EUROPEAN JOURNAL OF Haematology

European Journal of Haematology 89 (28–36)

ORIGINAL ARTICLE

Low-risk identification in multiple myeloma using a new 14-gene model

Tiehua Chen[1], Tamara Berno[2], Maurizio Zangari[2]

[1]Molecular Medicine Program, University of Utah, Salt Lake City, Utah, USA; [2]Myeloma Program, University of Utah, Salt Lake City, Utah, USA

Abstract

Identifying the best gene expression pattern associated with low-risk disease in patients with newly diagnosed multiple myeloma (MM) is important to direct clinical treatments. The MM Survival Index14 (MMSI14) was developed from GEP data sets of 22 normal plasma cells (NPC), 5 MM cell lines (MMCL), 44 monoclonal gammopathy of undetermined significance (MGUS), and 351 newly diagnosed MM patients. R/bioconductor and siggenes package were used to obtain heatmap, boxplot and histogram whose results were then analyzed by Kaplan–Meier analysis. Fourteen genes associated with low-risk disease in MM were identified. We validated the disease prognostic power of MMSI14 with an independent data set of other 214 newly diagnosed MM patients and also compared our model with the 70-gene, the 8-subgroup, IFM15, and HMCLs7 models. Survival analysis showed that a low MMSI14 signature was associated with longer survival. Applying MMSI14 to independent data sets, we were able to classify 39% of patients as low-risk, with a survival probability of more than 90% at 60 months. Multiple clinical parameters confirmed significant correlation between low- and high-risk subgroups defined by MMSI14. Comparing previously published models to the same data sets the MMSI14 model retained the best prognostic value. We have developed a new gene model (MMSI14) for defining low-risk, newly diagnosed MM. The multivariate comparative analysis confirmed that MMSI14 is the best available model to predict clinical outcome in MM patients.

Key words multiple myeloma; risk subgroups; 14-gene model

Correspondence Tiehua Chen, Molecular Medicine Program, University of Utah, Salt Lake City, UT 84132, USA. Tel: 801 585 1694; Fax: 801 585 0701; e-mail: tchen@u2m2.utah.edu

Accepted for publication 9 April 2012

doi:10.1111/j.1600-0609.2012.01792.x

Since the introduction of high-dose therapy combined with autologous stem cell transplantation, a significant improvement was observed in the multiple myeloma (MM) patients' overall survival (OS) (1, 2). With the introduction of novel agents (3–5), despite remarkable improvements in the outcome of MM treatment over the past two decades, the clinical management of this disease is still challenging (6–8). MM remains an incurable disease, showing a wide variety of OS from a few months to more than 10 yr. Current treatments are associated with significant morbidity and mortality with different patients' risk response duration to the same treatment protocol. In 2006, Zhan *et al.* (9) identified eight genetic subtypes of MM, and subsequently, Shaughnessy *et al.* (10) developed a 70-gene model able to classify 13% of the patients as high-risk. Later, Decaux *et al.* (11) established an IFM15 model, which defined 25% of the diseases as high-risk. More recently, Moreaux *et al.* (12) published a high-risk signature model (HMCLs7) that was able to define 5% as high-risk. None of these studies focused on low-risk disease.

Defining low-risk diseases is as crucial as defining high-risk diseases for setting up optimal approaches for a specific risk group. Despite earlier attempts (13), there is currently no valid available gene expression profile (GEP) model focusing on low-risk in MM. MMSI14 is the first model focused on both low-risk and high-risk diseases. Based on the assumption that multiple genes are involved in the proliferation and survival of MM cells (8, 14), we have hypothesized that some expressed genes would be associated with the best prognosis, and so, our effort was directed to define a gene expression model that was able to focus on both low- and high-risk diseases.

Design and methods

Patients and gene expression profiling (GEP) data

We analyzed GEP data sets from a public archive of Gene Expression Omnibus (GEO) including 636 plasma cell samples: 22 cases of normal plasma cells [NPC, GSE5900 (15)], 44 of MGUS [monoclonal gammopathy of undetermined significance, GSE5900 (15)], 565 newly diagnosed MM patients [TT2: 351, and TT3: 214, old data: GSE2658 (9), updated data: GSE24080 (16)], and five of MM cell lines [MMCL, GSE24522, GSE20540 (17)]. The heatmap, boxplot, histogram, Kaplan–Meier analysis, and clinical data were used to evaluate our 14-gene model (Multiple Myeloma Survival Index14: MMSI14). We first used TT2 as a training set, we then applied MMSI14 to analyze a data set obtained after a longer follow-up time (TT2) and we also verified MMSI14 using a completely independent data set (TT3). Both data sets were also correlated to 13 clinical parameters. Finally, we compared the same groups of patients with MMSI14 and the 70-gene, IFM15, HMCLs7, and the 8-subgroup models. To identify the expressed genes associated with low-risk disease, we used a comprehensive strategy. Initially, we used siggenes (False Discovery Rate: FDR < 0.01) to compare the differentially expressed genes between MGUS and NPC. From a total of 54675 probes, we identified 5061 genes differentially expressed between the two groups. Using the same approach, we identified 5140 genes commonly expressed in MM and in NPC samples, and 5070 genes differentially expressed in living and deceased patients. Finally, gene expression profiling patients with poor (survival $\leq$12 months, n = 28) or good clinical outcome (survival $\geq$60 months, n = 17) were compared. With this method, we were able to identify 469 differentially expressed genes. Intersecting the four above-mentioned sets of differentially expressed data, we finally identified 14 common genes: CHRDL1, DENND1B, FAM20B, HIST1H1C, IFI16, MAD2L1, NEK2, NOL11, PMS2L5, PPP3CC, RFC4, SGK3, TRIM25, and TYROBP (Table 1).

Significance analysis of differentially expressed genes

R (2.13.1) (18)/bioconductor (2.8) (19), siggenes (1.28.0) (20), and survival packages were used to identify the different expressed genes comparing MGUS vs. NPC, MM vs. NPC, survival status, and poor (survival $\leq$12 months) vs. good clinical outcome (survival $\geq$60 months). R or Python were used to cross all groups and to define the final common genes. Cluster and TreeView were used to view the heatmap of the genes. The score of MMSI14 was shown in boxplot and histogram, which were evaluated by Kaplan–Meier analysis and clinical parameters.

We used the mean log2 values of nine over-expressed genes divided by the mean log2 values of five down-regulated genes to report a score as MMSI14. Comparisons of survival probability were performed by the Kaplan–Meier survival analysis and log-rank tests of significance in R/survival package. Chi-square and t-test were used to determine the significant difference in R/Excel. Univariate and multivariate analysis were performed in Coxphf (21) package for 13 clinical parameters and previously described gene models.

Results

Two different expression patterns of fourteen genes

The expression level of the aforementioned 14 genes was tested in four clinical groups: NPC, MGUS, MM, and MMCL (Fig. 1A). Seven of nine over-expressed genes followed the expected expression pattern: NPC <

Table 1 The name, chromosome position, ID and functions of 14 genes

Gene name	Chromosome position	Gene ID	Functions
PMS2L5	7q11.23	242201_at	Postmeiotic segregation increased 2-like 5
FAM20B	1q25	202915_s_at	Family with sequence similarity20 member B
HIST1H1C	6p22.2	209398_at	Histone cluster 1, Hic
MAD2L1	4q27	203362_s_at	MAD2 mitotic arrest deficient-like 1
DENND1B	1 q31.3	1564164_at	DENN/MADD domain containing 1B
NEK2	1q32.3	204641_at	NIMA (neverin mitosis gene a)-related kinase 2
RFC4	3q27.3	204023_at	Replication factorC (activator 1) 4
IFI16	1q23.1	206332_s_at	Interferon, gamma-inducible protein 16
NOL11	17q24.2	221970_s_at	Nucleolar protein 11
TYROBP	19q13.1	204122_at	TYRO protein tyrosine kinase binding protein
SGK3	8q12	220038_at	Serum/glucocorticoid regulated kinase family.member3
TRIM25	17q23.2	206911_at	Tripartite motif-containing 25
PPP3CC	8p21.3	32540_at	Trotein phosphatase 3, catalytic subumt, gamma isozyme
CHRDL1	Xq23	209763_at	Chordin-like 1

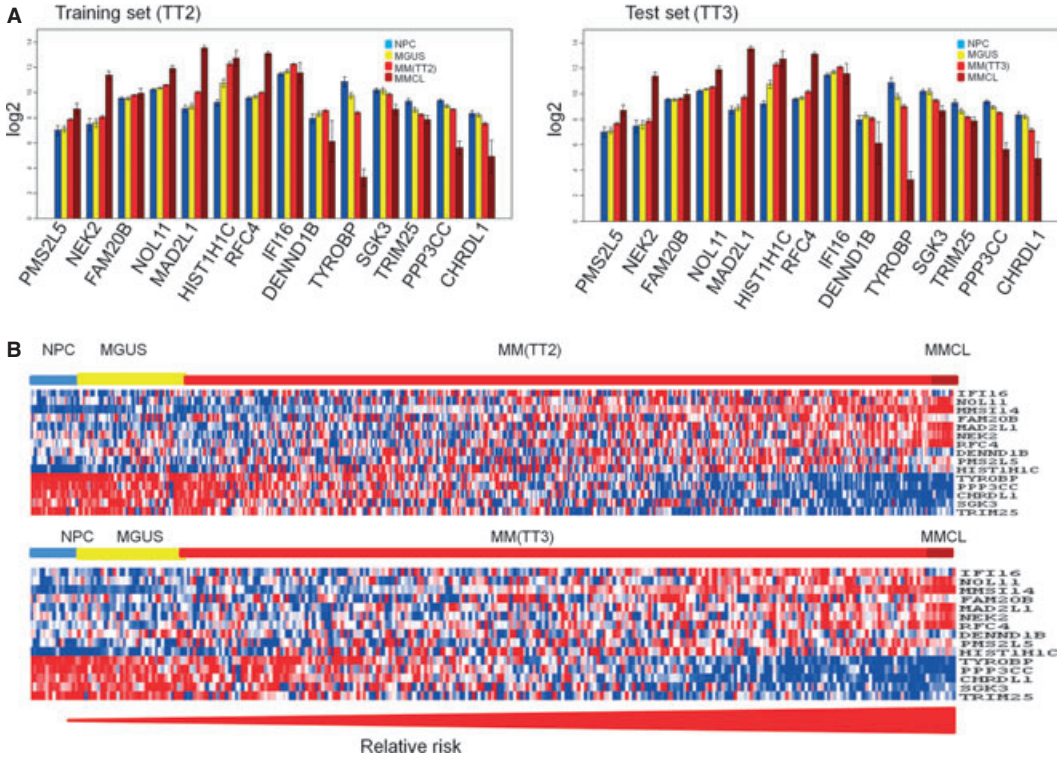

Figure 1 The expression level of 14 genes. (A) The histograms of expression level of 14 genes in normal plasma cells (NPC) ($n = 22$), MGUS ($n = 44$), multiple myeloma (MM) (TT2, training set, $n = 351$; TT3, test set, $n = 214$), and MMCL ($n = 5$). The 1st column is NPC, the 2nd is MGUS, the 3rd is MM (TT2/TT3), and the 4th one is MMCL. Plots represent mean ± confidence intervals. (B) The heat map of 14 genes in TT2/TT3. The gene order is unsupervised. The order of samples is supervised. All of those 14 genes are clustered correctly into two groups: OE (over-expression) and DR (down-regulated). Blue colors stand for DR and red color for OE. The relative risk is shown at bottom of the figure.

MGUS < MM (TT2; TT3) < MMCL. All five down-regulated genes showed a specular expression pattern among the different groups: NPC > MGUS > MM (TT2; TT3) > MMCL. To confirm that there are two different expression patterns among those 14 genes, we performed the half-supervised cluster analysis (supervised for samples, unsupervised for genes). With this methodology, we observed nine over-expressed genes distribution in one group and five down-regulated genes in the opposite group. Figure 1B shows the heatmaps of 14 genes for NPC, MGUS, TT2/TT3, and MMCL. The observed MGUS pattern appears similar to NPC, while MM GEP signature resembles MMCL. This distribution suggested that the identified 14 genes might be associated with the relative risk in MM.

MMSI14 and disease progression

We used the MMSI14 as a single score (the mean ratios of log2 of 9 over-expressed genes divided by that of five down-regulated genes) to compare the quartiles distributions among NPC, MGUS, MM (TT2, TT3), and MMCL (Fig. 2A). MGUS was found significantly different from both NPC ($P < 0.01$) and MM ($P < 0.01$) in both data sets (TT2, TT3). MMSI14 score was markedly increased in MMCL compared with all the other groups ($P < 0.01$, MMCL vs. MM). A remarkable trend of increment was observed comparing NPC, MGUS, MM (both TT2 and TT3) and MMCL, supporting the concept that MMSI14 score parallel the natural history of the disease.

MMSI14 model in the updated data set of TT2

The MMSI14 model was tested in a data set derived from a longer follow-up of TT2 patients. Using the cutoff values obtained from the original training data, the best cutoff value for low-risk patients was 1.131, while for high-risk was 1.288 ($P < 0.05$) (Fig. 3A). The Kaplan–Meier analysis showed a clear separation among low, intermediate, and high-risk subgroups. At 96 months, the survival probability for low-risk subgroup was 76%, for the intermediate-risk subgroup 46% and for the high-risk group 20%. We further calculated the rates of false-positive and false-negative errors (results). The false-positive rate in the high-risk subgroup was 13.9% (6/43), while the false-negative rate in the low-risk subgroup was 3.7% (5/135). These results confirmed the ability of the MMSI14 model to distinguish low-risk subgroups (39%) and high-risk subgroups (12%) from intermediate-risk subgroups (49%).

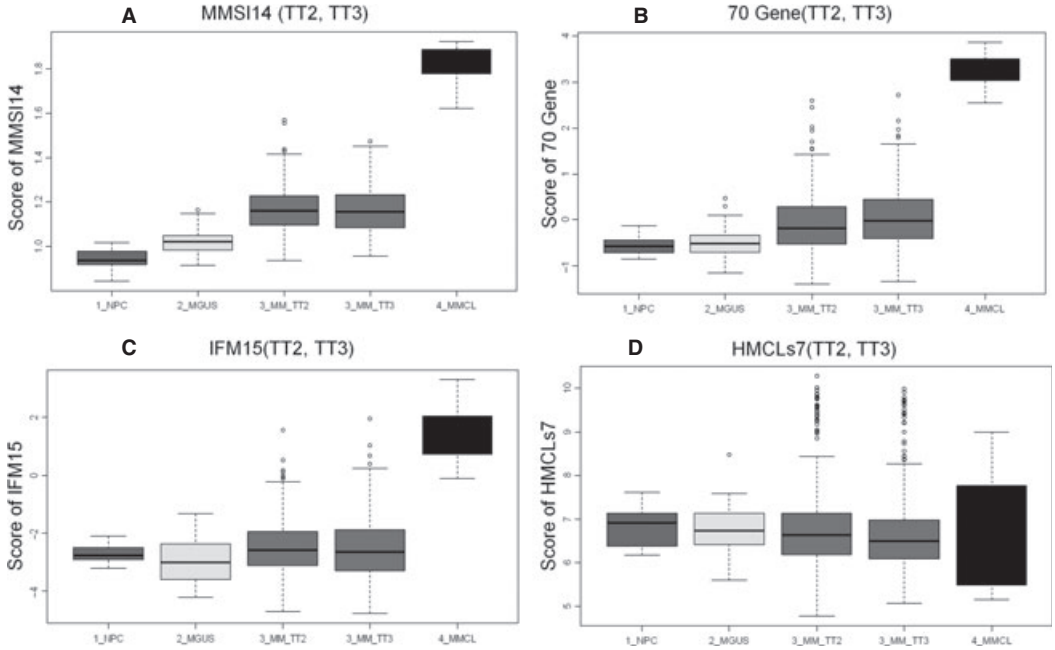

Figure 2 Score comparison of four gene models in different clinical subgroups. The box plots of MMSI14 (A), 70-gene score (B), IFM15 (C), and HMCls7 (D) in normal plasma cells (NPC) (n = 22), MGUS (n = 44), multiple myeloma (MM) (TT2, n = 351; TT3, n = 214), and MMCL (n = 5). Bottom, middle, and top lines of each box correspond to the 25th percentile, the 50th percentile (median), and the 75th percentile, respectively. The caps show 95%/5% confidential intervals. *Y*-axis represents the score of different gene models.

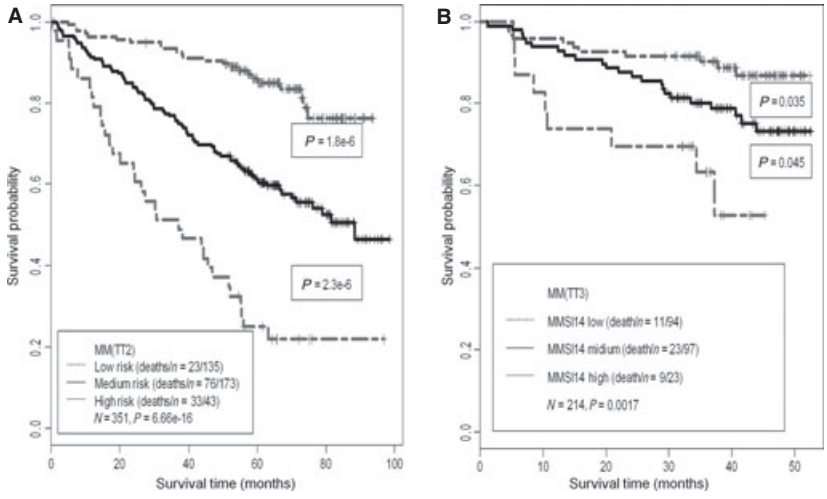

Figure 3 Kaplan–Meier analysis for overall survival (OS) by MMSI14 in two data set. (A) Overall survival in TT2 (updated data, n = 351). (B) Overall survival in TT3 (independent data set, n = 214).

MMSI14 model verification in TT3 data set

MMSI14 gene expression model was applied to a completely independent data set of newly diagnosed patients (TT3). MM patients enrolled in such trial were treated with similar drugs as TT2 except for the introduction of the proteasome inhibitor bortezomib. We performed the Kaplan–Meier analysis using the same cutoff values (1.131 and 1.288) derived for MMSI14 in TT2. The results are shown in Fig. 3C. MMSI14 was able to subdivide the entire population of Myeloma patients into three different subgroups. The low-risk subgroup included 94 individuals (44%) who experienced a survival probability of 88% at 50 months. The intermediate-risk subgroup comprised 97 subjects (45%) who achieved a survival rate of 73% at 50 months. Twenty-three patients were included in the high-risk subgroup (11%) with a survival probability of 52% at 50 months.

We also tested the MMSI14 model in predicting clinical parameters distribution among subgroups. Using 13 clinical parameters previously described, we observed different distribution among risk subgroups in TT2 patients. Cytogenetic abnormalities, expression of C-reactive protein, beta-2

microglobulin (β2M), lactate dehydrogenase (LDH), and number of lesions on magnetic resonance imaging (MRI) were significantly more expressed in the high-risk compared with the low-risk subgroup. The remaining parameters were equally distributed among three risk subgroups (Table 2). Analysis of the TT3 data set (Table 3) showed that six of the 13 clinical parameters had different distribution among risk subgroups. Cytogenetic abnormalities, β2M, and LDH were again significantly more distributed in high-risk subgroup compared with the low-risk.

To further verify the power of MMSI14, we performed a multivariate analysis on clinical parameters along with 70-gene, IFM15, HMCLs7, and MMSI14 models. Results are shown in Tables 4 and 5. Both hazard ratio (HR) and *P*-values showed that MMSI14 was the top predictor.

We applied the Kaplan–Meier analysis to cytogenetically subdivided groups both in TT2 and TT3 (Fig. 4A,C). The presence of any cytogenetic abnormalities could separate patients in two different risk subgroups in TT3 ($P < 0.01$) but not in TT2 ($P > 0.05$). MMSI14 was able to further subdivide the high-risk and low-risk cytogenetically identified risk groups into high-high, high-low, low-high, and low-low subgroups ($P < 0.05$) (Fig. 4B,D).

Table 2 Categorization of risk groups in TT2 ($n = 351$) data set

Clinic data	Low risk	Medium risk	High risk	P
Age, ≥65 yr	22%	23%	9%	NS
SEX (Male/Total)	54%	57%	63%	NS
RACE (White/T)	84%	92%	93%	NS
Albumin: <3.5 g/dL	1 0%	15%	23%	NS
ß2 microglobulin: >3.0 mg/L	33%	57%	67%	0.002**
C-reactive protein: ≥80 mg/L	37%	31%	47%	NS
Creatinine: ≥2.0 mg/dL	7%	10%	26%	0.004**
Cytogenetic abnormalities:	26%	39%	74%	<0.001TM
Hemoglobin: <10 g/dL	18%	31%	30%	0.07
Lactate dehydrogenase: ≥210 IU/L	13%	20%	56%	<0.001***
MRI ≥ 1	68%	65%	84%	0.04*
Plasma cells (aspirate): >33%	89%	95%	93%	NS
Plasma cells (bone marrow): >33%	86%	93%	95%	NS

NS, no significance. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Table 3 Categorization of risk groups in TT3 ($n = 214$) data set

Clinic data	Low risk	Medium risk	High risk	P
Age ≥65 yr	36%	26%	23%	NS
SEX (Male/Total)	72%	64%	51%	NS
RACE (White/T)	89%	91%	86%	NS
Albumin: <3.5 g/dL	7%	20%	20%	0.049*
β2 microglobulin: ≥3.0 mg/L	43%	73%	61%	0.03*
C-reactive protein: ≥8.0 mg/L	20%	39%	46%	0.02*
Creatinine: ≥2.0 mg/dL	3%	11%	14%	NS
Cytogenetic abnormalities:	23%	38%	60%	0.005**
Hemoglobin: <10 g/dL	20%	35%	51%	0.013*
Lactate dehydrogenase: ≥210 IU/L	11%	21%	34%	0.02*
MRI: ≥1	65%	66%	74%	NS
Plasma cells (aspirate): >33%	37%	52%	51%	NS
Plasma cells (bone marrow): >33%	48%	60%	77%	NS

NS, no significance. *$P < 0.05$; **$P < 0.01$.

Table 4 Multivariate analysis of OS in training data set (TT2, $n = 351$)

Variable	HR	P
Age, ≥55 yr	1.1	NS
SEX (Male/Total)	1.2	NS
RACE (White/T)	1.0	NS
Albumin: <3.5 g/dL	0.9	NS
β2 microglobulin: ≥3.0 mg/L	1.4	NS
C-reactive protein: >8.0 mg/L	1.1	NS
Creatinine: >2.0 mg/dL	1.2	NS
Cytogenetic abnormalities:	1.6	0.02*
Hemoglobin: <10 g/dL	1.1	NS
Lactate dehydrogenase: ≥210 IU/L	1.6	0.05
MRI: ≥1	2.0	0.003**
Plasma cells (aspirate): >33%	0.6	NS
Plasma cells (bone marrow): >33%	1.6	NS
70 -gene model	1.9	0.009**
IFM15	1.6	0.038*
HMCLs7	0.9	NS
MMSI14	2.8	0.00001***

NS, no significance. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Table 5 Multivariate analysis of OS in test data set (TT3, $n = 214$)

Variable	HK	P
Age, ≥65 yr	1.7	NS
Sex (Male/Total)	0.6	NS
RACE (White/T)	0.7	NS
Albumin: <3.5 g/dL	1.6	NS
β2 microglobulin: ≥3.0 mg/L	1.9	NS
C-reactive protein: >8.0 mg/L	1.5	NS
Creatinine: >2 0 mg/dL	2.0	NS
Cytogenetic abnormalities:	2.0	0.058
Hemoglobin: <10 g/dL	0.9	NS
Lactate dehydrogenase: ≥210 IU/L	1.0	NS
MRI: ≥1	1.2	NS
Plasma cells (aspirate): >33%	0.9	NS
Plasma cells (bone marrow): >33%	1.4	NS
70-gene model	1.5	NS
IFM15	1.2	NS
HMCLs7	1.4	NS
MMSI14	2.8	0.016*

NS, no significance. *$P < 0.05$.

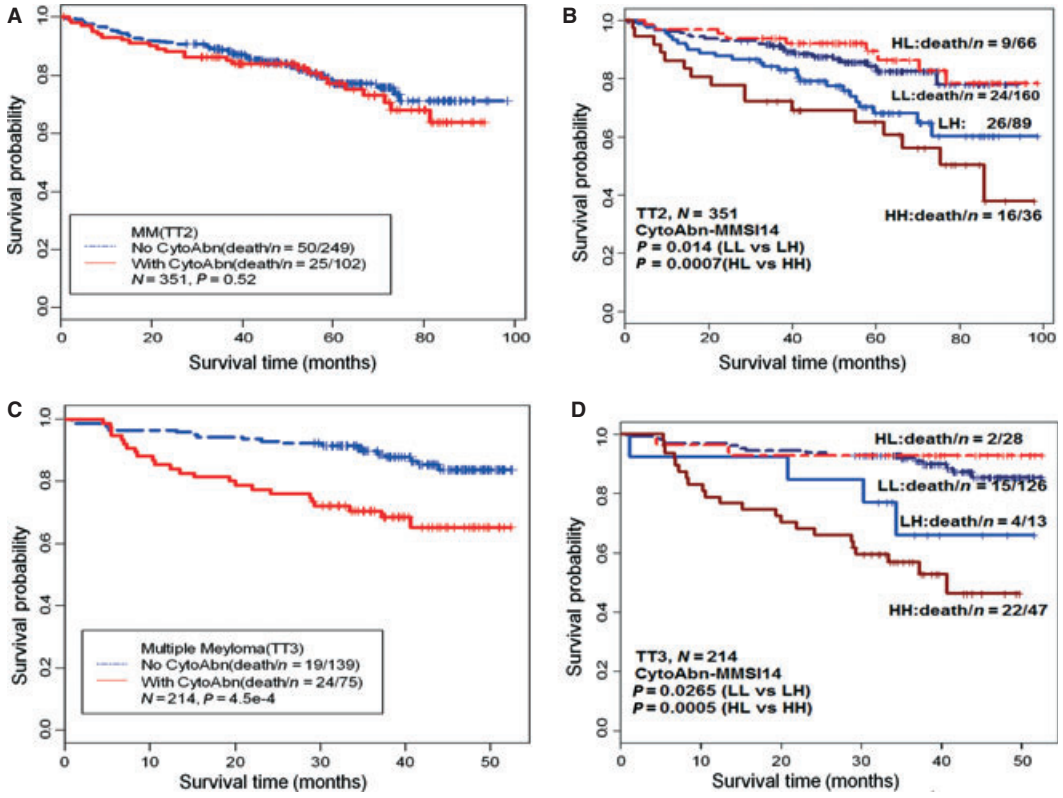

Figure 4 Effect of any cytogenetic abnormalities and MMSI14 score on survival. (A) TT2 training set; (C) TT3 cohort; (B) effect of MMSI14 (low- and high-risk groups) on TT2 training set by cytogenetic abnormalities; (D) effect of MMSI14 (low- and high-risk groups) on TT3 cohort by cytogenetic abnormalities.

Comparison of MMSI14 with 70-gene models

Only three shared genes were found in the two models (RFC4, IFI16, and NOL11). The mean ratios of log2 up/down-regulated genes of the 70-gene and MMSI14 shared similar pattern among NPC, MM, and MMCL groups (Fig. 2A,B). Both models showed correlation with the development of MM. The 70-gene model was not able to separate MGUS from NPC, while MMSI14 was able to define four different signatures.

A survival analysis of the two models applied to TT2 and TT3 data sets indicated that while the 70-gene was able to separate patients into two groups (13% high-risk and 87% low-risk groups), MMSI14 was able to subdivide the population into three different prognostic subgroups (12% high-risk, 49% intermediate-risk and 39% low-risk groups) (Fig. 3).

Comparison of MMSI14 with IFM15 and HMCLs7 models

No shared genes were observed between MMSI14 and IFM15 models. The mean ratios of log2 up/down-regulated genes of IFM15 showed no significant difference comparing NPC, MGUS, and MM (TT2, TT3); only MMCL groups showed a progressive mean increment (Fig. 2C).

A survival analysis of the two models in both data sets showed a similar ability of IFM15 and 70-gene models in dividing the patients of both data sets (TT2 and TT3) into two groups (25% high-risk and 75% low-risk groups). None of these two models could subdivide the group into three significantly different subgroups. The groups defined by IFM15 model could be further divided by MMSI14.

No common genes were found between HMCLs7 and MMSI14 models. The mean log2 values of HMCLs7 were similar among NPC, MGUS, and MM (TT2, TT3). In the MMCL group, HMCLs7 showed that the mean log2 values were similar to normal plasma cell group (Fig. 2D). For survival analysis, this model could not separate the TT3 patients into three subgroups.

Comparison of MMSI14 with 8-Subgroups

Applying the 8-subgroup model [CD-1(cyclin D1), CD-2 (cyclin D3), Hyperdiploidy (HY), Low bone disease (LB), MF (activation of c-MAF/MAFB proto-oncogenes), MS (spiked expression of MMSET), myeloid (MY), proliferation (PR)] to our data sets, we observed the definition of two

different risk categories (low-risk: CD1 + CD2 + HY + LB + MY and high-risk: PR + MS + MF). This model was not able to define a third risk subgroup no matter how we regroup them (such as low-risk: MY or MY + LB, intermediate-risk: CD1 + CD2 + HY + LB or CD1 + CD2 + HY + MF, and high-risk: PR + MS + MF or PR + MS, $P > 0.05$) (Fig. 5A,B). Regardless of regrouping, no statistically significant difference was observed between low-risk subgroups and intermediate-risk subgroups ($P > 0.05$). Comparing the 8-subgroup with MMSI14, except for MY and PR, we found considerable differences between the two models (Fig. 5C). MMSI14 defined 20–30% of low-risk patients in MS, MF, HY, CD1, and LB subgroups, 50–60% of low-risk in CD2 and MY subgroups, and no low-risk were identified within PR group. This analysis demonstrated the ability of the MMSI14 model to define three different patterns among those 8-subgroups.

We further performed a Kaplan–Meier analysis by regrouping TT2 patients into three subgroups (low-risk: MY + CD2; intermediate-risk: LB + CD1 + HY + MF + MS; and high-risk: PR). Within this subdivision, MMSI14 was able to separate low-risk ($P < 0.05$) and high-risk subgroups ($P < 0.01$) from the intermediate-risk population (Fig. 5D).

Discussion

To adapt optimal therapy to patients with MM, it is crucial to be able to identify different risk diseases (1). As early

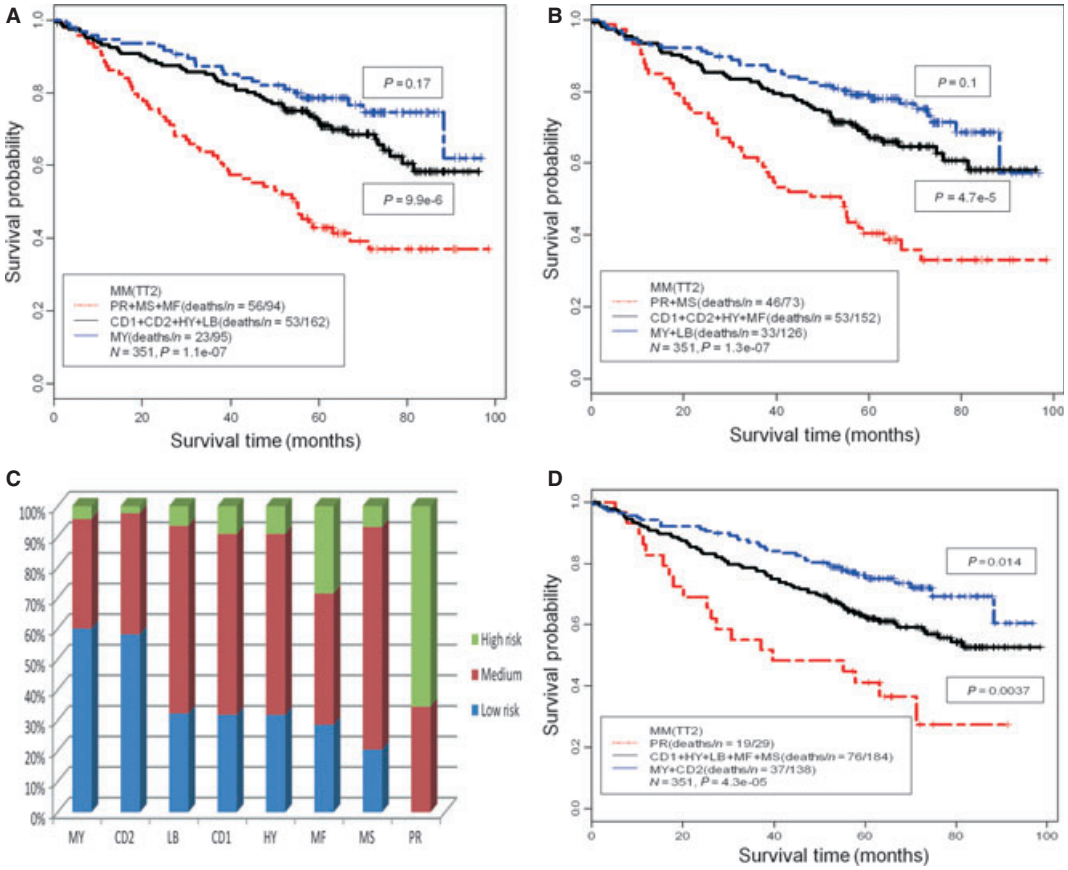

Figure 5 Regrouping by MMSI14 model on the eight subgroups. Based on the 8-subgroup model, there are two different ways to regroup. (A) The first: proliferation (PR) + MS + MF (high-risk), myeloid (MY) (low-risk) and CD1 + CD2 + HY + LB (intermediate-risk); (B) the second: PR + MS (high-risk), MY + LB (low-risk) and CD1 + CD2 + HY + MF (intermediate-risk). Regardless of regrouping, the high-risk subgroup is much different from others ($P < 0.01$), but there is no significant difference between low-risk and intermediate-risk subgroups ($P > 0.05$). (C) The *X*-axis represents the eight subgroups, and the *Y*-axis corresponds to the percentages of low, high, and intermediate-risk patients defined by MMSI14 model in each group. The blue color stands for low-risk, red for intermediate-risk, and grass green for high-risk. The combination of MMSI14 and the 8-subgroup molecular classification models [CD-1(cyclin D1), CD-2 (cyclin D3), Hyperdiploidy (HY), low bone disease (LB), MF (activation of c-MAF/MAFB proto-oncogenes), MS (spiked expression of MMSET), myeloid (MY), proliferation (PR)] shows three different patterns for low-risk >50% (MY, CD2), low-risk = 0 (PR), and low-risk ranged from 20% to 30% (LB, CD1, HY, MF, and MS). (D) The Kaplan–Meier analysis for regrouping among eight subgroups based on. The right panel (D) shows that regrouping among eight subgroups based on MMSI14 is able to separate low-risk subgroups ($P < 0.05$) and high-risk subgroups ($P < 0.01$) from the intermediate-risk subgroups.

as in the 1980s, the β2M and albumin levels were used to predict clinical outcome (22). Since then, four gene-based models have been developed for prediction of clinical outcome (9–12); all these models were not focused on low-risk diseases. MMSI14 is the first model focused on both low- and high-risk diseases, and it is more powerful than clinical parameters and the 70-gene model. All bar plots, heat maps, and box plots of MMSI14 genes showed that MMSI14 was strongly associated with the natural history of the disease. In a multivariate analysis of TT2 data set, MMSI14 was the strongest predictor for survival among multiple clinical parameters and the 70-gene model. Our survival analysis showed that MMSI14 distinguished three statistically different risk subgroups, and it is able to further separate the risk groups defined by the 70-gene model with both lower false-negative and false-positive rate. These results indicated that MMSI14 has superior prognostic power compared with the 70-gene model.

Our model was furthermore tested in an independent test data set TT3. Analyzing clinical parameters along with all four gene models, we found that MMSI14 is the best predictor for survival even in this data set. The box plot (Fig. 2) showed that the 70-gene model was similar to MMSI14 model among all groups. IFM15 model was able to separate only MMCL from our data sets, and HMCLs7 model showed no differences among all of four groups.

When we compared survival analysis from the different models, except for HMCLs7 that could not define any risk groups, the 70-gene, IFM15, and cytogenetic could define only two groups with different survival confirming that MMSI14 is the best predictor for survival among the all analyzed predictors.

Within the 8-subgroup model except for the PR subgroup MMSI14 separated three different risk groups among the remaining seven cohorts. The analysis obtained with the MMSI14 model on both cohorts indicates that at 96 months, the survival probability for low, intermediate and high-risk subgroups was 76%, 46%, and 20%, respectively. These results are clinically important for the treating physicians and suggest the necessity to introduce new and more effective compounds for the high-risk patients.

In conclusion, we have developed a new powerful GEP (Gene Expression Profiling) model (MMSI14) that is able to simultaneously define low- and high-risk diseases in newly diagnosed MM. Our analysis shows that MMSI14 has more prognostic power than clinical parameters and previously described GEP based models.

Acknowledgements

We thank Fenghuang Zhan for his financial support and great help, and Kenneth M Boucher for his critical comments.

References

1. McElwain TJ, Powles RL. High-dose intravenous melphalan for plasma-cell leukaemia and myeloma. *Lancet* 1983;**2**: 822–4.
2. Barlogie B, Hall R, Zander A, Dicke K, Alexanian R. High-dose melphalan with autologous bone marrow transplantation for multiple myeloma. *Blood* 1986;**67**:1298–301.
3. Cavo M, Zamagni E, Tosi P, *et al.* Superiority of thalidomide and dexamethasone over vincristine-doxorubicindexamethasone (VAD) as primary therapy in preparation for autologous transplantation for multiple myeloma. *Blood* 2005;**106**:35–9.
4. Harousseau JL, Attal M, Avet-Loiseau H, *et al.* Bortezomib plus dexamethasone is superior to vincristine plus doxorubicin plus dexamethasone as induction treatment prior to autologous stem-cell transplantation in newly diagnosed multiple myeloma: results of the IFM 2005-01 phase III trial. *J Clin Oncol* 2010;**28**:4621–9.
5. Richardson PG, Weller E, Lonial S, *et al.* Lenalidomide, bortezomib, and dexamethasone combination therapy in patients with newly diagnosed multiple myeloma. *Blood* 2010;**116**:679–86.
6. Barlogie B, Shaughnessy J, Tricot G, Jacobson J, Zangari M, Anaissie E, Walker R, Crowley J. Treatment of multiple myeloma. *Blood* 2004;**103**:20–32.
7. Moreau P, Avet-Loiseau H, Harousseau JL, Attal M. Current trends in autologous stem-cell transplantation for myeloma in the era of novel therapies. *J Clin Oncol* 2011;**29**:1898–906.
8. Shaughnessy JD. Cancer: an unexpected addiction. *Nature* 2008;**454**:172–3.
9. Zhan F, Huang Y, Colla S, *et al.* The molecular classification of multiple myeloma. *Blood* 2006;**108**:2020–8.
10. Shaughnessy JD Jr, Zhan F, Burington BE, *et al.* A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood* 2007;**109**:2276–84.
11. Decaux O, Lode L, Magrangeas F, *et al.* Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the Intergroupe Francophone du Myelome. *J Clin Oncol* 2008;**26**:4798–805.
12. Moreaux J, Klein B, Bataille R, *et al.* A high-risk signature for patients with multiple myeloma established from the molecular classification of human myeloma cell lines. *Haematologica* 2010;**96**:574–82.
13. Nair B, van Rhee F, Shaughnessy JD Jr, Anaissie E, Szymonifka J, Hoering A, Alsayed Y, Waheed S, Crowley J, Barlogie B. Superior results of Total Therapy 3 (2003-33) in gene expression profiling-defined low-risk multiple myeloma confirmed in subsequent trial 2006-66 with VRD maintenance. *Blood* 2010;**115**:4168–73.
14. Shaffer AL, Emre NC, Lamy L, *et al.* IRF4 addiction in multiple myeloma. *Nature* 2008;**454**:226–31.
15. Zhan F, Barlogie B, Arzoumanian V, *et al.* Gene-expression signature of benign monoclonal gammopathy evident in multi-

ple myeloma is linked to good prognosis. *Blood* 2007;**109**:1692–700.
16. Shi L, Campbell G, Jones WD, *et al.* The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* 2010;**28**:827–38.
17. McMillin DW, Delmore J, Weisberg E, *et al.* Tumor cell-specific bioluminescence platform to identify stroma-induced changes to anticancer drug activity. *Nat Med* 2010;**16**: 483–9.
18. R: Development Core Team. *R: A Language and Enviroment for Statistical Computing. R Foundation for Statistical Computing*. Vienna, Austria: R: Development Core Team, 2004. http://www.r-project.org (last accessed on 28 December 2011).
19. Gentleman RC, Carey VJ, Bates DM, *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;**5**:R80. http://www.bioconductor.org/install (last accessed on 28 December 2011).
20. Schwender H. *Siggenes: Multiple Testing Using SAM and Efron's Empirical Bayes Approaches*. R Package, version 1.28.0. http://www.bioconductor.org/packages/2.3/bioc/html/siggenes.html.2011 (last accessed on 28 December 2011).
21. Ploner M, Heinze G. *Coxphf: Cox Regression with Firth's Penalized Likelihood. R Package version 1.08*. http://CRAN.R-project.org/package=coxphf 2011 (last accessed on 28 December 2011).
22. Bataille R, Durie BG, Grenier J, Sany J. Prognostic factors and staging in multiple myeloma: a reappraisal. *J Clin Oncol* 1986;**4**:80–7.

