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], A. Keith Stewart[2,3], 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] & Todd R. Golub[1,4,8,9]

Multiple myeloma is an incurable malignancy of plasma cells, and its pathogenesis is poorly understood. Here we report the massively parallel sequencing of 38 tumour genomes and their comparison to matched normal DNAs. Several new and unexpected oncogenic mechanisms were suggested by the pattern of somatic mutation across the data set. 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 signalling was indicated 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 is an incurable malignancy of mature B-lymphoid cells, and its pathogenesis is only partially understood. About 40% of cases harbour chromosome translocations resulting in overexpression of genes (including *CCND1*, *CCND3*, *MAF*, *MAFB*, *WHSC1* (also called *MMSET*) and *FGFR3*) via their juxtaposition to the immunoglobulin heavy chain (IgH) locus[1]. Other cases exhibit hyperdiploidy. However, these abnormalities are probably insufficient for malignant transformation because they are also observed in the pre-malignant syndrome known as monoclonal gammopathy of uncertain significance. 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* (also called *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 tumour to normal from the same patient to identify the acquired somatic mutations. Recent reports have described the sequencing of whole genomes from a single patient[5–9]. Although 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 multiple myeloma mutations

We studied 38 multiple myeloma 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 analysed 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 tumour-specific mutations by comparing each tumour to its corresponding normal, using a series of algorithms designed to detect point mutations, small insertions/deletions (indels) and other rearrangements (Supplementary Fig. 1). On the basis of WGS, the frequency of tumour-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 varied greatly depending on base composition, with mutations at CpG dinucleotides occurring fourfold 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 ($P < 1 \times 10^{-16}$; Supplementary Fig. 3b), potentially owing to negative selective pressure against mutations disrupting coding sequences. There is also a

[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]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. [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 SW Rochester, Minnesota 55905, USA. [17]John Theurer Cancer Center at Hackensack University Medical Center, 92 Second Street, 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]Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. †Present address: Department of Haematology, Addenbrookes Hospital, Hills Road, Cambridge CB2 0QQ, UK.

Foresight EX1103
Foresight v Personalis
IPR2024-00170

©2011 Macmillan Publishers Limited. All rights reserved

lower mutation rate in intronic regions compared to intergenic regions ($P < 1 \times 10^{-16}$), 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 multiple myeloma 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). Ten 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 (on the basis of 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 ten significantly mutated genes ($P = 7.6 \times 10^{-14}$; 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 probably drivers of the pathogenesis of multiple myeloma, and are unlikely to simply be passenger mutations.

The significantly mutated genes include three previously reported to have point mutations in multiple myeloma: *KRAS* and *NRAS* (10 and 9 cases, respectively (50%), $P < 1 \times 10^{-11}$, $q < 1 \times 10^{-6}$), and *TP53* (3 cases (8%), $P = 5.1 \times 10^{-6}$, $q = 0.019$). Interestingly, we identified two 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 multiple myeloma, but for which point mutations have not been observed previously in cancer.

The remaining six genes have not previously been known to be involved in cancer, and indicate new aspects of the pathogenesis of multiple myeloma.

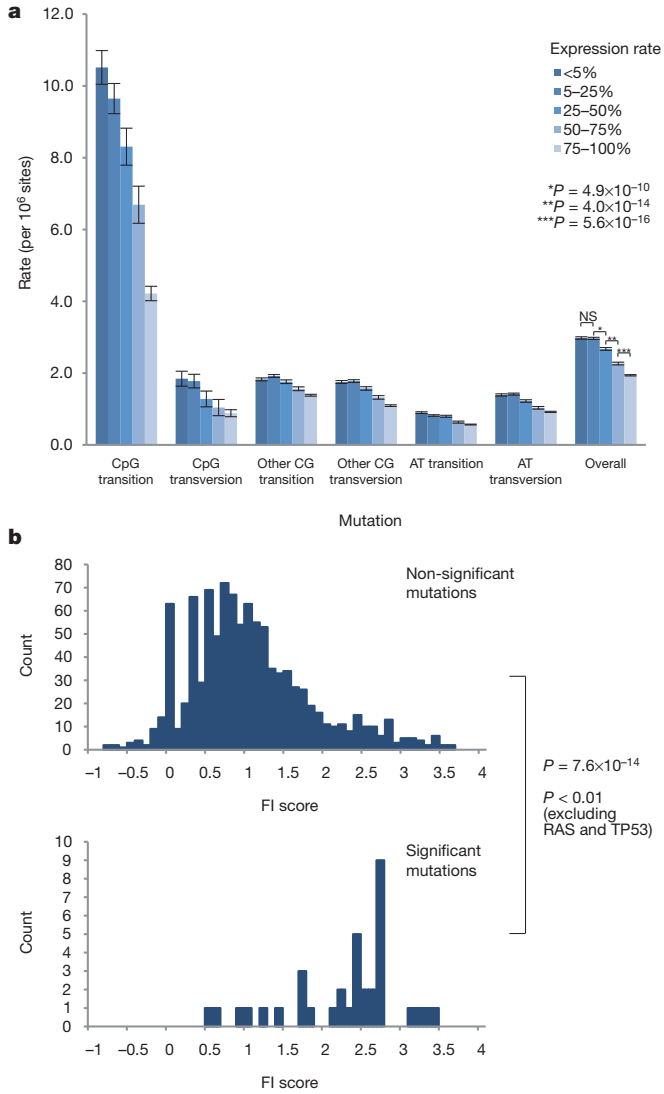

Figure 1 | Evidence for transcription-coupled repair and functional importance of statistically significant mutations. **a**, Intronic mutation rates subdivided by gene expression rates in multiple myeloma. Rates of gene expression were estimated by proportion of Affymetrix Present (P) calls in 304 primary multiple myeloma samples. Error bars indicate ±1 standard deviation. NS, not significant. **b**, Functional importance (FI) scores were generated for all point mutations and divided into distributions for nonsignificant mutations (top histogram; $n = 1{,}019$) and significant mutations (bottom; $n = 36$). Comparison of distributions is via the Kolmogorov–Smirnov statistic.

RNA processing and protein homeostasis mutations

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* (also called *RRP44*) gene harboured mutations in 4 out of 38 patients (11%, $P = 2.4 \times 10^{-6}$, $q = 0.011$). *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 indicate that the *DIS3* mutations result in loss of function. First, three of the four tumours 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 indicate that *DIS3* mutations may dysregulate protein translation as an oncogenic mechanism in multiple myeloma.

Further support for a role of translational control in the pathogenesis of multiple myeloma comes from the observation of mutations in the *FAM46C* gene in 5 out of 38 (13%) patients ($P = 1.8 \times 10^{-10}$, $q = 1 \times 10^{-6}$). 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 multiple myeloma 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 that are known to be tightly co-regulated[21]. Strong correlation with eukaryotic initiation and elongation factors involved in protein translation was similarly observed. Although 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. Consistent with this observation, FAM46C was recently shown to function as an mRNA stability factor (M. Fleming, manuscript submitted).

Foresight EX1103
Foresight v Personalis
IPR2024-00170

©2011 Macmillan Publishers Limited. All rights reserved

Table 1 | Statistically significant protein-coding mutations in multiple myeloma

Gene	*N*	*n*	Untreated *n*	CpG transition	Other C:G transition	C:G transversion	A:T mutation	Indel/ null	*P*-value	*q*-value
NRAS	20,711	9	3	0	0	3	6	0	$<1.0 \times 10^{-11}$	$<1.0 \times 10^{-6}$
KRAS	25,728	10	6	0	5	1	4	0	$<1.0 \times 10^{-11}$	$<1.0 \times 10^{-6}$
FAM46C	39,661	5	3	0	0	2	1	2	1.8×10^{-10}	1.0×10^{-6}
DIS3	89,758	4	1	0	1	1	2	0	2.4×10^{-6}	0.011
TP53	32,585	3	1	0	0	1	1	1	5.1×10^{-6}	0.019
CCND1	12,899	2	1	0	0	0	2	0	0.000027	0.086
PNRC1	19,621	2	2	0	1	0	0	1	0.000039	0.094
ALOX12B	40,369	3	0	1	0	1	1	0	0.000042	0.094
HLA-A	18,635	2	0	0	0	0	2	0	0.000045	0.094
MAGED1	53,950	2	1	0	0	0	0	2	0.000053	0.10

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

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

Of related interest, mutations of the *LRRK2* gene were observed in 3 out of 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 multiple myeloma because of the enormous rate of production of immunoglobulins by multiple myeloma cells[26–28]. The finding is also of clinical significance because of the success of the drug bortezomib (Velcade), which inhibits the proteasome and which shows remarkable activity in multiple myeloma compared to other tumour types[29].

Together, these results indicate that mutations affecting protein translation and homeostasis are extremely common in multiple myeloma (at least 16 out of 38 patients; 42%), thereby indicting 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 multiple myeloma showed that IRF4 was required for multiple myeloma survival, consistent with its role as a putative oncogene[30]. Genotyping for

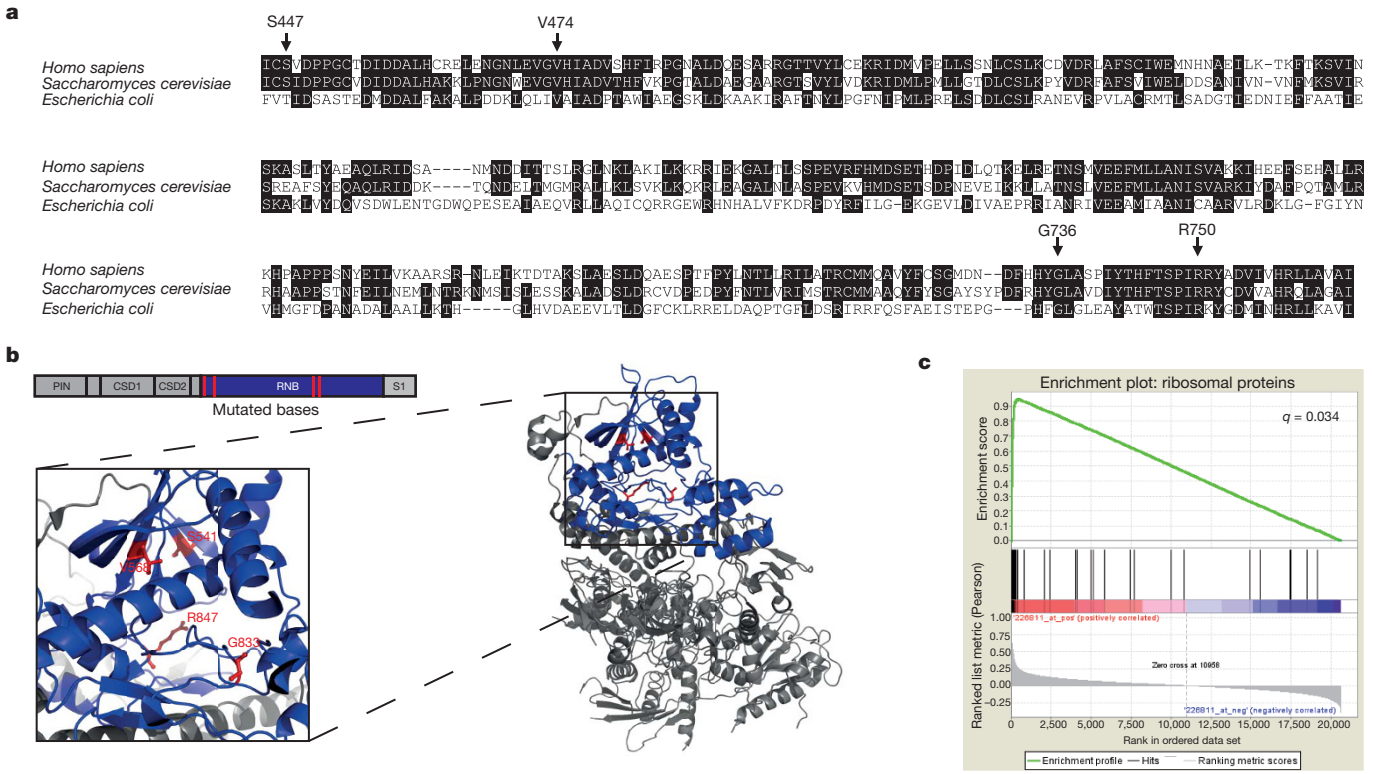

Figure 2 | Mutations likely to affect protein translation and/or homeostasis in multiple myeloma. **a**, Alignment of human, yeast and bacterial RNB domain of DIS3. Positions of observed mutations are indicated with respect to the human sequence. Yeast equivalents are, respectively, S541, V568, G833 and R847. **b**, One-dimensional and three-dimensional structures of yeast DIS3, with the RNB domain coloured in blue and mutations coloured in red. **c**, GSEA plot showing enrichment of ribosomal protein gene set among genes correlated with *FAM46C* expression in 414 multiple myeloma samples.

Foresight EX1103
Foresight v Personalis
IPR2024-00170

©2011 Macmillan Publishers Limited. All rights reserved

this mutation in 161 additional multiple myeloma samples identified two more patients with this mutation. IRF4 is a transcriptional regulator of PRDM1 (also called BLIMP1), 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 the thirty-eight patients harboured a BRAF kinase mutation (G469A). Although BRAF G469A has not previously been observed in multiple myeloma, this precise mutation is known to be activating and oncogenic[36]. We genotyped an additional 161 multiple myeloma patients for the 12 most common *BRAF* mutations and found mutations in 7 patients (4%). Three of these were K601N and four were V600E (the most common BRAF mutation in melanoma[37]). Our finding of common *BRAF* mutations in multiple myeloma has important clinical implications because such patients may benefit from treatment with BRAF inhibitors, some of which show marked clinical activity[38]. Our results also support the observation that inhibitors acting downstream of BRAF (for example, on MEK) may have activity in multiple myeloma[39].

Gene set mutations: NF-κB pathway

Another approach to identify biologically relevant mutations in multiple myeloma 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 multiple myeloma. For example, activation of the NF-κB pathway is known in multiple myeloma, 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 multiple myeloma.

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 multiple myeloma 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 polymerase chain reaction with reverse transcription (RT–PCR) analysis showed that *HOXA9* was in fact ubiquitously expressed in multiple myeloma, with most cases exhibiting biallelic expression consistent with dysregulation via an upstream HMT event (Supplementary Fig. 4b, c). Accordingly, we looked for mutations in genes known to regulate *HOXA9* directly. 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 trimethylation (H3K27me3) chromatin marks when cells differentiate beyond the haematopoietic stem-cell stage[40,41]. This repressive mark was weak or absent at the *HOXA9* locus in most multiple myeloma 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 multiple myeloma cells, we knocked down its expression with seven shRNAs (Supplementary Fig. 5). In 11 out of 12 multiple myeloma cell lines, HOXA9-depleted cells exhibited a competitive disadvantage (Fig. 3c and Supplementary Fig. 6).

These experiments indicate that aberrant *HOXA9* expression, caused at least in part by HMT-related genomic events, has a role

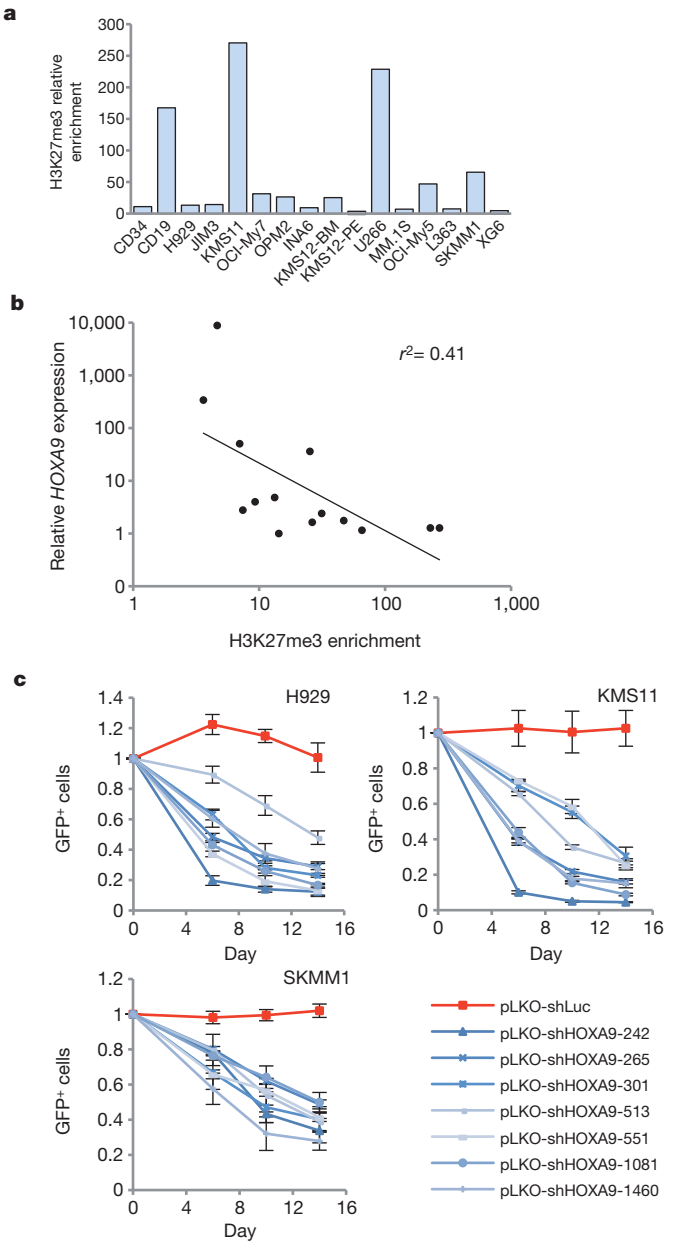

Figure 3 | *HOXA9* is a candidate oncogene in multiple myeloma. **a**, H3K27me3 enrichment at the *HOXA9* promoter in CD34 cells, CD19 cells and multiple myeloma cell lines relative to H3K27me3 methylation at the BC site, known to be hypomethylated in all cells. **b**, Relative *HOXA9* expression against H3K27me3 enrichment at the *HOXA9* locus. **c**, GFP competition assay in multiple myeloma cell lines. After 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 after infection (designated day 0). Error bars indicate standard error of the mean and represent a minimum of three independent experiments.

in multiple myeloma and may represent a new therapeutic target. Further supporting a role of *HOXA9* as a multiple myeloma 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 previous knowledge. Accordingly, we examined 616 gene sets in the MSigDB Canonical

Foresight EX1103
Foresight v Personalis
IPR2024-00170

©2011 Macmillan Publishers Limited. All rights reserved

Table 2 | Statistically significant mutated non-coding regions

Chr	Start	End	Length (nt)	Mut.	Samples	*P*-value	*q*-value	Separation (nt)	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	2,070	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	1,050	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	2,380	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	3,370	4	3	1.47×10^{-8}	0.0017	2055, 657, 295	*BCL7A* (promoter/intron)	–
14	68327320	68333190	5,870	4	4	7.05×10^{-6}	0.17	397, 156, 35	*ZFP36L1* (intron)	Indel
17	8106910	8111850	4,940	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

Regions of predicted regulatory potential showing mutation frequency beyond that expected by chance are shown ($q < 0.25$). Mut., mutations. 'Start' and 'End' columns indicate the first and last nucleotide of regions of regulatory potential according to hg18/NCBI36. 'Separation' column indicates the number of nucleotides within the regulatory region separating the observed mutations.

Pathways database. One top-ranking gene set was of particular interest because it did not relate to genes known to be important in multiple myeloma. This gene set encodes proteins involved in the formation of the fibrin clot in the blood coagulation cascade. There were 6 mutations, in 5 of 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 multiple myeloma cell lines (Supplementary Fig. 8). The coagulation cascade involves a number of extracellular proteases and their substrates and regulators, but their role in multiple myeloma 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. We defined 2.4×10^6 regulatory potential 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 1,000-fold higher than expected mutation frequency, as expected for post-germinal centre 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 five patients, indicating that somatic hypermutation probably occurs in regions beyond the 5′ UTR and first intron of *BCL6* (Table 2). Whether such non-coding *BCL6* mutations contribute to multiple myeloma 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 harboured coding mutations. Interestingly, we observed 7 mutations in 5 of 23 patients (22%) within non-coding regions of *BCL7A*, a putative tumour 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 multiple myeloma remains to be established.

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

Discussion

The analysis of multiple myeloma genomes reveals that mechanisms previously suspected to have a role in the biology of multiple myeloma (for example, NF-κB activation and HMT dysfunction) may have 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 multiple myeloma 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, most of these discoveries could not have been made by sequencing only a single multiple myeloma 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 multiple myeloma 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 multiple myeloma Genomics Portal (http://www.broadinstitute.org/mmgp) to support data analysis and visualization.

METHODS SUMMARY

Informed consent from multiple myeloma patients was obtained in line with the Declaration of Helsinki. DNA was extracted from bone marrow aspirate (tumour) 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 processed 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 artefacts. True positive mutation rates were estimated by Sequenom mass spectrometry genotyping of randomly selected mutations. *HOXA9* short hairpin

Foresight EX1103
Foresight v Personalis

©2011 Macmillan Publishers Limited. All rights reserved

(sh)RNAs were introduced into multiple myeloma cell lines using lentiviral infection using standard methods.

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

Received 11 June 2010; accepted 17 January 2011.

1. Bergsagel, P. L. & Kuehl, W. M. Molecular pathogenesis and a consequent classification of multiple myeloma. *J. Clin. Oncol.* **23,** 6333–6338 (2005).
2. Keats, J. J. *et al.* Promiscuous mutations activate the noncanonical NF-κB pathway in multiple myeloma. *Cancer Cell* **12,** 131–144 (2007).
3. Annunziata, C. M. *et al.* Frequent engagement of the classical and alternative NF-κB pathways by diverse genetic abnormalities in multiple myeloma. *Cancer Cell* **12,** 115–130 (2007).
4. van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41,** 521–523 (2009).
5. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465,** 473–477 (2010).
6. Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40,** 722–729 (2008).
7. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456,** 66–72 (2008).
8. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
9. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464,** 999–1005 (2010).
10. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).
11. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463,** 191–196 (2010).
12. Reva, B., Antipin, Y. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* **8,** R232 (2007).
13. Dziembowski, A. *et al.* A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature Struct. Mol. Biol.* **14,** 15–22 (2007).
14. Schmid, M. & Jensen, T. H. The exosome: a multipurpose RNA-decay machine. *Trends Biochem. Sci.* **33,** 501–510 (2008).
15. Schneider, C., Anderson, J. T. & Tollervey, D. The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. *Mol. Cell* **27,** 324–331 (2007).
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.* **284,** 20486–20498 (2009).
17. Ibrahim, H., Wilusz, J. & Wilusz, C. J. RNA recognition by 3′-to-5′ exonucleases: the substrate perspective. *Biochim. Biophys. Acta* **1779,** 256–265 (2008).
18. Zhan, F. *et al.* The molecular classification of multiple myeloma. *Blood* **108,** 2020–2028 (2006).
19. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
20. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34,** 267–273 (2003).
21. Tanay, A., Regev, A. & Shamir, R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl Acad. Sci. USA* **102,** 7203–7208 (2005).
22. Carrasco, D. R. *et al.* The differentiation and stress response factor XBP-1 drives multiple myeloma pathogenesis. *Cancer Cell* **11,** 349–360 (2007).
23. Zimprich, A. *et al.* Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44,** 601–607 (2004).
24. Paisán-Ruíz, C. *et al.* Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44,** 595–600 (2004).
25. Forman, M. S., Lee, V. M. & Trojanowski, J. Q. 'Unfolding' pathways in neurodegenerative disease. *Trends Neurosci.* **26,** 407–410 (2003).
26. Masciarelli, S. *et al.* CHOP-independent apoptosis and pathway-selective induction of the UPR in developing plasma cells. *Mol. Immunol.* **47,** 1356–1365 (2010).
27. Cenci, S. & Sitia, R. Managing and exploiting stress in the antibody factory. *FEBS Lett.* **581,** 3652–3657 (2007).
28. Todd, D. J., Lee, A. H. & Glimcher, L. H. The endoplasmic reticulum stress response in immunity and autoimmunity. *Nature Rev. Immunol.* **8,** 663–674 (2008).
29. Adams, J. The development of proteasome inhibitors as anticancer drugs. *Cancer Cell* **5,** 417–421 (2004).
30. Shaffer, A. L. *et al.* IRF4 addiction in multiple myeloma. *Nature* **454,** 226–231 (2008).
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* **18,** 568–579 (2010).
32. Pasqualucci, L. *et al.* Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J. Exp. Med.* **203,** 311–317 (2006).
33. Shaffer, A. L. *et al.* Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17,** 51–62 (2002).
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* **19,** 607–620 (2003).
35. Turner, C. A. Jr, Mack, D. H. & Davis, M. M. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* **77,** 297–306 (1994).
36. Wan, P. T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116,** 855–867 (2004).
37. Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417,** 949–954 (2002).
38. Flaherty, K. *et al. Phase I study of PLX4032: Proof of concept for V600E BRAF mutation as a therapeutic target in human cancer.* In *2009 ASCO Meeting* (American Society of Clinical Oncology, 2009).
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.* **149,** 537–549 (2010).
40. Lessard, J. & Sauvageau, G. Polycomb group genes as epigenetic regulators of normal and leukemic hemopoiesis. *Exp. Hematol.* **31,** 567–585 (2003).
41. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128,** 669–681 (2007).
42. Ruf, W. & Mueller, B. M. Thrombin generation and the pathogenesis of cancer. *Semin. Thromb. Hemost.* **32** (suppl. 1), 61–68 (2006).
43. Esumi, N., Fan, D. & Fidler, I. J. Inhibition of murine melanoma experimental metastasis by recombinant desulfatohirudin, a highly specific thrombin inhibitor. *Cancer Res.* **51,** 4549–4556 (1991).
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. USA* **92,** 12520–12524 (1995).
45. Zani, V. J. *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* **87,** 3124–3134 (1996).
46. Zhang, W. *et al.* Unravelling the hidden heterogeneities of diffuse large B-cell lymphoma based on coupled two-way clustering. *BMC Genomics* **8,** 332 (2007).
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 Chromosom. Cancer* **47,** 1067–1075 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This project was funded by a grant from the Multiple Myeloma Research Foundation. M.A.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.

Author Contributions 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. 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., A.K.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 analysed genetic material, including RNA/DNA extraction, fingerprinting, genotyping, data management, hybridizations, library preparation and sequencing.

Author Information Sequence data have been deposited in the dbGaP repository (http://www.ncbi.nlm.nih.gov/gap) under accession number phs000348.v1.p1. Additional data have been submitted to the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession numbers GSE26760 (GEP data) and GSE26849 (aCGH data); both data sets are also combined under accession code GSE26863. We have also created a multiple myeloma Genomics Portal (http://www.broadinstitute.org/mmgp) to support data analysis and visualization. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.R.G. (golub@broadinstitute.org) or G.G. (gadgetz@broadinstitute.org).

Foresight EX1103
Foresight v Personalis
IPR2024-00170

©2011 Macmillan Publishers Limited. All rights reserved