/29 (38%)**	0.055*
Albumin<3.5 g/dL	3/36 (8%)	5/52 (10%)	7/25 (28%)	4/12 (33%)	0/4 (0%)	3/18 (17%)	6/29 (21%)	0.133*
B2M (mg/L)	3.17	3.50	4.27	**8.73**	4.33	3.74	4.56	0.007#
CRP>=4 mg/L	19/36 (53%)	26/50 (52%)	14/25 (56%)	6/11 (55%)	1/4 (25%)	10/18 (56%)	14/29 (48%)	0.965*
Cytogenetics abnormality	9/27 (33%)	22/46 (48%)	9/23 (39%)	**6/9 (56%)**	**2/3 (67%)**	5/16 (31%)	**14/27 (52%)**	0.598
Hyperdiploidy	7/27 (26%)	**20/46 (43%)**	0/23 (0%)	2/9 (22%)	0/3 (0%)	3/16 (19%)	4/27 (15%)	0.004
Hypodiploidy	2/27 (7%)	2/46 (4%)	**9/23 (39%)**	3/9 (33%)	**2/3 (67%)**	2/16 (13%)	**10/27 (37%)**	0.0002
MRI>1	31/34 (91%)	44/52 (85%)	23/25 (92%)	11/12 (92%)	3/4 (75%)	**6/17 (35%)**	18/29 (62%)	1.70×10^{-5}
MRI>3	25/34 (74%)	28/52 (54%)	18/25 (72%)	11/12 (92%)	2/4 (50%)	**4/17 (24%)**	15/29 (52%)	0.002
Chr14>20	16/27 (59%)	15/45 (33%)	**17/20 (85%)**	6/10 (60%)	**3/3 (100%)**	5/15 (33%)	**23/23 (100%)**	4.30×10^{-7}
LDH(UI/L)	180.72	169.29	156.12	**235.33**	164.75	173.11	147.89	0.029#
Creatinine>=2.0 mg/dL	2/36 (6%)	4/52 (8%)	1/25 (4%)	3/12 (25%)	0/4 (0%)	1/18 (6%)	4/29 (14%)	0.403*
HGB<10 g/dL	6/36 (17%)	9/52 (17%)	9/25 (36%)	6/12 (50%)	2/4 (50%)	7/18 (39%)	7/29 (24%)	0.064*

* Fisher's exact test was used. # ANOVA test was used. The top significant group was highlighted with bold number.

TABLE 9. Comparison of Patient Classification Using 350 Genes And The 15-Gene Model

15-Gene Model / Subgroups / PAM with 350 Genes	1	2	3	4	5	6	7	Accuracy (%)
G1 (n=69)	59							85.51
G2 (n=85)		78						91.76
G3 (n=59)			54					91.53
G4 (n=31)				27				87.10
G5 (n=20)					18			90.00
G6 (n=36)						34		94.44
G7 (n=50)							48	96.00
Total								90.60

EXAMPLE 10

Identification of Genes with Similar Expression Between Multiple Myeloma and Cells at Different Stages of B Cell Development

Examples 10 and 11 describe the establishment of a B cell developmental stage-based classification of multiple myeloma using global gene expression profiling. To classify multiple myeloma with respect to genes that were significantly differentially expressed in tonsil B cell to tonsil plasma cell transition reffered to as "early differentiation genes" (EDG) and those that were differentially expressed in the tonsil plasma cell to bone marrow plasma cell transition reffered to as "late differentiation genes" (LDG), 74 newly diagnosed cases of multiple myeloma and 7 tonsil B cell, 7 tonsil plasma cell, and 7 bone marrow plasma cell samples were tested for variance across the 359 EDGs and 500 LDGs disclosed above. The top 50 EDGs that showed the most significant variance across all samples were defined as early differentiation genes for myeloma (EDG-MM); likewise, the top 50 LDGs showing the most significant variance across all samples were identified as late differentiation genes for myeloma-1 (LDG-MM1). Subtracting the LDG-MM1 from the 500 LDG and then applying one-way ANOVA test for variance to the remaining genes identified the top 50 genes showing similarities between bone marrow plasma cells and multiple myeloma. These genes were defined as LDG-MM2.

Within the top 50 EDG-MM (Table 10), 18 genes that showed up-regulation in the tonsil B cell to tonsil plasma cell transition showed down-regulation to levels at or below that seen in tonsil B cells. The remaining 32 EDG-MM showed a reverse profile, in that these genes were down-regulated in the tonsil B cell to plasma cell transition, but showed tonsil B cell-like expression in multiple myeloma. In Table 10, gene expression was described as being at 1 of 5 possible levels. An absent absolute call (AAC), indicating an undetectable or absent gene transcript, was defined as "–". For all the samples in a group, expression levels were defined as "+" if the gene transcript was present and the average difference (AD) was <1000, "++" for 1000 ≤ AD<5000, "+++" for 5000 ≤ AD < 10,000, and "++++" for AD≥10,000.

One of the most striking genes defining EDG-MM was the cyclin dependent kinase 8 (*CDK8*), which was found absent in tonsil B cells but up-regulated to extremely high levels in tonsil and bone marrow plasma cells and then shut down again in virtually all multiple myeloma cases. The mitotic cyclin showed a progressive loss in expression from tonsil B cell (++) to tonsil plasma cell (+) to bone marrow plasma cell (–), whereas multiple myeloma cases either showing bone marrow-like levels or tonsil B cell levels. Given that the tonsil B cells used in this study likely represent highly proliferative centroblasts, multiple myeloma cases with similar levels might be suggestive of a proliferative form of the disease. A total of 27 of the top 50 EDG-MM showed no variability in multiple myeloma, ie, all multiple myeloma and tonsil B cell samples showed similar levels of expression.

A majority (34 of 50) of the top 50 LDG-MM1 (Table 11) were genes that showed up-regulation from the transition of tonsil plasma cell to bone marrow plasma cell, but showed down-regulation to tonsil plasma cell levels in multiple myeloma. The overall pattern seen for LDG-MM1 was the reverse seen for the EDG-MM, where a majority of those genes showed down-regulation from tonsil B cell to plasma cell and up regulation to tonsil B cell-like levels in multiple myeloma. The most dramatically altered LDG-MM1 was seen in the massive up-regulation of the cxc chemokines *SDF1*, *PF4*, and *PPBP* in bone marrow plasma cells in contrast with complete absence of detectable transcripts in all multiple myeloma. These results are validated by the fact that two separate and distinct probe sets interrogating different region of *SDF1* (accession numbers L36033 and U19495) were found to show identical patterns. The *RB1* tumor suppressor gene showed a significant up-regulation in the tonsil plasma cell (+) to bone marrow plasma cell (++) transition with multiple

myeloma showing levels consistent with either cell type. Unlike with the EDG-MM, only 15 of the top 50 LDG-MM1 showed no variability within the multiple myeloma population.

The LDG-MM2 genes (Table ~~19~~ 12) showing similarities between bone marrow plasma cells and subsets of multiple myeloma revealed that all genes showed variability within multiple myeloma and that the variability could be dramatic, e.g. the apoptosis inhibitor *BIK*. Unlike those seen in EDG-MM and LDG-MM1, a large class of LDG-MM2 represented genes coding for enzymes involved in metabolism with a majority involved in glucose metabolism.

TABLE 10. EDG-MM: Tonsil B Cell-like Multiple Myeloma Genes

Accession	Symbol	Function	Quantitative Gene Expression			
			TBC	TPC	BPC	MM
D28364	*ANXA2*	annexin family	–	+	+	–/+
U81787	*WNT10B*	signaling; ligand	–	+	++	– /++
U88898	*LOC51581*[a]	unknown	–	+	+	–/+
X12451	*CTSL*	protease; cysteine	–	++	++	–
Z25347	*CDK8*	cell cycle; kinase	–	++++	++++	– /++
D38548	*KIAA0076*[a]	unknown	+	++	++	+/++
D86479	*AEBP1*	extracellular matrix	+	++	++	+
U04689	*OR1D2*	signaling; receptor	+	++	+	+
M31328	*GNB3*	signaling; G protein	+	++	++	+
U13395	*WWOX*	metabolism; oxidoreductase	+	++	++	+
X14675	*BCR*	signaling; GTPase for RAC	+	++	++	+
X16665	*HOXB2*	transcription; homeobox domain	+	++	++	–/+
Z11899	*POU5F1*	transcription; homeobox domain	+	++	++	+
Z36531	*FGL2*	secreted fibrinogen-like	+	++	++	+
X80907	*PIK3R2*	signaling; kinase adaptor	+	+++	+++	++
D31846	*AQP2*	aquaporin	++	+++	+++	++
L18983	*PTPRN*	phosphatase; membrane	++	++++	++++	++
M23323	*CD3E*	signaling; TCR partner	++	++++	++++	++
D83781	*KIAA0197*[a]	unknown	+	–	–	+
HT4824	*CBS*	metabolism; cystathionine-beta-synthase	+	–	–	–/++
S78873	*RABIF*	signaling; GTP releasing factor	+	–	–	+/++
U32645	*ELF4*	transcription; ets domian	+	–	–	–/+
X97630	*EMK1*	signaling; kinase; ELK domain	+	–	–	+

Accession	Symbol	Function	Quantitative Gene Expression TBC	TPC	BPC	MM
Z24724	*UNKNOWN*[a]	cell cycle	+	–	–	+/++
D16480	*HADHA*	mitochondrial oxidation	++	–	–	–/++
L77701	*COX17*	mitochondrial oxidation	++	–	–	–/++
M90356	*BTF3L2*	transcription; NAC domain	++	–	–	++
U08815	*SF3A3*	spliceosome	++	–	–	+/++
U53225	*SNX1*	intracellular trafficking	++	–	–	+/++
M25753	*CCNB1*	cell cycle	++	+	–	–/++
D87448	*TOPBP1*[a]	topoisomerase II binding protein	++	+	+	+/++
L38810	*PSMC5*	26S proteasome subunit 5	++	+	+	++
M29551	*PPP3CB*	signaling; calcium dependent p	++	+	+	+/++
M32886	*SRI*	signaling; calcium binding	++	+	+	++
U24704	*PSMD4*	26S proteasome subunit 4	++	+	+	++
U25165	*FXR1*	RNA binding protein	++	+	+	++
U37022	*CDK4*	cell cycle; kinase	++	+	+	++
U53003	*C21orf33*	unknown; highly conserved	++	+	+	++
X89985	*BCL7B*	actin crosslinking	++	+	+	++
D49738	*CKAP1*	tubulin folding	+++	+	+	++
D43950	*CCT5*	chaperonin	+++	++	++	+++
D82348	*ATIC*	metabolism; purine biosynthesis	+++	++	++	+++
D86550	*DYRK1A*	signaling; kinase	+++	++	++	+++
L06132	*VDAC1*	anion channel	+++	++	++	++/+++
L43631	*SAFB*	nuclear scaffold factor	+++	++	++	++/+++
M30448	*CSNK2B*	signaling; casein kinase regulation	+++	++	++	++/+++
X76013	*QARS*	metabolism; glutaminyl tRNA synthetase	+++	++	++	++/+++
D83735	*CNN2*	actin binding	++++	++	++	++/+++
M86667	*NAP1L1*	nucleosome assembly	+++ +	++	++	+++
X04828	*GNAI2*	signaling; G protein	++++	++	++	++/+++

Genes identified by one-way ANOVA analysis. Accession = GeneBank accession number. Symbol = HUGO approved gene symbol; unapproved symbol marked by [a]. TBC, tonsil B cells; TPC, tonsil plasma cells; BPC, bone marrow plasma cells; AC, absolute call; AD, average difference. Quantitative gene expression: –, AC absent; +, AC present and AD < 1,000; ++, AD = 1,000 to 5,000; +++, AD = 5,000 to 10,000; ++++, AD >10,000.

TABLE 11. LDG-MM1: Tonsil Plasma Cell-Like Multiple Myeloma Genes

Accession	Symbol	Function	Quantitative Gene Expression TPC	BPC	MM
U90902	*23612*[a]	unknown; related to *TIAM1*	–	+	–/++
D12775	*AMPD3*	metabolism; AMP deaminase	–	+	–/++
U37546	*BIRC3*	signaling; anti-apoptosis	–	++	–/++
Z11793	*SEPP1*	metabolism; selenium transport	–	+++	–/+

			Quantitative Gene Expression		
Accession	**Symbol**	**Function**	**TPC**	**BPC**	**MM**
L36033	*SDF1*	signaling; cxc chemokine	–	+++	–
U19495	*SDF1*	signaling; cxc chemokine	–	+++	–
M27891	*CST3*	protease inhibitor	–	++++	– / ++++
M26602	*DEFA1*	immunity	–	++++	– / ++++
M25897	*PF4*	signaling; cxc chemokine	–	++++	–
M54995	*PPBP*	signaling; cxc chemokine	–	++++	–
U79288	*KIAA051*[a]	unknown	+	++	+ / ++
M59465	*TNFAIP1*	signaling; anti-apoptosis	+	++	+ / ++++
X53586	*ITGA6*	adhesion	+	++	+ / ++
D50663	*TCTEL1*	dynein light chain	+	++	+ / ++
U40846	*NAGLU*	metabolism; hepran sulfate degradation	+	++	+ / ++
M80563	*S100A4*	Signaling; calcium binding	+	++	+ / ++++
X04085	*CAT*	metabolism; catalase	+	++	+ / ++
L02648	*TCN2*	metabolism; vitamin B12 transport	+	++	+
L35249	*ATP6B2*	lysosome; vacuolar proton pump	+	++	+
L09209	*APLP2*	amyloid beta precursor like	+	++	+
L41870	*RB1*	cell cycle	+	++	+ / ++
X76732	*NUCB2*	signaling; calcium binding	+	+++	+ / +++
D29805	*5B4GALT1*	adhesion	+	+++	+
M29877	*FUCA1*	lysosome; fucosidase	+	+++	+ / ++
M32304	*TIMP2*	metalloproteinase 2 inhibitor	+	+++	+ / ++++
D10522	*MACS*	actin crosslinking	+	++++	– / ++
L38696	*RALY*[a]	RNA binding	++	+++	++
U05875	*IFNGR2*	signaling; interferon gamma receptor	++	+++	++ / +++
U78095	*SPINT2*	protease inhibitor; blocks HGF	++	+++	– / +++
L13977	*PRCP*	lysosomal; angiotensinase C	++	++++	++
U12255	*FCGRT*	IgG Fc receptor	++	++++	– / +++
L06797	*CXCR4*	signaling; SDF1 receptor	++	++++	++ / ++++
D82061	*FABGL*	metabolism	++	++++	++ / +++
Y00433	*GPX1*	oxidation protection	+++	++++	++ / +++
M60752	*H2AFA*	histone; nucleosome	+	–	– / ++
U18300	*DDB2*	DNA repair	+	–	+
X63692	*DNMT1*	DNA methyltransferase	+	–	+
D11327	*PTPN7*	signaling; phosphatase	++	–	++
X54942	*CKS2*	cell cycle; kinase regulator	++	–	+ / +++
D14874	*ADM*	adrenomedullin	++	+	+/++++
D86976	*KIAA0223*[a]	minor histocompatability antigen	++	+	+/+++
X52979	*SNRPB*	mRNA splicing	++	+	+ / ++
Z49254	*MRPL23*	mitochondrial ribosomal protein	++	+	++
U66464	*HPK1*	signaling; kinase	++	+	+ / ++
U91903	*FRZB*	signaling; WNT antagonists	++	+	+ / ++
D87453	*MRPS27*	mitochondrial ribosomal protein	++	+	+ / ++
X59932	*CSK*	signaling; kinase	+++	++	++
L17131	*HMGIY*	transcription; high mobility group	++++	+	++ / ++++
L19779	*H2AFO*	histone; nucleosome	++++	++	++++
U70439	*SSP29*[a]	unknown	++++	+++	+++/+++

Accession	Symbol	Function	Quantitative Gene Expression TPC	BPC	MM
					+

Genes identified by one-way ANOVA analysis. Accession = GeneBank accession number. Symbol = HUGO approved gene symbol; unapproved symbol marked by [a]. TPC, tonsil plasma cells; BPC, bone marrow plasma cells; AC, absolute call; AD, average difference.Quantitative gene expression: –, AC absent; +, AC present and AD < 1,000; ++, AD = 1,000 to 5,000; +++, AD = 5,000 to 10,000; ++++, AD >10,000.

TABLE 12. LDG-MM2: Bone marrow Plasma Cell-like Multiple Myeloma Genes

Accession	Symbol	Function	Quantitative Gene Expression BPC	MM
U61145	*EZH2*	transcription; SET domain	–	–/+
HT4000	*SYK*	signaling; lymphocyte kinase	–	–/++
X89986	*BIK*	signaling; apoptosis inducer	–	–/++++
D85181	*SC5DL*	metabolism; sterol-C5-desaturase	+	–/+
M98045	*FPGS*	metabolism; folylpolyglutamate synthase	+	–/++
L41559	*PCBD*	transcription; enhances TCF1 activity	+	–/++
L25876	*CDKN2*	cell cycle; CDK inhibitor; phosphatase	+	+/++
U76638	*BRAD1*	transcription; BRCA1 heterodimer	+	+/++
L05072	*IRF1*	transcription; IRF family	+	+/++
D87440	*KIAA025*[a]	unknown	+	+/++
U02680	*PTK9*	tyrosine kinase	+	+/++
U28042	*DDX10*	oncogene; ATP-dependent RNA helicase	+	+/++
L20320	*CDK7*	cell cycle; kinase	+	+/++
X56494	*PKM2*	metabolism; pyruvate kinase	+	+/++++
M12959	*TCRA*	signaling; T cell receptor	++	–/++
HT3981	*INSL3*	signaling; insulin-like peptide; IGF family	++	–/++
U21931	*FBP1*	metabolism; fructose bisphophatase	++	–/++++
Z48054	*PXR1*	metabolism; peroxisome biogenesis	++	+/++
D84145	*WS-3*[a]	dynatin 6	++	+/++
D14661	*KIAA0105*[a]	transcription; WT1-associating protein	++	+/++
X77548	*NCOA4*	transcription; nuclear receptor coactivator	++	+/++
M90696	*CTSS*	cysteine protease	++	+/++
D11086	*IL2RG*	cytokine receptor	++	+/++
U70426	*RGS16*	signaling; GTPase activating protein	++	+/+++
X14850	*H2AX*	histone; required for antibody maturation	++	+/+++
M29927	*OAT*	metabolism; ornithine aminotransferase	++	+/+++
S74017	*NFE2L2*	transcription;	++	+/+++
HT4604	*GYG*	metabolism; glycogen biogenesis	++	+/+++
M55531	*SLC2A5*	metabolism; fructose transporter	++	+/++++

Accession	Symbol	Function	Quantitative Gene Expression BPC	MM
M60750	*H2BFL*	histone; nucleosome	++	+/++++
L19437	*TALDO1*	metabolism; transaldolase	++	++/+++
M10901	*NR3C1*	transcription; glucocorticoid receptor	++	++/+++
L41887	*SFRS7*	MRNA splicing factor	++	++/+++
M34423	*GLB1*	metabolism; galactosidase	++	++/++++
X15414	*AKR1B1*	metabolism; aldose reductase	+++	+/++++
J04456	*LGALS1*	signaling; inhibits CD45 phosphatase	+++	+/++++
X92493	*PIP5K1B*	signaling; kinase	+++	+/++++
U51478	*ATP1B3*	Na+, K+ transporter	+++	++/++++
X91257	*SARS*	seryl-tRNA synthetase	+++	++/++++
D30655	*EIF4A2*	translation initiation	+++	++/++++
D31887	*KIAA0062*[a]	unknown	+++	++/++++
X04106	*CAPN4*	cysteine protease; calcium dependent	+++	++/++++
D87442	*NCSTN*[a]	nicastrin	+++	++/++++
L76191	*IRAK1*	signaling; cytokine receptor kinase	+++	+++/++++
HT1428	*HBB*	hemoglobin	++++	–/++++
U44975	*COPEB*	oncogene; transcription factor	++++	–/++++
X55733	*EIF4B*	translation initiation	++++	+/++++
L09604	*PLP2*	signaling; colonic epithelium differentiation	++++	+/++++
HT1614	*PPP1CA*	signaling; phosphatase	++++	+++/++++
L26247	*SUI1*[a]	translation initiation; probable	++++	+++/++++

Accession = GeneBank accession number or TIGR database. Symbol = HUGO approved gene symbol; unapproved symbol marked by [a]. BPC, bone marrow plasma cells; AC, absolute call; AD, average difference. Quantitative gene expression: –, AC absent; +, AC present and AD < 1,000; ++, AD = 1,000 to 5,000; +++, AD = 5,000 to 10,000; ++++, AD >10,000.

EXAMPLE 11

Hierachical Cluster Analysis with EDG-MM, LDG-MM1, and LDG-MM2 Reveals Developmental Stage-Based Classification of Multiple Myeloma

To identify whether variability in gene expression seen in multiple myeloma (MM) might be used to discern subgroups of disease, hierarchical cluster analysis was performed on 74 newly diagnosed MM, 7 tonsil B cell, 7 tonsil plasma cell, and 7 bone marrow samples using the EDG-MM (Figure 11), LDG-MM1 (Figure 12), and LDG-MM2 (Figure 13). Hierarchical clustering was applied to all samples using 30 of the 50 EDG-MM. A total of 20 genes were filtered out with Max-Min < 2.5. This filtering was performed on this group because many of the top 50 EDG-MM showed no variability across MM and thus could not be used to distinguish MM subgroups. A similar clustering strategy was employed to cluster the samples using the 50 LDG-MM1 and 50 LDG-MM2.

The MM samples clustering with the tonsil B cell samples were then identified to determine whether the MM cases clustering with tonsil B cells, or tonsil and bone marrow plasma cells could be correlated with gene expression-defined MM subgroups (Table 13). This data showed that of the MM cases clustering tightly with the tonsil B cell samples, 13 of 22 were from the MM4 subgroup, accounting for a majority of all MM4 cases (13 of 18 MM4 samples). The LDG-MM defined cluster distribution of gene expression-defined MM subgroups was dramatically different in that 14 of the 28 MM samples clustering with the tonsil plasma cell samples were from MM3 subgroup (14 of 15 MM3 samples). LDG-MM2 again showed a strong correlation with the MM subgroups in that 14 of the 20 MM cases in this cluster were from the MM2 subgroup (14 of 21 MM2 cases). Thus, the MM4, MM3, and MM2 subtypes of MM have similarities to tonsil B cells, tonsil plasma cells, and bone marrow plasma cells respectively. MM1 represented the only subgroup with no strong correlations with normal cell counterparts tested here, suggesting that this class has unique characteristics yet to be uncovered.

The distribution of the four MM subgroups in the normal cell cluster groups was determined next (Table 14). The results demonstrate that whereas all MM3 cases were able to be classified, 6 MM1, 5 MM2, and 3 MM4 cases were not clustered with any normal cell group in any of the three cluster analyses. In all samples capable of being clustered, there were strong correlations between gene expression-defined subgroups and normal cell types with the exception of MM1. The

data also show that 3 MM1, 2 MM2, 4 MM3, and 1 MM4 cases were found to cluster in two groups. No samples were found in three groups and all cases clustering with two normal classes were always in an adjacent, temporally appropriate groups. P241 was an exception in that it was clustered in the bone marrow plasma cell and tonsil B cell groups.

Because one of the EDG-MMs was discovered to be cyclin B1 (*CCNB1*) (Table 10), it was determined if a panel of proliferation association genes recently discovered to be up-regulated in MM4 could be used to advance and validate the classification of MM4 as a so-called tonsil B cell-like form of MM. Box plots of the expression patterns of *CCNB1*, *CKS1*, *CKS2*, *SNRPC*, *EZH2*, *KNSL1*, *PRKDC*, and *PRIM1* showed significant differences across all the groups tested with strong significant correlation between tonsil B cells and MM4 (Figure 14). Several important observations were made in this analysis. For all the genes, with the exception of *SNRPC*, there was a progressive reduction in expression in the transition from tonsil B cells to tonsil plasma cells to bone marrow plasma cells. In addition, striking correlations were observed with *PRIM1* (Figure 14). Although *PRIM1* expression was significantly different across the entire group ($p = 4.25 \times 10^{-5}$), no difference exists between tonsil B cells and MM4 (Wilcoxon rank sum [WRS] p=0.1), or between tonsil plasma cells and MM3 (WRS p=0.6). Given the important function of several transcription factors in driving and/or maintaining plasma cell differentiation, it was determined if these factors showed altered expression across the groups under study. Although other factors showed no significant changes, *XBP1* (Figure 14) showed an enormous up-regulation between tonsil B cells and tonsil plasma cells as expected. However, the gene showed a reduction in bone marrow plasma cells and a progressive loss across the four MM subgroups with MM4 showing the lowest level (p=3.85×10^{-10}).

Based on conventional morphological features, plasma cells have been thought to represent a homogeneous end-stage cell type. However, phenotypic analysis and gene expression profiling disclosed herein demonstrated that plasma cells isolated from distinct organs can be recognized as belonging to distinct stages of development. Multiple myeloma plasma cells are derived from the bone marrow and are thought to represent a transformed counterpart of normal terminally differentiated bone marrow plasma cells. However, the dramatic differences in survival, which can range from several months to greater than 10 years, suggests that multiple myeloma

may represent a constellation of several subtypes of disease. Conventional laboratory parameters have not been particular useful in segregating distinct disease subtypes with sufficient robustness that would allow adequate risk stratification. In addition, unlike achievements in classifying leukemias and lymphomas based on similar nonrandom recurrent chromosomal translocations, the extreme karyotypic heterogeneity of multiple myeloma has made attempts at understanding the molecular mechanisms of the disease and classification prediction virtually impossible.

In studies presented here, it was identified that many EDGs and LDGs exhibit highly variable expression in multiple myeloma, suggesting that multiple myeloma might be amenable to a developmental stage-based classification. It appears from the results of this study that multiple myeloma can in fact be classified based on similarities in gene expression with cells representing distinct stages of B cell differentiation. This developmental based-system in conjunction with the gene expression-based system reported above represents a critical affirmation of the validity of the developmental-based system.

Recent studies provide support for the hypothesis that MM3 represents a tonsil plasma cell-like form of the disease. Microarray profiling with the U95Av2 GeneChip on 150 newly diagnosed patients (including the 74 described here) along with an analysis of chromosome 13 loss has revealed a significant link between reduced *RB1* transcripts with either monosomy or partial deletions of chromosome 13 (unpublished data). In these studies, it was observed that a number of multiple myeloma cases with or without chromosome 13 deletion had *RB1* transcripts at levels comparable to those seen in normal tonsil plasma cells. FISH analysis with a bacterial artificial chromosome BAC covering *RB1* demonstrated that these cases did not have interstitial deletions of the *RB1* locus. Given that RB1 was found to be a LDG-MM1, it was determined if the low levels of RB1 may be linked to tonsil plasma cell-like MM, i.e MM3. Of 35 multiple myeloma cases with RB1 AD values of <1100 (RB1 AD value not less than 1100 in 35 normal bone marrow plasma cell samples tested), 74% belonged to the MM3 class. In contrast, of 38 multiple myeloma cases lacking deletion 13 and having RB1 AD values greater than 1100, only 21% belonged to the MM3 subtype.

Although there is a significant link between the cell development-based classification and gene expression profiling-based classification disclosed herein, there are exceptions in that although as expected the majority of the MM4

cases belonged to the tonsil B cell-cluster subgroup, 5 MM3, 1 MM2, and 3 MM1 cases were also found in this cluster. The recognition that cases within one gene expression-defined subgroup could be classified in two normal cell defined clusters suggests these cases may have intermediate characteristics with distinct clinical outcomes. It is of interest to determine if the unsupervised gene expression-based system or developmental stage-based system alone or in combination will allow the creation of robust risk stratification system. This can be tested by allowing sufficient follow-up time on >150 uniformly treated multiple myeloma cases in which profiling has been performed at diagnosis.

MM1 was the only gene expression-defined subgroup lacking strong similarities to any of the normal cell types analyzed in this study. It is possible that MM1 has similarities to either mucosal-derived plasma cells or peripheral blood plasma cells which has recently been shown to represent a distinct type of plasma cells. Future studies will be aimed at providing a developmental stage position for this subtype.

The hypoproliferative nature of multiple myeloma, with labeling indexes in the clonal plasma cells rarely exceeding 1%, has lead to the hypothesis that multiple myeloma is a tumor arising from a transformed and proliferative precursor cell that differentiates to terminally differentiated plasma cells. It has been shown that there is a bone marrow B cell population transcribing multiple myeloma plasma cell-derived VDJ joined to IgM sequence in IgG- and IgA-secreting multiple myelomas. Other investigations have shown that the clonogenic cell in multiple myeloma originates from a pre-switched but somatically mutated B cell that lacks intraclonal variation. This hypothesis is supported by recent use of single-cell and *in situ* reverse transcriptase-polymerase chain reaction to detect a high frequency of circulating B cells that share clonotypic Ig heavy-chain VDJ rearrangements with multiple myeloma plasma cells. Studies have also implicated these precursor cells in mediating spread of disease and affecting patient survival.

Links of gene expression patterns between subsets of multiple myeloma and cells representing different late stages of B cell differentiation further support the above hypothesis in that MM4 and MM3 may have origins in a so called "multiple myeloma stem cell". This hypothesis can be tested by isolating B cells from tonsils or lymph nodes or peripheral blood of MM3 and MM4 patients, differentiating them into plasma cells *in vitro* using a new method described by Tarte

et al. (2002) and then testing for the presence of an multiple myeloma gene expression signature within the differentiated populations. Even if the multiple myeloma stem cell represents a minority population in the B cells, the multiple myeloma gene expression signature may be recognized, if not with conventional microarray, then by more sensitive quantitative real-time RT-PCR. A real time RT-PCR method is envisioned as expression profile models using at little as 20 genes that distinguish malignant multiple myeloma plasma cells from normal plasma cells at an accuracy of 99.5% have been developed. Regardless of the outcome of these experiments, it is clear that gene expression profiling has become an extremely powerful tool in evaluating the molecular mechanisms of plasma cell differentiation and how these events relate to multiple myeloma development and progression, which in turn should provide more rational means of treating this currently fatal disease.

TABLE 13. Distribution of Multiple Myeloma Subgroups in Hierarchical Clusters Defined by EDG-MM, LDG-MM1, and LDG-MM2 Genes

	Gene Expression-Defined MM Subgroups				
Normal Cell-Defined Cluster	MM1 (n = 20)	MM2 (n = 21)	MM3 (n =15)	MM4 (n = 18)	*p*
EDG-MM (n = 22)	3	1	5	**13**	.00005
LDG-MM1 (n = 29)	8	4	**14**	3	.000008
LDG-MM2 (n = 20)	6	**14**	0	0	.000001

TABLE 14. Distribution of Gene Expression-Defined Multiple Myeloma Subgroup Cases in Normal Cell Clusters defined by EDG-MM, LDG-MM1, and LDG-MM2

MM1	TBC	TPC	BPC	MM2	TBC	TPC	BPC	MM3	TBC	TPC	BPC	MM4	TBC	TPC	BPC
P026		Y	Y	P237		Y	Y	P052	Y	Y		P034	Y	Y	
P037		Y	Y	P241	Y		Y	P098	Y	Y		P051	Y		
P029	Y	Y		P079			Y	P107	Y	Y		P057	Y		
P061		Y		P083			Y	P158	Y	Y		P063	Y		
P066		Y		P121			Y	P119		Y		P065	Y		

MM1	TBC	TPC	BPC	MM2	TBC	TPC	BPC	MM3	TBC	TPC	BPC	MM4	TBC	TPC	BPC
P006		Y		P144			Y	P221		Y		P075	Y		
P120		Y		P157			Y	P030		Y		P084	Y		
P131		Y		P171			Y	P043		Y		P122	Y		
P002			Y	P176			Y	P053		Y		P127	Y		
P010			Y	P213			Y	P055		Y		P154	Y		
P067			Y	P215			Y	P138		Y		P187	Y		
P226			Y	P251			Y	P155		Y		P199	Y		
P025	Y			P250			Y	P163		Y		P255	Y		
P082	Y			P222		Y		P239		Y		P054		Y	
P085				P103		Y		P175	Y			P101		Y	
P099				P202		Y						P056			
P001				P015								*P091*			
P016				*P048*								*P168*			
P036				*P124*											
P118				*P212*											
				P249											

MM1, MM2, MM3, MM4, and PXXX represent gene expression-defined subgroups and patient identifiers, respectively. Y indicates that the case was found in the normal cell-defined cluster. Cases in italics were not found to cluster with any normal cell type. Some cases were found to cluster with two normal cell types. TBC, tonsil B cells; TPB, tonsil plasma cells; BPC, bone marrow plasma cells.

EXAMPLE 12

Gene Expression Profiling To Identify Genes That Could Be Useful As Diagnostic, Prognostic And Potential Targets In Myeloma.

As discussed earlier, global gene expression profiling identified genes directly involved in pathogenesis and classification of myeloma. Next, this profiling was used to identify genes whose abnormal expression may cause an aggressive phenotype of myeloma. 668 newly diagnosed patients with symptomatic or progressive multiple myeloma, which included 2 cycles of blood stem cell-supported high-dose melphalan (200mg/m^2) were enrolled in this study. A subset of 575 patients with available genetic measurements constituted sample for this analysis. Their median follow-up was 30 months. There were 185 progression or death vents and 128 deaths. Patient characteristic were as follows: 20% were 65 years or older, 31% had beta-2-microglobulin levels>= 4mg/L, 55% had C-reactive protein levels>=

4mg/L; 22% presented with hemoglobin values <=10g/dL, 10% with creatinine values >= 2mg/dL; LDH was elevated (>=190IU/L) in 15%; cytogenetic abnormalities were detected in 33%. Gene expression profiling was performed using U133Plus2.0 oligonucleotide microarrays as described previously (Tian, et al., 2003) for a consecutive 351 patient subset. The expression value 'Signal' for each probe set was derived from the MAS5.01 software (Affymetrix). Median follow-up for survival in this subset was 22 months and there were 98 events and 64 deaths.

Bacterial artificial chromosomes specific for *CKS1B* at 1q21 (RP11-307C12) and *ASPM* (RP11-32D17) at 1q31 were purchased from BAC/PAC Resources (Oakland, CA) and directly labeled with Spectrum green or Spectrum red (Vysis Inc, Downers Grove, IL). Metaphase fluorescence in situ hybridization was performed as previously described (Sawyer et al., 2005). The probes were confirmed to map to the 1q21 and 1q31 bands using metaphase spreads from normal human lymphocytes. Triple color interphase fluorescence in situ hybridization analyses of chromosomes 13 (RB) and 1q21 (CKS1B) copy number (Shaughnessy et al., 2000) were performed in a subset of 421 patients (145 events and 100 deaths, follow-up of 31 months); deletion 13q was scored positive when >=80% of clonal cells exhibited loss of a signal at 13q14 as described previously (McCoy et al., 2003). Of these 421 patients, 197 were among those with microarrays and 224 were not.

Nuclear protein was isolated from an aliquot of CD138 enriched plasma cells that were also analyzed by microarray. Western Blotting was carried out using the WesternBreeze® Chemiluminiscent Immunodetection protocol as described (Invitrogen, Carlsbad, CA). The antibodies to CKS1B and phospo-thr-187-$p27^{Kip1}$ were purchased from Zymed laboratories Inc., (South San Francisco, CA) and anti-Histone 1A was purchased from Upstate Biotechnology (Charlottesville, VA).

False discovery rate methods (Storey and Tibshirani, 2003) were used to adjust for multiple comparisons of the Affymetrix probe set Signals for the 351 microarrays: For each Affymetrix Signal, log rank tests for the equality of disease-related survival distributions were performed separately for quartile 1 vs quartiles 2 through 4 (to identify under-expressed genes) and quartile 4 vs quartiles 1 through 3 (to identify over-expressed genes). A false discovery rate cut-off of 2.5% was applied to each list and a total of 70 probe sets were retained. Extreme quartile membership (Q1 or Q4) was associated with a higher incidence of disease-related death in all retained probe sets. All other EFS and survival outcomes in this analysis were overall

(i.e. not disease-related). The Kaplan-Meier method was used to estimate event-free and overall survival distributions and log rank tests were used to test for their equality across groups. Chi-square tests and Fisher's exact test were used to test for the independence of categories. Proportional hazards regression was used to compare the effect of CKS1B amplification to other variables and the proportions of observed heterogeneity (i.e. R^2) were computed (O'Quigley and Xu, 2001). The statistical packages R version 2.0 (R Development Core Team, 2004) and SPLUS 6.1 (Insightful Corp., 2002) were used for this analysis.

Nearly 50% of the newly diagnosed myelomas contain one of five recurrent chromosomal translocations that result in the hyperactivation of *MAF, MAFB, FGFR3/MMSET, CCND3, CCND1* (Kuehl, W.M. et al., 2002) with divergent prognoses (Fonseca, R. et al., 2004), detectable as "spiked" expression by microarray analysis (Zhan, F. et al., 2003). Genetic subgroups were classified within the context of metaphase cytogentics as having normal karyotypes (originating in normal hematopoetic cells in case of hypoproliferative myeloma) or as having hyperdiploid, hypodiploid or "other" cytogenetic abnormalities. "Other" is defined as an abnormal metaphase karyotype, i.e. structural and/or numeric changes, with a diploid modal chromosome number. Thus, non-translocation entities were defined by metaphase rather than interphase cytogenetics.

A conventional, laboratory defined cutoff of 20% for the proportion of clonal plasma cells with 3 signals or >= 4 signals was used for tests of association between expression and amplificaion(Figure 16B and Table 17) and for validation of the association between amplification and overall survival (Figures 16C-D). Hypothesizing that 2 or more extra copies would confer additional risk compared to 1 extra copy, the multivariate analysis of overall survival (Table 19) estimated separate effect sizes for the 3 signal proportion and the >= 4 signal proportion. The index was defined as a weighted sum:(0.34*% 3copies+ 0.66*%>= 4 copies)/0.66, where the weights were the log-scale hazard ratio estimates for the two percentages, scaled to the effect of a unit increase in the proportion with >= 4 signals. The estimated log-scale hazard ratio corresponding to a ne unit difference in the proportion with >= 4 signals is nearly twice as large as that for 3 signals (i.e. 0.66/0.34= 1.94). The index is 0 for patients with <= 2signals in 100%f clonal cells and 100 for patients with >= 4 signals in 100% of clonal cells. The full range was observed in these patients. A cut-

off for the index of >= 46 minimized the unadjusted log rank P-value for survival in the 421 patient subset, however, all cutoffs between 3 and 88 had P< 0.003.

EXAMPLE 13

Results of global gene profiling

To molecularly define de novo high-risk multiple myeloma, the gene expression profiles in purified myeloma plasma cells were correlated with disease-related and overall survival in 351 newly diagnosed patients treated with 2 cycles of high-dose mephalan. Using log rank tests, 70 genes were identified for which fourth or first quartile membership was correlated with a high incidence of disease-related death (Table 15).

Table 15A. Quartile 4FDR 2.5% gene probe sets-rank correlations with 1q21 amplification index, CKS1B and PC labeling index and adjusted P-values for associations with overall survival.

Rank (Q4)	Chromosome	Probe set	Symbol	CKS1B Amplification Index r†	CKS1B r‡	PCLI r*	Adjusted Survival P-value[a]
1	8q21.13	202345_s_at	*NA*	0.20	0.22		0.001
2	**Xp22.2-p22.1**	**1555864_s_at**	***NA***	**0.34**	**0.47**		**0.007**
3	**5p15.33**	**204033_at**	***TRIP13***	**0.19**	**0.45**	**0.20**	**0.001**
4	1q22	206513_at	*AIM2*	0.15	0.13		0.089
5	2p24.1	1555274_a_at	*SELI*	0.28	0.31		0.001
6	21q22.3	211576_s_at	*SLC19A1*	0.17	0.23		0.007
7	3p21.3	204016_at	*LARS2*	-0.18			0.002
8	1q43	1565951_s_at	*OPN3*	0.36	0.36		0.007
9	**1q31**	**219918_s_at**	***ASPM***	**0.36**	**0.64**	**0.17**	**0.010**
10	**12q15**	**201947_s_at**	***CCT2***	**0.23**	**0.43**	**0.13**	**0.004**
11	16p13.3	213535_s_at	*UBE2I*		0.38		0.022
12	**20q13.2-q13.3**	**204092_s_at**	***STK6***	**0.31**	**0.51**	**0.19**	**0.044**
13	1p36.33-p36.21	213607_x_at	*FLJ13052*				0.150
14	xq12-q13	208117_s_at	*FLJ12525*		0.34		0.006
15	17q25	210334_x_at	*BIRC5*	0.20	0.36	0.14	0.110
16	**3q27**	**204023_at**	***NA***	**0.29**	**0.62**	**0.16**	**0.072**
17	**1q21.2**	**201897_s_at**	***CKS1B***	**0.50**	**1.00**	**0.15**	**0.007**
18	19q13.11-q13.12	216194_s_at	*CKAP1*	0.24	0.38		0.001
19	**1q21**	**225834_at**	***MGC5782 7***	**0.39**	**0.66**	**0.23**	**0.140**
20	19q13.12	238952_x_at	*DKFZp77 9O175*		0.11		0.009
21	**17p13.3**	**200634_at**	***PFN1***	**0.30**	**0.41**		**0.002**

22	19p13.2	208931_s_at	*ILF3*	0.22	0.22		0.220
23	1q22	206332_s_at	*IFI16*	0.30	0.32	0.13	0.003
24	7p14-p13	220789_s_at	*TBRG4*		0.13	0.17	0.009
25	10p11.23	218947_s_at	*PAPD1*	0.31	0.30		0.150
26	8q24	213310_at	*EIF2C2*	0.28	0.37		0.031
27	3q12.1	224523_s_at	*MGC4308*	0.17	0.24	0.14	0.038
28	1p36.3-p36.2	201231_s_at	*ENO1*		0.23		< 0.001
29	18q12.1	217901_at	*DSG2*	0.15			0.005
30	**6q22**	**226936_at**	***NA***	**0.15**	**0.52**	**0.17**	**0.027**
31	8q24.3	58696_at	*EXOSC4*		0.20		0.330
32	**1q21-q25**	**200916_at**	***TAGLN2***	**0.47**	**0.52**		**0.120**
33	3q21	201614_s_at	*RUVBL1*	0.16	0.14		0.023
34	16q22-q24	200966_x_at	*ALDOA*	0.21	0.28		0.001
35	2p25.1	225082_at	*CPSF3*		0.39		0.073
36	1q43	242488_at	*NA*	0.18	0.27	0.14	0.090
37	3q12.3	243011_at	*MGC15606*		0.27		0.004
38	22q13.1	201105_at	*LGALS1*		0.31		0.051
39	**3p25-p24**	**224200_s_at**	***RAD18***	**0.17**	**0.41**	**0.14**	**0.040**
40	20p11	222417_s_at	*SNX5*				0.085
41	**1q21.2**	**210460_s_at**	***PSMD4***	**0.58**	**0.59**	**0.13**	**0.067**
42	**12q24.3**	**200750_s_at**	***RAN***	**0.22**	**0.40**		**0.056**
43	**1pter-q31.3**	**206364_at**	***KIF14***	**0.41**	**0.57**	**0.25**	**0.019**
44	7p15.2	201091_s_at	*CBX3*	0.14	0.20	0.16	0.150
45	**12q22**	**203432_at**	***TMPO***	**0.32**	**0.59**	**0.18**	**0.007**
46	**17q24.2**	**221970_s_at**	***DKFZP586L0724***	**0.27**	**0.47**		**0.081**
47	**11p15.3-p15.1**	**212533_at**	***WEE1***	**0.20**	**0.54**	**0.13**	**0.056**
48	3p12	213194_at	*ROBO1*				0.150
49	5q32-q33.1	244686_at	*TCOF1*				0.120
50	8q23.1	200638_s_at	*YWHAZ*	0.26	0.23		0.012
51	**10q23.31**	**205235_s_at**	***MPHOSPH1***		**0.40**	**0.16**	**0.050**

† Correlation between gene expression signal and the CKS1B amplification index (N=197, all patients with both GEP and FISH 1q21. Blank cells correspond to insignificant correlations (nominal P>0.05). ‡ Correlation between each gene's log-scale expression and CKS1B log-scale expression (N=351, all patients with GEP). Rows with CKS1B |r| >=0.4 are formatted bold. * Correlation between each gene's log-scale expression and the PCLI (N=305, 46 patients are missing PCLI). [a] PH regression of overall survival on quartile 1 expression for each gene, adjusted for FISH 13 80% cytogenetic abnormalities, B2M>4, CRP>4, ALB<3.5 and PCLI (N=277, 74 patients are missing at least one measurement, 17 are missing FISH13, 4 are missing CA, 10 are missing one of B2M, CRP and ALB and 46 are missing PCLI. These P-values are not adjusted for the quartile 1 log rank significance testing that determined the ranks in column 1.

Table 15B. Quartile 1 gene probe sets satisfying FDR 2.5% cutoff

Rank (Q1)	Chromo-some	Probe set	Symbol	CKS1B Amplifica-tion Index r†	CKS1B r‡	PCLI r*	Adjusted Survival P-value[a]
1	9q31.3	201921_at	*GNG10*	-0.20	-0.30		0.600
2	1p13	227278_at	*NA*			-0.12	0.900
3	Xp22.3	209740_s_at	*PNPLA4*				0.029
4	20q11.21	227547_at	*NA*	-0.29	-0.28	-0.15	0.630
5	10q25.1	225582_at	*KIAA1754*	-0.21	-0.32		0.003
6	1p13.2	200850_s_at	*AHCYL1*			-0.13	0.019

7	1p13.3	213628_at	*MCLC*	-0.30	-0.28	-0.15	0.440
8	1p22	209717_at	*EVI5*	-0.33	-0.29	-0.16	0.870
9	1p13.3	222495_at	*AD-020*	-0.30	-0.24	-0.20	0.920
10	6p21.31	1557277_a_at	*NA*		-0.11		0.460
11	1p22.1	1554736_at	*PARG1*		-0.20	-0.11	0.280
12	1p22	218924_s_at	*CTBS*	-0.16	-0.11	-0.13	0.460
13	**9p13.2**	**226954_at**	***NA***	**-0.22**	**-0.40**		**0.090**
14	1p34	202838_at	*FUCA1*	-0.17	-0.23		0.066
15	13q14	230192_at	*RFP2*	-0.28	-0.18		0.880
16	12q13.11	48106_at	*FLJ20489*	-0.23	-0.23	-0.11	0.300
17	11q13.1	237964_at	*NA*	-0.16	-0.20		0.044
18	2p22-p21	202729_s_at	*LTBP1*	-0.24	-0.21		0.097
19	1p13.1	212435_at	*NA*	-0.21	-0.21	-0.11	0.034

† Correlation between each gene's log-scale expression and the CKS1B amplification index (N=197, all patients with both GEP and FISH 1q21). Blank cells correspond to insignificant correlations.‡ Correlation between each gene's log-scale expression and CKS1B log-scale expression (N = 351, all patients with GEP). Rows with CKS1B |r| >= 0.4 are formatted bold. * Correlation between each gene's log-scale expression and the PCLI (N=305, 46 patients are missing PCLI).[a] PH regression of overall survival on Quartile 1 expression for each gene, adjusted for FISH 13 80%, cytogenetic abnormalities, B2M>4, CRP>4, ALB<3.5 and PCLI (N=277, 74 patients are missing at least one measurement, 17 are missing FISH 13, 4 are missing CA, 10 are missing one of B2M,CRP and ALB, and 46 are missing PCLI: P=0.51 for a log rank test of the effect of exclusion due to missing measurements). These P-values are not adjusted for the quartile 1 log rank significance testing that determined the ranks in column 1.

Although 10% of the genes on the microarray were derived from chromosome 1, 30% of the retained genes were derived from chromosome 1, 30% of the retained genes were derived from this chromosome (*P<0.0001*) with 12 of 51 quartile 4 genes (23.5%) mapped to chromosome 1q and 9 of 19 quartile 1 genes (47%) mapped to chromosome 1p (Table 16).

Table 16. Chromosome Distribution of 2.5% FDR Probe Sets

	U133Plus2.0		Q1		Q4		Combined		
Chromosome	No of Genes	%	No. of Genes	%	No. of Genes	%	No. of Genes	%	P value*
1	3,659	**9.9**	9	47.4	12	23.5	21	**30.0**	< 0.0001
2	2,522	6.9	1	5.3	2	3.9	3	4.3	
3	2,116	5.8	0	0.0	7	13.7	7	10.0	
4	1,456	4.0	0	0.0	0	0.0	0	0.0	
5	1,718	4.7	0	0.0	2	3.9	2	2.9	
6	2,005	5.4	1	5.3	1	2.0	2	2.9	
7	1,798	4.9	0	0.0	2	3.9	2	2.9	
8	1,311	3.6	0	0.0	4	7.8	4	5.7	
9	1,463	4.0	2	10.5	0	0.0	2	2.9	
10	1,444	3.9	1	5.3	2	3.9	3	4.3	
11	2,069	5.6	1	5.3	1	2.0	2	2.9	
12	1,927	5.2	1	5.3	3	5.9	4	5.7	
13	730	2.0	1	5.3	0	0.0	1	1.4	

14	1,195	3.2	0	0.0	0	0.0	0	0.0	
15	1,152	3.1	0	0.0	0	0.0	0	0.0	
16	1,507	4.1	0	0.0	2	3.9	2	2.9	
17	2,115	5.7	0	0.0	3	5.9	3	4.3	
18	582	1.6	0	0.0	1	2.0	1	1.4	
19	2,222	6.0	0	0.0	3	5.9	3	4.3	
20	1,072	2.9	1	5.3	2	3.9	3	4.3	
21	468	1.3	0	0.0	1	2.0	1	1.4	
22	906	2.5	0	0.0	1	2.0	1	1.4	
X	1,273	3.5	1	5.3	2	3.9	3	4.3	
Y	80	0.2	0	0.0	0	0.0	0	0.0	
m	5	0.0	0	0.0	0	0.0	0	0.0	
	36,795		19		51		70		
Unknown	17,880								
	54,675								

*An exact test for binomial proportions was used to compare the proportion of retained probe sets mapping to chromosome 1 to the proportion for the entire array.

The log-scale expression levels of proliferation-associated genes tended to have high correlations with CKS1B (Table 15). In addition, 25 of 29 (86%) genes, significantly correlated with the plasma cell labeling index, were strongly correlated with CKS1B, suggesting that this gene participated in a proliferation signaling network in patients with high risk disease. CKS1B is an independent predictor of overall survival after adjustment for chromosome 13 deletion, cytogenetic abnormalities, clinical prognostic factors and labeling index (P=0.007, Table 15, last column, row 17). Few other chromosome 1 genes are both strong independent predictors of survival, proliferation and *CKS1B* gene amplification. Since amplification of 1q21 was associated with myeloma progression the over-representation of 1q genes among the list of 70 justified a focus on this region in search for a molecular basis of high-risk myeloma; 4 genes (*TAGNL2, PSMD4, MGC57827, CKS1B*) map to 1q21, among which *CKS1B* quartile 4 membership was most strongly associated with survival in unadjusted log rank tests (i.e. according to the list order of Table 15). DNA synthesis is mediated by the action of the cyclin E/CDK2 complex, which is in turn negatively regulated by the cyclin-dependent kinase inhibitor p27Kip1. CKS1 is required for SCFSkp2-mediated ubiquitinylation and proteasomal degradation of cyclin-dependent kinase inhibitor p27Kip. p27Kip1 degradation not only permits DNA replication but also ensures the correct progression of cells through S phase into mitosis and Cks proteins interact with the proteasome to control the proteolysis of mitotic cyclins by way of regulating the transcriptional

activity of CDC20, a regulatory subunit of the anaphase-promoting complex/cyclosome ubiquitin ligase. Thus, CKS1 and the SCFSkp2-p27Kip1-Cdk1/2 axis appear to be important for both DNA synthesis and mitosis. The low p27Kip1 protein levels in cancer cells, in the absence of gene mutations, has prompted speculation that hyper-activation of CKS1B and/or SKP2, may account for the low levels of p27Kip1. Given its well-documented role in regulating cell cycle progression, its map location, link to myeloma cell proliferation and patient survival, CKS1B was considered a candidate gene, the inappropriate expression of which may promote an aggressive phenotype.

CKS1B levels were strongly correlated with clinical outcome (Figures 15A-B): 25 deaths had occurred among 88 patients with quartile 4 expression levels compared to only 39 among the 263 patients with quartile 1-3 levels ($p < 0.0001$, false discovery rate, 2.5%); this was also true for event-free survival (34 of 88 in the quartile 4 cohort had experienced an event compared to 64 of 263 in the remainder; $p < 0.0001$). Levels of SKP2, the CKS1B partner gene, were not associated with survival ($P = 0.3$). Up to 8 copies of *CKS1B* were detected in metaphases of primary samples (Figure 16A). Interphase fluorescence *in situ* hybridization analysis revealed 3 or more copies of *CKS1B* in 46% among 197 cases with concurrent gene expression data. Expression levels were linked to *CKS1B* copy number (Figure 16B). Conversely, amplification increased in frequency as CKS1B expression levels increased from quartile 1 to quartile 4 ($P < 0.0001$, Table 17).

Table 17. Relationship between CKS1B gene expression quartiles and CKS1B amplification in newly diagnosed myeloma

CKS1B Expression[†]	*# AMPLIFIED*	*% AMPLIFIED*
quartile 1[‡] n = 44	9	20%
quartile 2 n = 43	12	28%
quartile 3 n = 51	26	51%
quartile 4 n = 59	44	75%
total 197	91	46%

[†] $P < 0.0001$. Amplification is defined as >= 20% of cells with 3 or >=4 CKS1B signals, for validation in conjunction with Figures 7c-d. Other tables use the CKS1B amplification index and its optimal cutoff.[‡] Quartile assignments based upon 351 patients with GE

Examination of *CKS1B* gene amplification in the context of expression levels of the 70 genes (Table 15) revealed, as expected, correlations with genes on chromosome 1q21 but, importantly, also with genes linked to cell proliferation not mapping to 1q21. The BAC clone used to evaluate *CKS1B* gene copy number also measured the copy number of *PBXIP1* (mapping centromeric to *CKS1B*) and *PB591, LENEP, ZFP67, FLJ32934, ADAM15* and *EFNA4* (all mapping telomeric to *CKS1B*). In examining the relationship between gene copy number and the expression levels of these genes (Table 18), RNA expression was most strongly correlated with DNA copy number in the case of *CKS1B*. Importantly, none of the other genes mapping to this BAC were among the 70 linked to short survival.

Table 18. Relationship of quartile 4 gene expression to amplification for genes located on bacterial artificial chromosome (BAC) used to measure 1q21 amplification.

	Not Amplified		**Amplified* (Amplification. Index. >= 46)**			**Amplification Index**	**Log Rank**
Symbol	**n / 129**	**(%)**	**n / 68**	**(%)**	**P-Value†**	**r‡**	**P-Value[a]**
PBXIP1	24	(18.6)	28	(41.2)	0.0012	0.29	0.5285
CKS1B	20	(15.5)	39	(57.4)	< 0.0001	0.50	0.0002
PB591	23	(17.8)	38	(55.9)	< 0.0001	0.43	0.0873
LENEP	31	(24.0)	18	(26.5)	0.8389	0.03	0.6507
ZFP67	27	(20.9)	29	(42.6)	0.0023	0.34	0.8717
FLJ32934	28	(21.7)	11	(16.2)	0.4606	-0.02	0.6207
ADAM15	23	(17.8)	29	(42.6)	0.0003	0.23	0.2808
EFNA4	26	(20.2)	23	(33.8)	0.0528	0.21	0.3212

* The 0-100 scale CKS1B amplification index is a weighted sum of the proportions of clonal cells with 3 copies of CKS1B and >=4 copies of CKS1B, defined by (.34 * % 3 copies + .66 * % >= 4 copies) / .66

† For a test of the independence of amplification and 4th quartile membership (N=197)

‡ Correlation between each gene's expression and the 0-100 scale CKS1B amplification index

[a] Log rank test for association of Q4 membership and overall survival (N = 351, 64 deaths)

Further, the association of *CKS1B* over-expression with survival and event-free survival was validated in a cohort of 224 patients lacking microarray data. CKS1B amplification levels were inversely correlated with both event-free survival (P< 0.0001) and overall survival (P<0.0001, Figure 16C). These effects were also observed when all 421 patients were considered (overall survival, P<0.0001, Figure 16D; event-free survival, P<0.0001)

Multivariate proportional hazards analyses were performed using the 369 patients with both CKS1B amplification samples and complete risk factor data (Table ~~10~~ 19). The 3 genetic risk factors (CKS1B amplification, chromosome 13q14 deletion, metaphase karyotype abnormalities) all independently conferred both inferior event-free and overall survival, whereas hypoalbuminemia was the only one of three standard prognostic factors that retained adverse implications for both endpoints examined. Collectively, these 6 variables accounted for 46% and 33% of variability in survival and event-free survival, respectively, with the 3 standard, non-genetic parameters contributing only an additional 7.2% and 7.4%. CKS1B amplification was an independent predictor of outcome both as a 0-100 scale index and a two-group category (Tables 19A and B), after adjustments for the variables mentioned above and for the plasma cell labeling index.

Table 19A. Multivariate proportional hazards analysis† (n = 369)

		Event-Free Survival			Survival		
	%	HR	P	Cumulative r^2	HR	P	Cumulative r^2
CKS1B Amplification Index (0-100)		1.009	0.002	0.160	1.011	0.002	0.219
FISH Chromosome 13 Deletion	25.5	1.786	0.006	0.224	1.879	0.014	0.308
Abnormal Karyotype	35.0	1.875	0.001	0.272	2.298	< 0.001	0.393
Beta-2-microglobulin >= 4 mg/L	35.8	1.478	0.046	0.305	1.396	0.170	0.422
C-reactive protein >= 4 mg/L	63.4	1.533	0.028	0.320	1.586	0.055	0.448
Albumin < 3.5 g/dL	16.5	1.660	0.019	0.336	1.698	0.044	0.461
Events / Deaths		127			84		

Table 19B. Multivariate proportional hazards analysis† (n = 369)

		Event-Free Survival			Survival		
	%	HR	P	Cumulative r^2	HR	P	Cumulative r^2
CKS1B Amplification Index >= 46	32.5	1.68	0.008	0.132	2.12	0.001	0.207
FISH Chromosome 13 Deletion	25.5	1.74	0.010	0.204	1.83	0.020	0.293
Abnormal Karyotype	35.0	1.94	< 0.001	0.257	2.33	< 0.001	0.383

Beta-2-microglobulin >= 4 mg/L	35.8	1.52	0.033	0.293	1.43	0.140	0.417
C-reactive protein >= 4 mg/L	63.4	1.49	0.038	0.312	1.56	0.060	0.443
Albumin < 3.5 g/dL	16.5	1.69	0.016	0.331	1.73	0.035	0.455
Events / Deaths	127				84		

Paired CKS1B expression data at diagnosis and relapse, available in 32 cases, revealed increased expression in 84% at relapse (P=0.0001, Figure 18), which was especially prominent in patients with quartile 1-3 expression levels at diagnosis. Paired CKS1B copy number data at diagnosis and relapse were available in 17 patients: of 7 lacking amplification at diagnosis, 4 acquired >= 3 copies at relapse; of 10 cases with 3 copies at diagnosis, 4 had acquired >= 4 copies at relapse but 2 cases with 4 or more copies at diagnosis exhibited no further amplification at relapse. These data suggested that CKS1B amplification/over-expression was also associated with disease progression.

The frequency of CKS1B quartile 4 expression varied among previously reported genetic subgroups (Fonseca, R. et al., 2004) (Table 20). With respect to gene expression-based translocations, nearly two-thirds of patients with *MAF* or *MAFB* activation, one-third each with *FGFR3/MMSET* and *CCND1* activation, and only 18% without translocations had *CKS1B* hyper-activation (P < 0.0001). When examined in the context of metaphase karyotypes, *CKS1B* quartile 4 expression was present in approximately 20% of cases with hyperdiploid or normal, i.e. uninformative, karyotypes, whereas this feature was seen in nearly 50% of patients with hypodiploid and other cytogenetic abnormalities (P = 0.0002).

In a separate multivariate analysis that adjusted for genetic subgroups, *CKS1B* quartile 4 expression remained an independent adverse outcome predictor (Table ~~11~~ 20); the gene expression-derived translocation category as a whole conferred inferior event-free (P = 0.034) but not of overall survival (P = 0.261); however, consistent with published data (Fonseca R. et al., 2004), *CCND1* activation impacted both endpoints favorably. While not adjusted for the multiple log rank tests that identified the 70 genes, this analysis, suggests that *CKS1B* expression retains explanatory power within relevant genetic subgroups.

Table 20A. Relationship between genetic abnormalities and *CKS1B* expression quartiles

		CKS1B		
Abnormality		Q4		
Category†	**n / 347 (%)**	**n**	**(%)**	**P-Value***
Expression-derived translocation				
t(11;14)	60 (17.3)	20	(33.3)	< 0.0001
t(4;14)	48 (13.8)	17	(35.4)	
t(14;16) & t(14;20)	14 (4.0)	9	(64.3)	
No Translocation Spike	225 (64.8)	41	(18.2)	
Metaphase karyotype				
Hyperdiploid	55 (15.9)	10	(18.2)	0.0002
Non-hyperdiploid	48 (13.8)	24	(50.0)	
Other Cytogenetics Abnormality	9 (2.6)	4	(44.4)	
No Cytogenetics Abnormality	235 (67.7)	49	(20.9)	

† Translocations were determined from the expression spikes t(11;14)=CCND1, t(4:14)=FGFR3/MMSET, t(14;16)=MAF and t(14;20)=MAFB. Aneuploidy and other cytogenetic abnormalities were determined from cytogenetics, for which 4 observations were missing.

* Fisher's exact test of the independence of each category and CKS1B 4th quartile membership. Under the null hypothesis, Q4 contains on average 25% of patients within each level, corresponding to a proportional distribution across Q1-3 and Q4.

Table 20B: Multivariate analysis of CKS1B quartile 4 expression and cytogenetic abnormalities†

	Event-Free Survival			Survival	
	HR	**P**		**HR**	**P**
CKS1B Q4	1.97	0.003		2.16	0.005
Expression-derived translocation					
t(11;14)	0.59	0.034		0.82	0.261
t(4;14)	1.67			1.77	
t(14;16) & t(14;20)	1.48			1.12	
Metaphase karyotype					
Hyperdiploid	1.75	0.006		1.84	0.013
Non-hyperdiploid	2.29			2.56	
Other Cytogenetics Abnormality	2.35			2.71	
r^2	0.218			0.223	
Events / Deaths	97			63	

† N = 347. Of 351 patients with expression data, 4 are missing cytogenetics.

‡ Partial likelihood ratio test for the overall effect of the category.

Additionally, Western blot analysis of nuclear protein from plasma cells from 27 newly diagnosed myeloma cases and 7 myeloma cell lines showed a

strong correlation between CKS1B mRNA and protein, but no correlation between mRNA and protein levels for $p27^{Kip1}$. However, CKS1B protein and $p27^{Kip1}$protein levels showed an inverse correlation (Figure 17). Cytoplasmic and non-phosphorylated-thr-187-$p27^{Kip1}$ levels were not altered in myeloma cell lysates with respect to CKS1B expression. Levels of $p27^{Kip1}$ protein were not correlated with the mRNA levels of SKP2.

Thus, in summary, fluorescence in situ hybridization (FISH) analysis for 1q21 amplification and deletion 13 represents a better method for risk assessment than any other that exists. Additionally, since FISH demonstrated that the strong correlation of expression of CKS1B with gene copy number and also that the risk of death increased with increase in gene copy number, FISH could be used to detect residual disease and predict recurrence and progression. Since Western blot analysis demonstrated that an increase in the levels of CKS1B protein, CKS1B or any pathway in which it works could be therapeutically targeted. Thus, the amplification of 1q21 could be a diagnostic, prognostic and potential target in cancer especially myeloma.

The following references are cited herein:

Attal M et al., *N Engl J Med* 349: 2495-2502 (2003).

Barlogie B. et al., *Williams Hematology* : 1279-1304 (2001).

Barlogie B. et al. *Blood* 103: 20-32 (2004).

Bullinger L et al., *N Engl J Med* 350: 1605-1616 (2004).

Eisen et al., *Proc Natl Acad Sci U S A* 95:14863-14868 (1998).

Fonseca R et al., *Cancer Res* 64: 1546-1558 (2004).

Ganoth D. et al., *Nat Cell Biol* 3: 321-324 (2001).

Guiterrez N.C. et al., *Blood* 104:607-618 (2004).

Hideshima et al., *Blood* 104: 607-618 (2004).

Kuehl W.M. and Bersagel P.L. *Nature Rev Cancer* 2: 175-187 (2001).

Le Baccon P. *Genes Chromosome Cancer* 32: 250-264 (2001).

McCoy J. et al. *Blood* 102: 2512a (2003).

Morris M.C. et al., *Nature* 423(6943): 1009-1013 (2003).

Nakayama K. et al., *Dev Cell* 6: 661-672 (2004).

Nilsson T et al., *Br J Haematol* 120: 960-969 (2003).

O'Quigley J and Xu R. Explained variation in proportional hazards. In: *Crowley J ed. Handbook of Statistics in Clinical Oncology*. New York, NY: Marcel Dekker, 397-410 (2001).

Pagano M. *Mol Cell* 14: 414-416 (2004).

Pagano M. and Benmaamar R. *Cancer Cell* 4:251-256 (2003).

Peters J.M. *Mol Cell* 95: 931-943 (2002).

R Development Core Team R: A language and environmanet for statistical computing. R Foundation for statistical Computing, Vienna, Austria 2004. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rosenwald A. et al. *N Engl J Med* 346: 1937-1947 (2002).

Santra et al., *Blood* 101:2374-2376 (2003).

Sawyer J.R. et al. *Genes Chromosomes Cancer* 42: 95-106 (2005).

Sawyer J.R. et al. *Blood* 91: 1732-1741 (1998).

Shaughnessy J and Barlogie B. *Immunol Rev* 94: 140-63 (2003).

Shaughnessy J et al., *Blood* 101: 3849-3856 (2003).

Shaughnessy et al., *Blood* 96:1505-1511 (2000).

Shaughnessy et al., *Br. J. Haematol.* 120:44-52 (2003).

Sherr C.J. and Roberts J.M. *Genes Dev* 13: 1501-1512 (1999).

Shipp M.A. et al. *Nat Med* 8: 68-74 (2002).

Slingerland J and Pagano M. *J Cell Physiol* 183: 10-17 (2000).

Spruck C. et al. *Mol Cell* 7: 639-650 (2001).

Storey and Tibshirani, *Proc Natl Acad Sci* 100(16): 9440-9445 (2003).

Tarte K. et al., *Blood* 102; 592-600 (2003).

Tian E. et al. N Engl J Med 349; 2483-2494 (2003).

Tibshirani et al., *Proc. Natl. Acad. Sci. U. S. A.* 99:6567-6572 (2002).

Valk P.J. et al. *N Engl J Med* 350: 1617-1628 (2004).

Yeoh E.J. et al. *Cancer Cell* 1: 133-143 (2002).

Zhan et al., *Blood* 99:1745-1757 (2002a).

Zhan et al., *Blood* 101:1128-1140 (2003).

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. Further, these patents and publications are incorporated by reference herein to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. A method of diagnosis for multiple myeloma, comprising the steps of:

(a) isolating biological samples from an individual;

(b) examining the expression levels of a group of genes comprising *CCND1, MAF, MAFB, FGFR3,* and *MMSET,* or determining chromosomal translocation represented by overexpression of a group of genes comprising *CCND1, MAF, MAFB, FGFR3,* and *MMSET*; and

(c) examining the expression levels of a group of genes comprising *CST6, RAB7L1, MAP4K3, HRASLS2, TRAIL, IG, FGL2, GNG11, MCM2* and *FLJ10709*, wherein results from steps (b) and (c) would classify said individual into one of 7 groups of myeloma (multiple myeloma groups 1-7).

2. The method of claim 1, wherein said biological samples are plasma cells.

3. The method of claim 1, wherein individual in said myeloma groups 1, 2, 3 and 6 would have poor prognosis compared to individual in said myeloma groups 4, 5 and 7.

4. The method of claim 1, wherein overexpression of *CCND1*, or translocation t(11;14)(q13;q32) would classify said individual into myeloma group 3.

5. The method of claim 1, wherein overexpression of *MAF* or *MAFB*, or translocation t(14;16)(32;q23) or t(14;20)(q32;q13) would classify said individual into myeloma group 5.

6. The method of claim 1, wherein overexpression of *FGFR3* or *MMSET*, or translocation t(4;14)(p21;q32) would classify said individual into myeloma group 7.

7. The method of claim 1, wherein overexpression of *CST6, RAB7L1, MAP4K3,* and *HRASLS2,* and downregulation of *TRAIL* would classify said individual into myeloma group 6.

8. The method of claim 1, wherein downregulation of *CST6, RAB7L1, MAP4K3, HRASLS2, GNG11, MCM2* and *FLJ10709*, and overexpression of *TRAIL, IG* and *FGFL2* would classify said individual into myeloma group 1.

9. The method of claim 1, wherein downregulation of *CST6, RAB7L1, MAP4K3, HRASLS2, IG, FGFL2, MCM2* and *FLJ10709*, and overexpression of *TRAIL* and *GNG11* would classify said individual into myeloma group 2.

10. The method of claim 1, wherein downregulation of *CST6, RAB7L1, MAP4K3, HRASLS2, IG, FGFL2,* and *GNG11*, and overexpression of *TRAIL, MCM2* and *FGFL2* would classify said individual into myeloma group 4.

11. A method of identifying genes associated with a phenotype of interest in a population, comprising:

isolating cells from individuals within the population;

profiling gene expression in the cells using oligonucleotide microarrays; and

correlating the gene expression profiles in the cells with the phenotype of interest, thereby identifying genes associated with the phenotype of interest in the population.

12. The method of claim 11, wherein the phenotype of interest is disease-related survival and response to a drug.

13. A method of diagnosing an individual with myeloma, comprising the steps of:

(a) obtaining a bone marrow sample from the individual; and

(b) determining an amplification of a gene on chromosome 1q21 either alone or in combination with a deletion of a gene on chromosome 13q14 in the sample, thereby diagnosing the individual with myeloma.

14. The method of claim 13, further comprising:

predicting relapse or progression of the disease in the individual based on the extent of amplification of the gene on chromosome 1q21, wherein an increase in the amplification of the gene on chromosome 1q21 indicates relapse or progression of the disease.

15. The method of claim 13, wherein the amplified gene is *CKS1B* and the deleted gene is *RFP2*.

16. A method of identifying an individual having an aggressive form of a cancer, comprising the steps of:

(a) obtaining a bone marrow sample from the individual; and

(b) determining an amplification of a gene on chromosome 1q21, wherein the amplification of the gene is associated with poor survival, thereby identifying an individual having the aggressive form of the cancer.

17. The method of claim 16, wherein the amplified gene is *CKS1B*.

18. The method of claim 16, wherein the cancer is breast cancer, colon cancer, prostate cancer or myeloma.

19. A method of screening for a drug useful in treating a cancer, comprising:

contacting a sample comprising CKS1B gene orCKS1B gene product with a test compound; and

determining the effect of the compound on amplification, over-expression or activity of the CKS1B gene or the CKS1B gene product, wherein an inhibitory effect of the test compound indicates that the test compound is the drug useful for treating the cancer.

20. The method of claim 19, wherein the cancer is an aggressive form of breast cancer, colon cancer, prostate cancer or myeloma.

21. A method of treating an individual having cancer, comprising:

administering to the individual a pharmacologically effective amount of a compound that inhibits amplification, over-expression or activity of CKS1B gene or CKS1B gene product.

22. The method of claim 21, wherein said compound with the inhibitory effect on the CKS1B gene is a peptide nucleic acid or RNA-mediated interference.

23. The method of claim 21, wherein said compound with the inhibitory effect on the CKS1B gene product is an antibody, a CKS1B antisense RNA or a small molecule inhibitor.

24. The method of claim 21, wherein the cancer is an aggressive form of breast cancer, colon cancer, prostate cancer or myeloma.

25. A method of treating an individual having high-risk myeloma, comprising:

administering to the individual pharmacologically effective amount of a compound that inhibits amplification, over-expression or activity of CKS1B gene or CKS1B gene product and a vector comprising DNA sequence encoding *RFP2* gene.

26. The method of claim 25, wherein said compound with an inhibitory effect on the CKS1B gene is a peptide nucleic acod or RNA-mediated interference.

27. The method of claim 25, wherein said compound with an inhibitory effect on the CKS1B gene product is an antibody, a CKS1B antisense RNA or a small molecule inhibitor.

28. A kit, comprising:

(a) probe specific for CKS1B gene;

(b) probe specific for RFP2 gene; or combinations thereof.

5

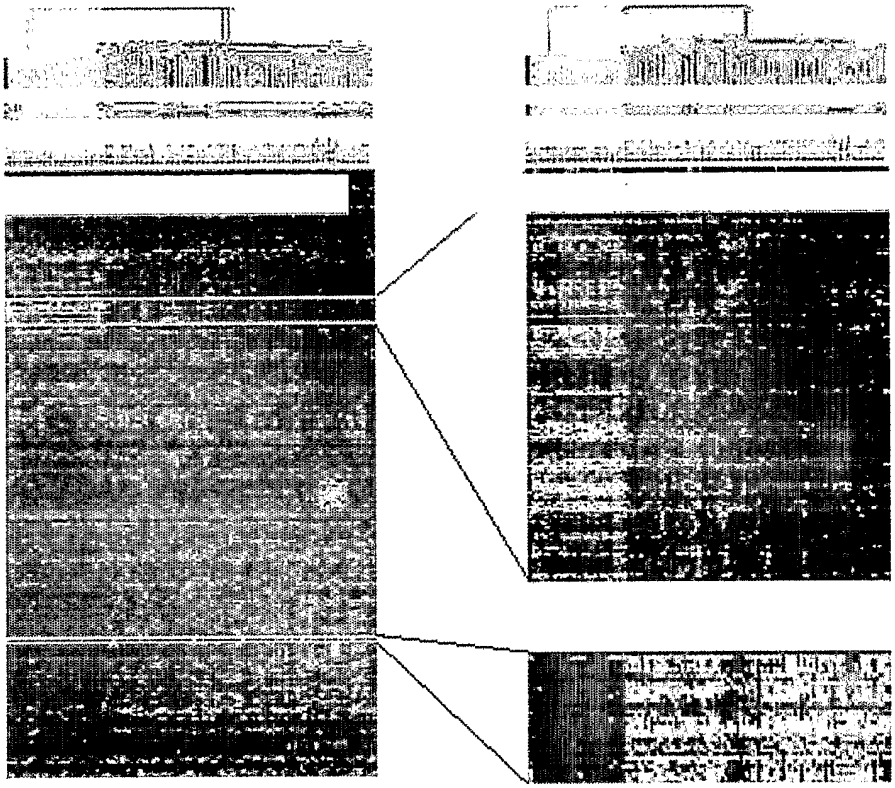


Fig. 1A

Fig. 1B

Fig. 1C

Fig. 1D

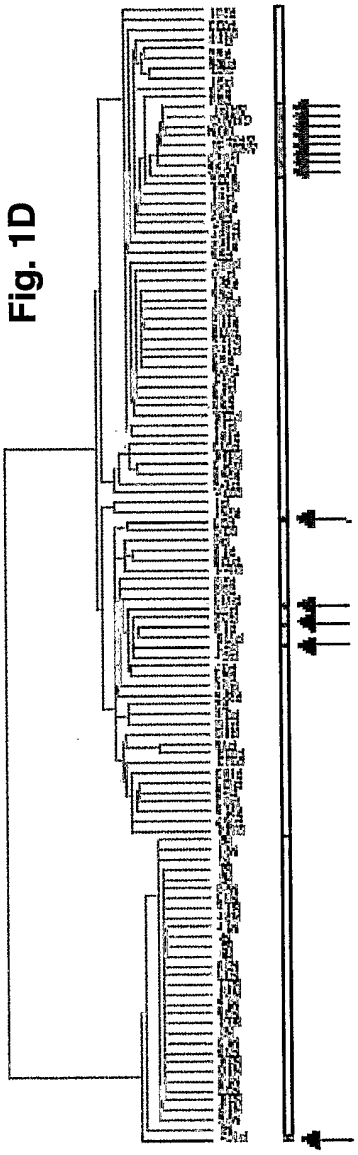
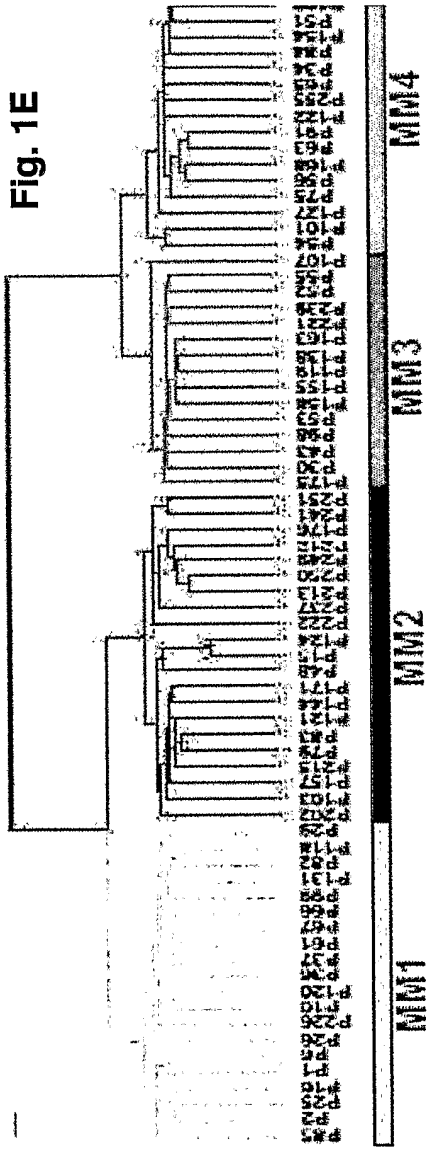


Fig. 1E



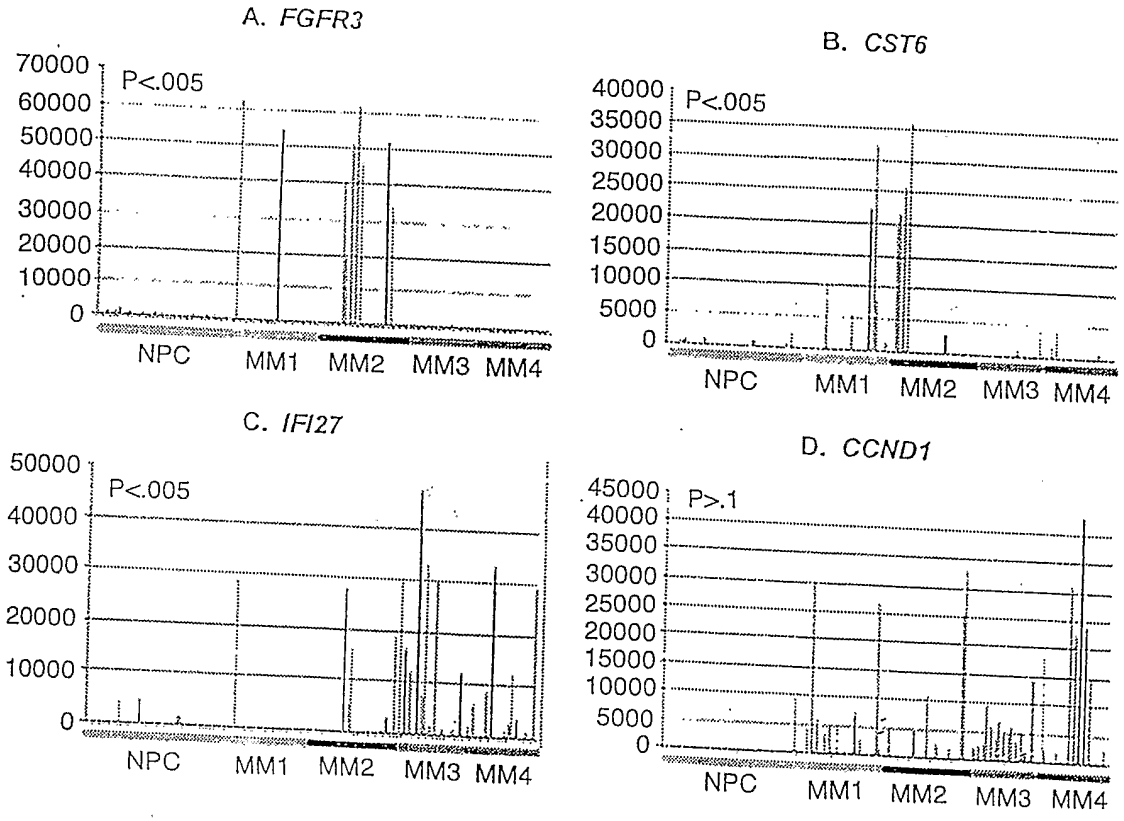

A. FGFR3
P<.005
70000
60000
50000
40000
30000
20000
10000
0
NPC
MM1
MM2
MM3
MM4
B. CST6
P<.005
40000
35000
30000
25000
20000
15000
10000
5000
0
NPC
MM1
MM2
MM3
MM4
C. IFI27
P<.005
50000
40000
30000
20000
10000
0
NPC
MM1
MM2
MM3
MM4
D. CCND1
P>.1
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
NPC
MM1
MM2
MM3
MM4

Fig. 2

Fig. 3A

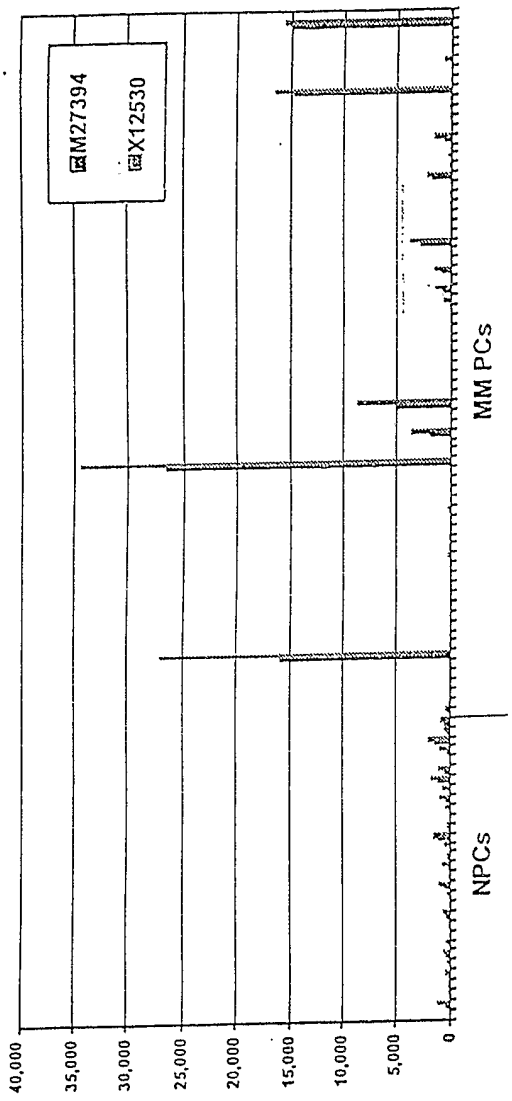
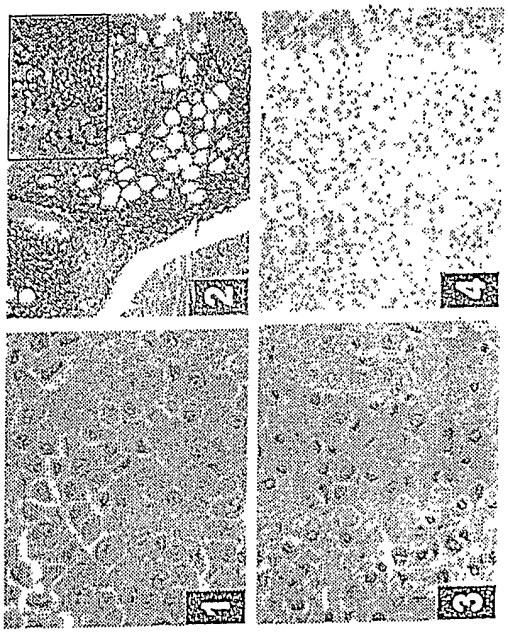

M27394
X12530
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
MM PCs
NPCs

Fig. 3B


1
2
3
4

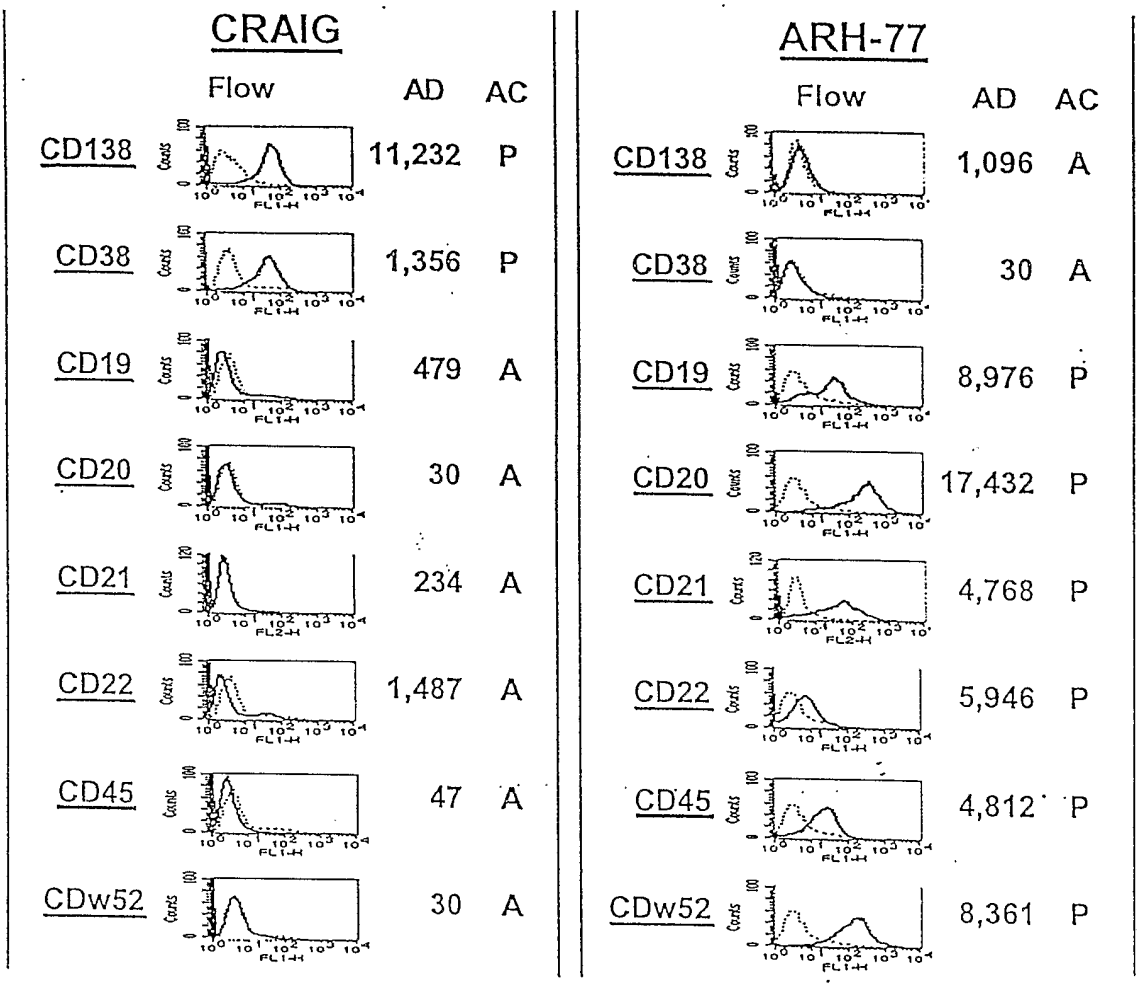

CRAIG
Flow
AD
AC
CD138
11,232
P
CD38
1,356
P
CD19
479
A
CD20
30
A
CD21
234
A
CD22
1,487
A
CD45
47
A
CDw52
30
A
ARH-77
Flow
AD
AC
CD138
1,096
A
CD38
30
A
CD19
8,976
P
CD20
17,432
P
CD21
4,768
P
CD22
5,946
P
CD45
4,812
P
CDw52
8,361
P

Fig. 4

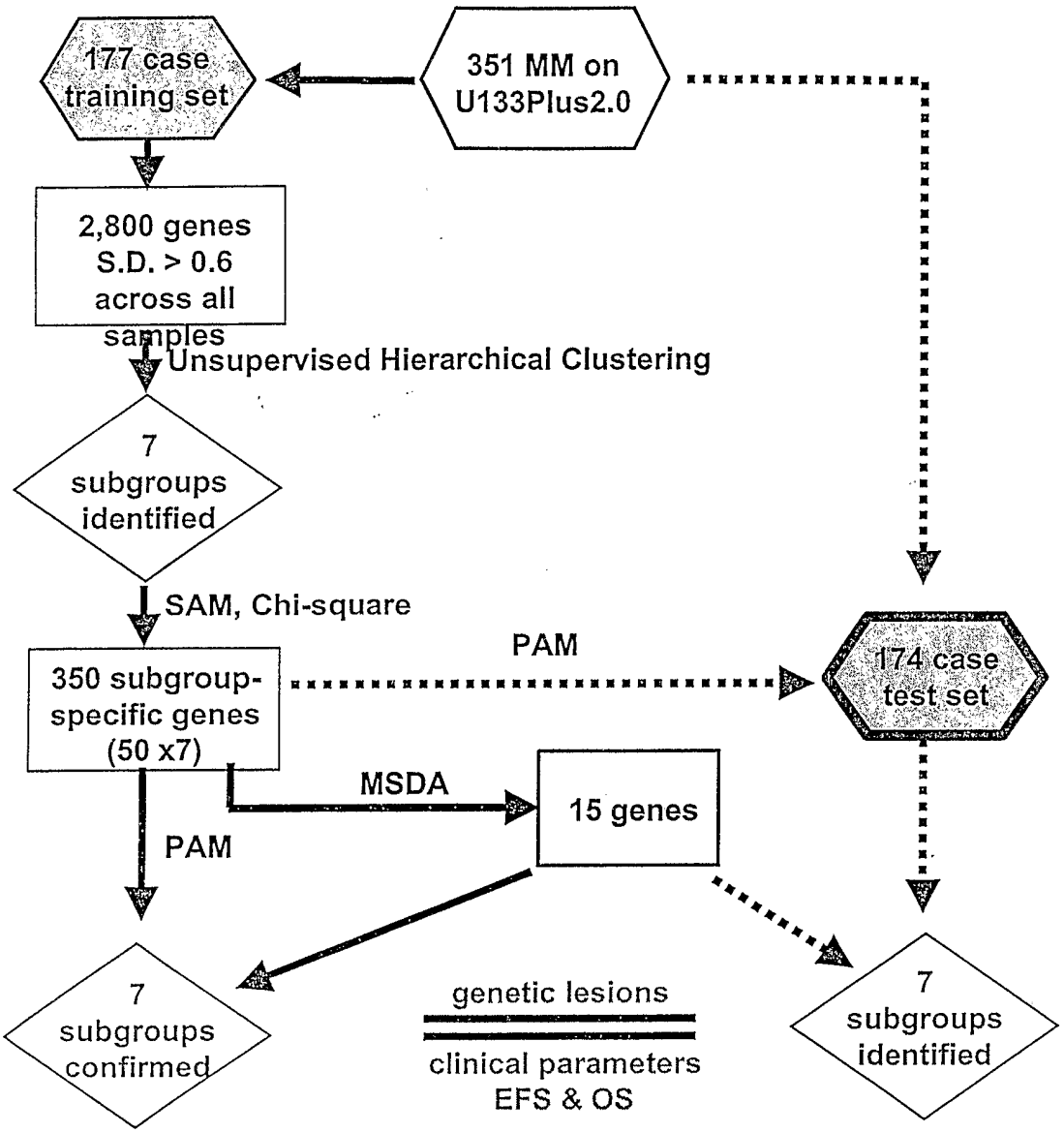

177 case training set
351 MM on U133Plus2.0
2,800 genes S.D. > 0.6 across all samples
Unsupervised Hierarchical Clustering
7 subgroups identified
SAM, Chi-square
350 subgroup-specific genes (50 x7)
PAM
174 case test set
MSDA
15 genes
PAM
7 subgroups confirmed
genetic lesions
clinical parameters
EFS & OS
7 subgroups identified

Fig. 5

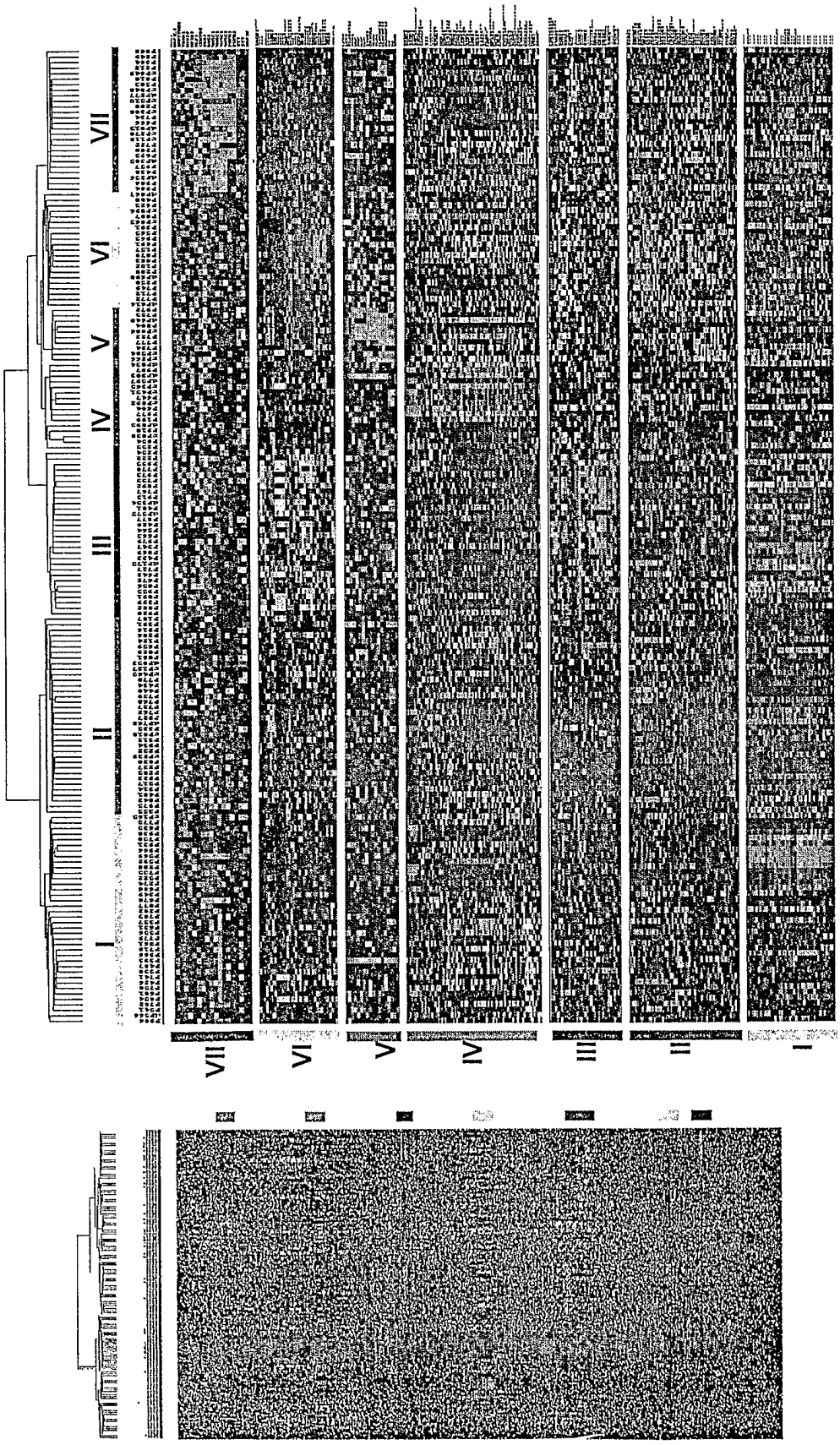


Fig. 6

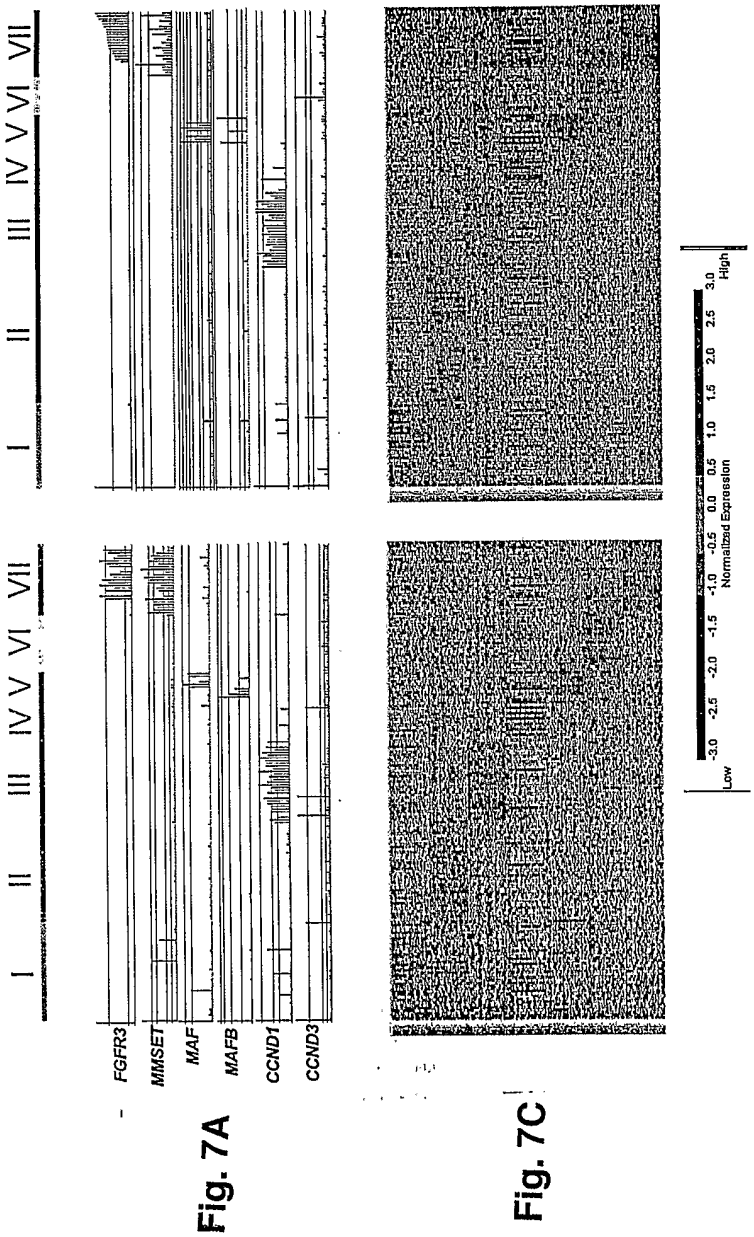


Fig. 7A

Fig. 7B

Fig. 7C

Fig. 7D

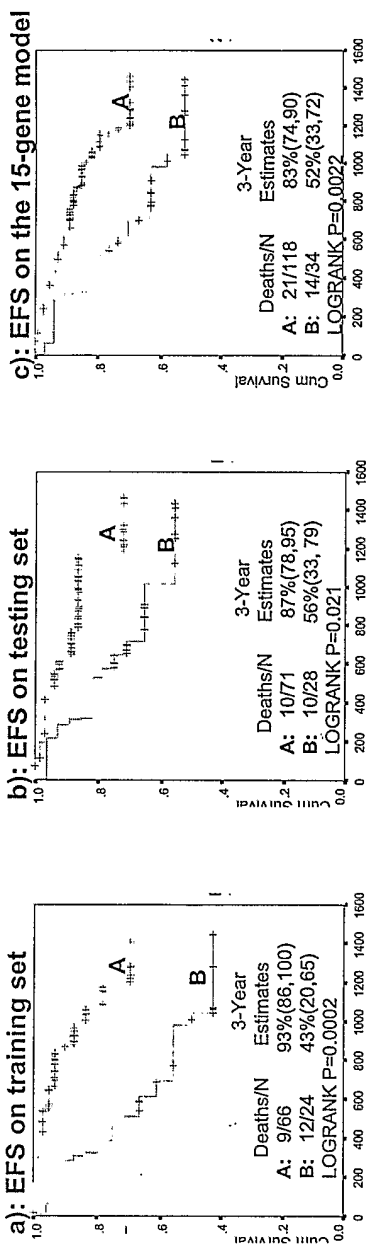


Fig. 8A

Fig. 8B

Fig. 8C

Fig. 8D

Fig. 8E

Fig. 8F

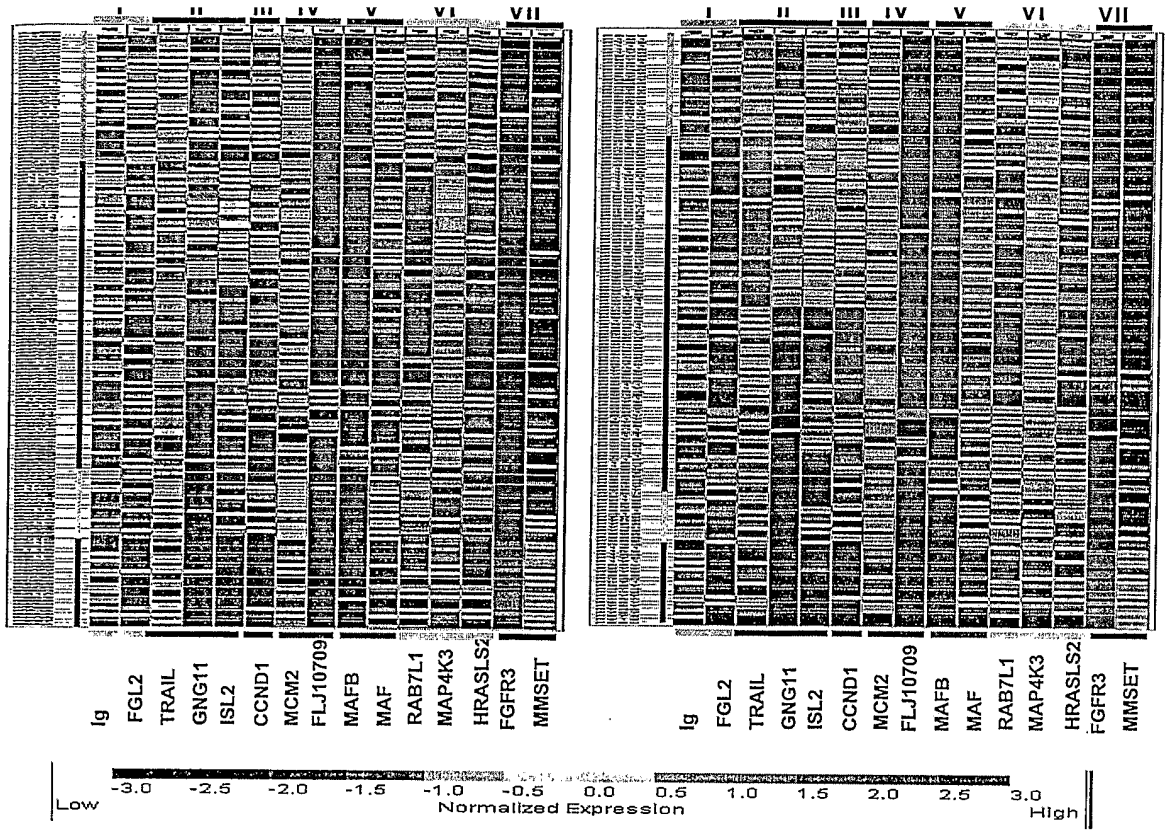

I
II
III
IV
V
VI
VII
Ig
FGL2
TRAIL
GNG11
ISL2
CCND1
MCM2
FLJ10709
MAFB
MAF
RAB7L1
MAP4K3
HRASLS2
FGFR3
MMSET
Low
-3.0
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
High
Normalized Expression

Fig. 9

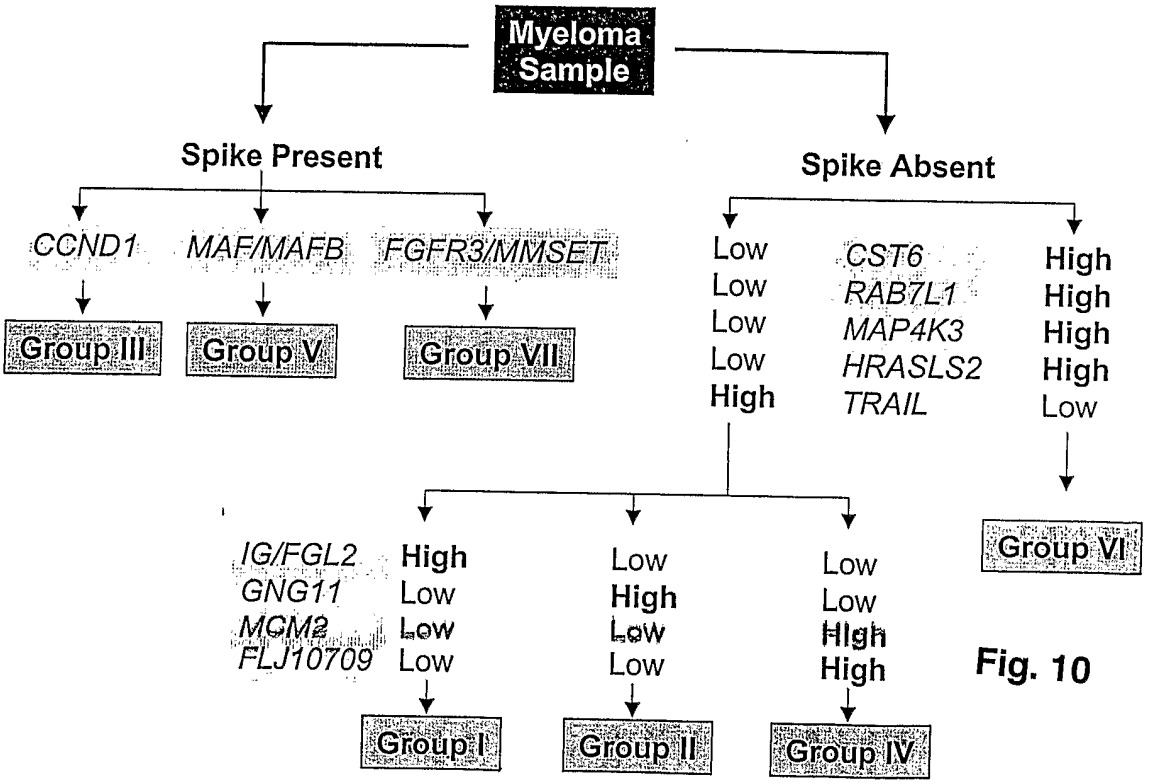

Myeloma Sample
Spike Present
Spike Absent
CCND1
MAF/MAFB
FGFR3/MMSET
Group III
Group V
Group VII
Low
Low
Low
Low
High
CST6
RAB7L1
MAP4K3
HRASLS2
TRAIL
High
High
High
High
Low
Group VI
IG/FGL2
GNG11
MCM2
FLJ10709
High
Low
Low
Low
Low
High
Low
Low
High
High
Group I
Group II
Group IV

Fig. 10

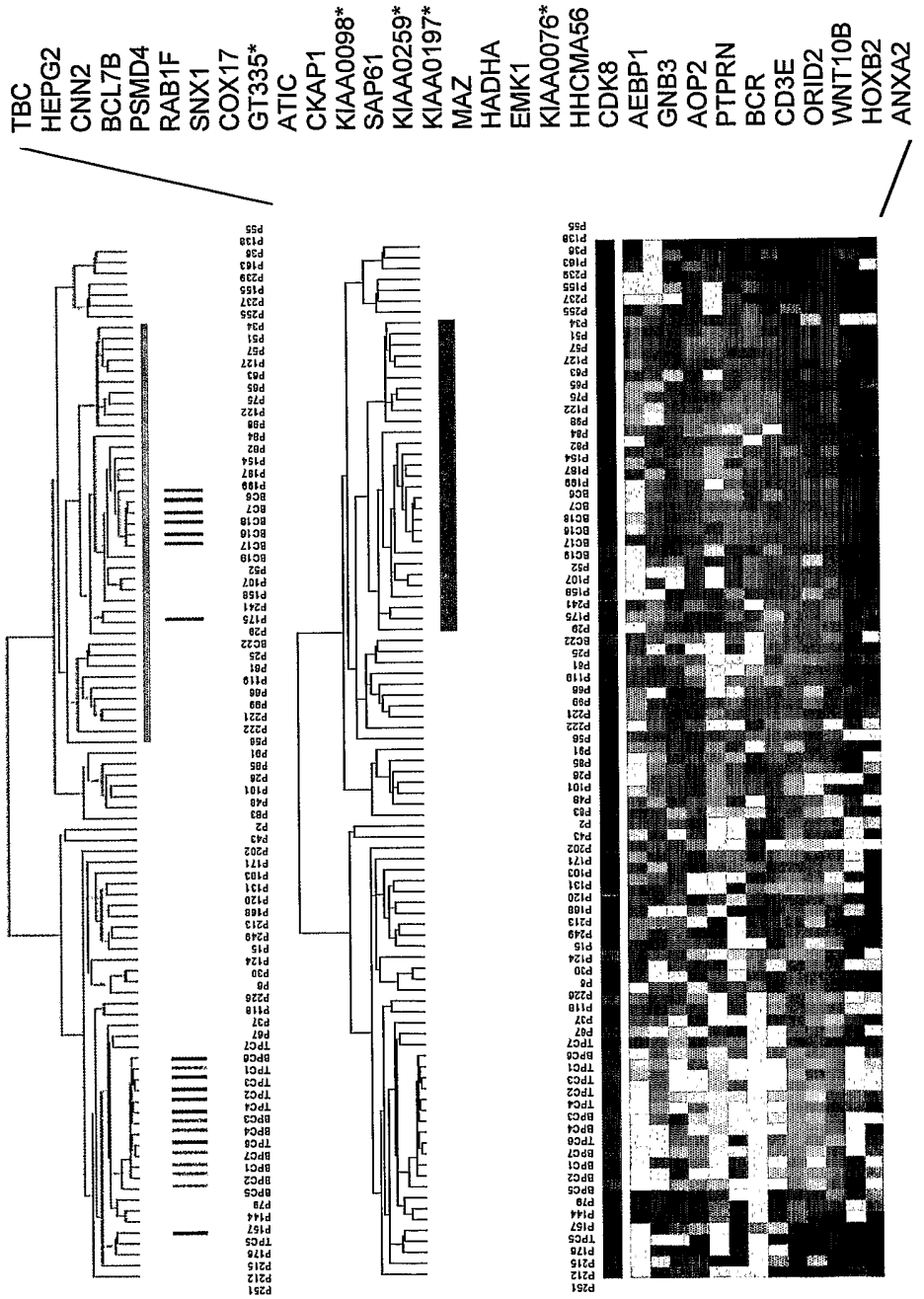

TBC
HEPG2
CNN2
BCL7B
PSMD4
RAB1F
SNX1
COX17
GT335*
ATIC
CKAP1
KIAA0098*
SAP61
KIAA0259*
KIAA0197*
MAZ
HADHA
EMK1
KIAA0076*
HHCMA56
CDK8
AEBP1
GNB3
AOP2
PTPRN
BCR
CD3E
ORID2
WNT10B
HOXB2
ANXA2

Fig. 11

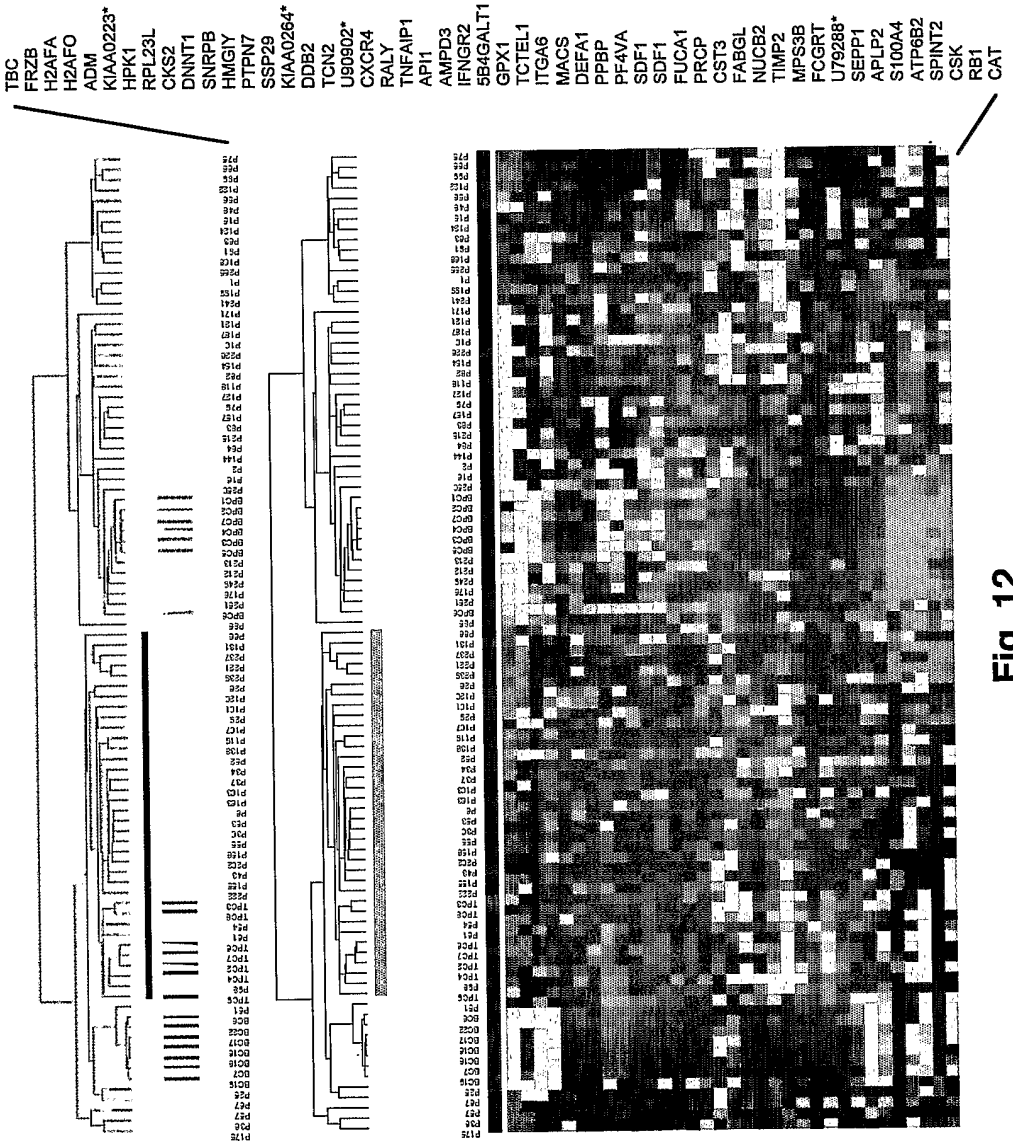

TBC
FRZB
H2AFA
H2AFO
ADM
KIAA0223*
HPK1
RPL23L
CKS2
DNNT1
SNRPB
HMGIY
PTPN7
SSP29
KIAA0264*
DDB2
TCN2
U90902*
CXCR4
RALY
TNFAIP1
API1
AMPD3
IFNGR2
5B4GALT1
GPX1
TCTEL1
ITGA6
MACS
DEFA1
PPBP
PF4VA
SDF1
SDF1
FUCA1
PRCP
CST3
FABGL
NUCB2
TIMP2
MPS3B
FCGRT
U79288*
SEPP1
APLP2
S100A4
ATP6B2
SPINT2
CSK
RB1
CAT

Fig. 12

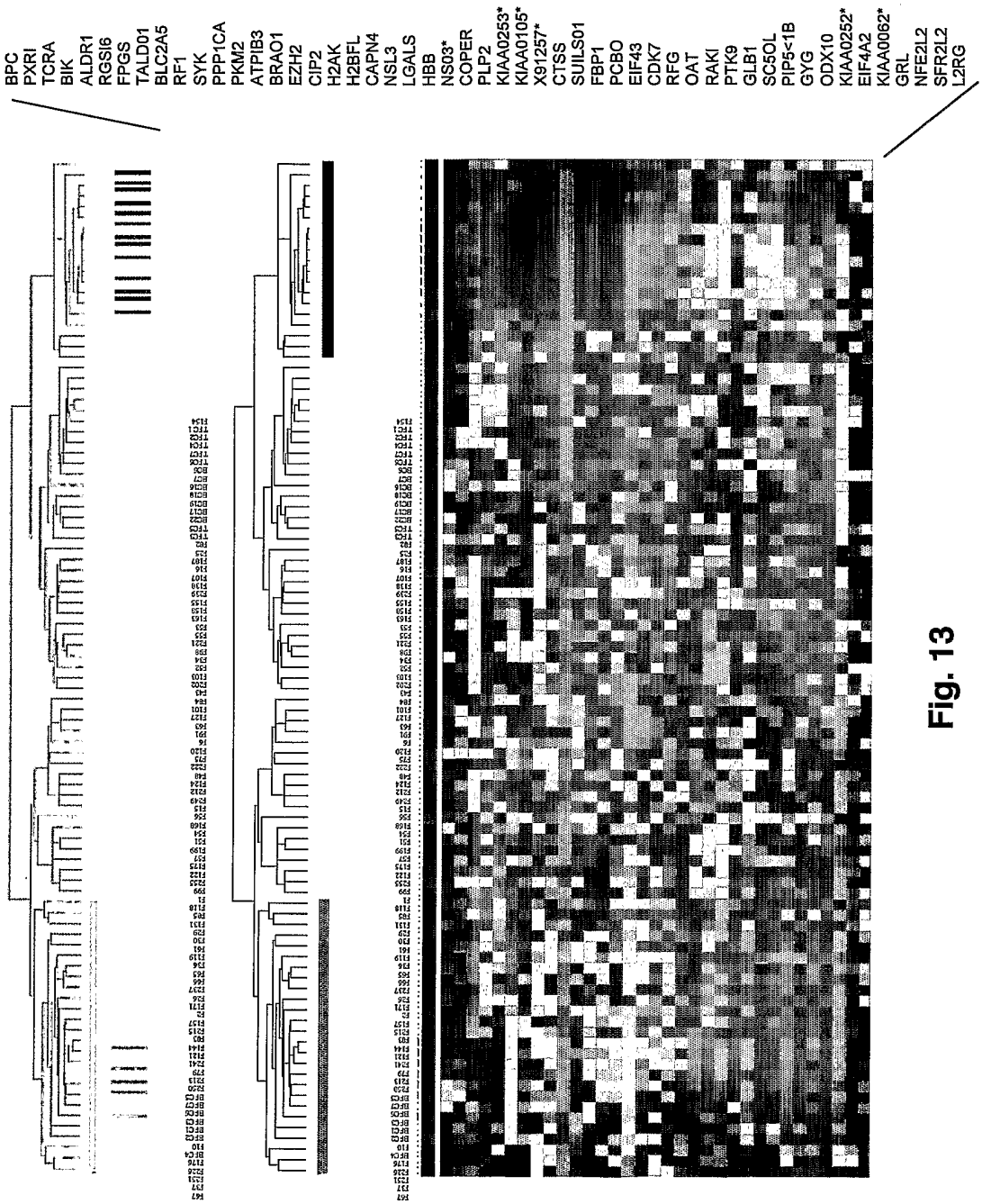

BPC
PXRI
TCRA
BIK
ALDR1
RGSI6
FPGS
TALD01
BLC2A5
RF1
SYK
PPP1CA
PKM2
ATPIB3
BRAO1
EZH2
CIP2
H2AK
H2BFL
CAPN4
NSL3
LGALS
HBB
NS03*
COPER
PLP2
KIAA0253*
KIAA0105*
X91257*
CTSS
SUILS01
FBP1
PCBO
EIF43
CDK7
RFG
OAT
RAKI
PTK9
GLB1
SC5OL
PIP5<1B
GYG
ODX10
KIAA0252*
EIF4A2
KIAA0062*
GRL
NFE2L2
SFR2L2
L2RG

Fig. 13

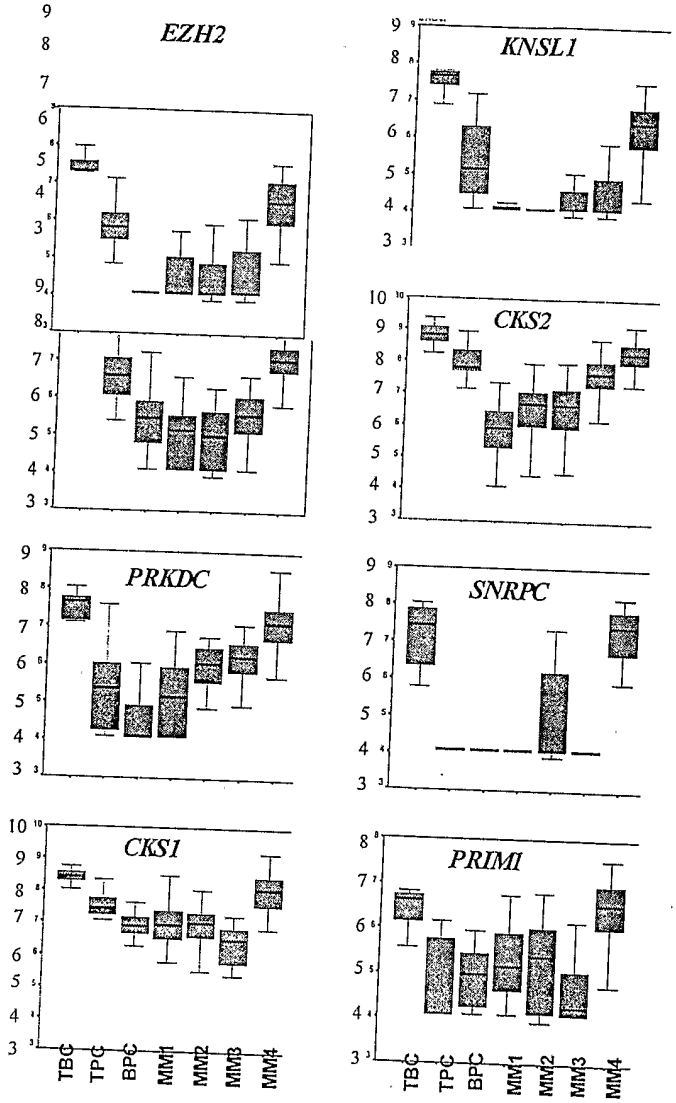

EZH2
KNSL1
CKS2
PRKDC
SNRPC
CKS1
PRIMI
TBC
TPC
BPC
MM1
MM2
MM3
MM4

Fig. 14

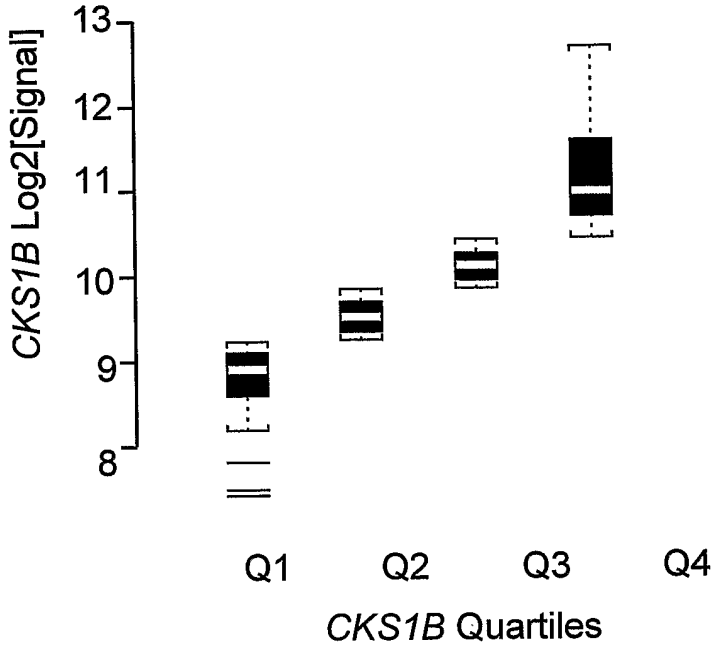


Fig. 15A

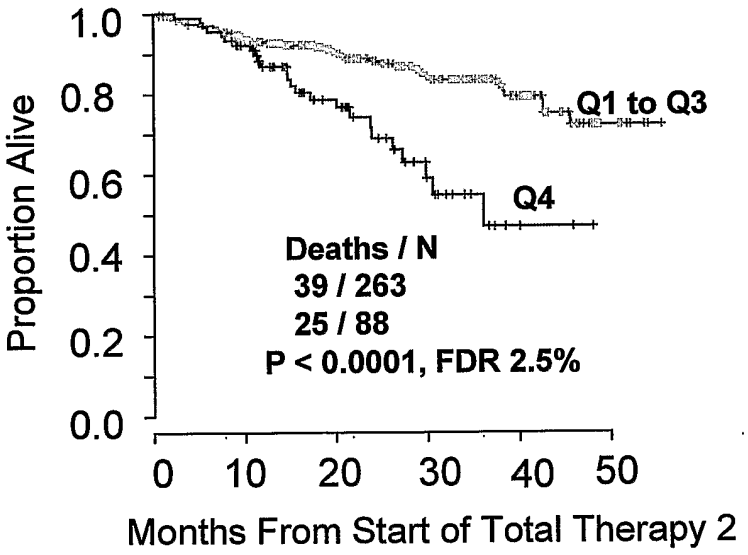


Fig. 15B

Fig. 16A

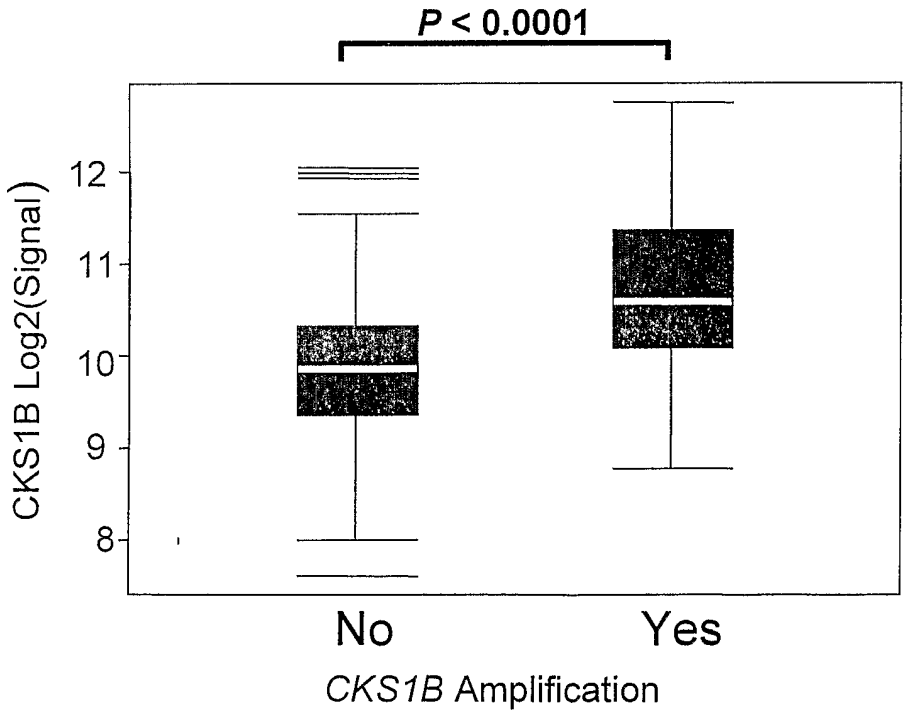


Fig. 16B

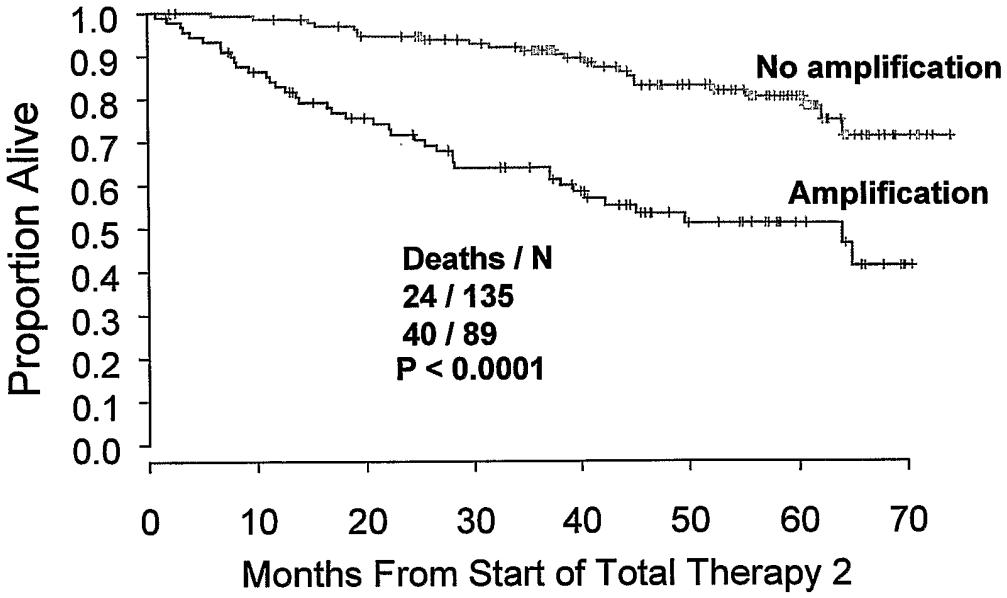


Fig. 16C

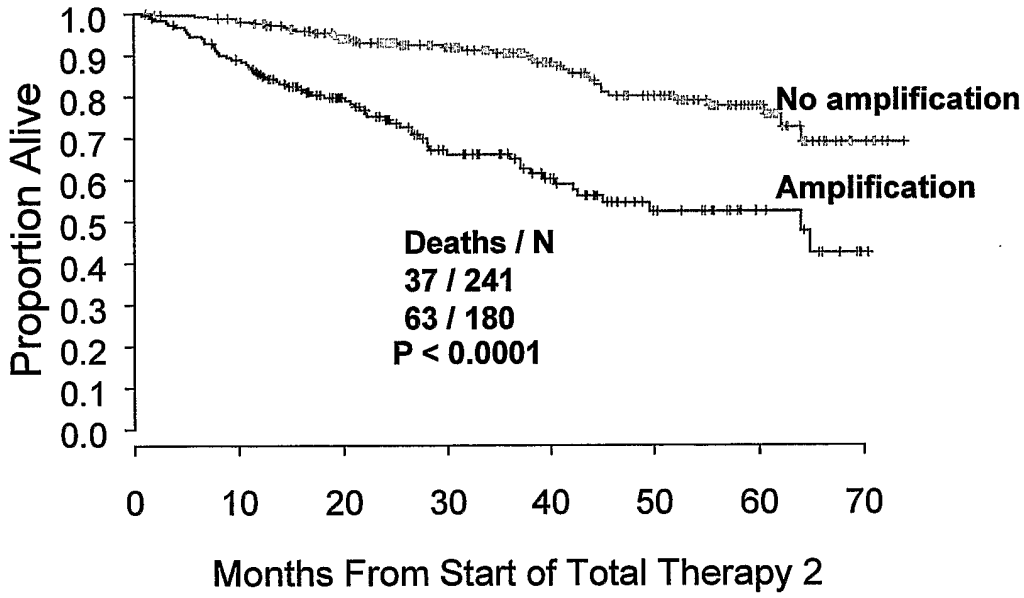


Fig. 16D

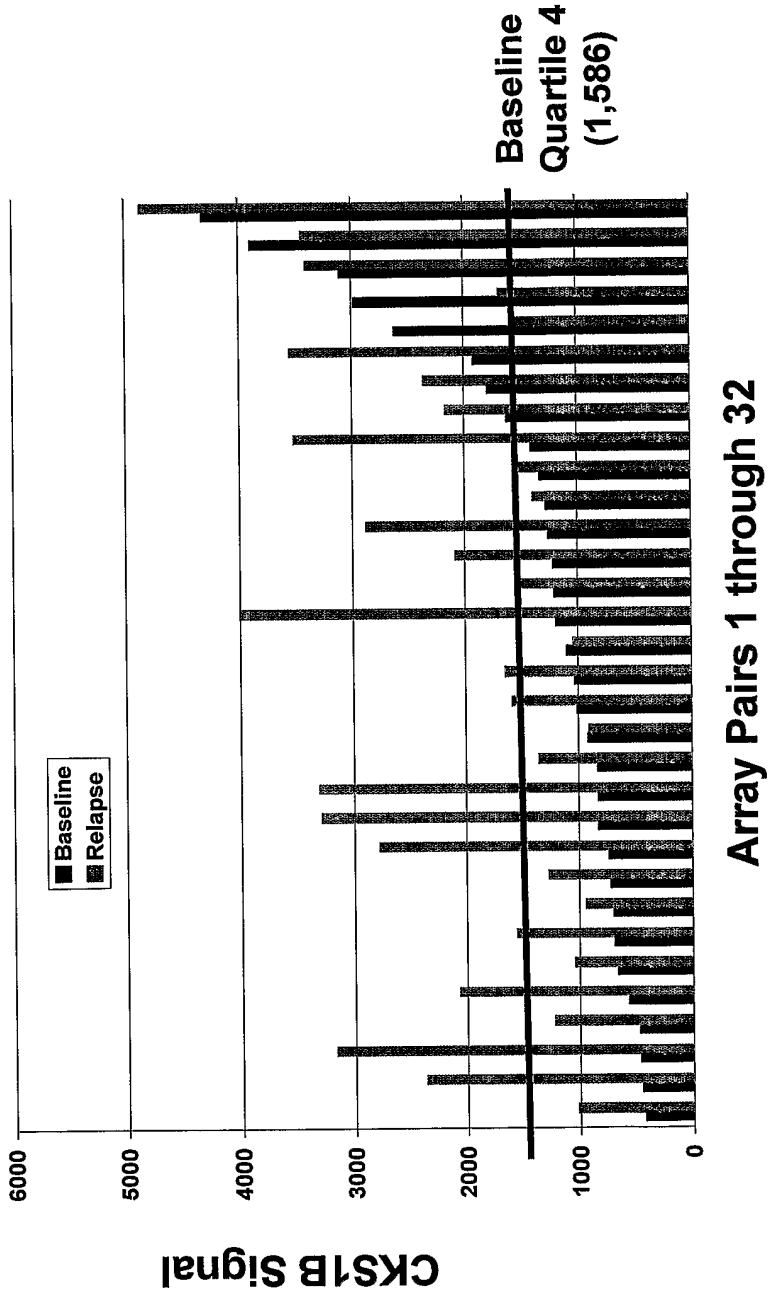

Baseline
Quartile 4
(1,586)
Baseline
Relapse
6000
5000
4000
3000
2000
1000
0
CKS1B Signal
Array Pairs 1 through 32

Fig. 17

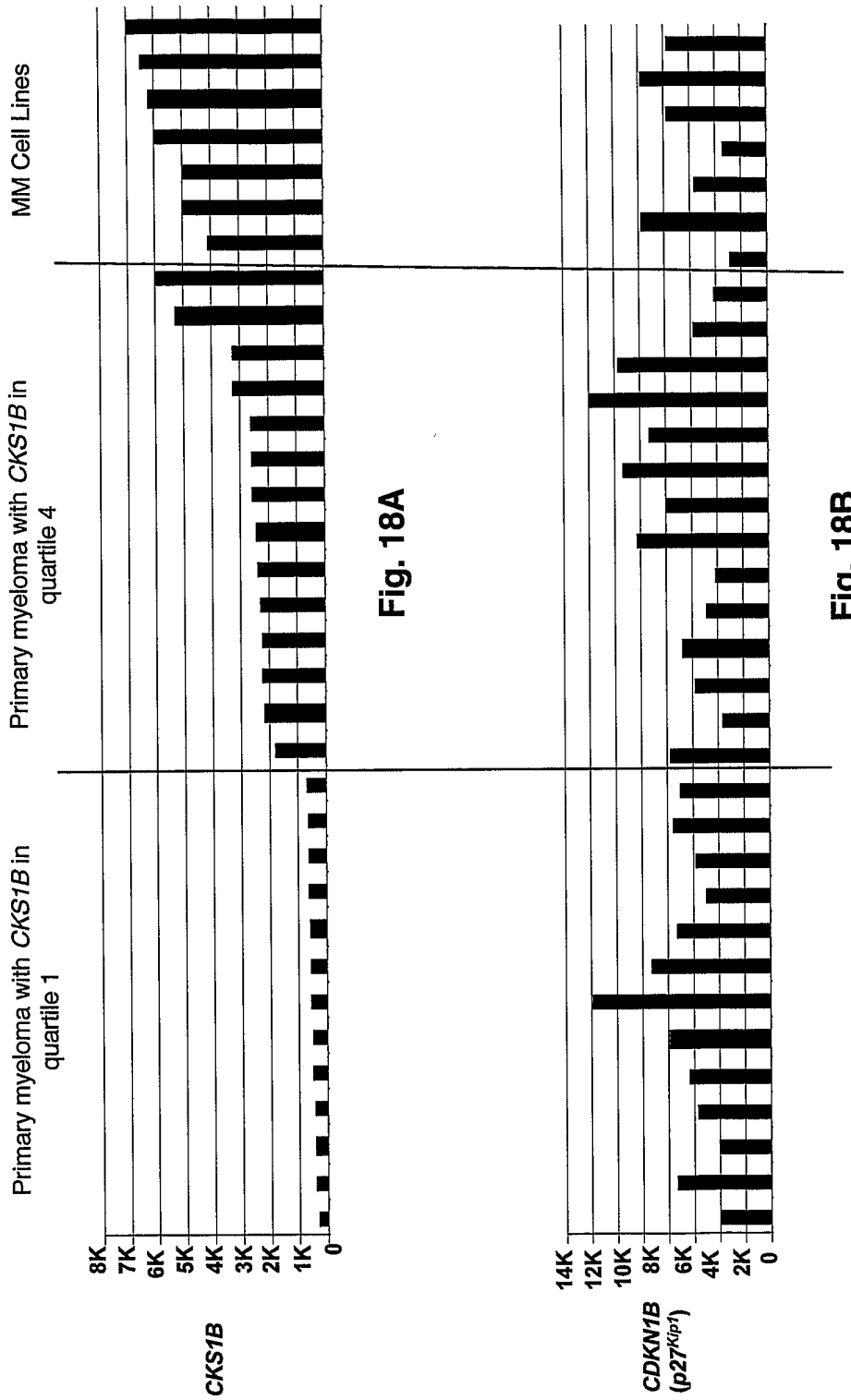


Fig. 18A

Fig. 18B

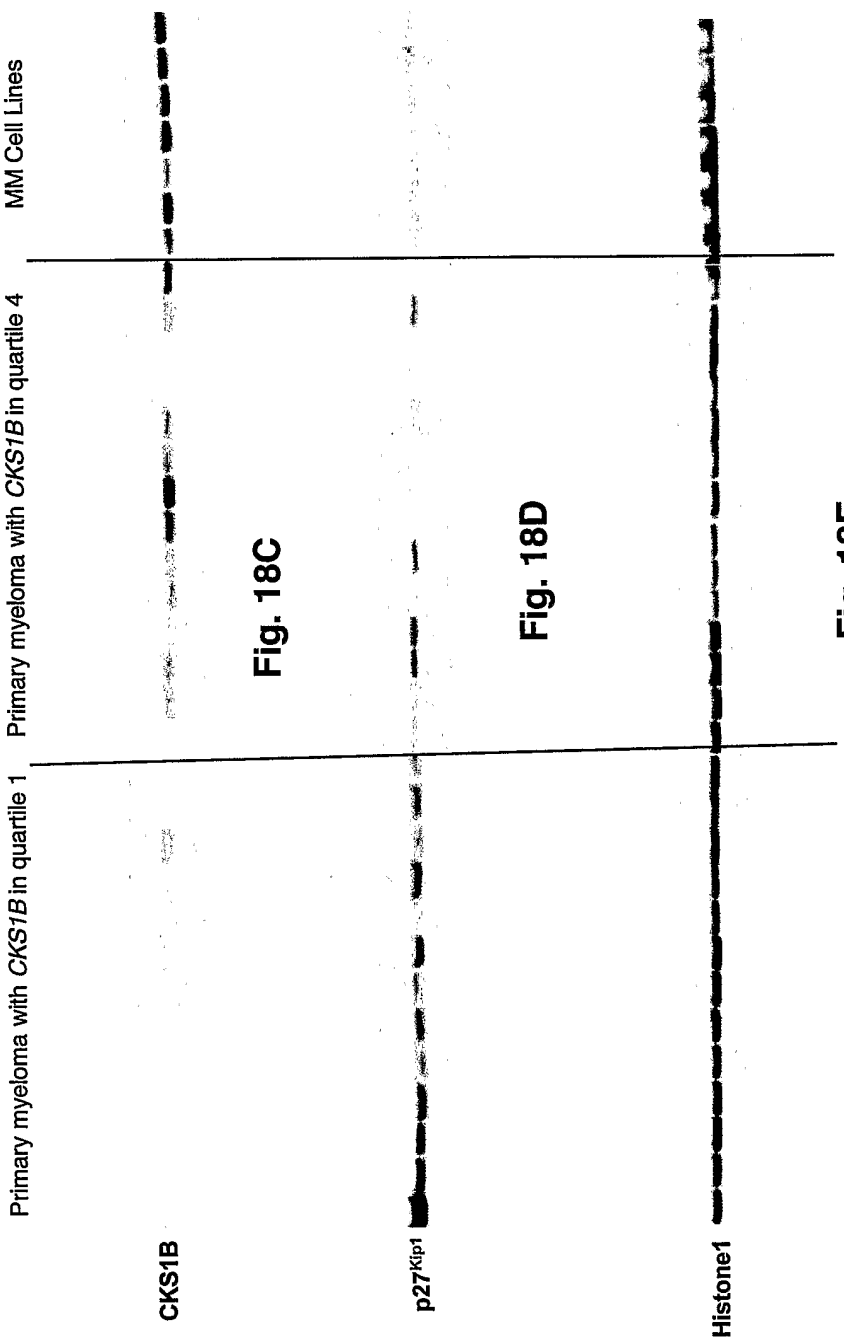


Fig. 18C

Fig. 18D

Fig. 18E

<110> Shaughnessy, John
Barlogie, Bart
Fenghuang, Zhan

<120> Diagnosis, Prognosis and Identification of Potential Therapeutic Targets of Multiple Myeloma Based on Gene Expression Profiling

<130> D6432CIP3PCT

<150> 10/931,780
<150> 11/110,209

<151> 2004-09-01
<151> 2005-04-20

<160> 2

<210> 1
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<221> primer_bind
<223> primer IGJH2

<400> 1
caatggtcac cgtctcttca 20

<210> 2
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<221> primer_bind
<223> primer MMSET

<400> 2
cctcaatttc ctgaaattgg tt 22