HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2013 January 31.

Published in final edited form as:

Nature. 2011 March 24; 471(7339): 467–472. doi:10.1038/nature09837.

Initial genome sequencing and analysis of multiple myeloma

Michael A. Chapman[1], **Michael S. Lawrence**[1], **Jonathan J. Keats**[2,3], **Kristian Cibulskis**[1], **Carrie Sougnez**[1], **Anna C. Schinzel**[4], **Christina L. Harview**[1], **Jean-Philippe Brunet**[1], **Gregory J. Ahmann**[2,3], **Mazhar Adli**[1,5], **Kenneth C. Anderson**[3,4], **Kristin G. Ardlie**[1], **Daniel Auclair**[3,6], **Angela Baker**[7], **P. Leif Bergsagel**[2,3], **Bradley E. Bernstein**[1,5,8,9], **Yotam Drier**[1,10], **Rafael Fonseca**[2,3], **Stacey B. Gabriel**[1], **Craig C. Hofmeister**[3,11], **Sundar Jagannath**[3,12], **Andrzej J. Jakubowiak**[3,13], **Amrita Krishnan**[3,14], **Joan Levy**[3,6], **Ted Liefeld**[1], **Sagar Lonial**[3,15], **Scott Mahan**[1], **Bunmi Mfuko**[3,6], **Stefano Monti**[1], **Louise M. Perkins**[3,6], **Robb Onofrio**[1], **Trevor J. Pugh**[1], **S. Vincent Rajkumar**[3,16], **Alex H. Ramos**[1], **David S. Siegel**[3,17], **Andrey Sivachenko**[1], **Suzanne Trudel**[3,18], **Ravi Vij**[3,19], **Douglas Voet**[1], **Wendy Winckler**[1], **Todd Zimmerman**[3,20], **John Carpten**[7], **Jeff Trent**[7], **William C. Hahn**[1,4,8], **Levi A. Garraway**[1,4], **Matthew Meyerson**[1,4,8], **Eric S. Lander**[1,8,21], **Gad Getz**[1,*], and **Todd R. Golub**[1,4,8,9,*]

[1]The Eli and Edythe L. Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02412, USA

[2]Mayo Clinic Arizona, 13400 East Shea Boulevard, Scottsdale, Arizona 85259, USA

[3]On behalf of the Multiple Myeloma Research Consortium, 383 Main Avenue, 5th Floor, Norwalk, Connecticut 06581, USA

[4]Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

[5]Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02114, USA

[6]Multiple Myeloma Research Foundation, 383 Main Avenue, 5th Floor, Norwalk, Connecticut 06581, USA

[7]The Translational Genomics Research Institute, 445 North Fifth Street, Phoenix, Arizona 85004, USA

[8]Harvard Medical School, Boston, Massachusetts 02115, USA

[9]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA

*To whom correspondence should be addressed: Correspondence and requests for materials should be addressed to T.R.G. (golub@broadinstitute.org) or G.G. (gadgetz@broadinstitute.org).

Author Information: The sequence data described here will be available from the dbGaP repository (http://www.ncbi.nlm.nih.gov/gap). We have also created a MM Genomics Portal (http://www.broadinstitute.org/mmgp) to support data analysis and visualization.

The authors have no competing financial interests to disclose.

Author Contributions: All authors contributed to the final manuscript. K.C.A., R.F., C.C.H., S.J., A.J.J., A.K., T.L., S.L., S.V.R., D.S.S., S.T., R.V., and T.Z. collected data and provided patient materials. J.J.K., C.S., G.J.A., K.G.A., D.A., A.B., P.L.B., S.B.G., J.L., T.L., S.M., B.M., L.M.P., R.O., W.W., and J.C. processed and analyzed genetic material, including RNA/DNA extraction, fingerprinting, genotyping, data management, hybridizations, library preparation, and sequencing.
M.A.C., J.J.K., A.C.S., C.L.H., M.A., and B.E.B. performed experimental work, including PCR, cloning, ChIP analyses, and RNAi experiments. M.A.C., M.S.L., J.J.K., K.C., J-P. B., Y.D., S.M., T.J.P., A.H.R., A.S., D.V., and G.G. performed data analyses. M.A.C., M.S.L., K.C., E.S.L., G.G., and T.R.G. produced the text and figures, including supplementary information. J.C., J.T., W.C.H., L.A.G., M.M., E.S.L., G.G., and T.R.G. provided leadership for the project.

[10]Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel

[11]The Ohio State University Medical Center, 320 West 10th Avenue, Columbus, Ohio 43210, USA

[12]St. Vincent's Comprehensive Cancer Center, 325 West 15th Street, New York, New York 11001, USA

[13]University of Michigan Comprehensive Cancer Center, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109, USA

[14]City of Hope Comprehensive Cancer Center, 1500 East Duarte Road, Duarte, California 91010, USA

[15]Winship Cancer Institute, Emory University, 1365-C Clifton Road, NE Atlanta, Georgia 30322, USA

[16]Mayo Clinic Rochester, 200 1st Street S.W. Rochester, Minnesota, 55905, USA

[17]Hackensack University Medical Center, 30 Prospect Avenue, Hackensack, New Jersey 07601, USA

[18]Princess Margaret Hospital, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada

[19]Washington University School of Medicine, Washington University in St. Louis, 660 South Euclid Avenue, St. Louis, Missouri 63110, USA

[20]University of Chicago Medical Center, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA

[21]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

Abstract

Multiple myeloma is an incurable malignancy of plasma cells, and its pathogenesis is poorly understood. Here we report the massively parallel sequencing of 38 tumor genomes and their comparison to matched normal DNAs. Several new and unexpected oncogenic mechanisms were suggested by the pattern of somatic mutation across the dataset. These include the mutation of genes involved in protein translation (seen in nearly half of the patients), genes involved in histone methylation, and genes involved in blood coagulation. In addition, a broader than anticipated role of NF-κB signaling was suggested by mutations in 11 members of the NF-κB pathway. Of potential immediate clinical relevance, activating mutations of the kinase BRAF were observed in 4% of patients, suggesting the evaluation of BRAF inhibitors in multiple myeloma clinical trials. These results indicate that cancer genome sequencing of large collections of samples will yield new insights into cancer not anticipated by existing knowledge.

Multiple myeloma (MM) is an incurable malignancy of mature B-lymphoid cells, and its pathogenesis is only partially understood. About 40% of cases harbor chromosome translocations resulting in over-expression of genes (including *CCND1*, *CCND3*, *MAF*, *MAFB*, *WHSC1*/*MMSET* and *FGFR3*) *via* their juxtaposition to the immunoglobulin heavy chain (IgH) locus[1]. Other cases exhibit hyperdiploidy. However, these abnormalities are

likely insufficient for malignant transformation because they are also observed in the pre-malignant syndrome known as *monoclonal gammopathy of uncertain significance* (MGUS). Malignant progression events include activation of *MYC*, *FGFR3*, *KRAS* and *NRAS* and activation of the NF-κB pathway[1-3]. More recently, loss-of-function mutations in the histone demethylase *UTX*/*KDM6A* have also been reported[4].

A powerful way to understand the molecular basis of cancer is to sequence either the entire genome or the protein-coding exome, comparing tumor to normal from the same patient in order to identify the acquired somatic mutations. Recent reports have described the sequencing of whole genomes from a single patient[5-9]. While informative, we hypothesized that a larger number of cases would permit the identification of biologically relevant patterns that would not otherwise be evident.

Landscape of MM mutations

We studied 38 MM patients (Supplementary Table 1), performing whole-genome sequencing (WGS) for 23 patients and whole-exome sequencing (WES; assessing 164,687 exons) for 16 patients, with one patient analyzed by both approaches (Supplementary Information). WES is a cost-effective strategy to identify protein-coding mutations, but cannot detect non-coding mutations and rearrangements. We identified tumor-specific mutations by comparing each tumor to its corresponding normal, using a series of algorithms designed to detect point mutations, small insertions/deletions (indels) and other rearrangements (Supplementary Fig. 1). Based on WGS, the frequency of tumor-specific point mutations was 2.9 per million bases, corresponding to approximately 7,450 point mutations per sample across the genome, including an average of 35 amino acid-changing point mutations plus 21 chromosomal rearrangements disrupting protein-coding regions (Supplementary Tables 2 and 3). The mutation-calling algorithm was found to be highly accurate, with a true positive rate of 95% for point mutations (Supplementary text, Supplementary Tables 4 and 5, and Supplementary Fig. 2).

The mutation rate across the genome rate varied greatly depending on base composition, with mutations at CpG dinucleotides occurring 4-fold more commonly than mutations at A or T bases (Supplementary Fig. 3a). In addition, even after correction for base composition, the mutation frequency in coding regions was lower than that observed in intronic and intergenic regions ($p < 1\times10^{-16}$; Supplementary Fig. 3b), potentially owing to negative selective pressure against mutations disrupting coding sequences. There is also a lower mutation rate in intronic regions compared to intergenic regions ($p < 1\times10^{-16}$), which may reflect transcription-coupled repair, as previously suggested[10, 11]. Consistent with this explanation, we observed a lower mutation rate in introns of genes expressed in MM compared to those not expressed (Fig. 1a).

Frequently mutated genes

We next focused on the distribution of somatic, non-silent protein-coding mutations. We estimated statistical significance by comparison to the background distribution of mutations (Supplementary Information). 10 genes showed statistically significant rates of protein-altering mutations (‘significantly mutated genes’) at a False Discovery Rate (FDR) of ≤0.10

(Table 1). To investigate their functional importance, we compared their predicted consequence (based on evolutionary conservation and nature of the amino acid change) to the distribution of all coding mutations. This analysis showed a dramatic skewing of functional importance (FI) scores[12] for the 10 significantly mutated genes ($p = 7.6\times10^{-14}$; Fig. 1b), supporting their biological relevance. Even after RAS and p53 mutations are excluded from the analysis, the skewing remained significant ($p < 0.01$).

We also examined the non-synonymous:synonymous (NS:S) mutation rate for the significantly mutated genes. The expected NS:S ratio was 2.82 ± 0.15, whereas the observed ratio was 39:0 for the significant genes ($p < 0.0001$), further strengthening the case that these genes are likely drivers of the pathogenesis of MM, and are unlikely to simply be passenger mutations.

The significantly mutated genes include three previously reported to have point mutations in MM: *KRAS* and *NRAS* (10 and 9 cases, respectively (50%), $p < 1\times10^{-11}$, $q < 1\times10^{-6}$), and *TP53* (3 cases (8%), $p = 5.1\times10^{-6}$, $q = 0.019$). Interestingly, we identified 2 point mutations (5%, $p = 0.000027$, $q = 0.086$) in *CCND1* (cyclin D1), which has long been recognized as a target of chromosomal translocation in MM, but for which point mutations have not been observed previously in cancer.

The remaining 6 genes have not previously been known to be involved in cancer, and suggest new aspects of the pathogenesis of MM.

Mutations affecting RNA processing and protein homeostasis

A striking finding of this study was the discovery of frequent mutations in genes involved in RNA processing, protein translation and the unfolded protein response. Such mutations were observed in nearly half of the patients.

The *DIS3/RRP44* gene harbored mutations in 4/38 patients (11%, $p = 2.4x10^{-6}$, $q = 0.011$). *DIS3* encodes a highly conserved RNA exonuclease which serves as the catalytic component of the exosome complex involved in regulating the processing and abundance of all RNA species[13, 14]. The four observed mutations occur at highly conserved regions (Fig. 2a) and cluster within the RNB domain facing the enzyme's catalytic pocket (Fig. 2b). Two lines of evidence suggest that the *DIS3* mutations result in loss of function. First, 3 of the 4 tumors with mutations exhibited loss of heterozygosity *via* deletion of the remaining *DIS3* allele. Second, two of the mutations have been functionally characterized in yeast and bacteria, where they result in loss of enzymatic activity leading to the accumulation of their RNA targets[15, 16]. Given that a key role of the exosome is the regulation of the available pool of mRNAs available for translation[17], these results suggest that *DIS3* mutations may dysregulate protein translation as an oncogenic mechanism in MM.

Further support for a role of translational control in the pathogenesis of MM comes from the observation of mutations in the *FAM46C* gene in 5/38 (13%) patients ($p < 1.8\times10^{-10}$, $q = 1\times10^{-6}$). There is no published functional annotation of *FAM46C*, and its sequence lacks obvious homology to known proteins. To gain insight into its cellular role, we examined its pattern of gene expression across 414 MM samples and compared it to the expression of 395

gene sets curated in the *Molecular Signatures Database* (MSigDB), using the GSEA algorithm[18-20]. The expression of *FAM46C* was highly correlated (q = 0.034 after multiple hypothesis correction; Fig. 2c) to the expression of the set of ribosomal proteins, which are known to be tightly co-regulated[21]. Strong correlation with eukaryotic initiation and elongation factors involved in protein translation was similarly observed. While the precise function of FAM46C remains unknown, this striking correlation provides strong evidence that FAM46C is functionally related in some way to the regulation of translation.

Notably, while not statistically significant on their own, we found mutations in 5 other genes related to protein translation, stability and the unfolded protein responses (Supplementary Table 6), further supporting a role of translational control in MM. Of particular interest, two patients had mutations in the unfolded protein response gene *XBP1*. Over-expression of a particular splice form of *XBP1* has been shown to cause a MM-like syndrome in mice, although no role of *XBP1* in the pathogenesis of human MM has been described[22].

Of related interest, mutations of the *LRRK2* gene were observed in 3/38 patients (8%; Supplementary Table 6). *LRRK2* encodes a serine-threonine kinase that phosphorylates translation initiation factor 4E-binding protein (4EBP). *LRRK2* is best known for its role in the predisposition to Parkinson's disease[23, 24]. Parkinson's disease and other neurodegenerative diseases such as Huntington's disease are characterized in part by aberrant unfolded protein responses[25]. Protein homeostasis may be particularly important in MM because of the enormous rate of production immunoglobulins by MM cells[26-28]. The finding is also of clinical significance because of the success of the drug bortezomib (Velcade) that inhibits the proteasome and which shows remarkable activity in MM compared to other tumor types[29].

Together, these results indicate that mutations affecting protein translation and homeostasis are extremely common in MM (at least 16/38 patients; 42%), thereby suggesting that additional therapeutic approaches that target these mechanisms may be worth exploring.

Identical mutations suggest gain-of-function oncogenes

Another way to recognize biologically significant mutations is to search for recurrence of identical mutations indicative of gain-of-function alterations in oncogenes. Two patients had an identical mutation (K123R) in the DNA-binding domain of the interferon regulatory factor IRF4. Interestingly, a recent RNA interference screen in MM showed that IRF4 was required for MM survival, consistent with its role as a putative oncogene[30]. Genotyping for this mutation in 161 additional MM identified two more patients with this mutation. IRF4 is a transcriptional regulator of PRDM1 (BLIMP-1), and two of 38 sequenced patients also exhibited *PRDM1* mutations. PRDM1 is a transcription factor involved in plasma cell differentiation, loss-of-function mutations of which occur in diffuse large B-cell lymphoma[31-35].

Clinically actionable mutations in BRAF

Some mutations deserve attention because of their clinical relevance. One of our 38 patients harboured a BRAF kinase mutation (G469A). While BRAF G469A has not previously been

observed in MM, this precise mutation is known to be activating and oncogenic[36]. We genotyped an additional 161 MM patients for the 12 most common *BRAF* mutations and found mutations in 7 patients (4%). Three of these were K601N and 4 were V600E (the most common BRAF mutation in melanoma[37]). Our finding of common *BRAF* mutations in MM has important clinical implications because such patients may benefit from treatment with BRAF inhibitors, some of which show dramatic clinical activity[38]. Our results also support the observation that inhibitors acting downstream of BRAF (e.g. MEK) may have activity in MM[39].

Gene set mutations: NF-κB pathway

Another approach to identify biologically relevant mutations in MM is to look not at the frequency of mutation of *individual* genes, but rather of *sets* of genes.

We first considered gene sets based on existing insights into the biology of MM. For example, activation of the NF-κB pathway is known in MM, but the basis of such activation is only partially understood [2, 3]. We observed 10 point mutations (p=0.016) and 4 structural rearrangements, affecting 11 NF-κB pathway genes (Supplementary Table 7): *BTRC*, *CARD11*, *CYLD*, *IKBIP*, *IKBKB*, *MAP3K1*, *MAP3K14*, *RIPK4*, *TLR4*, *TNFRSF1A*, and *TRAF3*. Taken together, our findings greatly expand the mechanisms by which NF-κB may be activated in MM.

Gene set mutations: histone modifying enzymes

We next looked for enrichment in mutations in histone-modifying enzymes. This hypothesis arose because of our observation that the homeotic transcription factor HOXA9 was highly expressed in a subset of MM patients, particularly those lacking known IgH translocations (Supplementary Fig. 4a). HOXA9 expression is regulated primarily by histone methyltransferases (HMT) including members of the MLL family. Sensitive RT-PCR analysis showed that HOXA9 was in fact ubiquitously expressed in MM, with most cases exhibiting biallelic expression consistent with dysregulation *via* an upstream HMT event (Supplementary Figs. 4b,c). Accordingly, we looked for mutations in genes known to directly regulate *HOXA9*. We found significant enrichment (p = 0.0024), with mutations in *MLL*, *MLL2*, *MLL3*, *UTX*, *WHSC1*, and *WHSC1L1*.

HOXA9 is normally silenced by histone-3 lysine-27 tri-methylation (H3K27me3) chromatin marks when cells differentiate beyond the hematopoietic stem cell stage[40, 41]. This repressive mark was weak or absent at the *HOXA9* locus in most MM cell lines (Fig. 3a). Moreover, there was inverse correlation between H3K27me3 levels and *HOXA9* expression (Fig. 3b), consistent with HMT dysfunction contributing to aberrant *HOXA9* expression.

To establish the functional significance of *HOXA9* expression in MM cells, we knocked down its expression with 7 shRNAs (Supplementary Fig. 5). In 11/12 MM cell lines, HOXA9-depleted cells exhibited a competitive disadvantage (Fig. 3c and Supplementary Fig. 6).

These experiments suggest that aberrant *HOXA9* expression, caused at least in part by HMT-related genomic events, plays a role in MM and may represent a new therapeutic target. Further supporting a role of *HOXA9* as a MM oncogene, array-based comparative genomic hybridization identified focal amplifications of the *HOXA* locus in 5% of patients (Supplementary Fig. 7).

Discovering new gene set mutations

We next asked whether it would be possible to discover pathways enriched for mutations in the absence of prior knowledge. Accordingly, we examined 616 gene sets in the MSigDB Canonical Pathways database. One top-ranking gene set was of particular interest because it did not relate to genes known to be important in MM. This gene set encodes proteins involved in the formation of the fibrin clot in the blood coagulation cascade. There were 6 mutations in 5/38 patients (16%, q = 0.0054), encoding 5 proteins (Supplementary Table 8). RT-PCR analysis confirmed expression of 4 of the 5 coagulation factors in MM cell lines (Supplementary Fig. 8). The coagulation cascade involves a number of extracellular proteases and their substrates and regulators, but their role in MM has not been suspected. However, thrombin and fibrin have been shown to serve as mitogens in other cell types[42], and have been implicated in metastasis[43]. These observations suggest that coagulation factor mutations should be explored more fully in human cancers.

Mutations in non-coding regions

Analyses of non-coding portions of the genome have not previously been reported in cancer. We focused on non-coding regions with highest regulatory potential (RP). We defined 2.4×10^6 RP regions (Supplementary Fig. 9), averaging 280 base pairs (bp). We then treated these regions as if they were protein-coding genes, subjecting them to the same permutation analysis used for exonic regions.

We identified multiple non-coding regions with high frequencies of mutation which fell into two classes (Table 2 and Supplementary Table 9). The first corresponds to regions of known somatic hypermutation. These have a 1000-fold higher than expected mutation frequency, as expected for post-germinal center B-cells (Supplementary Table 9). These regions comprise immunoglobulin-coding genes and the 5′-UTR of the lymphoid oncogene, *BCL6*, as reported[44]. Interestingly, we also found previously unrecognized mutations in the intergenic region flanking *BCL6* in 5 patients, indicating that somatic hypermutation likely occurs in regions beyond the 5′ UTR and first intron of *BCL6* (Table 2). Whether such non-coding *BCL6* mutations contribute to MM pathogenesis remains to be established.

The second class consisted of 18 non-coding regions with mutation frequencies beyond that expected by chance (q < 0.25) (Table 2 and Supplementary Table 10). Four of the 18 regions flanked genes that also harbored coding mutations. Interestingly, we observed 7 mutations in 5 of 23 patients (22%) within non-coding regions of *BCL7A*, a putative tumor suppressor gene discovered in the B-cell malignancy Burkitt lymphoma[45], and which is also deleted or hypermethylated in cutaneous T-cell lymphomas[46, 47]. The function of BCL7A is unknown, and the effect of its non-coding mutations in MM remains to be established.

Our preliminary analysis of non-coding mutations suggests that non-exonic portions of the genome may represent a previously untapped source of insight into the pathogenesis of cancer.

Discussion

The analysis of MM genomes reveals that mechanisms previously suspected to play a role in the biology of MM (e.g. NF-κB activation and HMT dysfunction) may in fact play broad roles by virtue of mutations in multiple members of these pathways. In addition, potentially new mechanisms of transformation are suggested, including mutations in the RNA exonuclease DIS3 and other genes involved in protein translation and homeostasis. Whether these mutations are unique to MM or are common to other cancers remains to be determined. Furthermore, frequent mutations in the oncogenic kinase BRAF were observed – a finding that has immediate clinical translational implications.

Importantly, the majority of these discoveries could not have been made by sequencing only a single MM genome – the complex patterns of pathway dysregulation required the analysis of multiple genomes. Whole-exome sequencing revealed the substantial majority of the significantly mutated genes. However, we note that half of total protein-coding mutations occurred *via* chromosomal aberrations such as translocations, most of which would not have been discovered by sequencing of the exome alone. Similarly, the recurrent point mutations in non-coding regions would have been missed with sequencing directed only at coding exons.

The analysis described here is preliminary. Additional MM genomes will be required to establish the definitive genomic landscape of the disease and determine accurate estimates of mutation frequency in the disease. The sequence data described here will be available from the dbGaP repository (http://www.ncbi.nlm.nih.gov/gap) and we have created a MM Genomics Portal (http://www.broadinstitute.org/mmgp) to support data analysis and visualization.

Methods Summary

Informed consent from MM patients was obtained in line with the Declaration of Helsinki. DNA was extracted from bone marrow aspirate (tumor) and blood (normal). WGS libraries (370-410 bp inserts) and WES libraries (200-350 bp inserts) were constructed and sequenced on an Illumina GA-II sequencer using 101 and 76 bp paired-end reads, respectively. Sequencing reads were procesed with the Firehose pipeline, identifying somatic point mutations, indels, and other structural chromosomal rearrangements. Structural rearrangements affecting protein-coding regions were then subjected to manual review to exclude alignment artifacts. True positive mutation rates were estimated by Sequenom mass spectrometry genotyping of randomly selected mutations. *HOXA9* shRNAs were introduced into MM cell lines using lentiviral infection using standard methods.

A complete description of the materials and methods are provided in the Supplementary Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was funded by a grant from the Multiple Myeloma Research Foundation. M.C. was supported by a Clinician Scientist Fellowship from Leukaemia and Lymphoma Research (UK). We are grateful to all members of the Broad Institute's Biological Samples Platform, Genetic Analysis Platform, and Genome Sequencing Platform, without whom this work would not have been possible.

References

1. Bergsagel PL, Kuehl WM. Molecular pathogenesis and a consequent classification of multiple myeloma. J Clin Oncol. 2005; 23:6333–8. [PubMed: 16155016]
2. Keats JJ, et al. Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. Cancer Cell. 2007; 12:131–44. [PubMed: 17692805]
3. Annunziata CM, et al. Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. Cancer Cell. 2007; 12:115–30. [PubMed: 17692804]
4. van Haaften G, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet. 2009; 41:521–3. [PubMed: 19330029]
5. Lee W, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature. 2010; 465:473–7. [PubMed: 20505728]
6. Campbell PJ, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet. 2008; 40:722–9. [PubMed: 18438408]
7. Ley TJ, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008; 456:66–72. [PubMed: 18987736]
8. Shah SP, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009; 461:809–13. [PubMed: 19812674]
9. Ding L, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010; 464:999–1005. [PubMed: 20393555]
10. Pleasance ED, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 463:184–90. [PubMed: 20016488]
11. Pleasance ED, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 463:191–6. [PubMed: 20016485]
12. Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. Genome Biol. 2007; 8:R232. [PubMed: 17976239]
13. Dziembowski A, et al. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol. 2007; 14:15–22. [PubMed: 17173052]
14. Schmid M, Jensen TH. The exosome: a multipurpose RNA-decay machine. Trends Biochem Sci. 2008; 33:501–10. [PubMed: 18786828]
15. Schneider C, Anderson JT, Tollervey D. The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. Mol Cell. 2007; 27:324–31. [PubMed: 17643380]
16. Barbas A, et al. Determination of key residues for catalysis and RNA cleavage specificity: one mutation turns RNase II into a SUPER-ENZYME. J Biol Chem. 2009; 284:20486–98. [PubMed: 19458082]
17. Ibrahim H, Wilusz J, Wilusz CJ. RNA recognition by 3′-to-5′ exonucleases: the substrate perspective. Biochim Biophys Acta. 2008; 1779:256–65. [PubMed: 18078842]
18. Zhan F, et al. The molecular classification of multiple myeloma. Blood. 2006; 108:2020–8. [PubMed: 16728703]

19. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–50. [PubMed: 16199517]
20. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34:267–73. [PubMed: 12808457]
21. Tanay A, Regev A, Shamir R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc Natl Acad Sci U S A. 2005; 102:7203–8. [PubMed: 15883364]
22. Carrasco DR, et al. The differentiation and stress response factor XBP-1 drives multiple myeloma pathogenesis. Cancer Cell. 2007; 11:349–60. [PubMed: 17418411]
23. Zimprich A, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron. 2004; 44:601–7. [PubMed: 15541309]
24. Paisan-Ruiz C, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron. 2004; 44:595–600. [PubMed: 15541308]
25. Forman MS, Lee VM, Trojanowski JQ. 'Unfolding' pathways in neurodegenerative disease. Trends Neurosci. 2003; 26:407–10. [PubMed: 12900170]
26. Masciarelli S, et al. CHOP-independent apoptosis and pathway-selective induction of the UPR in developing plasma cells. Mol Immunol. 47:1356–65. [PubMed: 20044139]
27. Cenci S, Sitia R. Managing and exploiting stress in the antibody factory. FEBS Lett. 2007; 581:3652–7. [PubMed: 17475256]
28. Todd DJ, Lee AH, Glimcher LH. The endoplasmic reticulum stress response in immunity and autoimmunity. Nat Rev Immunol. 2008; 8:663–74. [PubMed: 18670423]
29. Adams J. The development of proteasome inhibitors as anticancer drugs. Cancer Cell. 2004; 5:417–21. [PubMed: 15144949]
30. Shaffer AL, et al. IRF4 addiction in multiple myeloma. Nature. 2008; 454:226–31. [PubMed: 18568025]
31. Mandelbaum J, et al. BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. Cancer Cell. 2010; 18:568–79. [PubMed: 21156281]
32. Pasqualucci L, et al. Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. J Exp Med. 2006; 203:311–7. [PubMed: 16492805]
33. Shaffer AL, et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. Immunity. 2002; 17:51–62. [PubMed: 12150891]
34. Shapiro-Shelef M, et al. Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. Immunity. 2003; 19:607–20. [PubMed: 14563324]
35. Turner CA Jr, Mack DH, Davis MM. Blimp-1 a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. Cell. 1994; 77:297–306. [PubMed: 8168136]
36. Wan PT, et al. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell. 2004; 116:855–67. [PubMed: 15035987]
37. Davies H, et al. Mutations of the BRAF gene in human cancer. Nature. 2002; 417:949–54. [PubMed: 12068308]
38. Flaherty, K.; Puzanov, I.; Sosman, J.; Kim, K.; Ribas, A.; McArthur, G.; Lee, RJ.; Grippo, JF.; Nolop, K.; Chapman, P. 2009 ASCO Meeting. American Society of Clinical Oncology; 2009. Phase I study of PLX4032: Proof of concept for V600E BRAF mutation as a therapeutic target in human cancer.
39. Kim K, et al. Blockade of the MEK/ERK signalling cascade by AS703026, a novel selective MEK1/2 inhibitor, induces pleiotropic anti-myeloma activity in vitro and in vivo. Br J Haematol. 2010; 149:537–49. [PubMed: 20331454]
40. Lessard J, Sauvageau G. Polycomb group genes as epigenetic regulators of normal and leukemic hemopoiesis. Exp Hematol. 2003; 31:567–85. [PubMed: 12842702]
41. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. Cell. 2007; 128:669–81. [PubMed: 17320505]

42. Ruf W, Mueller BM. Thrombin generation and the pathogenesis of cancer. Semin Thromb Hemost. 2006; 32(1):61–8. [PubMed: 16673267]

43. Esumi N, Fan D, Fidler IJ. Inhibition of murine melanoma experimental metastasis by recombinant desulfatohirudin, a highly specific thrombin inhibitor. Cancer Res. 1991; 51:4549–56. [PubMed: 1873799]

44. Migliazza A, et al. Frequent somatic hypermutation of the 5′ noncoding region of the BCL6 gene in B-cell lymphoma. Proc Natl Acad Sci U S A. 1995; 92:12520–4. [PubMed: 8618933]

45. Zani VJ, et al. Molecular cloning of complex chromosomal translocation t(8;14;12)(q24.1;q32.3;q24.1) in a Burkitt lymphoma cell line defines a new gene (BCL7A) with homology to caldesmon. Blood. 1996; 87:3124–34. [PubMed: 8605326]

46. Zhang W, et al. Unravelling the hidden heterogeneities of diffuse large B-cell lymphoma based on coupled two-way clustering. BMC Genomics. 2007; 8:332. [PubMed: 17888167]

47. Carbone A, et al. Array-based comparative genomic hybridization in early-stage mycosis fungoides: recurrent deletion of tumor suppressor genes BCL7A, SMAC/DIABLO, and RHOF. Genes Chromosomes Cancer. 2008; 47:1067–75. [PubMed: 18663754]

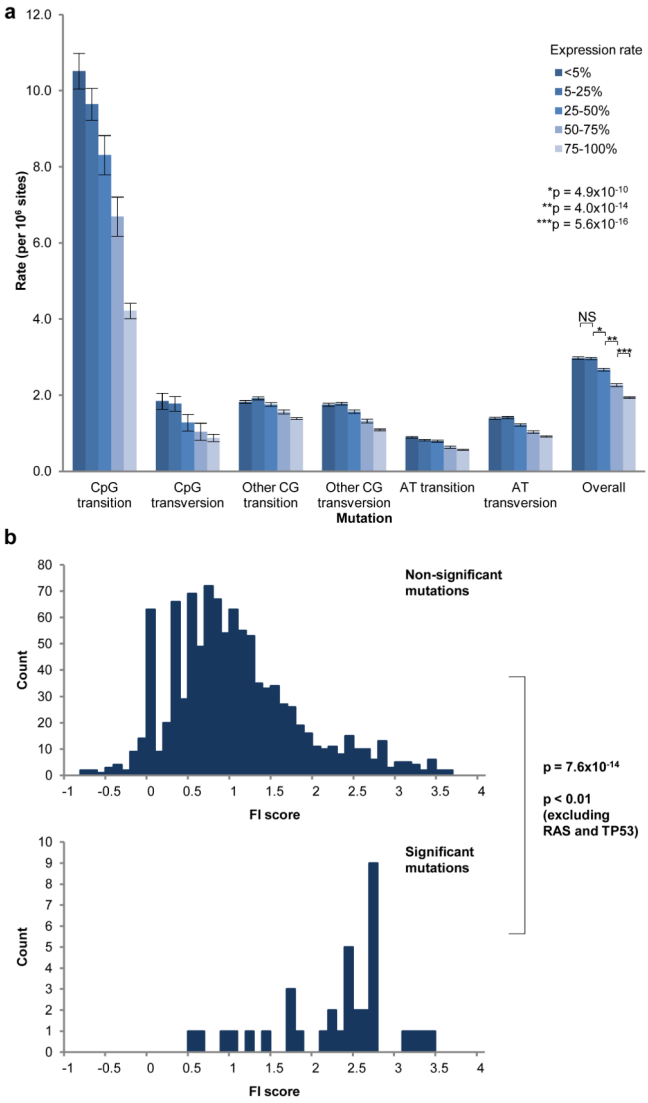

Figure 1. Evidence for transcription-coupled repair and functional importance (FI) of statistically significant mutations

(a) Intronic mutation rates subdivided by gene expression rates in MM. Rates of gene expression were estimated by proportion of Affymetrix Present (P) calls in 304 primary MM samples. Error bars indicate standard deviation. (b) FI scores were generated for all point mutations and divided into distributions for non-significant mutations (upper histogram) and significant mutations (lower). Comparison of distributions is *via* the Kolmogorov-Smirnov statistic.

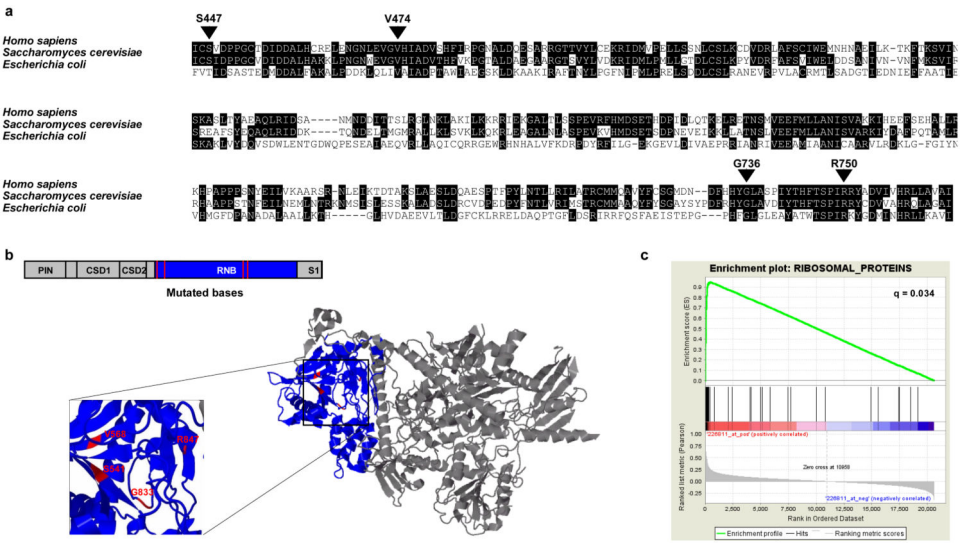

Figure 2. Mutations likely to affect protein translation and/or homeostasis in MM

(a) Alignment of human, yeast, and bacterial RNB domain of DIS3. Positions of observed mutations are indicated with respect to the human sequence. Yeast equivalents are, respectively, S541, V568, G833, and R847. (b) 2D and 3D structures of yeast DIS3, with RNB domain colored in blue and mutations colored in red. (c) GSEA plot showing enrichment of ribosomal protein gene set amongst genes correlated with *FAM46C* expression in 414 MM samples.

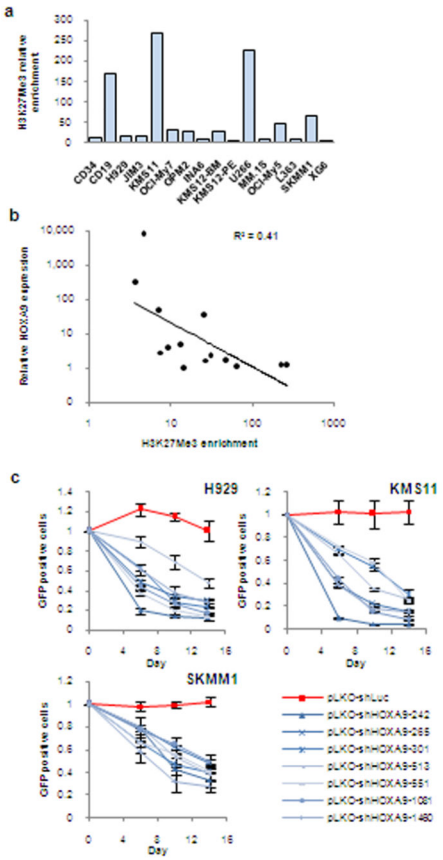

Figure 3. *HOXA9*is a candidate oncogene in MM

(a) H3K27Me3 enrichment at the *HOXA9* promoter in CD34 cells, CD19 cells, and MM cell lines relative to H3K27Me3 methylation at the BC site, known to be hypomethylated in all cells. (b) Relative *HOXA9* expression vs. H3K27Me3 enrichment at the *HOXA9* locus. (c) GFP competition assay in MM cell lines. Following lentiviral infection with seven *HOXA9* shRNAs or a control shRNA targeting luciferase, GFP-positive cells were monitored by flow cytometry and compared to the proportion of GFP-positive cells present in the population 3 days post-infection (designated day 0). Error bars indicate standard error of the mean and represent a minimum of 3 independent experiments.

Table 1

Statistically significant protein-coding mutations in MM

Territory (N) refers to total covered territory in bp across 38 sequenced samples. Total numbers of mutations (n) and numbers of mutations occurring in therapy-naïve disease (Untreated n) are shown for each gene

Gene	N	n	Untreated n	CpG transition	Other C:G transition	C:G transversion	A:T mutation	Indel/null	p-value	q-value
NRAS	20711	9	3	0	0	3	6	0	$<1.0\times10^{-11}$	$<1.0\times10^{-6}$
KRAS	25728	10	6	0	5	1	4	0	$<1.0\times10^{-11}$	$<1.0\times10^{-6}$
FAM46C	39661	5	3	0	0	2	1	2	1.8×10^{-10}	1.0×10^{-6}
DIS3	89758	4	1	0	1	1	2	0	2.4×10^{-6}	0.011
TP53	32585	3	1	0	0	1	1	1	5.1×10^{-6}	0.019
CCND1	12899	2	1	0	0	0	2	0	0.000027	0.086
PNRC1	19621	2	2	0	1	0	0	1	0.000039	0.094
ALOX12B	40369	3	0	1	0	1	1	0	0.000042	0.094
HLA-A	18635	2	0	0	0	0	2	0	0.000045	0.094
MAGED1	53950	2	1	0	0	0	0	2	0.000053	0.10

Table 2

Statistically significant mutated non-coding regions

Regions of predicted regulatory potential showing mutation frequency beyond that expected by chance are shown ($q < 0.25$).

Chr	Start	End	Length	Muts	Samples	p-value	q-value	Separation	Gene	Coding events
1	554350	555310	960	3	3	3.86×10^{-6}	0.11	494, 44	AK125248 (intron)	
1	82793220	82793300	80	2	2	8.39×10^{-6}	0.19	8	TTLL7/LPHN2 (IGR)	
1	147333070	147335140	2070	4	3	2.47×10^{-6}	0.09	350, 1, 85	NBPFA (intron)	
2	40865560	40865630	70	2	2	4.99×10^{-6}	0.14	2	SLC8A1/PKDCC (IGR)	
3	149273920	149274010	90	2	2	4.80×10^{-6}	0.14	78	ZIC4/AGTR1 (IGR)	
3	189142550	189143600	1050	8	5	5.55×10^{-14}	3.9×10^{-8}	298, 8, 17, 26, 26, 80, 1	BCL6/LPP (IGR)	
3	189440810	189441310	500	3	3	2.64×10^{-6}	0.09	1, 291	LPP (intron)	
4	7819430	7819530	100	2	2	8.01×10^{-6}	0.18	26	AFAP1 (intron)	Missense mutation
4	39875900	39876610	710	3	2	5.88×10^{-6}	0.16	109, 412	RHOH (intron)	
4	62180540	62181370	830	3	3	1.05×10^{-5}	0.22	211, 432	LPHN3 (intron)	
4	157902080	157904460	2380	4	4	6.95×10^{-6}	0.17	996, 423, 443	PDGFC (3′UTR/intron)	
7	92754250	92754270	20	2	2	2.03×10^{-7}	0.02	1	CCDC132 (intron)	
9	16564360	16565100	740	3	2	8.65×10^{-6}	0.19	250, 76	BNC2 (intron)	
12	120943010	120943460	450	3	3	6.99×10^{-7}	0.04	17, 9	BCL7A (promoter)	
12	120943580	120946950	3370	4	3	1.47×10^{-8}	0.0017	2055, 657, 295	BCL7A (promoter/intron)	
14	68327320	68333190	5870	4	4	7.05×10^{-6}	0.17	397, 156, 35	ZFP36L1 (intron)	Indel
17	8106910	8111850	4940	4	2	4.85×10^{-6}	0.14	1483, 389, 83	PFAS (intron)	Complex rearrangement
20	60328960	60329510	550	2	2	1.42×10^{-6}	0.06	120	LAMA5 (intron)	Missense mutation