Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

Imaging, Diagnosis, Prognosis

Clinical Cancer Research

Homozygous Deletion Mapping in Myeloma Samples Identifies Genes and an Expression Signature Relevant to Pathogenesis and Outcome

Nicholas J. Dickens[1], Brian A. Walker[1], Paola E. Leone[1], David C. Johnson[1], José L. Brito[1], Athanasia Zeisig[1], Matthew W. Jenner[1], Kevin D. Boyd[1], David Gonzalez[1], Walter M. Gregory[2], Fiona M. Ross[3], Faith E. Davies[1], and Gareth J. Morgan[1]

Abstract

Purpose: Myeloma is a clonal malignancy of plasma cells. Poor-prognosis risk is currently identified by clinical and cytogenetic features. However, these indicators do not capture all prognostic information. Gene expression analysis can be used to identify poor-prognosis patients and this can be improved by combination with information about DNA-level changes.

Experimental Design: Using single nucleotide polymorphism–based gene mapping in combination with global gene expression analysis, we have identified homozygous deletions in genes and networks that are relevant to myeloma pathogenesis and outcome.

Results: We identified 170 genes with homozygous deletions and corresponding loss of expression. Deletion within the "cell death" network was overrepresented and cases with these deletions had impaired overall survival. From further analysis of these events, we have generated an expression-based signature associated with shorter survival in 258 patients and confirmed this signature in data from two independent groups totaling 800 patients. We defined a gene expression signature of 97 cell death genes that reflects prognosis and confirmed this in two independent data sets.

Conclusions: We developed a simple 6-gene expression signature from the 97-gene signature that can be used to identify poor-prognosis myeloma in the clinical environment. This signature could form the basis of future trials aimed at improving the outcome of poor-prognosis myeloma. *Clin Cancer Res; 16(6); 1856–64. ©2010 AACR.*

Multiple myeloma is an incurable clonal plasma cell malignancy that accounts for around 10% of all hematologic cancers (1). Although there have been significant improvements in the treatment of myeloma, there remains a clinical need to identify patients with a poor prognosis for whom alternate treatment strategies can be explored. The International Staging System can achieve this for groups of patients and is based on clinical factors that are surrogates for disease biology. The integration of chromosomal analysis into the assessment of patients offers an additional strategy that has the potential to group patients into biologically relevant and prognostically important groups (2–5). A number of chromosomal abnormalities have been identified that can be used in this fashion, including a group of poor prognostic IgH translocations comprising t(4;14), t(14;20), and t(14;16), together with the other poor prognostic markers del(1p), gain(1q), and del(17p). Deletion of chromosome 13 has also been considered important prognostically when detected by cytogenetics. However, within each of the groups, there is variability in clinical outcome, such that on their own they lack specificity in capturing all poor-prognosis cases. More recently, global gene expression analysis has been used for risk stratification, and it is postulated that combining this information with cytogenetic changes associated with disease progression will improve prognostic grouping. Using high-resolution array technologies, increasingly small regions of copy number alteration could be identified. We and others have used these technologies to characterize regions relevant to the pathogenesis of myeloma (2, 4, 6, 7). Based on the initial results of these studies, we reasoned that identification of the full range of genes inactivated by the loss of genetic material during disease pathogenesis could be used as a means to identify expression signatures and specific genes with

Authors' Affiliation: [1]Section of Haemato-Oncology, The Institute of Cancer Research, London, United Kingdom; [2]Clinical Trials Research Unit, University of Leeds, Leeds, United Kingdom; and [3]Leukaemia Research Fund UK Myeloma Forum Cytogenetics Group, Wessex Regional Genetics Laboratory, Salisbury, United Kingdom

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

N.J. Dickens and B.A. Walker contributed equally to this project.

Corresponding Author: Brian A. Walker, Section of Haemato-Oncology, The Institute of Cancer Research, 15 Cotswold Road, London SM2 5NG, United Kingdom. Phone: 44-20-8722-4132; Fax: 44-20-8722-4432; E-mail: brian.walker@icr.ac.uk.

doi: 10.1158/1078-0432.CCR-09-2831

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

Translational Relevance

The research discussed in this article provides new insights into prognosis in myeloma. Poor-prognosis risk is currently identified by clinical and cytogenetic features. However, these indicators do not capture all prognostic information. Gene expression analysis can be used to identify patients with poor prognosis, and this can be improved by combination with information about DNA-level changes. The biological information about the homozygous deletions discussed here has been integrated with gene expression and clinical data. The 97-gene expression signature discussed in this article has also been further developed into a 6-gene signature that is clinically applicable. This six-gene signature is discussed and is aimed at producing a test that could be used directly with patients. In the future, this test and others like it have the potential to be used in the tailor treatment of individual myeloma patients.

prognostic significance. Homozygous deletions (HZD) are particularly relevant in this respect because, by definition, they contain genes that are inactivated on both alleles (8, 9).

In this work, we used high-density single nucleotide polymorphism (SNP) arrays (median of 2.5-kb resolution) to identify the location and frequency of HZD in samples collected from presenting patients with myeloma in a randomized clinical trial. In combination with global gene expression data, the number of target genes was reduced with the aim of identifying key pathologically relevant signatures and networks dysregulated in myeloma pathogenesis. Deletions affecting genes in the cell death network were able to identify a group with poor-prognosis myeloma [i.e., shorter progression-free survival (PFS) and overall survival (OS)], and we were able to use these data derived at the DNA level to define a more generally applicable cell death expression signature. The prognostic value of this expression signature was further validated in two additional large trial data sets that used different treatment regimens. To translate the signature into a readily applicable test, the signature was developed such that, by using the expression ratio of only three pairs of genes, it is possible to identify cases with poor-prognosis myeloma with high specificity.

Materials and Methods

Patient samples. Bone marrow aspirates were obtained from newly diagnosed patients with multiple myeloma entered into the MRC Myeloma IX study, after informed consent. The trial recruited 1,966 patients and consisted of two arms, the first for older and less fit patients and the second for younger fitter patients. All younger patients received autologous transplantation following induction with cyclophosphamide, thalidomide, and dexamethasone or cyclophosphamide, vincristine, Adriamycin, and dexamethasone. Older patients were treated with either attenuated cyclophosphamide, thalidomide, and dexamethasone, or melphalan and prednisolone. All patients were then randomized to thalidomide maintenance or no thalidomide maintenance. Patients in the analysis presented here were representative of the trial in general (Supplementary Table S1). The trial was approved by the MRC Leukaemia Data Monitoring and Ethics committee (MREC 02/8/95, ISRCTN68454111).

Expression analysis. DNA and RNA were prepared for hybridization to the GeneChip Mapping 500K Array set and the U133 Plus 2.0 expression GeneChip, respectively (Affymetrix), according to the instructions of the manufacturer and have previously been described (4, 6, 10). There were 114 samples for which good-quality DNA samples were available (of which 84 had matched nontumor DNA) and 258 samples for which good-quality RNA samples were available. All DNA samples had corresponding RNA samples. The samples with microarrays were selected randomly based only on the availability of the sample. Analysis of mapping array data was done as previously described using GCOS, GTYPE, dChip, and CNAG (6). The expression levels were generated with dChip using the default perfect match/mismatch calculations and median normalization. The microarray data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (11) and are accessible through Gene Expression Omnibus Series accession number GSE15695. All 258 expression samples were used to derive both the 97-gene and the 6-gene signatures.

Copy number analysis and homozygous deletions. SNP genotypes and copy numbers were obtained using the Affymetrix GCOS software (version 1.4.0) to obtain raw feature intensities, which were then processed using the Affymetrix software to derive SNP genotypes. The paired tumor-control copy number data were inferred using dChip (12). The dChip-inferred copy number was used to identify HZDs. These HZDs were further filtered by integration with expression array data to identify those HZDs that ablate the expression of the genes with deletions (see Supplementary Methods).

Expression signatures. Cancer-associated changes at the DNA level are known to be associated with both pathologically and prognostically important effects and thus can be used to define prognostically relevant gene expression signatures. To do this, we identified genes that were differentially expressed ($P < 0.05$) between samples with HZD at the DNA level in the cell death network and those lacking such deletions. This analysis gave a list of genes that was subsequently used to perform hierarchical clustering of all 258 samples for which gene expression information was available. This list of genes was then used to derive a six-gene test using a filtering process based on Cox regression, the independence and significance of expression ratios, and specificity (Supplementary Methods).

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

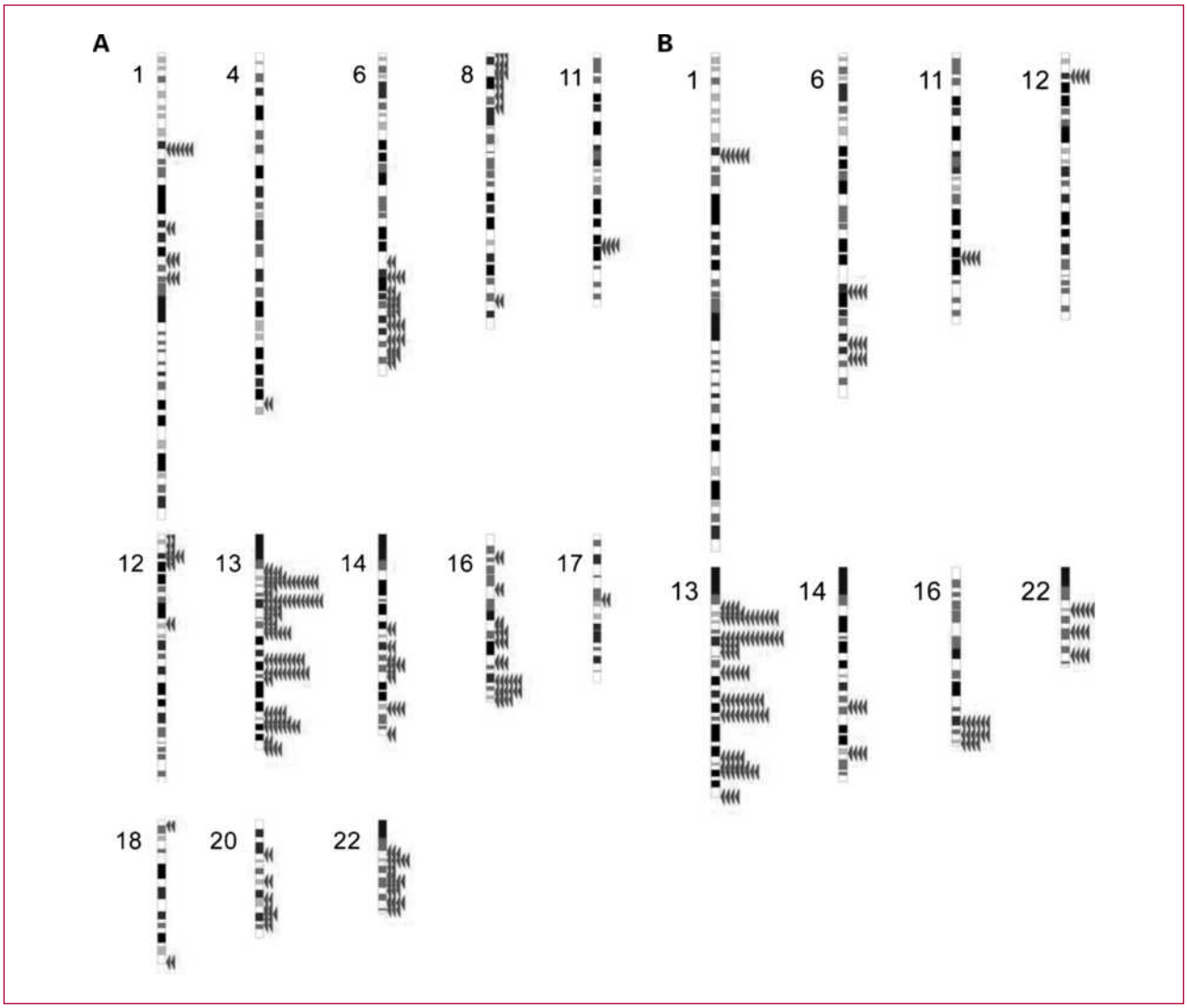

Fig. 1. Positions of homozygously deleted genes on the genome. A, genes with HZD in at least two samples with loss of expression in all HZD samples. B, genes with HZD in at least 5% of all samples.

Network analysis. Some networks of genes are well described, and databases such as KEGG (13) and BioCarta[4] provide well-curated sources of information. However, automatically identifying all biological relationships between genes is an intensive task and detailed manual annotation is not possible for large lists of genes. Ontologies such as the Gene Ontology (GO) project (14) and associated tools provide one method by which we can automatically identify a larger set of relationships between genes. GO terms are a hierarchical structure of semiautomatic annotations of genes. Broad annotation categories are subdivided into informational categories, and as the GO tree is descended, they become more descriptive. However, at the more detailed levels, there are fewer members of the annotation groups. The GO term "cell death" (GO term GO:008219) is a level 5 biological process; it includes both cytolysis and programmed cell death. The programmed cell death referred to by GO is an umbrella term that includes both apoptotic and nonapoptotic mechanisms. We will refer to this as the cell death network.

Survival analysis. Kaplan-Meier survival curves were generated using R, Bioconductor, and the survival package (15). The log-rank test, using R, was done to check for differences between curves. The effects of the prognostic factors β_2-microglobulin, age, t(4;14), t(11;14), hyperdiploidy, and del(13) were analyzed using univariate Cox regression. The independence of these factors and the prognostic signature were compared using multivariate Cox regression in R (Supplementary Methods).

[4] http://www.biocarta.com

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

Results

Homozygous deletions in patient samples. HZDs, by definition, inactivate the genes contained within them and have been shown to affect genes important in tumor progression and clinical outcome (16–18). Using the smoothing algorithm within dChip, single examples of HZD of at least 100 kb were identified in 114 cases presenting with myeloma. These included *FAM46C* (1p), *TSPYL4* (6q), *PARK2* (6q), *TLR4* (9q), *RB1* (13q), *WWOX* (16q), *CDH1* (16q), keratin locus (17q), *GSK3A*, and neighboring genes (19q), *UTX* (Xp), *CNKSR2* (Xp), and *HDHD1A* (Xp). Frequently occurring HZDs were located at 1p32.3 (*FAF1*/*CDKN2C*), 11q (*BIRC2* and *BIRC3*), 14q (*TRAF3* and *AMN*), and 16q (*CYLD*). The majority of these genes are relevant in myeloma biology: *CDKN2C* and *RB1* (cell cycle regulation); *TRAF3*, *BIRC2*, *BIRC3*, and *CYLD* (NF-κB regulation; refs. 6, 19, 20); *WWOX* (apoptosis; refs. 6, 21); *GSK3A* (Wnt signaling; refs. 22, 23); and *CDH1* (frequently methylated; refs. 24, 25).

However, this manual approach is not efficient, and a more sensitive bioinformatic approach was designed to identify these deletions in a standardized manner (for details, see Supplementary Methods). In this bioinformatic approach, a custom algorithm was designed to maximize the detection of deletions within and including genes. Using this algorithm on the dChip-processed mapping data from 84 newly diagnosed myeloma patients with paired peripheral blood controls (Supplementary Table S1) identified 783 genes that contained an acquired HZD in more than one tumor sample, and of these, 170 were associated with loss of expression of that gene and represent true HZD events (Fig. 1A). Only 29 genes had HZD in 5% or more samples and could be considered recurrent (Fig. 1B; Table 1). All of the deletions identified by the manual dChip method were included in the list generated by the custom algorithm, cross-validating the two approaches.

HZDs tend to occur in the genomic regions in which hemizygous deletions occur: 1p, 6q, 8p, 12p, 13q, 14q, 16q, 20p, and 22 (Fig. 1B); however, recurrent HZDs are also seen at 11q, which is not affected by loss of heterozygosity (16, 18). The median size of the HZDs was 37.5 kb, with the smallest deletion identified in more than one patient being 13 bases and the largest 14.9 Mb. We show that the frequency of HZD is variable, with 56 of 84 cases having at least one HZD (median of six HZDs per case), making it important to define which are driver and which are passenger lesions.

A subset of the HZDs were confirmed by DNA-based quantitative PCR to validate the method (Supplementary Methods). All samples tested, which were expected to have HZDs, were validated by quantitative PCR. Of the 29 homozygously deleted genes, 14 (48%) were located on chromosome 13, which has frequent hemizygous deletions. However, these genes are still altered in the myeloma tumor clone relative to their germ line by HZD and may be relevant to disease pathology. Inherited copy number variations could affect the copy number at any locus but the copy numbers inferred by dChip take this into account by comparison to matched control genomic DNA. In the list of genes with HZD, there are 20 that are present in regions of copy number variation[5] (Table 1); however, these are not consistently either gains or losses and the HZD are not present in the germ line DNA and thus are potentially relevant to myeloma pathology.

Survival analyses were carried out on the 29 genes with HZD in 5% or more cases. Of these genes, deletion of *CDKN2C*/*FAF1* at 1p32 was found to be associated with a shorter OS ($P = 0.029$; median OS, 20 versus 46 months). These genes were grouped together because the deletion detected spanned both genes in all of our samples. There were five genes deleted in at least 10% of cases: *DCLK1*, *ATP8A2*, *KLF12*, *PCDH9*, and *FGF14*. Of these, only samples with *ATP8A2* HZD had a significantly shorter OS ($P = 0.018$; median OS, 14 versus 46 months; Supplementary Fig. S1).

Biological interpretation of the spectrum of HZD genes. GO annotations were used to interpret the identified genes and define network-specific abnormalities present within the filtered list of 170 genes. There were 43 terms at GO level 5 assigned to the list, and after consideration of the annotations, we tested the prognostic significance of the following GO terms: regulation of progression through cell cycle probe set (GO:0000074), negative regulation of progression through cell cycle (GO:0045786), protein transport (GO:0015031), intracellular transport (GO:0046907), and the cell death network (GO: 008219), all of which were overrepresented in the annotations of genes with HZD.

Within the list of genes with HZD, we identified a significant enrichment of genes defined as the cell death network by GO. The genes significantly enriched included genes important in cell cycle regulation (*CDKN2C*, *EMP1*, and *PLAGL1*), apoptosis (*CTSB*, *BIRC2*, *BIRC3*, *TNFRSF10B*, *TNFRSF10D*, *FAF1*, *FGF14*, and *SGK*), and regulation of transcription (*ESR1*, *FOXO1*, and *TFDP1*). Although there were 15 genes in this cell death network, there were only 11 distinct genetic regions; this difference was due to juxtaposed pairs of genes being deleted in the same cases: *CDKN2C* and *FAF1* on 1p, *SGK* and *ESR1* on 6q, *TNFRSF10B* and *TNFRSF10D* on 8p, and *BIRC2* and *BIRC3* on 11q. Deletion affecting any of the genes within the GO-defined cell death network identified 25% of all cases of myeloma. Samples with a deletion of any gene annotated as cell death ($n = 24$) had a significantly shorter OS ($P = 0.015$; median OS, 32 versus 48 months) and shorter PFS ($P < 0.001$; Fig. 2).

We examined other prognostic events associated with HZD and showed that t(4;14), t(14;16), t(14;20), high β_2-microglobulin, del(16q), and del(17p) were not overrepresented. Examination of a prognostic expression–based proliferation index (26) within the cell death–deleted cases showed that there was an overrepresentation of samples with a high proliferation index ($P = 0.019$). However, removal of cases with a high proliferation index reveals that the residual cases still have a significantly shorter OS ($P = 0.029$), indicating that the cell death pathway is an independent prognostic factor.

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

To link alterations of the cell death network detected at the DNA level to associated changes at the level of gene expression, an analysis of differential gene expression between the samples with and those without HZD of cell death genes was carried out. To do this, mapping samples were divided into those with any cell death HZD and into those without; a *t* test was carried out (two-tailed, unequal variance) with a significance threshold of $P < 0.001$ (uncorrected). This analysis generated a list of 97 genes annotated as cell death by GO (Supplementary Table S2). Unsupervised hierarchical clustering (centroid based) was then done on the 258 samples with expression data using the 97-gene list. This analysis revealed two clusters with distinct expression patterns. The samples that clustered in the same branch as the majority of cell death HZD samples had a significantly shorter OS ($P < 0.001$; median, 32 versus 48 months) and shorter PFS ($P = 0.019$). The samples in this cell death expression class were checked for an overrepresentation of other known factors that would affect survival and they were representative of the set as a whole.

To validate this signature, we repeated the expression-based hierarchical clustering analysis using the 97-gene list in two additional data sets [GSE2658 (ref. 27) and GSE9782 (ref. 28)]. The raw data for this series was not available, but

Table 1. Genes with homozygous deletions in 5% or more of cases

Chromosome	Band	Ensembl ID	Gene	Description	No. of cases
1	p33, p32.3	ENSG00000185104	*FAF1*	FAS-associated factor 1 (protein FAF1) (hFAF1)	6
1	p33, p32.3	ENSG00000123080	*CDKN2C*	Cyclin-dependent kinase 6 inhibitor (p18-INK6) (cyclin-dependent kinase 4 inhibitor C) (p18-INK4c)	6
6	q22.1, q22.31, q22.2	ENSG00000196376	*SLC35F1*	Solute carrier family 35, member F1	4
6	q24.2	ENSG00000118491	*C6orf94*	UPF0418 protein C6orf94.	4
6	q25.1	ENSG00000131016	*AKAP12*	A-kinase anchor protein 12	4
11	q22.1, q22.3, q22.2	ENSG00000023445	*BIRC3*	Baculoviral IAP repeat-containing protein 3	4
12	p13.2	ENSG00000139083	*ETV6*	Transcription factor ETV6	4
13	q12.11	ENSG00000150457	*LATS2*	Large tumor suppressor homologue 2	4
13	q12.12	ENSG00000027001	*MIPEP*	Mitochondrial intermediate peptidase	5
13	q12.13	ENSG00000132932	*ATP8A2*	Probable phospholipid-transporting ATPase IB	12
13	q13.3	ENSG00000133083	*DCLK1*	Doublecortin-like and CAM kinase-like 1	13
13	q13.3, q14.11	ENSG00000183722	*LHFP*	Lipoma HMGIC fusion partner precursor	4
13	q14.11	ENSG00000133106	*EPSTI1*	Epithelial stromal interaction 1 isoform 2	4
13	q14.3	ENSG00000102837	*OLFM4*	Olfactomedin 4 precursor	6
13	q21.32	ENSG00000184226	*PCDH9*	Protocadherin-9 precursor	9
13	q22.1	ENSG00000118922	*KLF12*	Krueppel-like factor 12	10
13	q32.1	ENSG00000125257	*ABCC4*	Multidrug resistance-associated protein 4	5
13	q32.3	ENSG00000125246	*CLYBL*	Citrate lyase β subunit-like protein, mitochondrial precursor	5
13	q32.3	ENSG00000175198	*PCCA*	Propionyl-CoA carboxylase α chain	6
13	q33.1	ENSG00000102466	*FGF14*	Fibroblast growth factor 14	8
13	q34	ENSG00000185989	*RASA3*	Ras GTPase-activating protein 3	4
14	q24.2	ENSG00000198732	*SMOC1*	SPARC-related modular calcium-binding protein 1	4
14	q32.13, q32.12	ENSG00000100600	*LGMN*	Legumain	4
16	q23.1	ENSG00000186153	*WWOX*	WW domain-containing oxidoreductase	6
16	q24.1	ENSG00000103187	*COTL1*	Coactosin-like protein	6
16	q24.3	ENSG00000129993	*CBFA2T3*	Myeloid translocation gene-related protein 2 isoform MTG16b	4
22	q11.22	ENSG00000211666	*IGLV2-14*	Immunoglobulin λ light chain V gene segment	5
22	q12.3	ENSG00000133424	*LARGE*	Glycosyltransferase-like protein	4
22	q13.31	ENSG00000093000	*NUP50*	Nucleoporin 50 kDa	4

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

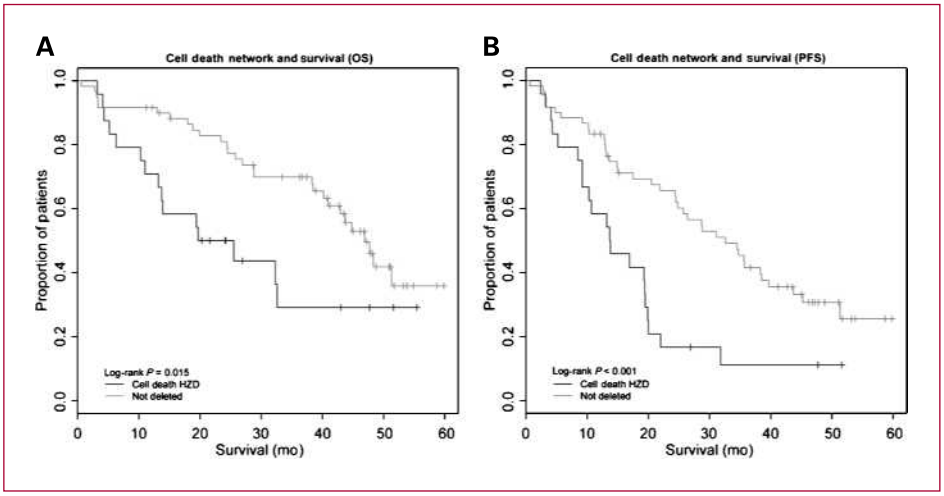

Fig. 2. Kaplan-Meier survival curves for samples with deletions of the cell death genes against those without these deletions in samples from Myeloma IX, Overall Survival (A), and Progression free survival (B).

using the same 97-gene list and performing hierarchical clustering in the same way again produced a distinct cluster with higher expression of the same genes as identified in our data. In each of these data series, the samples in the higher-expressing cell death cluster also had a significantly shorter OS than the other samples: in GSE2658, median OS was 46 months in the cell death cluster and >70 months (median not reached) outside of the cell death cluster ($P < 0.001$); in GSE9782, median OS was 10 months in the cell death cluster and 22 months outside of the cluster ($P < 0.001$). The higher-expressing genes are identified in Supplementary Table S2. The GSE2658 data set consisted of 559 patients from the total therapy programs TT2 and TT3, the first using thalidomide and the second thalidomide plus bortezomib. The GSE9782 data set consisted of relapsed patients treated with bortezomib from the CREST, SUMMIT, and APEX studies including the 040 companion study.

We went on to examine the associations with the 97-gene signature and showed it to be independent of other known prognostic factors, including β_2-microglobulin, serum albumin, and cytogenetic factors [del(13), t(4;14), t(11;14), del(17p), and del(16q); Table 2A]. This suggests that the prognostic implication of dysregulation of the cell death network is independent of these cytogenetic changes and that identifying patients with this signature may be clinically useful.

The 97-gene signature identifies 34% of Myeloma IX patients as having a poor prognosis, but because it is expression microarray–based, its regular use in the clinical setting is currently impractical. As a consequence, a more readily applicable six-gene cell death signature that identifies a similar set of patients with poor prognosis was derived. This was achieved using univariate and multivariate regression analyses, comparing ratios of expression of genes from within the 97-gene signature (Supplementary Methods). As a result of this analysis, we identified three pairs of genes whose relative expression provides a robust marker of prognosis, identifying 12% of patients as poor prognosis. This approach is based on the expression ratios of *BUB1B* (budding uninhibited by benzimidazoles 1 homologue β; involved in spindle checkpoint) versus *HDAC3* (histone deacetylase 3; involved in transcriptional repression), *CDC2* (cell division cycle 2; G_1-S to G_2-M transition) versus *FIS1* (fission 1 homologue; mitochondrial fission), and *RAD21* (rad21 homologue; double-strand DNA break repair) versus *ITM2B* (integral membrane protein 2B; Alzheimer's disease); if any one of these pairs had a ratio of ≥1, then the test was positive for poor prognosis. There is a clear divide between genes potentially involved in myeloma/cancer pathogenesis (*BUB1B*, *CDC2*, and *RAD21*) and those that are either potential tumor suppressor genes or uninvolved in cancer pathogenesis (*HDAC3*, *FIS1*, and *ITM2B*). Cases identified by these six genes have a median OS of 13 months with the signature and 45 months without; they also have a shorter PFS of 11 months in the six-gene positive cases and 22 months in the negative cases in the Myeloma IX data (Fig. 3A). The six-gene signature identifies 12% of cases in the Myeloma IX data set but an average of 20% across all of the analyzed data sets, and in each case, the patients with the six-gene signature had a significantly shorter prognosis (Fig. 3B and C). The specificity of this signature for identifying cases previously identified as being poor prognosis by the 97-gene signature is 100%. The 6-gene signature is not intended to replace the biologically relevant 97-gene signature; it is used to provide a more specific signature of poor prognosis that could be used as a prognostic test. Multivariate Cox regression was used to compare the six-gene signature as a prognostic factor versus other published signatures (Table 2B) and other conventional prognostic factors (Table 2C). This showed that the six-gene signature is independent of the other prognostic factors.

Discussion

We have carried out a comprehensive analysis of HZD in myeloma using Affymetrix SNP-based mapping data. In

[5] http://projects.tcag.ca/variation/downloads/variation.hg18.v8.aug.2009.txt

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

Table 2. Multivariate Cox analyses of the classification of the Myeloma IX data comparing the 97-gene signature with other prognostic factors and the 6-gene signature with existing signatures and other prognostic factors

Prognostic factor	Multivariate Cox regression *P* value
(A) 97-gene expression signature and conventional prognostic factors	
97-gene signature	<0.001
β_2-Microglobulin	<0.001
Serum albumin	0.350
Del(13q)	0.490
t(4;14)*	0.940
t(11;14)	0.063
Del(17p)	0.210
Del(16q)	0.250
(B) 6-gene signature and other gene signatures	
6-gene signature	<0.001
Arkansas 70-gene	<0.001
IFM 15-gene	0.790
Prognostic index	0.210
(C) 6-gene signature and conventional prognostic factors	
6-gene signature	<0.001
β_2-Microglobulin	<0.001
Serum albumin	0.610
Del(13q)	0.300
t(4;14)*	0.180
t(11;14)	0.021
Del(17p)	0.420
Del(16q)	0.260

*t(4;14) is a well-known poor prognostic factor; however, it is not independent of del(13q) or t(11;14) and so does not appear as an independent factor in the multivariate analyses.

myeloma, in contrast to lymphoid tumors, there is marked genomic instability with a particular focus on a number of defined regions where these events are concentrated and within which HZDs occur (29, 30). An exception to this is 11q, which is frequently amplified by chromosomal gain and where HZDs are also seen. Using a standard manual approach based on dChip, we have identified a number of genes affected by HZD. These data provide further support for our previous findings and those of others showing that the recurrent loss of regions containing negative regulators of the NF-κB pathway is an important recurrent event in myeloma. In addition, we have also gone on to identify novel deleted genes that also affect cell death pathways, consistent with the importance of the upregulation of survival pathways leading to the immortalization of the myeloma clone. We have also identified genes affecting chromatin methylation status such as *UTX*, which has recently been shown to be mutated in myeloma cell lines (31).

We have examined the prognostic significance of the frequently occurring HZDs but, because of limited numbers,

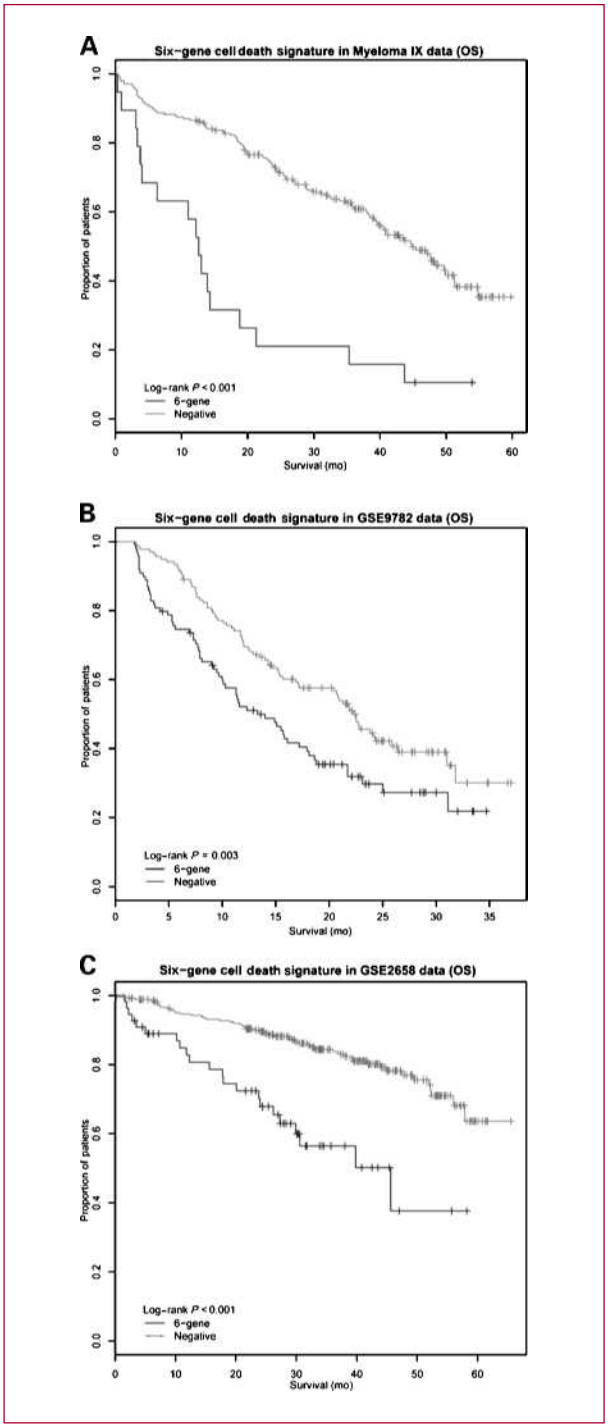

Fig. 3. Kaplan-Meier survival curves for samples with the six-gene signature against those without in Myeloma IX (A), GSE2658 (B), and GSE9782 (C).

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

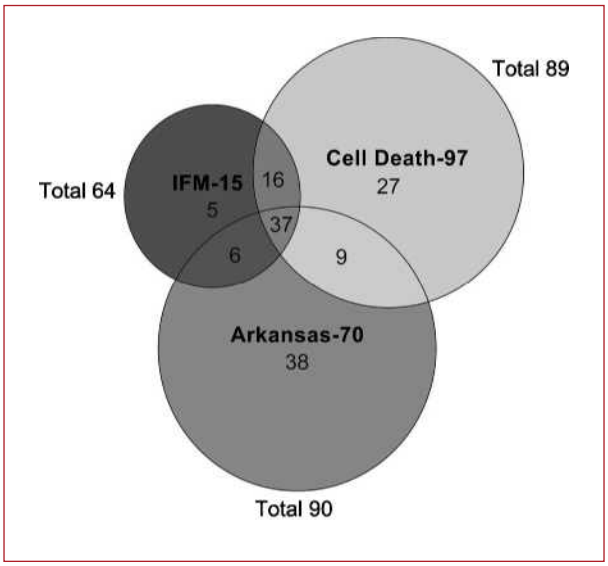

Fig. 4. Venn diagram showing the overlap of predictions between the Arkansas-70, IFM-15, and cell death-97 signatures.

failed to identify relevant effects. We reasoned that rather than individual lesions being important, the deregulation of pathways that are important defines outcome. Thus, to define the prognostic effect of pathway-associated changes, we wished to identify all potential HZDs affecting gene expression. We therefore designed a sensitive and standardized approach to global HZD detection using a custom algorithm, with which we identified 170 genes with HZD and loss of expression. These genes were enriched for genes involved in a cell death network. Pathway analysis showed that deregulation of the cell death pathway at the DNA level was significantly associated with impaired clinical outcome. Detecting such changes in new patients could be important to define prognosis, but detecting changes at the DNA level in clinical samples is difficult. In this context, it is readily possible to detect changes at the RNA level in routine diagnostic laboratories.

Consequently, we postulated that if we could develop an RNA-based expression signature focused on the cell death network, then we would be able to define a biologically relevant and prognostically important set of genes. Thus, the cases with a dysregulated cell death network at the DNA level were used to identify associated expression changes within cell death genes. This analysis generated a signature consisting of 97 genes associated with poor outcome (Table 1). The genes included in this signature were genes important in both the intrinsic and the extrinsic apoptotic pathways as well as in the tumor necrosis factor (TNF)/TNF-related apoptosis inducing ligand (TRAIL)/NF-κB signaling and phosphatidylinositol-3-kinase (PI3K) pathways. We went on to test the signature for its validity at presentation, relapse, and by treatment including thalidomide or bortezomib using publicly available data. The results of this analysis show that being in the poor-prognosis group, as identified using hierarchical clustering by the expression of the 97 genes, is independent of either prognostic lesion or treatment used. Currently, the International Staging System, the fluorescence *in situ* hybridization–based abnormalities del(1p) and del(17p), and the poor-prognosis immunoglobulin translocations could be used to define prognosis, but even these variables do not capture all of the clinical variability. Thus, more variables are required to define risk for an individual patient, and expression signatures are able to address at least some of these requirements. The most widely reported signature is the 70-gene signature reported by the University of Arkansas group, which they have incorporated into their treatment strategies. More recently, the Intergroupe Francophone du Myélome (IFM) group have reported a signature, and we have tried to understand how the three signatures (including ours) compare.

To understand how this expression signature anchored in changes within cell death genes at the DNA level compared with other prognostically important expression signatures, we applied two other published prognostic signatures [Arkansas 70-gene (ref. 32) and IFM 15-gene (ref. 33) signatures] to our data set using their published methods. The genes in each of the signatures are different with no overlap apart from *BIRC5*. Using each of the signatures to predict poor prognosis in our data set, we identified a core group of 37 cases (42% of poor-prognosis cases and 14% of all myelomas) identified as having a poor prognosis by all of the signatures. However, our signature identifies a distinct set of 27 poor-prognosis cases. In total, we identify 89 cases with poor prognosis, but the IFM 15-gene signature identifies 64 of our cases. Our signature is almost equally as sensitive to identifying poor-prognosis cases as the Arkansas 70-gene signature, which identifies 90 cases compared with 89 using our signature (Fig. 4).

The use of global gene expression arrays and data is mostly impractical within the clinical environment, and consequently, we developed a more limited signature that would give similar information. This test used information from three pairs of genes suitable for quantitative PCR analysis, removing the need to derive a set of control genes for the analysis because each member of the pair acts as a control for the other gene. Whereas some of the expression ratio pairs from the 97-gene list were more sensitive to identifying poor prognosis than those finally selected for the 6-gene signature, those selected were highly specific such that patients who were not of poor-prognosis risk would not be incorrectly classified as such. The six-gene signature is a significant independent variable when other prognostic factors, including International Staging System, del(17p), and poor-prognosis IgH translocations, are taken into account (Table 2). The six genes can all be described as biologically relevant, but because they were specifically selected for cell death annotations, they are not necessarily involved in the biology of poor prognosis. However, they are demonstrable markers of poor prognosis, and the six-gene signature that we developed here has high specificity for the identification of patients with poor-prognosis myeloma, at both presentation and relapse, who are suitable candidates for alternate treatment approaches.

Published OnlineFirst March 14, 2010; DOI: 10.1158/1078-0432.CCR-09-2831

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

Myeloma UK, Cancer Research UK, Leukaemia Research Fund UK, the Bud Flanagan Leukaemia Fund, the Kay Kendall Leukaemia Fund, and the Royal Marsden Hospital National Institute for Health Research Centre.

Data and samples were collected as part of the MRC Myeloma IX phase III trial—clinical trial registration (ISRCTN68454111). Fluorescence *in situ* hybridization and cytogenetics were carried out by Leukaemia Research Fund UK Myeloma Forum Cytogenetics Group, Wessex Regional Genetics Laboratory, Salisbury, United Kingdom. Overall trial statistics were carried out by the Clinical Trials Research Unit, University of Leeds, Leeds, United Kingdom.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 10/23/2009; revised 01/06/2010; accepted 01/21/2010; published OnlineFirst 03/09/2010.

References

1. Kyle RA, Rajkumar SV. Multiple myeloma. N Engl J Med 2004;351:1860–73.
2. Carrasco DR, Tonon G, Huang Y, et al. High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. Cancer Cell 2006;9:313–25.
3. Ross FM, Ibrahim AH, Vilain-Holmes A, et al. Age has a profound effect on the incidence and significance of chromosome abnormalities in myeloma. Leukemia 2005;19:1634–42.
4. Walker BA, Leone PE, Jenner MW, et al. Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms and genes important in the pathogenesis of multiple myeloma. Blood 2006;108:1733–43.
5. Zandecki M, Lai JL, Facon T. Multiple myeloma: almost all patients are cytogenetically abnormal. Br J Haematol 1996;94:217–27.
6. Jenner MW, Leone PE, Walker BA, et al. Gene mapping and expression analysis of 16q loss of heterozygosity identifies WWOX and CYLD as being important in determining clinical outcome in multiple myeloma. Blood 2007;110:3291–300.
7. Walker BA, Morgan GJ. Use of single nucleotide polymorphism-based mapping arrays to detect copy number changes and loss of heterozygosity in multiple myeloma. Clin Lymphoma Myeloma 2006;7:186–91.
8. Minna JD, Roth JA, Gazdar AF. Focus on lung cancer. Cancer Cell 2002;1:49–52.
9. Yokota J, Kohno T. Molecular footprints of human lung cancer progression. Cancer Sci 2004;95:197–204.
10. Leone PE, Walker BA, Jenner MW, et al. Deletions of CDKN2C in multiple myeloma: biological and clinical implications. Clin Cancer Res 2008;14:6033–41.
11. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207–10.
12. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 2004;20:1233–40.
13. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000;28:27–30.
14. Blake JA, Harris MA. The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. Current protocols in bioinformatics/editoral board, Andreas D Baxevanis et al. 2002;Chapter 7:Unit 7.2.
15. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
16. Gorringe KL, Jacobs S, Thompson ER, et al. High-resolution single nucleotide polymorphism array analysis of epithelial ovarian cancer reveals numerous microdeletions and amplifications. Clin Cancer Res 2007;13:4731–9.
17. Nagayama K, Kohno T, Sato M, Arai Y, Minna JD, Yokota J. Homozygous deletion scanning of the lung cancer genome at a 100-kb resolution. Genes Chromosomes Cancer 2007;46:1000–10.
18. Stark M, Hayward N. Genome-wide loss of heterozygosity and copy number analysis in melanoma using high-density single-nucleotide polymorphism arrays. Cancer Res 2007;67:2632–42.
19. Annunziata CM, Davis RE, Demchenko Y, et al. Frequent engagement of the classical and alternative NF-κB pathways by diverse genetic abnormalities in multiple myeloma. Cancer Cell 2007;12:115–30.
20. Keats JJ, Fonseca R, Chesi M, et al. Promiscuous mutations activate the noncanonical NF-κB pathway in multiple myeloma. Cancer Cell 2007;12:131–44.
21. Bednarek AK, Laflin KJ, Daniel RL, Liao Q, Hawkins KA, Aldaz CM. WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. Cancer Res 2000;60:2140–5.
22. Davies FE, Dring AM, Li C, et al. Insights into the multistep transformation of MGUS to myeloma using microarray expression analysis. Blood 2003;102:4504–11.
23. Tian E, Zhan F, Walker R, et al. The role of the Wnt-signaling antagonist DKK1 in the development of osteolytic lesions in multiple myeloma. N Engl J Med 2003;349:2483–94.
24. Galm O, Wilop S, Reichelt J, et al. DNA methylation changes in multiple myeloma. Leukemia 2004;18:1687–92.
25. Seidl S, Ackermann J, Kaufmann H, et al. DNA-methylation analysis identifies the E-cadherin gene as a potential marker of disease progression in patients with monoclonal gammopathies. Cancer 2004;100:2598–606.
26. Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 1999;96:9212–7.
27. Zhan F, Huang Y, Colla S, et al. The molecular classification of multiple myeloma. Blood 2006;108:2020–8.
28. Mulligan G, Mitsiades C, Bryant B, et al. Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. Blood 2007;109:3177–88.
29. Walter MJ, Payton JE, Ries RE, et al. Acquired copy number alterations in adult acute myeloid leukemia genomes. Proc Natl Acad Sci U S A 2009;106:12950–5.
30. Schiffman JD, Wang Y, McPherson LA, et al. Molecular inversion probes reveal patterns of 9p21 deletion and copy number aberrations in childhood leukemia. Cancer Genet Cytogenet 2009;193:9–18.
31. van Haaften G, Dalgliesh GL, Davies H, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet 2009;41:521–3.
32. 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.
33. 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.

AACR American Association for Cancer Research

Clinical Cancer Research

Homozygous Deletion Mapping in Myeloma Samples Identifies Genes and an Expression Signature Relevant to Pathogenesis and Outcome

Nicholas J. Dickens, Brian A. Walker, Paola E. Leone, et al.

Clin Cancer Res 2010;16:1856-1864. Published OnlineFirst March 14, 2010.