High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients

Daniel R. Carrasco,[1,2,8] Giovanni Tonon,[1,8] Yongsheng Huang,[3,8] Yunyu Zhang,[1] Raktim Sinha,[1] Bin Feng,[1] James P. Stewart,[3] Fenghuang Zhan,[3] Deepak Khatry,[1] Marina Protopopova,[5] Alexei Protopopov,[5] Kumar Sukhdeo,[1] Ichiro Hanamura,[3] Owen Stephens,[3] Bart Barlogie,[3] Kenneth C. Anderson,[1,4] Lynda Chin,[1,7] John D. Shaughnessy, Jr.,[3,9] Cameron Brennan,[6,9] and Ronald A. DePinho[1,5,9,*]

[1] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115
[2] Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115
[3] The Donna and Donald Lambert Laboratory of Myeloma Genetics, Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205
[4] The Jerome Lipper Multiple Myeloma Center, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115
[5] Center for Applied Cancer Science, Belfer Institute for Innovative Cancer Science, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts
[6] Neurosurgery Service, Memorial Sloan-Kettering Cancer Center, Department of Neurosurgery, Weill Cornell Medical College, New York, New York 10021
[7] Department of Dermatology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, 02115
[8] These authors contributed equally to this work.
[9] Cocorresponding authors: cbrennan@mskcc.org (C.B.); shaughnessyjohn@uams.edu (J.D.S.); ron_depinho@dfci.harvard.edu (R.A.D.)
*Correspondence: ron_depinho@dfci.harvard.edu

Summary

To identify genetic events underlying the genesis and progression of multiple myeloma (MM), we conducted a high-resolution analysis of recurrent copy number alterations (CNAs) and expression profiles in a collection of MM cell lines and outcome-annotated clinical specimens. Attesting to the molecular heterogeneity of MM, unsupervised classification using non-negative matrix factorization (NMF) designed for array comparative genomic hybridization (aCGH) analysis uncovered distinct genomic subtypes. Additionally, we defined 87 discrete minimal common regions (MCRs) within recurrent and highly focal CNAs. Further integration with expression data generated a refined list of MM gene candidates residing within these MCRs, thereby providing a genomic framework for dissection of disease pathogenesis, improved clinical management, and initiation of targeted drug discovery for specific MM patients.

Introduction

Multiple myeloma (MM) is characterized by clonal proliferation of plasma cells in the bone marrow, usually with elevated serum and urine monoclonal paraprotein and associated end organ sequelae. MM is the second most frequent hematological cancer in the US after non-Hodgkin's lymphoma. MM is typically preceded by an age-progressive condition termed monoclonal gammopathy of undetermined significance (MGUS), a condition present in 1% of adults over the age of 25 that progresses to MM at a rate of 0.5%–3% per year (Bergsagel et al., 2005; Kyle and Rajkumar, 2004; Mitsiades et al., 2004). MM remains largely incurable despite high-dose chemotherapy with stem cell support. Novel agents such as thalidomide, the immunoregulator Revlimid, and the proteasome inhibitor bortezomib can achieve responses in patients with relapsed and refractory MM; however, the median survival remains at 6 years, with only 10% of patients surviving at 10 years (Barlogie et al., 2004; Rajkumar and Kyle, 2005; Richardson et al., 2005).

Significant effort has been directed toward the identification of the molecular genetic events in this malignancy with the goals of improving early detection and providing new therapeutic targets. Unlike most hematological malignancies and more similar to solid tissue neoplasms, MM genomes are typified by numerous structural and numerical chromosomal aberrations (Kuehl and Bergsagel, 2002). Reflecting the increasing genomic

SIGNIFICANCE

MM is the second most common hematological malignancy and remains incurable. Effective targeted therapies require detailed knowledge of the spectrum of genetic lesions governing MM pathogenesis. This study provides a comprehensive and integrated view of recurrent amplifications and deletions and their associated gene expression alterations. This integration, coupled with this high-resolution aCGH platform, delimited a tractable number of candidate genes for enlistment into functional validation and ultimately drug discovery. Importantly, unsupervised NMF-based classification of the MM genome profiles stratified MM into specific subgroups in which the classical hyperdiploid MM was further subdivided into two subgroups with distinct clinical outcomes. Thus, MM is a molecularly heterogeneous disease in which distinct loci can be linked to clinical behavior and prognosis.

instability that characterizes disease progression, metaphase chromosomal abnormalities can be detected in only one-third of newly diagnosed patients but are evident in the majority of patients with end-stage disease (Fonseca et al., 2004). Yet, applying DNA content or interphase fluorescence in situ hybridization (FISH) analyses, aneuploidy and translocations are detectable in virtually all subjects with MM and even MGUS (Bergsagel and Kuehl, 2001; Chng et al., 2005). Extensive molecular (Kuehl and Bergsagel, 2002; Shaughnessy and Barlogie, 2003), cytogenetic (Bergsagel and Kuehl, 2001; Debes-Marun et al., 2003; Sawyer et al., 1998), and chromosomal CGH (Avet-Loiseau et al., 1997; Cigudosa et al., 1998) analyses have uncovered a number of recurrent genetic alterations in MM and its precursor MGUS, some of which have been linked to disease pathogenesis and clinical behavior.

Chromosomal translocations involving the IgH locus are seen in most MM cell lines, consistent with MM's origin from antigen-driven B cells in postgerminal centers (Kuehl and Bergsagel, 2002). Five recurrent loci/genes are commonly juxtaposed to the powerful IgH enhancer locus elements, including 11q13 (*CCND1*), 4p16 (*FGFR3/WHSC1*), 6p21 (*CCND3*), 16q23 (*MAF*), and 20q11 (*MAFB*), resulting in deregulated expression of these target genes in neoplastic plasma cells (Bergsagel and Kuehl, 2005). Such translocations, present in MGUS, appear central to MM genesis, whereas progression is associated with mutational activation of *NRAS* or *KRAS* oncogenes and inactivation of *CDKN2A*, *CDKN2C*, *CDKN1B*, and/or *PTEN* tumor suppressor genes (TSGs). Late mutational events involve inactivation of *TP53* and secondary translocations involving *MYC* (Kuehl and Bergsagel, 2002).

Two oncogenic pathways have been hypothesized for MGUS/MM pathogenesis. Hyperdiploid MM involves multiple trisomies of chromosomes 3, 5, 7, 9, 11, 15, 19, and 21, whereas nonhyperdiploid MM is associated with prevalent IgH translocations (Bergsagel et al., 2005; Cremer et al., 2005; Fonseca et al., 2004). Ploidy level also impacts prognosis: nonhyperdiploidy imparts short survival (Fonseca et al., 2004) that can be counteracted by the presence of trisomies involving chromosomes 6, 9, 11, and 17. Deletion of chromosome 13, especially band 13q14, is commonly observed in nonhyperdiploid MM and confers high risk (Fonseca et al., 2004). Expression profiling has been used to define molecular subgroups of MM with clinical correlates, and a novel TC (translocation/cyclin D) classification of MM has been proposed (Bergsagel et al., 2005). Gene expression profiles of 511 outcome-annotated MM cases have pointed to amplification at 1q21 as an independent predictor of outcome (F.Z. and J.D.S., unpublished data). While these antecedent efforts have led to important insights into the pathogenesis and clinical behavior of MM, these numerous recurrent genomic alterations point to many undefined genetic elements that may prove relevant to disease initiation, progression, and drug responsiveness. However, while recurrent chromosomal gains have been mapped to 1q, 3q, 9q, 11q, 12q, 15q, 17q, and 22q and recurrent losses to 6q, 13q, 16q, Xp, and Xq, the presumed cancer-relevant targets in these loci are not known.

This study sought to develop a more thorough molecular view of MM pathogenesis through an integrated analysis of the cancer-associated alterations on the DNA and RNA levels in clinically annotated material. This MM oncogenomic analysis identified genetic elements with strong biological and clinical correlates as evidenced by their common targeting in other tumor types, correlation with survival, and presence of known oncogenes and TSGs or their homologs not previously implicated in MM pathogenesis. The high-resolution view afforded here also provides an efficient entry point for the discovery of new MM-relevant genes.

Results

Unsupervised classification of MM genomic features identifies distinct patient subgroups

High-resolution aCGH methodology was used to catalog copy number alterations (CNAs) in the genomes of $CD138^+$-enriched plasma cells (>90% $CD138^+CD45^-$) derived from 67 newly diagnosed MM patients prior to treatment (Table S1 in the Supplemental Data available with this article online). This data set revealed a highly rearranged MM genome, harboring large numbers of distinct CNAs. Since genomic alterations hold the potential to identify genetic events playing direct roles in disease pathogenesis and progression (Aguirre et al., 2004; Pollack et al., 2002; Tonon et al., 2005), we sought to determine whether MM genomic profiles could specify meaningful genetic and clinical subgroups by unsupervised methodologies. To this end, we developed an algorithm based on nonnegative matrix factorization (NMF) (Brunet et al., 2004) (Experimental Procedures; C.B. and L.C., unpublished data), designed to extract distinctive genomic features from aCGH profiles, hereafter designated as gNMF. Briefly, gNMF was performed on aCGH data after transforming to nonnegative values, and gNMF consensus matrices were generated (Figure S1). Ranks K = 2, 3, and 4 generated matrices showing stable cluster assignments suggesting the existence of up to four distinct genomic patterns among the 67 MM samples.

The rank K = 2 classification divided these 67 samples into a "kA" subgroup (n = 38) characterized by odd-chromosome gains and a "kB" subgroup (n = 29) characterized by loss of chromosomes 1p, 8p, 13, and 16q and amplification of ch1q (Figure 1A). Thus, the K = 2 classification yields a grouping reminiscent of the well-recognized hyperdiploid (e.g., odd-chromosome gain pattern) and nonhyperdiploid subclasses. However, when correlated with clinical outcome data, the kA (hyperdiploid) subgroup showed only a trend toward improved survival over kB (nonhyperdiploid) subgroup (Figure 1B; see below), prompting assessment of further genomic subclasses within kA and kB. Indeed, with rank K = 4 matrix, these 67 MM samples were subdivided by gNMF into four distinct molecular subclasses, k1–k4 (Figure 2A). All 21 k1 and 16 k2 samples belonged to the kA subgroup, while all 13 k3 and 16 of 17 k4 samples were in the kB subgroup. In other words, unbiased genome-wide classification of genomic alterations provided molecular evidence that MM is a heterogeneous disease, and that the traditional hyperdiploid MM class consists of two molecular subclasses.

Definition of a subclass of hyperdiploid MM with poor prognosis

Although considered a relatively good prognosis group, hyperdiploid MM patients (kA subgroup) did not uniformly survive longer in our cohort (Figure 1B; log rank test p = 0.25 and 0.10 for event-free survival [EFS] and overall survival [OS], respectively). Since gNMF classification with K4 rank showed that kA was

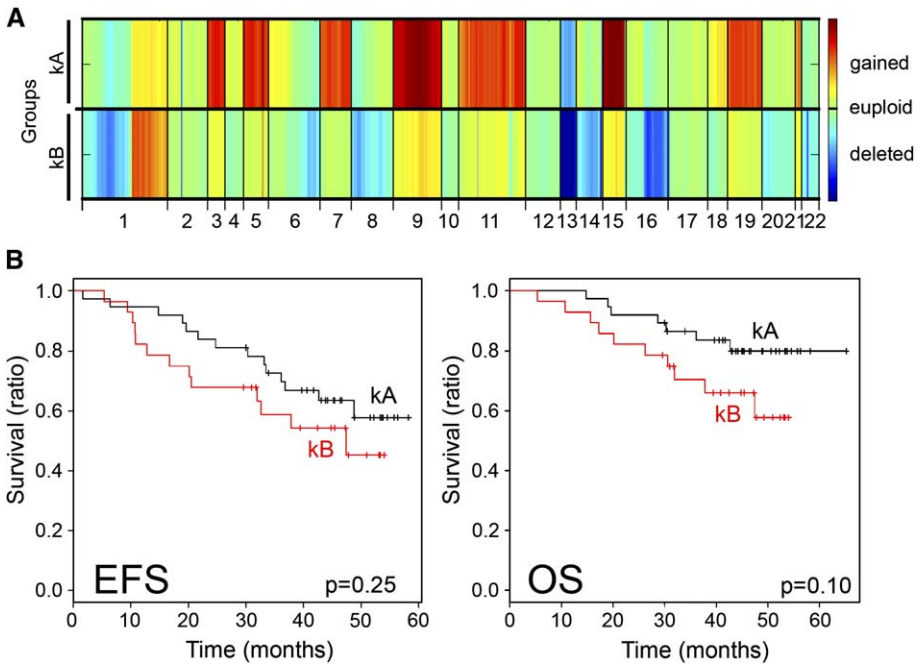

Figure 1. gNMF classification with rank K = 2 identifies two subgroups, kA and kB, reminiscent of the MM hyperdiploid/nonhyperdiploid subgroups

A: aCGH profiles of 67 clinically annotated primary tumors were subjected to NMF analyses (1000 repetitions). With rank K = 2, two distinct subgroups, kA and kB, were identified (y axis), and centroids of each group are shown. x axis represents genomic map order (from ch1 to ch22). The colors denote gained (red), euploid (yellow/green), or deleted (blue) chromosome material.

B: KM event-free survival (EFS; left) and overall survival (OS; right) curves for 64 MM patients demonstrating no significant difference in survival (p = 0.25 and 0.1, respectively) when divided into subgroups kA versus kB.

composed of k1 and k2 subclasses, we evaluated outcomes of patients in k1 and k2. Here, Kaplan-Meier (KM) curves for k1 (n = 21) and k2 (n = 16) patients were generated for EFS and OS with median follow-up of 43.4 months post-TT2 regimen (Figure 2B). Clearly, k1 possessed a significant EFS advantage with a trend for better OS over k2 (log rank test p = 0.012 for EFS and p = 0.12 for OS), indicating that the genomic heterogeneity in hyperdiploid MM holds biological relevance and embedded within it are molecular events dictating clinical behavior of the disease.

As the first step toward elucidating genetic determinants of outcome in hyperdiploid MM, we exploited the high degree of relatedness between k1 and k2 genomes to uncover clinically relevant CNAs (Figure 2A). Comparison of k1 versus k2 genomic

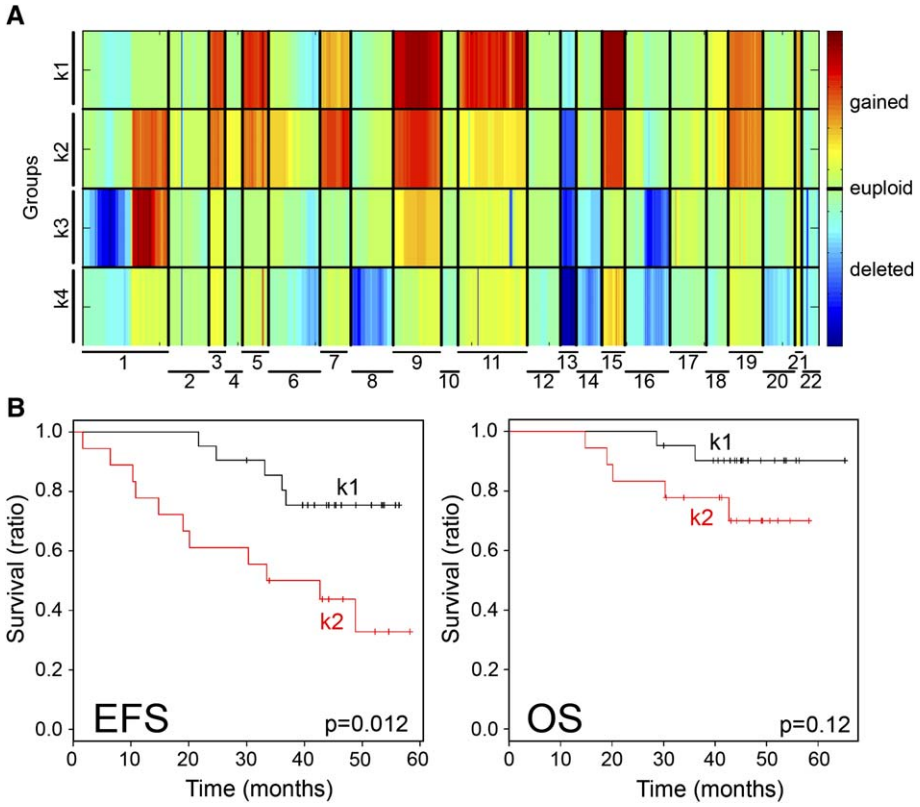

Figure 2. gNMF classification with rank K = 4 identifies four distinct subgroups

A: aCGH profiles of 67 clinically annotated primary tumors were subjected to NMF analyses (1000 repetitions). y axis indicates the four subgroups identified by NMF. The x axis coordinates represent genomic map order (from ch1 to ch22). The colors denote gained (red), euploid (yellow/green), or deleted (blue) chromosome material.

B: KM event-free survival (EFS; left) and overall survival (OS; right) curves for 64 MM patients for k1 and k2 subgroups. k1 shows significantly better event-free survival than k2 (p = 0.012), while OS did not reach statistical significance (p = 0.12).

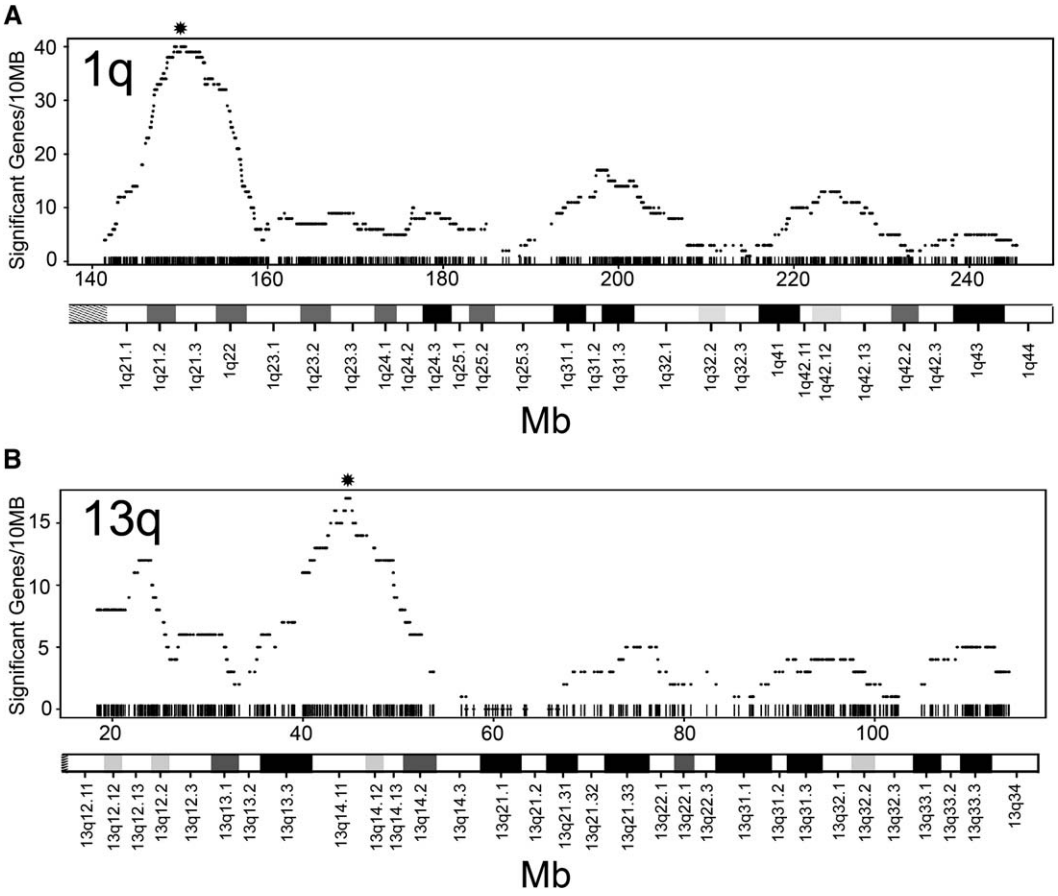

Figure 3. Distribution of genes differentially expressed between k1 and k2 subgroups and residing on chromosomes 1q or 13q

Expression probe sets are mapped to their respective genomic positions and are shown as vertical hash marks along the bottom of each plot. Black hash marks denote genes found to be differentially expressed between k1 and k2 subgroups by SAM. Count of significant genes within a moving 10 Mb window is shown (y axis), and asterisks indicate the center of regions of significant clustering ($p < 0.05$ by permutation testing). The significant region spans approximately 143–158 Mb for ch1 (**A**) and 38–50 Mb for ch13 (**B**), though boundaries are approximates based on the moving window width.

patterns (Experimental Procedures) identified several prominent features, which include the following: (1) ch11 gain in 20/21 k1 versus only 6/16 k2 samples ($p < 0.001$, χ^2 test); (2) gains of ch1q in 9/16 k2 versus 0/21 k1 samples ($p < 0.001$); and (3) ch13 loss in 10/16 k2 versus only 4/21 k1 samples ($p = 0.019$). Consistently, these three features were among the major components defined by gNMF in the K4 classification (Figure 2A). Together, our analysis suggests that, among hyperdiploid MM, ch11 gain confers a more favorable outcome, whereas ch1q gain and ch13 loss drive poor outcome.

To mine further for poor survival ch1q and ch13 genes in k2, the significance analysis of microarrays (SAM) approach (Tusher et al., 2001) was applied to corresponding transcriptome profiles. Significantly increased expression in k2 versus k1 was noted for 111/2210 probes (95 genes) mapping to ch1q (FDR = 15%) (Table S2), and decreased expression in k2 versus k1 for 48/1163 probes (46 genes) mapping to ch13 (FDR = 15%) (Table S3). In both ch1q and ch13 cases, SAM-significant probes clustered in specific chromosomal bands. We therefore tested whether subregions of ch1q and ch13 were significantly enriched for differentially expressed genes by counting significant genes in a 10 Mb moving window, and testing for significance by permutation of gene position (Tonon et al., 2005). This approach revealed a marked enrichment in overexpressed genes residing at 1q21–q23, spanning approximately 143–158 Mb (Figure 3A; $p < 0.05$), a region that is particularly notable, as it encompasses two of our high-priority minimal common regions (MCRs) associated with poor survival (see below). The 1q genes upregulated in the poor prognosis k2 subgroup included several encoding cancer-relevant activities (Table S2; Discussion). Analogous studies of ch13 showed significant enrichment of underexpressed genes residing at 13q14 between 38 and 50 Mb (Figure 3B; $p < 0.05$), a region known to sustain the highest frequency of LOH on ch13 in MM and including *RB1* (Elnenaei et al., 2003). Within this region, comparison between gene expression in k1 and k2 groups identifies additional candidate TSGs (Table S3; Discussion).

In summary, molecular classification based on genomic features provided clear evidence that hyperdiploid MM is a genetically heterogeneous disease, and that a poor prognosis subset of hyperdiploid MM patients can be identified by the presence of ch1q gain and/or ch13 loss. In addition to the translational implications, the stratification of hyperdiploid MM into two outcome-correlated subclasses itself serves as strong validation of application of this gNMF algorithm for molecular classification of aCGH profiles.

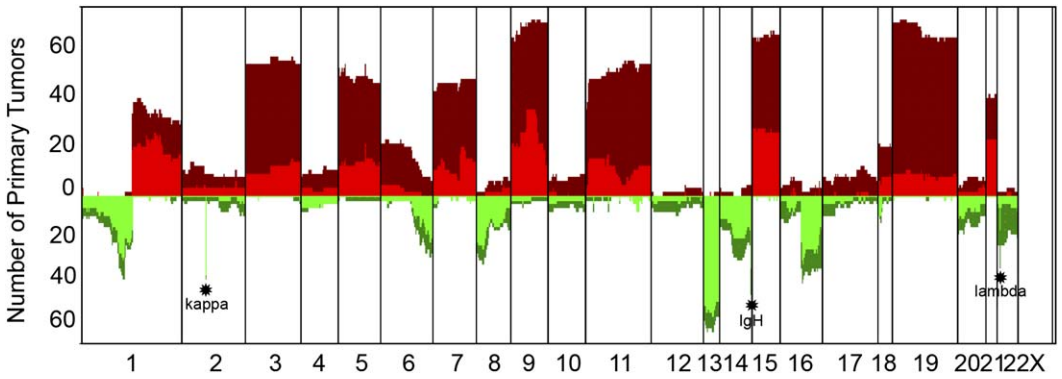

Figure 4. Summary of genomic profiles of MM and recurrence of chromosomal alterations in primary MM tumors

Integer-value recurrence of CNAs across the samples in segmented data (y axis) is plotted for each probe evenly aligned along the x axis in chromosomal order. Dark red or green bars denote the number of samples with gain or loss of chromosome material, and bright red or green bars represent the number of samples showing amplification or deletion (Experimental Procedures). Asterisks show focal deletions of the *kappa* (2p12), *IgH* (14q32), and *lambda* chain (22q11) loci physiological in B cell postgerminal center neoplasms.

Recurrent and focal CNAs in the MM genome with potential biological and clinical significance

In addition to genomic classification and identification of prognosis-associated loci, the high-definition picture of the MM genome enabled definition of recurrent CNAs with strong involvement in MM pathogenesis. The performance features of the aCGH platforms (Brennan et al., 2004) and segmentation analysis (Aguirre et al., 2004; Tonon et al., 2005) readily identified large regional changes. The skyline recurrence plot (Figure 4) mirrors well the frequencies of previously reported chromosomal gains of 1q, 3, 5, 7, 9, 11, 15, 19, and 21, and losses of 1p and 13 (Avet-Loiseau et al., 1997; Cigudosa et al., 1998; Cremer et al., 2005; Fonseca et al., 2004). In addition, these surveys also captured focal, recurrent, high-amplitude CNAs present in both cell lines and primary tumor specimens (Figure S2A). Analysis of 358 distinct CNAs across the MM tumor collection (n = 67) and MM cell lines collection (n = 43) delimited 298 MCRs (Figure S2B and data not shown), which were further filtered down to 87 prioritized MCRs based on the criteria of presence in primary tumors and occurrence of at least one high-amplitude event ($\log_2$ ratio > 0.8). These 87 MCRs comprised 47 amplifications (Table 1) and 40 deletions (Table 2), spanning a median size of 0.89 Mb with an average of 12 known genes. That these "high-priority" MCRs possess high disease relevance is reflected by (1) consistent verification by real-time quantitative PCR (QPCR) in all randomly assayed high-priority MCRs and by FISH in selected cases (Tables 1 and 2, asterisks; Figure S3 and data not shown), and (2) inclusion of all signature MM loci of known pathogenetic relevance such as deletion of a region including the *TP53* tumor suppressor and focal amplifications of areas including the hepatocyte growth factor (*HGF*) and the *MYC* and *ABL1* oncogenes.

Integrated copy number and expression analysis identified known and potential cancer genes in MM pathogenesis

As copy number alterations function to alter expression of resident genes (Aguirre et al., 2004; Platzer et al., 2002; Tonon et al., 2005), we conducted an integrated RNA expression analysis by the gene weight measure as described (Aguirre et al., 2004; Hyman et al., 2002). First, for each gene residing within an amplified MCR, we asked whether its expression showed a copy number-correlated pattern by comparing the mRNA levels in tumors with and without CNAs in the region of interest. In addition, modeling after bona fide oncogenes, such as *MYC*, whose expressions are known to be dysregulated by mechanisms other than gene dosage alteration, we also compared expression of the gene in tumors with or without CNAs, relative to normal plasma cells, respectively (Experimental Procedures). Genes showing this "oncogene-like" expression pattern—namely copy number-correlated expression *and* significant overexpression in tumors without amplification versus normal plasma cells—were considered high-probability candidates targeted for amplification in these MCRs during MM development.

By such stringent criteria, approximately 30% of the 2151 genes residing in the high-priority MCRs were considered strong candidates. These included genes with credentialed roles in MM pathogenesis such as *MYC*, *MCL1*, *IL6R*, *HGF*, and *ABL1* (Table 1 and data not shown), as well as many functionally diverse genes with no known link to MM development. Particularly noteworthy are several E3 ubiquitin ligase genes, including the anaphase promoting complex subunit 2 (*ANAPC2*), F box protein 3 (*FBXO3*), F box protein 9 (*FBXO9*), SMAD-specific E3 ubiquitin protein ligase 2 (*SMURF2*), and huntingtin interacting protein 2 (*HIP2*). Amplification/expression was observed for molecular chaperones including chaperonin containing TCP1, subunit 3 (γ) (*CCT3*), Der1-like domain family, member 1 (*DERL1*), *DNAJ* (*HSP40*) homolog, subfamily C, member 1 (*DNAJC1*), and the cochaperone adaptor (*CDC37*), which has been shown to be oncogenic in transgenic mice (Stepanova et al., 2000). There was deregulation of many ribosome biogenesis and protein synthesis genes, including ribosomal protein-encoding genes *RBM8A* and *RPL18*, and translational control gene *EEF2* (Table 1). Thus, we conclude that integrated copy number and expression analysis provides an efficient first-pass means of distinguishing bystanders from potential cancer gene(s) target(s) within any given CNAs; as previously observed (Aguirre et al., 2004; Pollack et al., 2002; Tonon et al., 2005).

High-priority MCRs define loci with biological and prognostic relevance

To determine whether any high-priority MCRs (Tables 1 and 2) possessed prognostic relevance, we compared the survival

Table 1. High-confidence MCRs in multiple myeloma: Gains and amplifications

					MCR recurrence							
	MCRs				Gains/losses			Amps/dels				
Cytogenetic band	Position (Mb)	Size (Mb)	Max/min value	Known genes	%	Tumors	Cell lines	Tumors	Cell lines	Candidate genes	miRNA	Integration sites
1q21.1–1q22*,◆	142.60–152.10	9.50	2.46	228	55.5	26	35	14	27	MCL1, IL6R, PSMD4, PSMB4, UBE2Q, UBAP2L, RBM8A, RPS27, PIAS3, POLR3C, HIST2H2AA, LASS2, MRPL9, JTB, HAX1, SHC1, APH-1A, BCL9, ZNF364		Dkmi7, Rorc, NKI-6, Dkmi8
1q22	152.73–153.27	0.53	1.44	21	53.6	24	35	13	28	SSR2, MAPBPIP, CCT3, MEF2D, C1orf85	hsa-mir-9-1	Evi53
1q43-1q44	236.73–245.42	8.69	0.93	94	35.5	20	19	12	12	FH, HNRPU, TFB2M		
2p25.1	11.35–11.68	0.33	1.48	3	13.6	7	8	2	5			
2p16.1	55.68–56.01	0.32	1.16	4	11.8	8	5	3	2	SMEK2, MGC15407		
2q35	219.08–219.44	0.36	1.02	11	9.09	7	3	4	2	BCS1L	hsa-mir-26b	
3q27.1–3q27.2	185.53–187.13	1.60	0.90	19	47.3	35	17	9	5	POLR2H, EIF4G1		
4p14–4p12	39.60–48.73	9.12	0.88	45	6.36	7	0	2	0	HIP2, SCC-112, OCIAD1, RHOH, UCHL1, TMEM33		
4q22.3–4q24	95.84–104.04	8.20	1.83	35	9.09	8	2	2	1	PPP3CA, H2AFZ, SLC39A8		Evi157
5p12	42.76–43.16	0.41	0.87	5	43.6	35	13	10	6			Evi124
5q13.2–5q13.3*	73.12–76.06	2.94	1.26	19	36.4	31	9	8	4	TINP1, HMGCR, COL4A3BP, POLK		Evi19
5q14.1	77.48–79.97	2.49	1.26	20	35.5	31	8	8	4	SCAMP1, ARSB, JMY, MTX3, PAPD4		
5q31.3	140.55–140.66	0.11	2.30	10	33.6	35	2	14	1	PCDHB10		
6p21.32–6p21.32	32.04–32.20	0.16	0.82	11	20.9	14	9	3	6	RDBP, STK19		
6p21.1	45.99–53.24	7.25	0.92	53	18.2	13	7	1	2	CD2AP, MUT, FBXO9, TFAP2B, TMEM14A, ICK, ELOVL5	hsa-mir-206, hsa-mir-133b	
6q13	70.81–73.68	2.87	0.89	11	15.5	10	7	1	3	SMAP1	hsa-mir-30c-2, hsa-mir-30a-5p	
7p14.3	30.89–33.74	0.26	1.36	15	42.7	30	17	10	5	LSM5, FKBP9, NT5C3, PTHB1, KBTBD2, RP9		
7p12.2*	50.25–50.53	0.28	2.06	3	43.6	30	18	7	6			
7q21.11	79.93–81.24	1.30	1.32	4	43.6	30	18	6	6	HGF		
7q21.12	86.03–87.18	1.15	1.85	11	44.5	30	19	6	5	MCFP, DMTF1, ASK, TP53AP1		
7q32.1–7q32.2	126.07–129.78	3.70	1.44	48	50	31	24	13	10	CALU, TNPO3, UBE2H, ARF5, RBM28, IRF5, METTL2, CDC26	hsa-mir-129-1, hsa-mir-182, hsa-mir-96, hsa-mir-183, hsa-mir-335	
8q24.12–8q24.13◆	120.50–126.52	6.02	0.84	37	20.9	4	19	2	8	DEPDC6, MRPL13, DERL1, ZHX1, TATDN1, RNF139, FBXO32		
8q24.21–8q24.3◆	128.50–146.20	17.70	2.50	127	24.5	6	21	3	10	MYC, PVT1	hsa-mir-30b, hsa-mir-30d, hsa-mir-151	Si22, Dkmi21
9q34.11–9q34.3*	127.35–137.60	10.26	1.61	202	54.5	47	13	14	3	ABL1, CIZ1, NUP188, SET, ENDOG, SURF1, SURF2, SURF4, CAMSAP1, UBADC1, SDCCAG3, EDF1, DPP7, ANAPC2, COBRA1	hsa-mir-199b, hsa-mir-219-2, hsa-mir-126	Notch1, Ppp2r4, Gfi1b, Evi43
9q34.3	137.62–138.29	0.67	0.82	9	50.9	46	10	11	1	MRPL41, ZMYND19, EHMT1		
10p12.33–10p12.31	17.69–22.09	4.40	1.08	15	14.5	6	10	2	6	DNAJC1		
11p13	33.30–34.10	0.81	0.80	8	34.5	32	6	10	3	CD59, FBXO3, M11S1, HIPK3		Lmo2
11p11.2	46.72–47.34	0.62	1.01	11	41.8	33	13	10	7	ACP2, CKAP5, DDB2		Sfpi1

Table 1. *Continued*

	MCRs				MCR recurrence							
					Gains/losses			Amps/dels				
Cytogenetic band	Position (Mb)	Size (Mb)	Max/min value	Known genes	%	Tumors	Cell lines	Tumors	Cell lines	Candidate genes	miRNA	Integration sites
11q13.4–11q14.1	73.40–77.71	4.30	1.27	54	51.8	36	21	6	12	SPCS2, RPS3	hsa-mir-326	
11q14.1	77.01–82.85	0.00	1.27	24	51.8	36	21	7	10	PCF11, HBXAP		
11q23.3	112.74–122.23	3.47	1.47	112	50.9	35	21	8	12	UBE4A, ATP5L, RPS25, HMBS, CBL, DDX6, TRAPPC4, HYOU1, RNF26		
12q24.33	131.00–131.07	0.07	1.23	2	5.45	2	4	1	2			
14q11.2	19.41–19.51	0.11	1.33	3	5.45	3	3	3	0			
15q24.2	73.45–73.72	0.27	1.49	4	47.3	44	8	18	2	IMP3		
16q24.2	86.43–87.13	0.70	0.98	6	8.18	4	5	2	0			
17p11.2*	17.62–17.71	0.09	1.13	4	12.7	6	8	1	2	RPL13, PAIP1		
17q23.2	53.94–54.69	0.75	0.89	11	17.3	8	11	1	6	MTMR4, FLJ11029, RAD51C	hsa-mir-301	
18q12.1–18q12.2*	30.53–31.95	1.43	1.12	12	26.4	13	16	5	6	ZNF271, P15RS, SLC39A6	hsa-mir-187	Evi135
18q21.2–18q21.33	51.41–59.71	8.30	1.39	42	29.1	13	19	5	10	MALT1, BCL2, FVT1	hsa-mir-122a	Evi36
19p13.3–19p13.2	0.24–11.19	10.96	0.91	320	66.4	47	26	7	14	CDC34, POLRMT, WDR18, ATP5D, MUM1, NDUFS7, RPS15, UQCR, REXO1, CSNK1G2, MOBKL2A, MKNK2, INSR, TIMM13, MRPL54, EEF2, MAP2K2, HDGF2, TICAM1, RPL36, NDUFA11, CLPP, SH2D3A, NDUFA7, RPS28, RAB11B, EIF3S4, MRPL4, ICAM1, ICAM3, CDC37, APG4D, ILF3, QTRT1, SMARCA4	hsa-mir-7-3, hsa-mir-199a-1	Tcfe2a, Evi103, Evi158
19q13.11	37.57–38.16	0.59	0.97	9	47.3	44	8	7	0	PDCD5, ANKRD27		
19q13.12	41.33–41.90	0.57	0.85	11	47.3	43	9	6	0	ZNF146		
19q13.2	44.51–44.66	0.15	0.85	8	47.3	43	9	6	0	ZFP36, RPS16, SUPT5H		Evi24
19q13.31	49.11–49.85	0.47	0.83	21	46.4	42	9	5	0	ZNF227		
19q13.32–19q13.43	53.31–63.07	9.76	0.89	346	49.1	42	12	6	3	FLT3LG, AKT1S1, PSCD2, RPL18, BAX, RUVBL2, SNRP70, RPL13A, RPS11, TBC1D17, NUP62, NDUFA3, PRPF31, RPS9, HSPBP1, RPL28, UBE2S, U2AF2	hsa-mir-150, hsa-mir-99b, hsa-let-7e, hsa-mir-125a, hsa-mir-371, hsa-mir-372, hsa-mir-373	
20q12–20q13.12*,◆	39.14–43.39	4.25	1.30	43	15	5	11	3	1	TDE1, SFRS6, YWHAB, PPIA		
21q22.3	45.54–46.91	1.37	1.16	17	37.3	27	14	15	6	MCM3AP, C21orf106, HRMT1L1		

For each MCR, the number of NCBI known genes is reported. Number of transcripts is based on Build 35 of the NCBI. Within the amplified MCR, some of the genes showing both copy number-driven expression and overexpression also in the absence of amplification (Experimental Procedures) are reported (candidate genes). MCRs validated by QPCR (asterisks) and/or FISH (black diamonds) identify PS-MCRs. MCR recurrence is denoted as percentage of the total data set. Only the known genes within the boundaries have been included. Known hotspots for proviral integration residing within MCRs and MCR-containing microRNAs are indicated. The MCR in 1q was subjected to further fine mapping (Figure S3).

for MM cases with or without specific MCRs against KM analyses (data not shown). In other words, we asked whether presence or absence of a particular MCR was correlated with outcome, without taking into consideration the gNMF subclass assignment of the MM samples. This straightforward correlation identified 14 MCRs associated with poor survival (hereafter designated as PS-MCR for "poor-survival" MCR) (Table 1, diamonds). This PS-MCR list included an amplification on ch8 (including *MYC*) and a deletion on ch17 (including *TP53*)—both genetic events have been linked to poor prognosis in MM (Kuehl and Bergsagel, 2002).

Of the remaining amplified PS-MCRs, one resided in the critical ch1q region and two resided in novel MM loci. The first novel amplified PS-MCR defined a ch8q24 region spanning 6 Mb with 37 genes (distinct from *MYC*) and was notable for its association with a poor clinical outcome and tumor recurrence in other human cancer types (Tonon et al., 2005; van Duin et al., 2005). Among the seven genes showing oncogene-like expression

Table 2. High-confidence MCRs in multiple myeloma: Losses and deletions

	MCRs				MCR recurrence							
					Gains/losses			Amps/dels				
Cytogenetic band	Position (Mb)	Size (Mb)	Max/min	Known genes	%	Tumors	Cell lines	Tumors	Cell lines	Candidate genes	miRNA	Integration sites
1p36.22◆	10.413–10.625	0.21	−0.98	4	12	6	7	4	0	DFFA		
1p35.2◆	30.858–31.042	0.18	−0.98	4	14	7	8	5	1	LAPTM5		
1p32.3–1p32.2◆	54.231–57.125	2.89	−0.87	23	25	12	16	8	9	SSBP3, USP24		
1p13.3–1p12 *,◆	111.486–118.215	6.73	−1.00	71	41	15	30	13	21	DENND2D, DDX20, ST7L, PPP2CZ, LRIG2, PTPN22, CD58, IGSF2		Rap1a, Nras
3p21.31	46.876–47.033	0.16	−1.61	5	11	1	11	1	3	CCDC12, HYPB		
4q11–4q13.2	52.587–68.443	15.86	−0.93	48	19	3	18	3	3	CENPC1, BRDG1		Dkmi9
4q13.3–4q21.1	71.101–76.940	5.84	−0.81	51	19	3	18	3	3	RASSF6, MOBKL1A		Crlz1
4q22.1*	88.469–89.085	0.62	−1.16	10	17	2	17	2	4	IBSP, SPARCL1		
4q23	99.844–100.589	0.75	−1.70	9	18	3	17	3	3	TM4SF9, METAP1		
6p25.3	0.295–0.356	0.06	−1.12	2	5	4	1	4	0	DUSP22		
8p23.3–8p21.3	0.187–20.123	19.94	−0.89	135	28	16	15	13	11	NAT2, RP1L1, PINX1, TUSC3, DLC1, MTUS1	hsa-mir-124a-1	
8p21.3–8p12	22.078–33.573	11.49	−0.97	87	33	18	18	14	14	TNFRSF10B, PPP2R2A, BNIP3L	hsa-mir-320	
10q25.1–10q25.2	111.644–112.353	0.71	−0.90	6	18	4	16	1	4	MXI1		
10q26.2–10q26.3*	128.900–135.241	6.34	−1.27	40	18	4	16	2	6	PTPRE, PPP2R2D, BNIP3, INPP5A		Evi168
11p15.4	5.733–5.799	0.07	−3.44	6	15	4	12	2	2	OR52N4, OR52N5, OR52N1, OR52N2		
11q11	55.097–55.176	0.08	−3.86	2	7	4	4	2	0	OR4C16, OR4C11, OR4P4, OR4S2		
11q22.1–11q22.2	100.460–102.151	1.69	−3.12	14	9	4	6	4	5	PORIMIN, YAP1, MMP8		
11q23.2–11q23.3	112.818–114.552	1.73	−0.94	15	5	2	4	2	3	IGSF4		
11q24.2–11q24.3*	125.668–128.280	2.61	−2.10	12	8	1	8	1	4	ST3GAL4, FLI1, LOC387820		Fli1 | Ets1
12q13.11–12q13.12	46.814–47.702	0.89	−0.80	30	10	4	7	1	2	ANP32D		
12q13.2	54.824–54.893	0.07	−1.71	5	7	5	3	1	1	SMARCC2		
12q14.1–12q14.2	61.013–62.674	1.66	−1.02	9	8	3	6	1	1	PPM1H	hsa-let-7i	
12q23.2–12q23.3	101.334–102.980	1.65	−0.90	12	12	3	10	1	4	STAB2		Evi12
13q34	112.239–113.013	0.77	−1.01	12	49	32	22	28	16	CUL4A		
14q12	29.291–31.655	1.23	−0.80	11	32	9	26	7	8	STRN3, HECTD1		
14q32.13–14q32.2	93.920–95.630	1.71	−0.96	22	33	13	23	9	11	BCL11B		Evi151
15q15.1	39.460–39.583	0.12	−1.41	5	8	1	8	1	1	NUSAP1		
16p13.3	2.102–2.236	0.13	−2.07	10	14	7	8	1	2	PKD1, TRAF7, E4F1		
16q11.2–16q12.2◆	45.173–52.875	7.70	−1.71	45	31	23	11	20	6	DNAJA2, SIAH1, PAPD5, NKD1, CARD15, RBL2, FTS, CYLD		
16q13◆	56.234–56.570	0.14	−1.13	9	29	20	12	14	7	GPR56, KATNB1, KIFC3, CNGB1		
16q24.1–16q24.2	83.456–85.894	2.44	−1.29	20	31	19	15	13	8	FOXC2, FOXL1		Evi96, Zdhhc7
16q24.3◆	88.232–88.627	0.39	−1.29	19	31	19	15	9	8	GAS8		
17p13.2*,◆	3.306–4.791	1.49	−1.31	43	23	5	20	2	14	TAX1BP3, GSG2		
17p13.1–17p12*,◆	6.454–13.921	7.47	−1.03	125	30	6	27	2	19	TP53, TNFSF12, TNFSF13	hsa-mir-195, hsa-mir-324	Trp53
19q13.33	55.578–55.678	0.10	−1.14	5	12	1	12	1	2	POLB1, SPIB		
20p12.1◆	13.083–17.411	4.33	−0.87	21	15	10	7	6	3	OTOR		
20q11.21	29.745–30.246	0.50	−1.64	13	7	6	2	3	1	FKHL18, DUSP15, HCK		Bcl2l, Evi47, Cri07
20q13.12◆	43.967–44.107	0.14	−1.08	6	12	8	5	4	1	PCIF1		
21p11.2–21p11.1	9.932–10.167	0.24	−1.11	3	6	1	6	1	1	TPTE		
22q11.23	22.638–23.065	0.43	−1.59	10	25	12	15	2	6	DKFZP434P211		

For each MCR, the number of NCBI known genes is reported. Number of transcripts is based on Build 35 of the NCBI. MCRs validated by QPCR (asterisks) and/or FISH (black diamonds) identify PS-MCRs. Potential tumor suppressor genes residing within the MCRs are indicated. Known hotspots for proviral integration residing within MCRs and MCR containing microRNAs are indicated. The MCR at 11q11 has recently been shown by Sebat et al. (2004) to be a copy number polymorphism (ORF511, ch11q11).

pattern are *DEPDC6* and *FBXO32*. The second novel amplified PS-MCR targets a ch20q region (43 genes) that is associated with disease progression and metastases in prostate cancer (Wullich et al., 2004), esophageal squamous cell carcinoma, and gastric and colorectal adenocarcinoma (Fujita et al., 2003). In this amplicon resides *PPIA* (Cyclophilin A), which functions as a paracrine and autocrine modulator of endothelial cells proliferation, migration, and invasive capacity (Kim et al., 2004).

Among the ten deleted PS-MCRs, four highly discrete regions mapped to ch1p within a much larger, less defined region implicated in MM prognosis based on FISH analyses (Panani et al., 2004). Two loci contain only four genes including *DFFA*, a key caspase-3 substrate that triggers DNA fragmentation during apoptosis (Table 2). The remaining deletion PS-MCRs include three on ch16, two on ch17—one of them harboring *TP53* (Table 2)—and two on ch20. Among these, the 16q12 region (45 genes) includes the tumor suppressor *CYLD*. The 16q13 region contains only five genes: G protein-coupled receptor 56, G protein-coupled receptor 97, two genes encoding the centrosomal proteins *KATNB1* and *KIFC3*, and the hypothetical protein encoding gene *C16orf50*. Finally, the ch17p13.3–13.2 locus spanning 1.49 Mb with 43 genes includes several prime tumor suppressor candidates such as *TAX1BP3*, a Wnt/β-catenin signaling pathway inhibitor (Kanamori et al., 2003), and *GSG2*, a kinase required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment (Dai et al., 2005). In summary, the triangulation of aCGH, expression profiles, and clinical annotation has identified genomic regions and candidate oncogenes and TSGs with plausible links to MM disease progression and clinical outcome.

Discussion

In this study, unsupervised classification of MM by gNMF defined four genomically distinct subclasses, revealing a level of molecular heterogeneity not previously appreciated for this disease. Additionally, a compendium of focal and recurrent genetic lesions, a subset of which is linked to poor prognosis, has been identified. Finally, integration of expression data with genomic and clinical information has further delimited a list of prime oncogene candidates residing within the amplified MCRs.

Distinct molecular subclasses linked to disease pathogenesis and prognosis

This study demonstrates that embedded within the aCGH profiles are biologically significant patterns of genomic alterations. gNMF with rank of K = 2 divided the MM primary tumors into kA versus kB subgroups corresponding to the traditional hyperdiploid and nonhyperdiploid groups, respectively (Figure 1). Although the lack of a significant outcome difference between these two subgroups may relate to sample size and/or time of clinical follow up (median 43 months), the fact that further stratification with rank K = 4 resulted in distinct subclasses with differences in clinical outcome (Figure 2 and data not shown) indicates that the more likely explanation relates to inherent heterogeneity within the previously defined hyperdiploid and nonhyperdiploid groups. Indeed, when hyperdiploid MM cases were stratified into two subclasses, k1 and k2, a clear survival advantage was observed among patients in the k1 group (Figure 2), which was characterized by presence of ch11 gain and absence of ch1q gain or ch13 loss. Importantly, by reducing the k1 versus k2 classifier to presence/absence of ch1q and ch13 alterations determined by FISH in hyperdiploid MM (i.e., tumor samples with odd-chromosome gain pattern as determined by average of gene expression values across chromosomal arms), we were able to validate the prognostic difference of k1 and k2 in an independent cohort of 135 outcome-annotated TT2-treated MM cases (Barlogie et al., 2006) (Figure S4). A further validation of the biological significance of k1 and k2 subclasses was provided by gene set enrichment analysis (GSEA) (Monti et al., 2005) of the transcriptomes, revealing perturbation of distinct cancer-relevant pathways in each subclass. While *TP53*, *KRAS*, *FRAP1*, and the proteasome pathways were altered in both subgroups (data not shown), dysregulation of additional cancer-relevant pathways, such as sonic hedgehog and *RAC1* (Mitsiades et al., 2004; Qiang et al., 2005), was observed only in k2 samples, consistent with a more advanced evolution of MM. However, it should be emphasized that, although hyperdiploid MM stratified into good or poor prognosis subgroups based on well-known genomic features bearing obvious immediate translational implications, the broad prognostic relevance of these gNMF subclasses will require stringent validation in prospective clinical trials, where the impact of sampling bias, therapy-specific outcome, and/or other confounding factors can be investigated.

To identify key genes within ch1q or ch13 that may dictate clinical behavior of hyperdiploid MM, we integrated gene expression data with copy number profiles to generate a short list of strong candidates (Tables S2 and S3). Among the 1q genes upregulated in k1 versus k2 are many that have not been previously linked to any cancer and that affect survival and proliferation, hypoxia, and cell adhesion and motility (Table S2). Regarding ch13 deletion, most reported cases show loss of the entire chromosome, although ~15% of MM patients sustain more focal deletion or LOH targeting ch13q14 (Elnenaei et al., 2003). Our integrated RNA-DNA analysis was able to narrow the region of interest to a 10 Mb region on ch13q14, encompassing some interesting candidates. *RB1* was included in this region; however, its role in MM has been questioned even in advanced MM and cell lines (Kuehl and Bergsagel, 2002; Shaughnessy et al., 2003). Among the downregulated genes within this region are several putative TSGs (Table S3). While integration of expression and copy number provides an effective means to convert genomic features into candidate genes, it should be emphasized that such genome-anchored analysis will preferentially capture genes whose dysregulation and involvement in MM are driven by copy number aberrations, while missing other MM-relevant genes that are predominantly regulated by other mechanisms.

MCRs harbor candidate genes of biological and clinical relevance

Most highly recurrent focal MCRs identified in this study are novel, and several possess clinical relevance (Figure 4 and Tables 1 and 2). It should be noted that our genomic approach primarily emphasized regions and genes within defined MCRs, and therefore genes residing outside of MCRs that might be related to prognosis would not be captured by our analysis. Interestingly, the MM genome shared many CNAs with other histologically unrelated cancers such as pancreas, lung, breast, and ovary, suggesting common mechanisms of disease pathogenesis (Aguirre et al., 2004; Cheng et al., 2004; Tonon et al., 2005). Also, several known hotspots for proviral integration and/or chromosomal rearrangement targeting pathogenetic loci were located within high-priority MCRs (Tables 1 and 2) (Collier et al., 2005).

The integration of copy number changes and gene expression information (using bone marrow-derived plasma cells as reference) enabled significant reductions in the number of candidate oncogenes residing within the high-priority MCR amplicons. The reliability of this approach is confirmed by the ability of this

algorithm to successfully sift established oncogenes from surrounding, likely bystander genes. For instance, *HGF* was located within a small MCR including four genes and was found to be the only gene with such an oncogene-like expression pattern. It should be noted that microRNAs embedded within the MCRs could also exert an oncogenic role, as recently demonstrated (He et al., 2005). Indeed, 20% of the 87 high-priority MCRs contain microRNAs that could therefore represent the key genetic element of the CNA.

Among candidates from other MCRs exhibiting oncogene-like expression are components of prominent cancer pathways (Table 1). The amplification and overexpression of E3 ubiquitin ligases and two DEAD box proteins on ch11 and ch17 are intriguing given the critical role of the ubiquitin-proteasome pathway in auditing the levels of key cell cycle and apoptosis control proteins (Nakayama and Nakayama, 2005) and the proposed role of human RNA helicases in various types of cancer (Abdelhaleem, 2004). Interestingly, in the amplified PS-MCR on ch20, among the genes showing an oncogene-like expression pattern, there was *PPIA* (Cyclophilin A) which functions as a paracrine and autocrine modulator of endothelial cell proliferation, migration, and invasive capacity (Kim et al., 2004).

Among the TSGs, *TP53* resided in an MCR that was also linked to poor prognosis, confirming previous reports (Bergsagel and Kuehl, 2005). Interestingly, *TP53* was generally overexpressed (data not shown), a finding consistent with upregulation of mutant p53 (Furubo et al., 1999). *CDKN2C* and *CDKN1B*, often deleted in MM (Bergsagel and Kuehl, 2005; Filipits et al., 2003), were not included in the high-priority MCR list, since the segmentation algorithm discards CNAs identified by single probes in exchange for an improved false positive rate (Aguirre et al., 2004); however, they were recurrently lost in the raw CGH profiles (Figure S3 and data not shown). As expected, *CDKN2A* and *PTEN* did not show copy number losses, as they are usually inactivated in MM through promoter methylation and point mutations, respectively (Chang et al., 2006; Urashima et al., 1997). In addition, several bona fide and putative TSGs not previously linked to MM were identified (Table 2).

In conclusion, the application of gene-specific CGH platforms, custom bioinformatic tools, and expression profiles has identified many recurrent amplifications and deletions that significantly alter expression of specific subsets of genes. The high degree of MM genomic complexity and the common targeting of the same loci in other tumor types support the view that a large number of important oncogenes and TSGs remain to be identified. This integrated view of the MM genome is consistent with the concept of widespread changes in the expression of genes with cancer activity in the pathogenesis of MM, a concept with therapeutic and diagnostic implications for MM in particular and cancer in general.

Experimental procedures

Primary tumor genomic DNA and MM cell lines

Genomic DNA from primary tumors derived from the Donna and Donald Lambert Laboratory of Myeloma Genetics, Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences (n = 67) (Barlogie et al., 2006). Patients were treated with the TT2 Protocol (median follow-up 43 months, range: 5 months to 65 months) (Barlogie et al., 2006) (Table S1). MM cell lines were collected at Genetics Branch, NCI (Michael Kuehl), Weill Medical College and Graduate School of Medical Sciences of Cornell University (Leif Bergsagel), and Jerome Lipper Multiple Myeloma Center, the Department of Medical Oncology, DFCI (K.C.A.) (G.T. and R.A.D., unpublished data). The Institutional Review Board of the University of Arkansas for Medical Sciences approved the research studies, and all subjects provided written informed consent approving use of their samples for research purposes.

aCGH profiling

Primary tumor DNA was obtained from CD138-enriched cell populations using CD138 magnetic microbeads (Miltenyi Biotec Inc.). Genomic DNA from cell lines and primary tumors were extracted according to manufacturer instructions (Gentra System Inc.). Genomic DNA was fragmented and random-prime labeled as previously described (Tonon et al., 2005) and hybridized to either human cDNA or oligo microarrays. All MM cell lines were analyzed using a cDNA array platform (Aguirre et al., 2004), while the 67 primary tumors were analyzed using the oligonucleotide array platform (Brennan et al., 2004; Tonon et al., 2005). The cDNA microarray contains 14,160 cDNA clones (Agilent Technologies, Human 1 clone set) with 13,281 genome-mappable clones, for which approximately 11,211 unique map positions were defined (NCBI, Build 35). The median interval between mapped elements is 0.73 Mb, 94.1% of intervals are less than 1 Mb, and 98.9% are less than 3 Mb. The oligo array contains 22,500 elements designed for expression profiling (Agilent Technologies, Human 1A V2), for which 16,097 unique map positions were defined (NCBI, Build 35). The median interval between mapped elements is 0.55 Mb, 96.7% of intervals are less than 1 Mb, and 99.5% are less than 3 Mb. Fluorescence ratios of scanned images of the arrays were calculated as the average of two paired arrays (dye swap), and raw aCGH profiles were processed to identify statistically significant transitions in copy number using a segmentation algorithm, as described (Aguirre et al., 2004; Olshen et al., 2004). The data are centered by the tallest mode in the distribution of the segment values. After mode-centering, we defined gains and losses for the oligonucleotide data set as $\log_2$ ratios $\geq +0.11$ or -0.11 (± 4 SD of the middle 50% quantile of data) and amplification and deletion as a ratio >0.4 or <-0.4, respectively. For the cDNA data set, gains and losses were defined as $\log_2$ ratios of $\geq +0.13$ or -0.13 (± 4 SD of the middle 50% quantile of data) and amplification and deletion as a ratio >0.5 or <-0.5, respectively. The segmented $\log_2$ ratio distributions and thresholds chosen are shown in Figure S4D. The narrow central peak sharply defines "normal" copy number for a sample (most common chromosomal copy number, not necessarily diploid). Thus "abnormal" (relative gain/loss) is well defined by tight thresholds (± 4 SD of the middle 50% quantile). The aCGH data discussed in this publication are available as Supplemental Data and at http://genomic.dfci.harvard.edu/array_cgh_db.htm.

Adaptation of the NMF algorithm to aCGH analysis and Fisher's exact test

The algorithm used for aCGH analysis is a modification of NMF that entails conversion of aCGH data followed by NMF-based classification (C.B. and L.C., unpublished data). NMF is a method designed to reduce data dimensionality based on matrix decomposition by parts, and it has recently been shown to be effective in elucidating meaningful structure inherent in gene expression data sets: organizing both the genes and samples to provide biologically or clinically relevant correlations (Brunet et al., 2004). Only segmented aCGH data and the 67 primary tumor samples were analyzed. Briefly, the aCGH data set was first dimension reduced by eliminating redundant probes, defined as two or more probes showing identical segmented values in all samples. In this manner, 16,084 mapped probes were reduced to 942 genomic regions of unique segmented values. The reduced aCGH data (67 samples by 942 regions) were converted to nonnegative values by assigning two dimensions to each of the regions: a "gain" dimension for $\log_2$ ratios greater than zero, and a "loss" dimension for the absolute value of $\log_2$ ratios less than zero. The resultant data set is a nonnegative matrix of dimension 67×1884, which is subject to NMF using the software package (Brunet et al., 2004) and run in MATLAB (The MathWorks, Inc.). For each factor level two through six, NMF is repeated 1000 times to build a consensus matrix, and this is used to assign samples to clusters based on the most common consensus. As a measure of stability, the cluster assignments were repeated with a random 10% of the samples excluded on each iteration, and a nearly identical clustering was observed.

Fisher's exact test was used to identify genomic regions presenting significantly different occurrence of copy number gains or losses between k1 and

k2 groups. Briefly, for each sample, each of the 942 genomic regions was classified as having copy number normal, gained, or lost based on $\log_2$ ratio thresholds of ±0.13. Then, a 2 × 2 contingency table was tested for k1 versus k2 samples gained versus normal; a second matrix tested lost versus normal. Fisher's exact test p values were corrected for multiple testing ("qvalue" function, R package "qvalue," http://cran.r-project.org/).

Expression profiling on Affymetrix GeneChip

Detailed protocols for RNA purification, cDNA synthesis, cRNA preparation, and hybridization to Affymetrix H133A and H133Plus2.0 GeneChip microarrays were performed as described (Zhan et al., 2003). As control, RNA was derived from CD138-selected bone marrow-derived plasma cells from 12 healthy donors, as described (Zhan et al., 2003). The genomic positions for each gene were mapped based on NCBI Build 35 of human genome. SAM was performed as described (Tusher et al., 2001). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE4452.

Integrated copy number and expression analysis

For each gene probe set falling within an amplified MCR, two different analyses of gene expression were conducted: (1) expression in tumors with CNA compared to normal plasma cells, and (2) expression in tumors without CNA compared to normal plasma cells. Because of the widely varying proportion of copy number-altered samples in each MCR, SAM was impractical to apply generally to the two measures. Thus, all expression comparisons were instead performed using the same "gene weight" measure, similar to T score, as described below and in previous publications (Aguirre et al., 2004; Hyman et al., 2002). For each gene probe set, gene weight (GW) of expression values for test set "T" compared to reference set "R" is calculated by

$$GW_{T\ versus\ R} = \frac{\overline{T} - \overline{R}}{\sigma^T + \sigma^R}$$

Significance was determined by permuting sample labels for expression data (1000 permutations, p value $\leq$ 0.001).

GSEA

GSEA has been described elsewhere (Mootha et al., 2003). Briefly, the method requires two inputs: (1) a list of genes that have been ranked according to expression difference between two states and (2) a priori defined gene sets (e.g., pathways), each consisting of members drawn from this list. The 95 pathways were derived from Monti et al. (2005) and Mootha et al. (2003). No normalization was applied to the data. The ranking metric used was Signal2Noise, with a weighted ES scoring scheme. The phenotype was permuted with 1000 permutations.

QPCR verification and FISH

PCR primers were designed to amplify products of 100–150 bp within target and control sequences as described (Aguirre et al., 2004). Metaphase spread slides were prepared following standard protocols (Protopopov et al., 1996). The BACs RPC11 HS (ch1, spanning *BCL9*), RP11-180M15 (ch12, spanning *CDKN1B*), and RP11-237C13 were used for hybridizations. The probes for FISH were labeled using nick translation, according to manufacturer's instructions (Roche Molecular Biochemicals) with either biotin-14-dATP or digoxigenin-11-dUTP. Biotinylated probes were detected using Cy3-conjugated avidin (Accurate Chemical). For digoxigenin-labeled probes, antidigoxigenin-FITC Fab fragments (Enzo Life Sciences) were used. Slides were counterstained with 5 μg/ml DAPI (Merck) and mounted in Vectashield antifade medium (Vector Laboratories). FISH signal acquisition and spectral analysis were performed using filter sets and software by Applied Spectral Imaging.

Automated MCR definition

Loci of amplification and deletion are evaluated across samples with an effort to define MCRs targeted by overlapping events in two or more samples.

An algorithmic approach has been previously described (Aguirre et al., 2004; Tonon et al., 2005). It is applied to the segmented data as follows:

(1) Segments with values >0.4 or <−0.4 (0.5 and −0.5 for cDNA) are identified as altered.
(2) If two or more similarly altered segments are adjacent in a single profile or separated by <0.5 Mb, the entire region spanned by the segments is considered to be an altered span.
(3) Altered segments or spans <20 Mb are retained as "informative spans" for defining discrete locus boundaries. Longer regions are not discarded, but are not included in defining locus boundaries.
(4) Informative spans are compared across samples to identify overlapping amplified or deleted regions (informative spans only); each is called an "overlap group."
(5) Overlap groups are divided into separate groups wherever the recurrence rate falls <25% of the peak recurrence for the whole group. Recurrence is calculated by counting the number of samples with alteration at high threshold (±0.4, or 0.5 and −0.5 for cDNA).
(6) MCRs are defined as contiguous spans within an overlap group, having at least 75% of the peak recurrence. If there are more than three MCRs in a locus, the whole region is reported as a single complex MCR. In cases where MCRs are defined by two overlapping CNAs, MCR inclusion in the final list and boundary definition is subjected to individual review.

Identification of prognostic MCRs

As described above for NMF analysis, for each unique genomic region in the aCGH (segmented data), the samples are divided into those that have copy gain versus the rest, and those that have copy loss versus the rest. This is based on a low threshold of $\log_2 = 0.13$. For each genomic region, KM survival was calculated for each pair of altered/unaltered groups. The survival curves were tested for significant difference via log rank test (survdiff function, Survival package, http://cran.r-project.org/). The regions showing a p value of $\leq$0.05 were then mapped to the MCRs.

Supplemental data

The Supplemental Data include Supplemental Experimental Procedures, five supplemental figures, and three supplemental tables and can be found with this article online at http://www.cancercell.org/cgi/content/full/9/4/313/DC1/.

Acknowledgments

We wish to thank John Quackenbush for critical reading and advice. This work was supported by The Fund to Cure Myeloma; NCI grants RO1 CA93947, RO1 CA099041, and P50 CA93683 to L.C.; and the Robert A. and Renee E. Belfer Foundation Institute for Innovative Science at the DFCI. R.A.D. is an ACS Research Professor and Ellison Medical Foundation Senior Scholar. D.R.C. is a supported by a Mentored Clinician Scientist Award (K08AG0103) and is a Sidney Kimmel Foundation Scholar. G.T. is supported by a SPORE MM Career Development Award. Additional support was provided by grants CA84313 (R.A.D.), CA55819 (J.D.S. and B.B.), CA55819 (J.D.S.), PO1 CA78373 (K.C.A.), and SPORE P50 CA100707 (K.C.A.).

Received: November 21, 2005
Revised: February 1, 2006
Accepted: March 20, 2006
Published: April 10, 2006

References

Abdelhaleem, M. (2004). Do human RNA helicases have a role in cancer? Biochim. Biophys. Acta *1704*, 37–46.

Aguirre, A.J., Brennan, C., Bailey, G., Sinha, R., Feng, B., Leo, C., Zhang, Y., Zhang, J., Gans, J.D., Bardeesy, N., et al. (2004). High-resolution characterization of the pancreatic adenocarcinoma genome. Proc. Natl. Acad. Sci. USA *101*, 9067–9072.

Avet-Loiseau, H., Andree-Ashley, L.E., Moore, D., II, Mellerin, M.P., Feusner, J., Bataille, R., and Pallavicini, M.G. (1997). Molecular cytogenetic abnormalities in multiple myeloma and plasma cell leukemia measured using comparative genomic hybridization. Genes Chromosomes Cancer *19*, 124–133.

Barlogie, B., Shaughnessy, J., Tricot, G., Jacobson, J., Zangari, M., Anaissie, E., Walker, R., and Crowley, J. (2004). Treatment of multiple myeloma. Blood *103*, 20–32.

Barlogie, B., Tricot, G., Anaissie, E., Shaughnessy, J., Rasmussen, E., van Rhee, F., Fassas, A., Zangari, M., Hollmig, K., Pineda-Roman, M., et al. (2006). Thalidomide and hematopoietic-cell transplantation for multiple myeloma. N. Engl. J. Med. *354*, 1021–1030.

Bergsagel, P.L., and Kuehl, W.M. (2001). Chromosome translocations in multiple myeloma. Oncogene *20*, 5611–5622.

Bergsagel, P.L., and Kuehl, W.M. (2005). Molecular pathogenesis and a consequent classification of multiple myeloma. J. Clin. Oncol. *23*, 6333–6338.

Bergsagel, P.L., Kuehl, W.M., Zhan, F., Sawyer, J., Barlogie, B., and Shaughnessy, J., Jr. (2005). Cyclin D dysregulation: an early and unifying pathogenic event in multiple myeloma. Blood *106*, 296–303.

Brennan, C., Zhang, Y., Leo, C., Feng, B., Cauwels, C., Aguirre, A.J., Kim, M., Protopopov, A., and Chin, L. (2004). High-resolution global profiling of genomic alterations with long oligonucleotide microarray. Cancer Res. *64*, 4744–4748.

Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. USA *101*, 4164–4169.

Chang, H., Qi, X.Y., Claudio, J., Zhuang, L., Patterson, B., and Stewart, A.K. (2006). Analysis of PTEN deletions and mutations in multiple myeloma. Leuk. Res. *30*, 262–265.

Cheng, K.W., Lahad, J.P., Kuo, W.L., Lapuk, A., Yamada, K., Auersperg, N., Liu, J., Smith-McCune, K., Lu, K.H., Fishman, D., et al. (2004). The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. Nat. Med. *10*, 1251–1256.

Chng, W.J., Van Wier, S.A., Ahmann, G.J., Winkler, J.M., Jalal, S.M., Bergsagel, P.L., Chesi, M., Trendle, M.C., Oken, M.M., Blood, E., et al. (2005). A validated FISH trisomy index demonstrates the hyperdiploid and nonhyperdiploid dichotomy in MGUS. Blood *106*, 2156–2161.

Cigudosa, J.C., Rao, P.H., Calasanz, M.J., Odero, M.D., Michaeli, J., Jhanwar, S.C., and Chaganti, R.S. (1998). Characterization of nonrandom chromosomal gains and losses in multiple myeloma by comparative genomic hybridization. Blood *91*, 3007–3010.

Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., and Largaespada, D.A. (2005). Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature *436*, 272–276.

Cremer, F.W., Bila, J., Buck, I., Kartal, M., Hose, D., Ittrich, C., Benner, A., Raab, M.S., Theil, A.C., Moos, M., et al. (2005). Delineation of distinct subgroups of multiple myeloma and a model for clonal evolution based on interphase cytogenetics. Genes Chromosomes Cancer *44*, 194–203.

Dai, J., Sultan, S., Taylor, S.S., and Higgins, J.M. (2005). The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. Genes Dev. *19*, 472–488.

Debes-Marun, C.S., Dewald, G.W., Bryant, S., Picken, E., Santana-Davila, R., Gonzalez-Paz, N., Winkler, J.M., Kyle, R.A., Gertz, M.A., Witzig, T.E., et al. (2003). Chromosome abnormalities clustering and its implications for pathogenesis and prognosis in myeloma. Leukemia *17*, 427–436.

Elnenaei, M.O., Hamoudi, R.A., Swansbury, J., Gruszka-Westwood, A.M., Brito-Babapulle, V., Matutes, E., and Catovsky, D. (2003). Delineation of the minimal region of loss at 13q14 in multiple myeloma. Genes Chromosomes Cancer *36*, 99–106.

Filipits, M., Pohl, G., Stranzl, T., Kaufmann, H., Ackermann, J., Gisslinger, H., Greinix, H., Chott, A., and Drach, J. (2003). Low p27Kip1 expression is an independent adverse prognostic factor in patients with multiple myeloma. Clin. Cancer Res. *9*, 820–826.

Fonseca, R., Barlogie, B., Bataille, R., Bastard, C., Bergsagel, P.L., Chesi, M., Davies, F.E., Drach, J., Greipp, P.R., Kirsch, I.R., et al. (2004). Genetics and cytogenetics of multiple myeloma: a workshop report. Cancer Res. *64*, 1546–1558.

Fujita, Y., Sakakura, C., Shimomura, K., Nakanishi, M., Yasuoka, R., Aragane, H., Hagiwara, A., Abe, T., Inazawa, J., and Yamagishi, H. (2003). Chromosome arm 20q gains and other genomic alterations in esophageal squamous cell carcinoma, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization. Hepatogastroenterology *50*, 1857–1863.

Furubo, S., Harada, K., Shimonishi, T., Katayanagi, K., Tsui, W., and Nakanuma, Y. (1999). Protein expression and genetic alterations of p53 and ras in intrahepatic cholangiocarcinoma. Histopathology *35*, 230–240.

He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J., and Hammond, S.M. (2005). A microRNA polycistron as a potential human oncogene. Nature *435*, 828–833.

Hyman, E., Kauraniemi, P., Hautaniemi, S., Wolf, M., Mousses, S., Rozenblum, E., Ringner, M., Sauter, G., Monni, O., Elkahloun, A., et al. (2002). Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res. *62*, 6240–6245.

Kanamori, M., Sandy, P., Marzinotto, S., Benetti, R., Kai, C., Hayashizaki, Y., Schneider, C., and Suzuki, H. (2003). The PDZ protein tax-interacting protein-1 inhibits β-catenin transcriptional activity and growth of colorectal cancer cells. J. Biol. Chem. *278*, 38758–38764.

Kim, S.H., Lessner, S.M., Sakurai, Y., and Galis, Z.S. (2004). Cyclophilin A as a novel biphasic mediator of endothelial activation and dysfunction. Am. J. Pathol. *164*, 1567–1574.

Kuehl, W.M., and Bergsagel, P.L. (2002). Multiple myeloma: evolving genetic events and host interactions. Nat. Rev. Cancer *2*, 175–187.

Kyle, R.A., and Rajkumar, S.V. (2004). Multiple myeloma. N. Engl. J. Med. *351*, 1860–1873.

Mitsiades, C.S., Mitsiades, N., Munshi, N.C., and Anderson, K.C. (2004). Focus on multiple myeloma. Cancer Cell *6*, 439–444.

Monti, S., Savage, K.J., Kutok, J.L., Feuerhake, F., Kurtin, P., Mihm, M., Wu, B., Pasqualucci, L., Neuberg, D., Aguiar, R.C., et al. (2005). Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. Blood *105*, 1851–1861.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

Nakayama, K.I., and Nakayama, K. (2005). Regulation of the cell cycle by SCF-type ubiquitin ligases. Semin. Cell Dev. Biol. *16*, 323–333.

Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557–572.

Panani, A.D., Ferti, A.D., Papaxoinis, C., Raptis, S.A., and Roussos, C. (2004). Cytogenetic data as a prognostic factor in multiple myeloma patients: involvement of 1p12 region an adverse prognostic factor. Anticancer Res. *24*, 4141–4146.

Platzer, P., Upender, M.B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J.K., Mack, D., Ried, T., and Markowitz, S. (2002). Silence of chromosomal amplifications in colon cancer. Cancer Res. *62*, 1134–1138.

Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani, R., Botstein, D., Borresen-Dale, A.L., and Brown, P.O. (2002). Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc. Natl. Acad. Sci. USA *99*, 12963–12968.

Protopopov, A.I., Gizatullin, R.Z., Vorobieva, N.V., Protopopova, M.V., Kiss, C., Kashuba, V.I., Klein, G., Kisselev, L.L., Graphodatsky, A.S., and Zabarovsky, E.R. (1996). Human chromosome 3: high-resolution fluorescence in situ hybridization mapping of 40 unique NotI linking clones homologous to genes and cDNAs. Chromosome Res. *4*, 443–447.

Qiang, Y.W., Walsh, K., Yao, L., Kedei, N., Blumberg, P.M., Rubin, J.S., Shaughnessy, J., Jr., and Rudikoff, S. (2005). Wnts induce migration and invasion of myeloma plasma cells. Blood *106*, 1786–1793.

Rajkumar, S.V., and Kyle, R.A. (2005). Multiple myeloma: diagnosis and treatment. Mayo Clin. Proc. *80*, 1371–1382.

Richardson, P.G., Mitsiades, C.S., Hideshima, T., and Anderson, K.C. (2005). Novel biological therapies for the treatment of multiple myeloma. Best Pract. Res. Clin. Haematol. *18*, 619–634.

Sawyer, J.R., Lukacs, J.L., Munshi, N., Desikan, K.R., Singhal, S., Mehta, J., Siegel, D., Shaughnessy, J., and Barlogie, B. (1998). Identification of new nonrandom translocations in multiple myeloma with multicolor spectral karyotyping. Blood *92*, 4269–4278.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. Science *305*, 525–528.

Shaughnessy, J.D., Jr., and Barlogie, B. (2003). Interpreting the molecular biology and clinical behavior of multiple myeloma in the context of global gene expression profiling. Immunol. Rev. *194*, 140–163.

Shaughnessy, J., Jacobson, J., Sawyer, J., McCoy, J., Fassas, A., Zhan, F., Bumm, K., Epstein, J., Anaissie, E., Jagannath, S., et al. (2003). Continuous absence of metaphase-defined cytogenetic abnormalities, especially of chromosome 13 and hypodiploidy, ensures long-term survival in multiple myeloma treated with Total Therapy I: interpretation in the context of global gene expression. Blood *101*, 3849–3856.

Stepanova, L., Finegold, M., DeMayo, F., Schmidt, E.V., and Harper, J.W. (2000). The oncoprotein kinase chaperone CDC37 functions as an oncogene in mice and collaborates with both c-myc and cyclin D1 in transformation of multiple tissues. Mol. Cell. Biol. *20*, 4462–4473.

Tonon, G., Wong, K.K., Maulik, G., Brennan, C., Feng, B., Zhang, Y., Khatry, D.B., Protopopov, A., You, M.J., Aguirre, A.J., et al. (2005). High-resolution genomic profiles of human lung cancer. Proc. Natl. Acad. Sci. USA *102*, 9625–9630.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

Urashima, M., Teoh, G., Ogata, A., Chauhan, D., Treon, S.P., Sugimoto, Y., Kaihara, C., Matsuzaki, M., Hoshi, Y., DeCaprio, J.A., and Anderson, K.C. (1997). Characterization of p16(INK4A) expression in multiple myeloma and plasma cell leukemia. Clin. Cancer Res. *3*, 2173–2179.

van Duin, M., van Marion, R., Watson, J.E., Paris, P.L., Lapuk, A., Brown, N., Oseroff, V.V., Albertson, D.G., Pinkel, D., de Jong, P., et al. (2005). Construction and application of a full-coverage, high-resolution, human chromosome 8q genomic microarray for comparative genomic hybridization. Cytometry A *63*, 10–19.

Wullich, B., Riedinger, S., Brinck, U., Stoeckle, M., Kamradt, J., Ketter, R., and Jung, V. (2004). Evidence for gains at 15q and 20q in brain metastases of prostate cancer. Cancer Genet. Cytogenet. *154*, 119–123.

Zhan, F., Tian, E., Bumm, K., Smith, R., Barlogie, B., and Shaughnessy, J., Jr. (2003). Gene expression profiling of human plasma cell differentiation and classification of multiple myeloma based on similarities to distinct stages of late-stage B-cell development. Blood *101*, 1128–1140.

Accession numbers

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE4452.