APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma

Brian A Walker[1], **Christopher P Wardell**[1], **Alex Murison**[1], **Eileen M Boyle**[1], **Dil B Begum**[1], **Nasrin M Dahir**[1], **Paula Z Proszek**[1], **Lorenzo Melchor**[1], **Charlotte Pawlyn**[1], **Martin F Kaiser**[1], **David C Johnson**[1], **Ya-Wei Qiang**[2], **John R Jones**[1], **David A Cairns**[3], **Walter M Gregory**[3], **Roger G Owen**[4], **Gordon Cook**[4], **Mark T Drayson**[5], **Graham H Jackson**[6], **Faith E Davies**[1,2], and **Gareth J Morgan**[1,2]

[1]Division of Molecular Pathology, The Institute of Cancer Research, London, United Kingdom

[2]University of Arkansas for Medical Sciences, Little Rock, Arkansas

[3]Clinical Trials Research Unit, University of Leeds, Leeds, United Kingdom

[4]St James's University Hospital, University of Leeds, Leeds, UK

[5]Clinical Immunology, School of Immunity & Infection, University of Birmingham, Birmingham, United Kingdom

[6]Department of Haematology, Newcastle University, Newcastle-Upon-Tyne, United Kingdom

Abstract

We have sequenced 463 presenting cases of myeloma entered into the UK Myeloma XI study using whole exome sequencing. Here we identify mutations induced as a consequence of misdirected AID in the partner oncogenes of *IGH* translocations, which are activating and associated with impaired clinical outcome. An APOBEC mutational signature is seen in 3.8% of cases and is linked to the translocation mediated deregulation of *MAF* and *MAFB*, a known poor prognostic factor. Patients with this signature have an increased mutational load and a poor prognosis. Loss of *MAF* or *MAFB* expression results in decreased *APOBEC3B* and *APOBEC4* expression, indicating a transcriptional control mechanism. Kataegis, a further mutational pattern associated with APOBEC deregulation, is seen at the sites of the *MYC* translocation. The

APOBEC mutational signature seen in myeloma is, therefore, associated with poor prognosis primary and secondary translocations and the molecular mechanisms involved in generating them.

Introduction

Humans have evolved with the need to evade infection and as a consequence have developed DNA rearranging mechanisms to generate high affinity antibodies. An inevitable consequence of these processes is the generation of abnormal recombination events leading to oncogene activation or tumour suppressor gene inactivation. Myeloma is a malignancy of plasma cells that develops later in life in which abnormal rearrangements at the Ig loci have been shown to be important initiating events. Consequently, studying the mechanisms underlying the development of translocations and their downstream effects can provide major insights into the aetiology of the disease[1].

Aberrant chromosomal translocations are seen in ~40% of presenting patients and predominantly involve the *IGH* locus at 14q32 [2]. There are five main partners to the *IGH* locus which are 4p, 6q, 11q, 16q and 20q all of which are considered as being classical myeloma translocations that are seen in close to 100% of the tumour population. These translocations result in the over-expression of an oncogene on the partner chromosome which can be categorised using a translocation/cyclin (TC) classification[3]. The expression of the partner oncogene has a strong influence on the cell, resulting in changes to the genome and in a characteristic clinical behaviour of the disease. For example, the t(4;14) results in over-expression of *FGFR3* and *MMSET* (*WHSC1*)[4]. This translocation group is associated with a poor prognosis,[5] which is abrogated to some extent by the use of bortezomib treatment[6]. The over-expression of *MMSET*, which encodes a histone methyltransferase, results in gene specific DNA hypermethylation, which is distinct from the other translocation groups[7]. The t(11;14) results in over-expression of *CCND1* and the occurrence of this translocation is more likely in individuals carrying the G allele of the cyclin D1 SNP, rs9344, which affects its splicing pattern[8]. Both the t(4;14) and t(11;14) are relatively frequent events (12% and 15%, respectively) with the other three translocations occurring less frequently. The t(6;14) (1%) results in over-expression of *CCND3* and the t(14;16) (4%) and t(14;20) (1%) result in over-expression of the transcription factors *MAF* and *MAFB*, respectively[3]. The 'maf' translocations are associated with a poor prognosis[9].

In initial sequencing studies of myeloma it has been noted that the spectrum of mutations fall into two groups, one of which is characterised by an APOBEC signature[10]. This signature comprises of C>T, C>G and C>A mutations in a TpC context and it has also been described in several cancers such as breast, pancreatic, CLL, and B cell lymphoma. This mutational signature comprises only a subset of samples, with the rest having a rather generic mutation signature with enrichment of C>T transitions at CpG dinucleotides, representing an intrinsic mutational process occurring as a result of the spontaneous deamination of methylated cytosine to thymine. Here we have performed whole exome sequencing on 463 patients with the addition of the Ig and *MYC* loci, to capture translocation breakpoints, copy number abnormalities and somatic mutations to determine how these affect mutation patterns in the plasma cell. We find that MYC translocations are indicative

of poor prognosis and are associated with kataegis, translocation partner genes are mutated by AID in 15% of samples and mutation of *CCND1* results in poor prognosis, and that APOBEC mutational signatures are associated with the t(14;16) and t(14;20) translocations and a high mutational load.

Results

Rearrangements at *MYC* are associated with a poor outcome

Whole exome sequencing was performed on 463 presentation patients enrolled into the Myeloma XI trial. Sequencing metrics are presented in Supplementary Table 1. In addition to capturing the exome, extra baits were added covering the *IGH*, *IGK*, *IGL* and *MYC* loci in order to determine the breakpoints associated with translocations in these genes. These combined data allow us to examine the effect of translocations on the mutational spectra in myeloma. Using a combination of targeted capture and expression-based classification we identified the five main translocation partners and those surrounding the *MYC* locus. Translocations were detected in 232 (50.1%) patients of which 59 patients (12.7%) had a t(4;14), 86 patients (18.6%) a t(11;14), 17 patients (3.7%) a t(14;16), 5 patients (1%) a t(6;14) and 4 patients (0.9%) a t(14;20). The remainder had translocations involving 8q24 with 21 (4.5%) patients having both a classical translocation and an 8q24 translocation and 62 (13.4%) having only an 8q24 translocation (of which 54 were hyperdiploid (HRD) and 8 were neither HRD nor had a classical translocation). Breakpoints are shown in Supplementary Fig. 1 and the distribution agrees with previously published results[11].

MYC translocations were found in 85 patients (18.4%). The most common partner loci were *IGH@* (14 patients), *IGL@* (14 patients), *FAM46C* (8 patients), *IGK@* (5 patients), *FOXO3* (5 patients) and *BMP6* (3 patients). Several other genes of interest in B cells were identified as partners to *MYC* including *RB1*, *XBP1*, *TXNDC5*, *CCND3* and *CCND1*. It has been suggested previously that super enhancers located at these loci upregulate their partner genes[12, 13]. We report a significant negative correlation of *MYC* translocations with the t(4;14) (Bayesian inference correlation=-0.13, BF=2.11) and positive correlation with the hyperdiploid (HRD) group (Bayesian inference correlation=0.13, BF=1.55). Based on these results we describe *MYC* translocations as being the most common translocation in myeloma and they are associated with impaired clinical outcomes (Figure 1A-D). *MYC* translocations are an independent poor prognostic marker when analysed in multivariate analyses.

We used copy number data, generated from multiplex ligation-dependent probe amplification, to identify additional abnormalities at the *MYC* locus. There was a translocation involving the *MYC* locus in 18.4% of patients. There was a gain in 81 of the patients (19%), which can be subdivided into those also with a translocation (31 patients) and those with focal or chromosomal gain alone (50 patients). Combined, there was a *MYC* translocation or gain in 135 patients (32%) of patients. Amplification of *MYC* (>4 copies) was not detected in this dataset.

We examined the expression of *MYC* in these samples and found a significant increase in expression in samples with a *MYC* translocation compared to those with no abnormality (Figure 1E). Gains and deletions were not associated with significant increases in expression

of *MYC*. There was no difference in the expression of *MYC* between samples where the translocation partner was an Ig locus and those which were non-Ig loci (Supplementary Fig. 2), but there was a significant difference in OS between those with an Ig locus partner and those with a non-Ig locus partner (log rank p=0.04, Supplementary Fig. 2). If *MYC* translocated samples are removed from the dataset, expression of *MYC* alone is not sufficient to result in a significant difference in PFS or OS.

Kataegis co-localises with *MYC* translocation breakpoints

We were also able to detect a mutational signature, kataegis, where regional clustering of mutations can be indicative of somatic genomic rearrangements[14]. It is difficult to detect kataegis using exome sequencing but the tiled regions surrounding the *IGH*, *IGK*, *IGL* and *MYC* loci could be used to detect it. By creating rainfall plots we were able to discern samples with regional hypermutation. As expected, the Ig loci contained clusters of hypermutation, but these were not enriched for C>T or C>G mutations within a TpCpH trinucleotide context and as such are not caused by kataegis. We found the hallmarks of kataegis in 15 samples (3.2%), where there was enrichment for TpCpH mutations with an inter-mutational distance <1 kb (Figure 2). Of these 15, 9 were found in the tiled region surrounding *MYC* and others were detected on chromosomes 1, 10, 11, 16, and 17. Kataegis was co-localised with copy number abnormalities in 12 of the samples. Two of the samples with kataegis surrounding *MYC* also had an inter-chromosomal translocation at *MYC* involving either *IGK* or *IGL*. Interestingly, the partner chromosomes also showed signs of kataegis e.g. in the t(2;8) kataegis was found at *IGK* and *MYC* and in the t(8;22) kataegis was found at *MYC* and *IGL* (Figure 2). The pattern of mutations clustered around the translocation breakpoint and according to the cancer clonal fraction (CCF) were present in all cells, indicating that kataegis most likely occurs at the same time as the translocation. APOBECs are thought to be involved in the generation of kataegis and as such this co-localisation is indicative of APOBEC involvement in the generation of *MYC* breakpoints.

Hypermutation of Translocation Partner Oncogenes

In agreement with previous studies[15] we find that *CCND1* is mutated, and this was seen in 10 patients, all of whom have a t(11;14) (Supplementary Table 3). All mutations occurred in the first exon of *CCND1* and all but one were located outside of the cyclin box fold domains. Five patients had multiple mutations and the K50R mutation was detected in three samples, the K46N mutation was seen in two samples but all of the other mutations we describe were unique. There was no association of mutation in *CCND1* with type of breakpoint (RAG or switch mediated) or allele of the variant associated with the t(11;14). There was a significant association with mutation of *CCND1* and the distance to the translocation breakpoint, where samples with a breakpoint closer to *CCND1* were more likely to have mutations within *CCND1* (p=0.03; Supplementary Fig. 3). There was also an association of mutated *CCND1* and a poor prognosis when compared with non-mutated t(11;14) patients (Overall survival median of 20.2 months versus not reached, log rank p=0.005; Figure 3C) in the Myeloma XI trial. This effect was also seen when comparing patients with mutated *CCND1* to all other patients in the cohort and was an independent prognostic factor for both PFS and OS. To determine the significance of this result we sequenced the first exon of *CCND1* in 102 t(11;14) samples from the Myeloma IX trial and

found mutations in a further 10 samples (9.8%; Supplementary Fig. 4). There was no effect on survival in the Myeloma IX dataset with mutations in *CCND1*. We also examined the allelic frequency of the variant associated with the t(11;14), rs9344, in the Myeloma XI dataset and found that in agreement with our previous observations[8] the G allele is significantly associated with the translocation (Supplementary Table 2).

Given the association of mutated *CCND1* in the t(11;14) we examined the other translocation groups and their associated partner chromosome oncogenes. We found that *FGFR3*, *MAF* and *MAFB* all had mutations which were restricted to the relevant translocation group (16.9% mutated *FGFR3* in t(4;14); 12.5% mutated *MAF* in t(14;16); and 25% mutated *MAFB* in t(14;20); Figure 3A). There were no mutations in *CCND3* (in 6 t(6;14) patients) and the mutations in *MMSET* were not in t(4;14) patients. In contrast to the poor prognosis associated with mutations of *CCND1* in the t(11;14), no poor prognosis was associated with mutations in *FGFR3*, *MAF* or *MAFB*. However, the latter two sample sets may be too small to address this question adequately.

The mutated oncogenes were all on the der(14) and most likely reflect somatic hypermutation events mediated by AID, an member of the APOBEC family, which would normally affect the V(D)J rearrangement upstream of the *IGH* constant regions. After the translocation event, *MMSET* is on the der(4) and is, therefore, unlikely to be mutated by this mechanism. To determine that AID may be responsible we examined the transition/ transversion ratio and the proximity of the mutations to the AID consensus motif, WRC. In agreement with previous studies[16] the mutations in the *IGH* translocation partner genes are 58.8% in A:T bases, of which 50% are transversions (Supplementary Fig. 5), and in close proximity to WRC motifs (Supplementary Table 3) indicating that AID is involved in this mutational process. As AID is associated with hypermutation of the V(D)J and switch regions of the highly expressed *IGH* locus we examined the expression of the translocation partner oncogene between those with or without a mutation, but found no significant differences (Supplementary Fig. 6).

The CCF in which the mutation was found differs by translocation group (Figure 3B). In the t(11;14) the mutations were founder events, present in all of the cells whereas in the t(14;16) and t(14;20) the mutations were in only ~50% of the cells indicating they are obtained later than the translocation themselves. In the t(4;14) the mutations in *FGFR3* can be either clonal or subclonal indicating that these mutations can develop at the same time as the translocation or at a later time point.

The mutation patterns for each of the der(14) oncogenes differ. In the t(11;14) mutations in *CCND1* occur solely near the N-terminus of the protein. In *MAF* and *MAFB* the mutations are constrained in the basic-leucine zipper domain at the C-terminal of the protein. The focussed mutation profile seen in *CCND1*, *MAF* and *MAFB* are indicative of activating mutations. The mutations in *FGFR3* are dispersed through several domains but have also been described as mutations that can activate the RAS/MAPK pathway in urothelial cancer and myeloma (Supplementary Fig. 7) [17, 18].

APOBEC Mutations are enriched in 'maf' translocation groups

It has previously been shown that mutations can be described in a specific trinucleotide context and that in myeloma there are two signatures that predominate[10]. We are able to recapitulate these two different signatures, which consist of a generic signature comprised of enrichment of C>T transitions in a CpG context, Signature A (Figure 4A) and a second signature, Signature B, in which there is enrichment for C>G and C>T, especially in a TpCpA context. Signature B is hypothesised to result from aberrant APOBEC activity, where the APOBECs enzymatically modify single-stranded DNA. AID is a member of the APOBEC family and is involved in class-switch recombination and somatic hypermutation in B cells.

We noted that the t(14;16) and t(14;20) have a statistically significant higher number of mutations per sample compared to the other translocation sub-groups (2-sided t-test $p=1.65\times10^{-5}$), Figure 4B, and that the Signature B (APOBEC) related context of mutations in the t(14;16) and t(14;20) was significantly higher than other translocation groups (T(C>T)A, $p=9.1\times10^{-6}$; T(C>T)T, $p=0.0014$; T(C>G)A, $p=0.001$; T(C>G)T, $p=0.0064$), Figure 4C. Collectively, mutations in these trinucleotide contexts comprise a mean of 28.7% and 21.3% of mutations in t(14;16) and t(14;20) samples, respectively (compared to t(4;14) 6.5%, t(6;14) 6.2%, t(11;14) 5.8%). We examined the proportion of each Signature present in the translocation sub-groups and found that there is a significant enrichment for Signature B (APOBEC) mutations in the t(14;16) and t(14;20) (0.56, $p=2\times10^{-16}$; 0.44, $p=8.26\times10^{-11}$, respectively) compared to the t(4;14), t(6;14), t(11;14) and hyperdiploid samples (0.094, 0.096, 0.074, 0.098 and 0.078, respectively) (Figure 4D). These data indicate that the 'maf' translocation groups are largely characterised by APOBEC signature mutations and have a higher mutation load than the other translocation groups.

In order to determine if there are some samples in the other translocation groups which also have an APOBEC signature we assigned each sample to either Signature A or Signature B depending on the proportion of mutation type in each sample. This generated clusters of samples whose mutations are either mostly Signature A (cluster A) or Signature B (cluster B), Figure 5A. Here we find that the t(14;16) and t(14;20) cases comprise 66.6% of cluster B but only 1.3% of cluster A. In line with the proportion of 'maf' samples in Cluster B the number of mutations in this cluster is significantly higher than in Cluster A (mean 295.44 vs. 127.22, two-sided t-test $p=1.18\times10^{-15}$; Figure 5B). Cluster A is comprised of 445 patients and cluster B 18 patients, indicating that Signature B only affects 3.9% of patients. However, when we performed survival analysis of these patient clusters we find that there is a significant effect on overall survival (log rank $p=0.02$; Figure 5C). Due to the low number of samples with 'maf' translocations and the interaction of the translocation, APOBEC signature and mutational load this effect on survival is not significant in a multivariate analysis as it is not possible to delineate whether this effect on survival is due to any single one of these markers alone and it remains more likely that the impact of three abnormalities is linked.

Knockdown of *MAF* and *MAFB* decreases APOBEC expression

Both the t(14;16) and t(14;20) result in over-expression of a maf transcription factor. As these translocation groups are enriched for APOBEC signature mutations we sought to determine if there is a link between maf gene expression and APOBEC expression. We examined two well-characterised gene expression datasets from myeloma patients (UAMS, GSE4581; MRC Myeloma IX, GSE15695) for characteristic expression patterns of APOBEC genes in t(14;16) and t(14;20) groups. We found that there is significantly increased expression of *APOBEC3A* and *APOBEC3B* in t(14;16) and t(14;20) cases in both the UAMS and Myeloma IX datasets (Figure 6).

To investigate the possibility that maf expression controls APOBEC expression we knocked down either *MAF* or *MAFB* in RPMI-8226 (t(14;16)) or Sachi (t(14;20)) cell lines, respectively. Quantitative gene expression analysis of the cell lines indicated that only *APOBEC3B* was expressed in RPMI-8226 and only *APOBEC4* was expressed in Sachi. In the knockdowns we show that decreased expression of *MAF* results in decreased *APOBEC3B* expression and knockdown of *MAFB* results in decreased expression of *APOBEC4* (Figure 7), indicating a causative link between the two sets of genes.

Analysis of ENCODE data indicates a MAFK binding site in the promoter of *APOBEC3A* and *APOBEC3B*. Although MAFK is a different class of maf transcription factor it shares a binding motif with MAF and MAFB, which could explain the over-expression of APOBECs in the t(14;16) samples. Previously, APOBEC3B has been associated with C>T transitions in breast cancer,[19, 20] and given that *APOBEC3B* is significantly over-expressed in t(14;16) or MAF group samples in both datasets this is the most likely causative candidate for the C>T transitions observed in our Signature B cases.

Discussion

Translocations involving the Ig loci in myeloma are recognised as primary events, being present in all cells, whereas copy number abnormalities and somatic mutations tend to be present in sub-clones[21, 22]. Here we identify translocations involving *MYC* as being the most common structural chromosomal abnormalities in myeloma, bringing the total translocated group to 50.1% of presenting patients and in addition *MYC* translocations are associated with adverse clinical outcomes. Furthermore, we show that when copy number abnormalities are taken into account, the percentage of myeloma cases where *MYC* is deregulated is 55%, making it the most common genetic event in myeloma, being even more common than mutational activation of the RAS pathway. This observation on presenting clinical cases is consistent with previous work using myeloma cell lines and patients which suggested that *MYC* deregulation is more or less ubiquitous and is mediated by non-physiological DNA damage and repair pathways[13]. In contrast, we show that in clinical samples the *MYC* translocation partner is an Ig locus in 30% of cases and in the remainder it is mediated by translocations to genes expressed at this stage of B cell differentiation. While we were able to demonstrate frequent copy number change at the *MYC* locus, using exome sequencing we were unable to gain additional mechanistic information as to why these events were occurring. From both a clinical and aetiological standpoint the positive association of *MYC* translocations with HRD is important because it not only reduces the

size of the group where HRD is the primary aetiological factor but it also removes poor prognostic cases from this group.

Tiling of the *MYC* and IG loci enabled us to identify of a second mutational signature, kataegis, when translocations occurred between the two loci. This signature is distinct from the APOBEC signature and results in closely spaced mutations often surrounding DNA damage breakpoints. We identified two samples with the kataegis signature which also had *MYC* translocations and in these cases the kataegis signature was present on both partner chromosomes, indicating that the mutational signature and the translocation co-occurred. The mechanism resulting in kataegis is not known but may also be related to the APOBECs[14]. The presence of kataegis in the *MYC* region is of interest because it suggests a relationship between it and the development of translocations and copy number abnormalities at this site.

As initiating events, translocations are poised to control the fate of the cell and ultimately of the patient, and as such it is not surprising to find that they drive the mutational pathogenesis of the disease. We find that the translocation partner oncogene is mutated in 11-25% of samples, depending on the translocation. We show that *MMSET* is not mutated following translocation in the t(4;14), presumably because the site of the breakpoint results in this gene being carried on the der(4) chromosome and not the der(14) as is the case with *FGFR3*, *CCND1*, *MAF* and *MAFB*. There is evidence to show that the mutations on the partner oncogenes are mediated by AID (distance to WRC motif and A:T mutation rate) and so they are most likely mediated via aberrant somatic hypermutation as a consequence of AID, which would normally mutate the functional V(D)J rearrangement on chromosome 14. AID is a member of the APOBEC family that deaminates C to U in actively transcribed immunoglobulin variable and switch regions resulting in somatic hypermutation through recruitment of error prone polymerases that mutate the surrounding bases through mismatch repair and base excision repair. It is not clear why only a proportion of the translocation partner genes are mutated. However, a similar situation is seen in mantle cell lymphoma (MCL) which also harbours a t(11;14). In this B cell malignancy, mutation of *CCND1* is also detected in a subset (35%) of cases and is associated with SOX11-negative and *IGHV*-mutated MCL, suggesting their acquisition in the germinal centre[23]. These mutations are also restricted to the first exon of *CCND1*, consistent with them developing as a consequence of a similar mechanism as that seen in myeloma. Interestingly, in terms of the timing of development, the mutations are clonal in t(11;14) and in some t(4;14) myeloma samples, but are sub-clonal in t(14;16) and t(14;20), indicating that they were acquired subsequent to the translocation.

As the mutations in the partner oncogenes are mediated by AID it raises the question as to whether they are simple passenger variants or whether they are driver events providing a selection advantage. Importantly the mutations are not randomly distributed within the partner oncogenes. In *CCND1* they are restricted to the 5′ end of the gene, outside the cyclin domain. In *MAF* and *MAFB* the mutations are only seen in the basic-leucine zipper domain whereas in *FGFR3* they are scattered throughout the gene. Interestingly in terms of their pathological relevance, although the mutations in *FGFR3* look random there is evidence that these are involved in activation of FGFR3 and downstream signalling of the RAS/MAPK

pathway[18]. Many of the mutations in *FGFR3* result in acquisition of a cysteine residue, often involved in di-sulphide bond formation, which may result in homo-dimerization and activation of the molecule[24]. Mutations in *FGFR3* also included the K650 mutation, which constitutively activates the receptor[25]. Importantly, several of these mutations (Y373C and K650E) are present in myeloma cell lines and have been shown to result in constitutive activation of the RAS/MAPK pathway[17]. We conclude, therefore, that these mutations have been positively selected for their importance in the pathogenesis of the disease and could be targeted therapeutically. The major difference between the *CCND1* and *FGFR3* mutations is the effect on patient survival. However, this may not be surprising as *FGFR3* activates the RAS/MAPK pathway and no effect on survival has been shown with RAS/MAPK pathway mutations (*NRAS*, *KRAS*, *BRAF*) with current treatments even though they are present in ~50% of patients.

Mutations in cancer samples can be sub-divided based on the context of the surrounding bases. Two signatures have previously been identified in myeloma, the second of which is an APOBEC signature, and here we have been able to identify a specific genomic subgroup affected by this signature. APOBECs are a family of DNA editing enzymes, which mostly act on single-stranded DNA through deamination of cytosine to uracil[26]. As such, they have a characteristic pattern of mutation which results in enrichment from C>G and C>T mutations in a TpCpH context. Here we find this signature in 18 of the 463 samples and these samples are highly enriched for the maf translocations, t(14;16) and t(14;20). These samples in turn also have a higher mutation load than the other samples and the samples with this signature have an adverse progression free (PFS) and overall survival (OS). While the t(14;16) and t(14;20) have been shown to be adverse prognostic previously,[27, 28, 29] the mechanism by which this is mediated has been unclear. It is known that the t(14;16) and t(14;20) as well as the t(4;14) result in the indirect upregulation of *CCND2*[3] but the relationship of this to adverse outcomes is undetermined. Here, we show that the t(14;16) and the t(14;20) have more mutations than other cytogenetic groups of myeloma, these mutations have an APOBEC signature, and they over-express APOBEC genes, specifically *APOBEC3A* and *APOBEC3B*. The common mechanism between the two different translocation groups is that they result in over-expression of a maf transcription factor (*MAF* or *MAFB*). Interestingly, examination of ENCODE data shows maf binding sites in the promoter regions of *APOBEC3A* and *APOBEC3B*, giving a link between the translocations and the increase in mutation load and type. APOBEC3B has also been implicated in the APOBEC mutational signature in breast, ovarian and multiple other human cancers, consistent with these data[19, 20, 30].

It is interesting to note that the presence of t(14;16) or t(14;20) in pre-malignant MGUS and asymptomatic myeloma is associated with a favourable prognosis and a long time to progression to myeloma in contrast to when they are seen in myeloma[9]. Given the high mutational burden of this subgroup it will be important to determine if the mutational signature seen in the maf sub-groups is also present at the MGUS stage, or whether the signature only manifests when the disease has progressed to myeloma.

Here we show three different mutational signatures mediated by the APOBEC family: translocation partner mutation by AID, Signature B (APOBEC) mediated by APOBEC3A or

APOBEC3B, and kataegis mediated by an unknown APOBEC family member. We also show for the first time a clinical impact of APOBEC mutations and their association with a poor prognosis. The poor prognosis of this mutational signature is inextricably linked to a high mutation load and the adverse t(14;16) and t(14;20) translocation subgroups, making it not currently possible to disentangle the individual impact of these markers on prognosis.

Methods

Patient Samples

Samples were taken, following informed consent, from patients newly diagnosed with symptomatic myeloma and enrolled in the Myeloma XI trial (NCT01554852). The study was approved by the NHS Health Research Authority, National Research Ethics Service Committee and by local review committees at all participating centres. Experimental procedures were approved by the NHS Health Research Authority National Research Ethics Service Committee London-Surrey Borders (CCR3019). This was a large, phase III, open label trial where patients, with no age limit were randomised between triplet immunomodulatory drug induction of either cyclophosphamide, thalidomide, dexamethasone (CTD) or cyclophosphamide, lenalidomide, dexamethasone (CRD). Patients with a suboptimal response (<very good partial response) were randomised to pre-transplant treatment with a proteasome inhibitor triplet (cyclophosphamide, bortezomib, dexamethasone, CVD). Older or less fit patients had appropriate dose reductions to therapy and did not receive an autologous stem cell transplant (ASCT). All patients subsequently underwent a maintenance randomisation to either no maintenance, lenalidomide maintenance or lenalidomide and vorinostat maintenance. Exclusion criteria include renal failure requiring dialysis and solitary plasmocytoma as well as the standard criterion. Progression-free and overall survival was measured from initial randomisation and the median follow up for this analysis was 25 months (95% 24.3-26.2). The patient demographics are presented in Table 1 and a diagram of the different study arms in Supplementary Fig. 8. The median PFS was 26.6 months (95% CI 23.6-29.9) and OS was not reached but the 3 years survival rate was 66% (95% CI 60%-73).

CD138+ cells were isolated from the bone marrow of patients using CD138+ MACSorting (Miltenyi Biotech, Bisley, UK) to a purity >95%. Cells were lysed in RLT+ buffer and DNA and RNA extracted using the AllPrep kit (Qiagen, Manchester, UK). Peripheral blood was also isolated from patients, white blood cells purified by Ficoll-Pacque and DNA extracted from the cell pellet using the QIAamp DNA mini kit (Qiagen). RNA and DNA were quantified using the 2200 Tapestation (Agilent Technologies, Santa Clara, CA) and Pico-green (Life Technologies, Paisley, UK).

Multiplex Ligation-Dependent Amplification

50 ng of sample DNA was subjected to MLPA reactions using the SALSA P425-B1 probemix developed by MRC-Holland (Amsterdam, The Netherlands).[31] The probemix contained 46 probes for copy number analysis of the major prognostic markers in myeloma including: 1p32-31 (*FAF1, CDKN2C, PPAP2B* and *DAB1*), 1p21, 1p12 (*FAM46C*), 1q21.3 (*CKS1B*), 1q23.3, 5q31.3, 12p13.31 (*CD27, CHD4*), 13q14 (*RB1, DLEU1, DLEU2, DIS3*),

16q12 (*CYLD*), 16q23 (*WWOX*), and 17p13 (*TP53*). In addition, 11 reference probes were also included, analyzing various autosomal chromosome loci which are only infrequently involved in CNA in myeloma. A second panel was used to interrogate the *MYC* locus, which contained probes for exons 1 and 3. The MLPA reaction, including internal quality controls and negative controls, was performed according to manufacturer's instructions. The PCR-products were analyzed using an ABI 3730 DNA analyzer (Life Technologies, Paisley, UK) and Coffalyser software (MRC-Holland, Netherlands). After intra-sample and inter-sample normalizations, copy number at each locus was estimated.[32] In summary, values 1.2; between 1.19 and 0.71; and 0.7 were considered as gain, normal, and deletion, respectively. If multiple probes per gene were examined a consistent result between them was required (2 probes with same result or majority of probes with same result depending on locus/chromosome examined).[33]

Gene Expression

Translocations were defined using the TC classification based on expression of the Ig-partner oncogene. In addition, *MYC* expression was assayed (Hs00153408; Life Technologies, Paisley, UK) using qRT-PCR and relative quantitative (RQ) values were compared with the endogenous control (*GAPDH*).

Exome Sequencing and Analysis

DNA from both tumor and peripheral blood samples were used in the exome capture protocol as previously described.[34] Briefly, 200 ng DNA was fragmented on a Covaris E-Series using the settings suggested in the Sureselect[XT] Target Enrichment System for Illumina Paired-End Sequencing Library protocol (version 1.5). DNA was end-repaired, A-tailed and adaptors ligated using the NEBNext DNA library prep master mix set for Illumina (New England Biolabs, Hitchin, UK). Adaptor-ligated DNA was amplified using the NEBNext High-fidelity PCR master mix (New England Biolabs) using 7 PCR cycles and 750 ng DNA hybridized against RNA baits overnight (SureSelect Human All Exon V5, Agilent Technologies). The RNA baits were designed against the human exome with the addition of custom baits that tiled the *IGH*, *IGK*, *IGL* and *MYC* loci in order to detect the major translocations in myeloma. After hybridisation the captured DNA was indexed and amplified using Herculase II fusion DNA polymerase (Agilent Technologies) for 8 PCR cycles. Four exome samples were pooled and run on one lane of a HiSeq 2000 (Illumina, Hinxton, UK) using 76-bp paired end reads.

Data Quality Metrics and Processing

FastQC (v0.10.0) was used for basic quality control of Illumina paired-end sequencing data. These files were aligned to the reference genome (GRCh37), using BWA (v.0.5.9) followed by Stampy (v.1.0.20) to improve gapped alignment. BAM files were recalibrated using GATK (v2.3.9) and deduplicated using Picard (v.1.85). Tumour and normal samples were realigned as pairs using the GATK indel realigner to improve indel call rates. Calls from a panel of 6668 highly variable SNPs within the exome capture were compared between the tumour and normal samples to confirm that they were correctly paired.

Somatic Mutation Calling

Single nucleotide variants (SNVs) were called using MuTect (v1.1.4). They were further filtered using the following criteria: minimum of 10x depth at that site in both tumour and normal BAMs, a minimum of 1 non-reference base call in both directions, a mean Phred quality score of greater than 26 for that base in the tumour sample, a mean mapping quality score of equal to or greater than 50 across all reads at that site in the tumour sample, and the site must be uniquely alignable according to the CRG alignablity tracks available from the UCSC genome browser and created using the GEM mappability tool (Fast Computation and Applications of Genome Mappability by Derrien et al, 2012). C>A|G>T SNVs likely to be oxidation artefacts created during library preparation were removed.[35]

Short indels were called using the GATK Indelocator and furthered filtered according to the following criteria: a minimum of 1 non-reference base call in both directions, a mean Phred quality score of greater than 26 for that base in the tumour sample, a mean mapping quality score of equal to or greater than 50 across all reads at that site in the tumour sample, and the Site must be uniquely alignable. No more than 2 reads covering that site in the normal sample could contain any indel. A window of 21 base pairs centred on the first base of the indel was taken and had to conform to the following rules which remove low-complexity sequences: no homopolymers of greater than 6 base pairs and no dinucleotide may occur more than 5 times.

Deletions of whole exons (windows defined as the regions used in the Agilent exome capture) were found by comparing the read depth between the tumour and normal samples. The mean depth across the window was required to be greater than 0.2 of the median depth in the normal sample and less than 0.06 in the tumour sample, with the normal value being at least 8x greater than the tumour value.

Copy number across the exome was determined using Control-FREEC[36] utilising 500 base pair bins, each overlapping with the subsequent and previous 250 bp. A minimum average read depth of 50 was required in the control samples, with at least two neighbouring bins required to show copy number aberration to call a region as gained or lost.

All somatic events were annotated using both SnpEff (v3.1) and Oncotator (v0.4.2.2) with SnpEff providing the most deleterious interpretation regardless of transcript and Oncotator annotating only the canonical transcript. Significantly mutated genes were detected by providing all SNV and short indels to the MutSigCV (v1.4) algorithm. A q-value cut-off of 0.1 was used.

The proportion of tumour cells containing an SNV was calculated using the following equation: $CCF = \min\left(1, \dfrac{CN.r}{R}\right)$

CCF = cancer cell fraction (proportion of cells containing the mutation)

CN = copy number at that site

r = number of reads containing the mutation at that site

R = total number of reads at that site

Translocations were called using Delly[37] and manually curated in the Integrative Genomics Viewer (IGV). Translocations were considered real when there were at least 10 supporting reads and if no translocations were also found in the peripheral blood sample. Simultaneously, translocations were determined by qRT-PCR using defined cut-offs for expression of each partner oncogene.[38] Results from both techniques were compared and a consensus translocation call generated. There was concordance between assays in 93.4% of cases. When assays were not concordant the data were examined and a decision made.

Nonnegative Matrix Factorization (NMF)

Nonnegative matrix factorization (NMF) using the NMF package in R was used to factorize a matrix of frequency of trinucleotide mutation contexts per sample, in order to identify underlying mutation signatures among all mutations in the 463 samples.[10, 14] The algorithm was run for between two and 18 underlying signatures, with 10 runs at each number. Two signatures were selected as the optimum number of signatures based around a number of quality control metrics. Samples were clustered using the consensus hierarchical clustering procedure implemented in the NMF package.

Correlation studies

To determine correlation between cytogenetic abnormalities a probabilistic approach based on Bayesian interference was determined in R [39] using the program "JAGS"[40] and its the R interface Bayesmed [41] as previously described [42]. The probability of the observed data under the null hypothesis versus the alternative hypothesis or Bayes factor (BF) was computed. BF>1 was considered significant.

Survival analysis

Survival analysis was performed in R[39] using package survival[43, 44] and coin[45, 46] Differences between curves were tested for statistical significance using the log-rank test, with p<0.05 taken as the level of significance.

Cell lines and Cell culture

Sachi is a cell line derived from a t(14;20) primary plasma cell leukemia with increased levels of *MAFB* transcripts and was derived by Hanamura and colleagues (2001).[47, 48] RPMI8226 is a cell line obtained from the ATCC with a t(14;16) translocation and increase levels of *MAF* transcripts.[49] RPMI8226 and Sachi were cultured in RPMI1640 (Invitrogen, Carlsbad, CA) containing 10% heat-inactivated FBS, and 4mM L-glutamine, as described previously.[50] 293T cells were cultured in Dulbecco Modified Eagle Medium (DMEM) containing 10% heat-inactivated FBS, penicillin (100 U/mL), streptomycin (100 mg/mL), 4 mM L-glutamine, and 1 mM sodium pyruvate. The Cells were maintained at 37°C and humidified with 95% air and 5% CO_2 for cell culture.

Silencing c-maf and mafB expression by short hairpin RNA

The knockdown of c-Maf and MafB genes were performed in myeloma cells.[51] Briefly, the sequence for shRNA specific to human c-MAF gene (5′-CCGGTGGAAGACTACTACTGGATGACTCGAGTCATCCAGTAGTAGTCTTCCATT TTT-3′) and to MafB (5′-CCGGGCCTTGTCTTATGGTCAAATTCTCGAGAATTTGACCATAAGACAAGGCTTT TT-3′) in lentiviral vector were transfected into 293T cells and supernatant from culture media were harvested and concentrated. A control oligonucleotide sequence not matching any sequence in the human genome (5′-GATCCCCGACACGCGACTTGTACCACTTCAAGAGAGTGGTACAAGTCGGTCGTC TTTTTA −3′) was used as a control shRNA sequence (designated as shCont). The myeloma cells were infected with lentivirus supernatants for 48 hours and total RNA was isolated using TRIzol (Life Technologies) according to the manufacturer's instructions.

Real-time quantitative PCR

The gene expression was detected by qRT-PCR analysis. Briefly, one microgram of total RNA was reverse transcribed into cDNA. Quantitative PCR (qPCR) was performed using an ABI Prism 7600 sequence detection system (Applied Biosystems, Foster City, CA). The primers specific for human c-Maf (Hs04185012_s1), MafB (forward primer 5′-gcccgaccgaacagaagac-3′, reverse primer 5′-ctcgggcgtcaggttgag-3′, probe 5′-FAM-agcagatgaacccc-MGB-3′), Apobec-3a (Hs02572821_s1),-3b (Hs00358981_m1) and −4 (Hs01941751_s1) were purchased from Applied Biosystems (Paisley, UK). A reaction mixture contained 400 ng cDNA, dedicated buffers with specific primers and probes (5′-labled by 6-carboxyfluorescein and 3′-labled by 1-carboxy-teteramethyrhdamine), and DNA polymerase in a total 20 μl volume. Following 2 min incubation at 50°C and 10 min incubation at 95°C for denaturing, the reaction was subjected to 40-cycle amplification at 95°C for 15 s to denature and at 60°C for 1 min for annealing/extension. Each cDNA sample was analyzed in triplicate in parallel with GAPDH as a control. Changes in mRNA concentration were determined by subtracting the CT (threshold cycle) of target gene from the CT of GAPDH (=CT gene-CT GAPDH). The mean of control was subtracted from the target gene reaction (mean control- target-gene=e). The difference was calculated as 2^e by the $2^{-\Delta CT}$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Morgan GJ, Walker BA, Davies FE. The genetic architecture of multiple myeloma. Nat Rev Cancer. 2012; 12:335–348. [PubMed: 22495321]

2. Kuehl WM, Bergsagel PL. Multiple myeloma: evolving genetic events and host interactions. NatRevCancer. 2002; 2:175–187.

3. Zhan F, et al. The molecular classification of multiple myeloma. Blood. 2006; 108:2020–2028. [PubMed: 16728703]

4. Chesi M, Nardini E, Lim RS, Smith KD, Kuehl WM, Bergsagel PL. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood. 1998; 92:3025–3034. [PubMed: 9787135]

5. Walker BA, et al. A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. Blood. 2010; 116:e56–65. [PubMed: 20616218]

6. Pineda-Roman M, et al. Sustained complete remissions in multiple myeloma linked to bortezomib in total therapy 3: comparison with total therapy 2. Br J Haematol. 2008; 140:625–634. [PubMed: 18302711]

7. Walker BA, et al. Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. Blood. 2011; 117:553–562. [PubMed: 20944071]

8. Weinhold N, et al. The CCND1 c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. Nat Genet. 2013; 45:522–525. [PubMed: 23502783]

9. Ross FM, et al. The t(14;20) is a poor prognostic factor in myeloma but is associated with long-term stable disease in monoclonal gammopathies of undetermined significance. Haematologica. 2010; 95:1221–1225. [PubMed: 20410185]

10. Alexandrov LB, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–421. [PubMed: 23945592]

11. Walker BA, et al. Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells. Blood. 2013; 121:3413–3419. [PubMed: 23435460]

12. Walker BA, et al. Translocations at 8q24 juxtapose MYC with genes that harbor superenhancers resulting in overexpression and poor prognosis in myeloma patients. Blood cancer journal. 2014; 4:e191. [PubMed: 24632883]

13. Affer M, et al. Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. Leukemia. 2014

14. Nik-Zainal S, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012; 149:979–993. [PubMed: 22608084]

15. Chapman MA, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011; 471:467–472. [PubMed: 21430775]

16. Peled JU, et al. The biochemistry of somatic hypermutation. Annual review of immunology. 2008; 26:481–511.

17. Ronchetti D, et al. Deregulated FGFR3 mutants in multiple myeloma cell lines with t(4;14): comparative analysis of Y373C, K650E and the novel G384D mutations. Oncogene. 2001; 20:3553–3562. [PubMed: 11429702]

18. Foth M, et al. Fibroblast growth factor receptor 3 activation plays a causative role in urothelial cancer pathogenesis in cooperation with Pten loss in mice. J Pathol. 2014

19. Burns MB, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013; 494:366–370. [PubMed: 23389445]

20. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet. 2013; 45:977–983. [PubMed: 23852168]

21. Melchor L, et al. Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. Leukemia. 2014

22. Gabrea A, Leif Bergsagel P, Michael Kuehl W. Distinguishing primary and secondary translocations in multiple myeloma. DNA repair. 2006; 5:1225–1233. [PubMed: 16829212]

23. Bea S, et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc Natl Acad Sci U S A. 2013; 110:18250–18255. [PubMed: 24145436]

24. Adar R, Monsonego-Ornan E, David P, Yayon A. Differential activation of cysteine-substitution mutants of fibroblast growth factor receptor 3 is determined by cysteine localization. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2002; 17:860–868.

25. Bellus GA, et al. Distinct missense mutations of the FGFR3 lys650 codon modulate receptor kinase activation and the severity of the skeletal dysplasia phenotype. American journal of human genetics. 2000; 67:1411–1421. [PubMed: 11055896]

26. Bacolla A, Cooper DN, Vasquez KM. Mechanisms of base substitution mutagenesis in cancer genomes. Genes. 2014; 5:108–146. [PubMed: 24705290]

27. Boyd KD, et al. A novel prognostic model in myeloma based on co-segregating adverse FISH lesions and the ISS: analysis of patients treated in the MRC Myeloma IX trial. Leukemia. 2012; 26:349–355. [PubMed: 21836613]

28. Avet-Loiseau H. Ultra high-risk myeloma. Hematology Am Soc Hematol Educ Program. 2010; 2010:489–493. [PubMed: 21239841]

29. Mateos MV, et al. Bortezomib, melphalan, and prednisone versus bortezomib, thalidomide, and prednisone as induction therapy followed by maintenance treatment with bortezomib and thalidomide versus bortezomib and prednisone in elderly patients with untreated multiple myeloma: a randomised trial. Lancet Oncol. 2010; 11:934–941. [PubMed: 20739218]

30. Leonard B, et al. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res. 2013; 73:7222–7231. [PubMed: 24154874]

31. Alpar D, et al. Multiplex ligation-dependent probe amplification and fluorescence in situ hybridization are complementary techniques to detect cytogenetic abnormalities in multiple myeloma. Genes Chromosomes Cancer. 2013; 52:785–793. [PubMed: 23720363]

32. Schwab CJ, et al. Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia. Genes Chromosomes Cancer. 2010; 49:1104–1113. [PubMed: 20815030]

33. Boyle EM, et al. A molecular diagnostic approach able to detect the recurrent genetic prognostic factors typical of presenting myeloma. Genes Chromosomes Cancer. 2015; 54:91–98. [PubMed: 25287954]

34. Kozarewa I, et al. A modified method for whole exome resequencing from minimal amounts of starting DNA. PLoS ONE. 2012; 7:e32617. [PubMed: 22403682]

35. Costello M, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013; 41:e67. [PubMed: 23303777]

36. Boeva V, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012; 28:423–425. [PubMed: 22155870]

37. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012; 28:i333–i339. [PubMed: 22962449]

38. Kaiser MF, et al. A TC classification-based predictor for multiple myeloma using multiplexed real-time quantitative PCR. Leukemia. 2013; 27:1754–1757. [PubMed: 23318961]

39. R development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. , editor. R Foundation for statistical Computing; 2013.

40. JAGS 3.4.0. , editor. JAGS 3.4.0. Just Another Gibbs Sampler. 2013 edn

41. Nuijten, M.; Wetzels, R.; Matzke, D.; Dolan, CV.; Wagenmakers, EJ. Default Bayesian hypothesis tests for correlation, partial correlation, and mediation. 1.0 edn. 2014.

42. Wetzels R, Wagenmakers EJ. A default Bayesian hypothesis test for correlations and partial correlations. Psychonomic bulletin & review. 2012; 19:1057–1064. [PubMed: 22798023]

43. Therneau, T.; Grambsch, P. Modeling Survival Data: Extending the Cox Model. Springer; 2000.

44. Therneau, T. A Package for Survival Analysis in S. R package version 2.37-7 edn2014.

45. Hothorn T, Hornik K, van de Wiel MA, Zeileis S. Implementing a Class of Permutation Tests: The {coin} Package. Journal of Statistical Software. 2008; 28:1–23.

46. Hothorn T, Hornik K, van de Wiel MA, Zeileis A. A Lego System for Conditional Inference. The American Statistician. 2006; 60:257–263.

47. Ohbayashi K, et al. A xeno-transplantable plasma cell leukemia line with a split translocation of the IgH gene. Cancer Genet Cytogenet. 2003; 144:31–35. [PubMed: 12810253]

48. Hanamura I, et al. Ectopic expression of MAFB gene in human myeloma cells carrying (14;20) (q32;q11) chromosomal translocations. Japanese journal of cancer research : Gann. 2001; 92:638–644. [PubMed: 11429052]

49. Chesi M, et al. Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. Blood. 1998; 91:4457–4463. [PubMed: 9616139]

50. Qiang YW, et al. Wnts induce migration and invasion of myeloma plasma cells. Blood. 2005; 106:1786–1793. [PubMed: 15886323]

51. Qiang YW, et al. Myeloma-derived Dickkopf-1 disrupts Wnt-regulated osteoprotegerin and RANKL production by osteoblasts: a potential mechanism underlying osteolytic bone lesions in multiple myeloma. Blood. 2008; 112:196–207. [PubMed: 18305214]
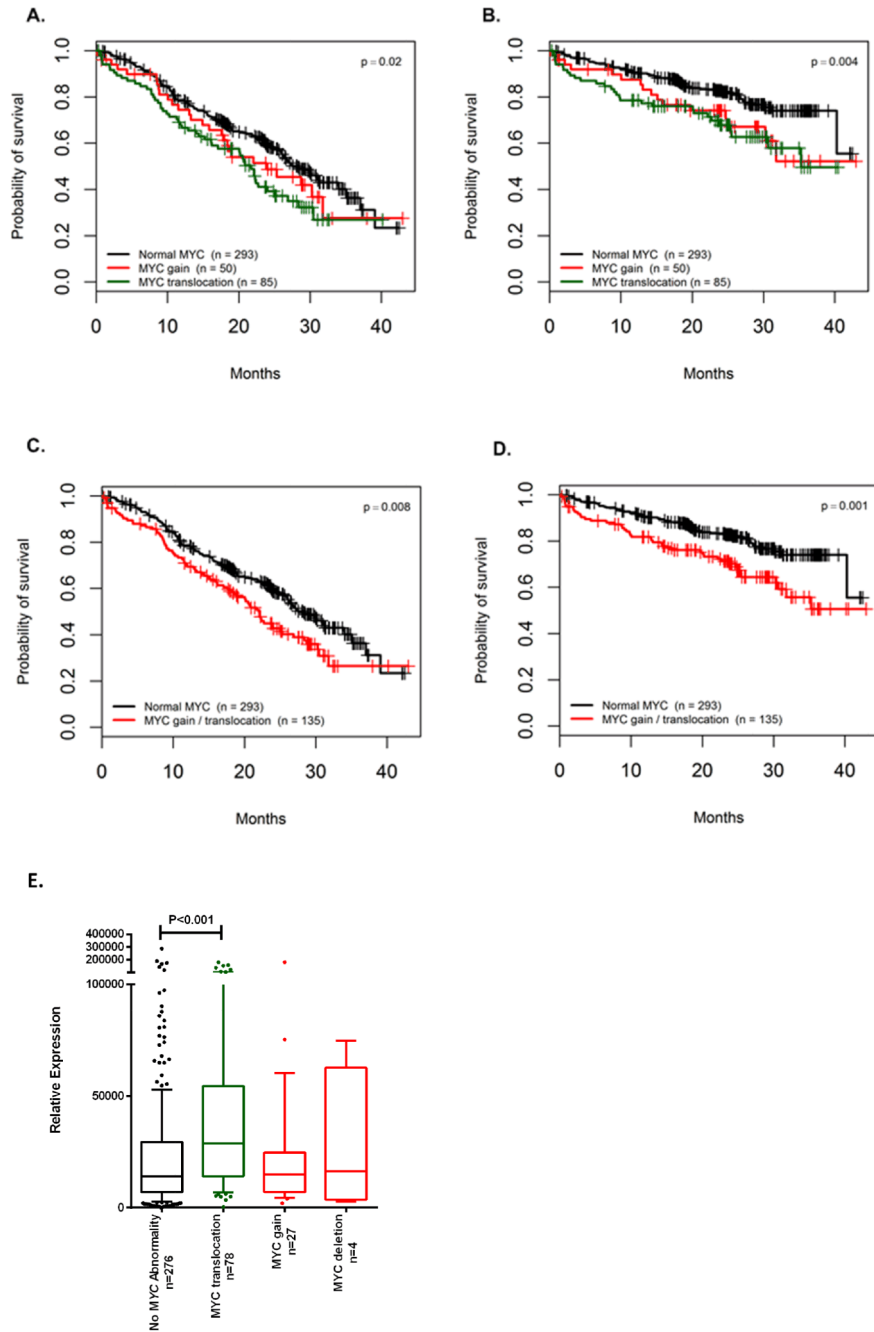
Figure 1. MYC abnormalities affect survival

Progression free (A) and overall survival (B) of patients with *MYC* translocation, gain or no abnormality. Progression free (C) or overall survival (D) of patients comparing no *MYC* abnormality to those with either a translocation or gain of the locus. E, Relative expression of *MYC* in samples with no *MYC* abnormality, a translocation at 8q24 or gain or loss of the *MYC* locus. The box and whisker plot shows the 10 and 90 percentiles (whiskers) and the median and 25 and 75 percentiles (box). P-values determined by log rank test.
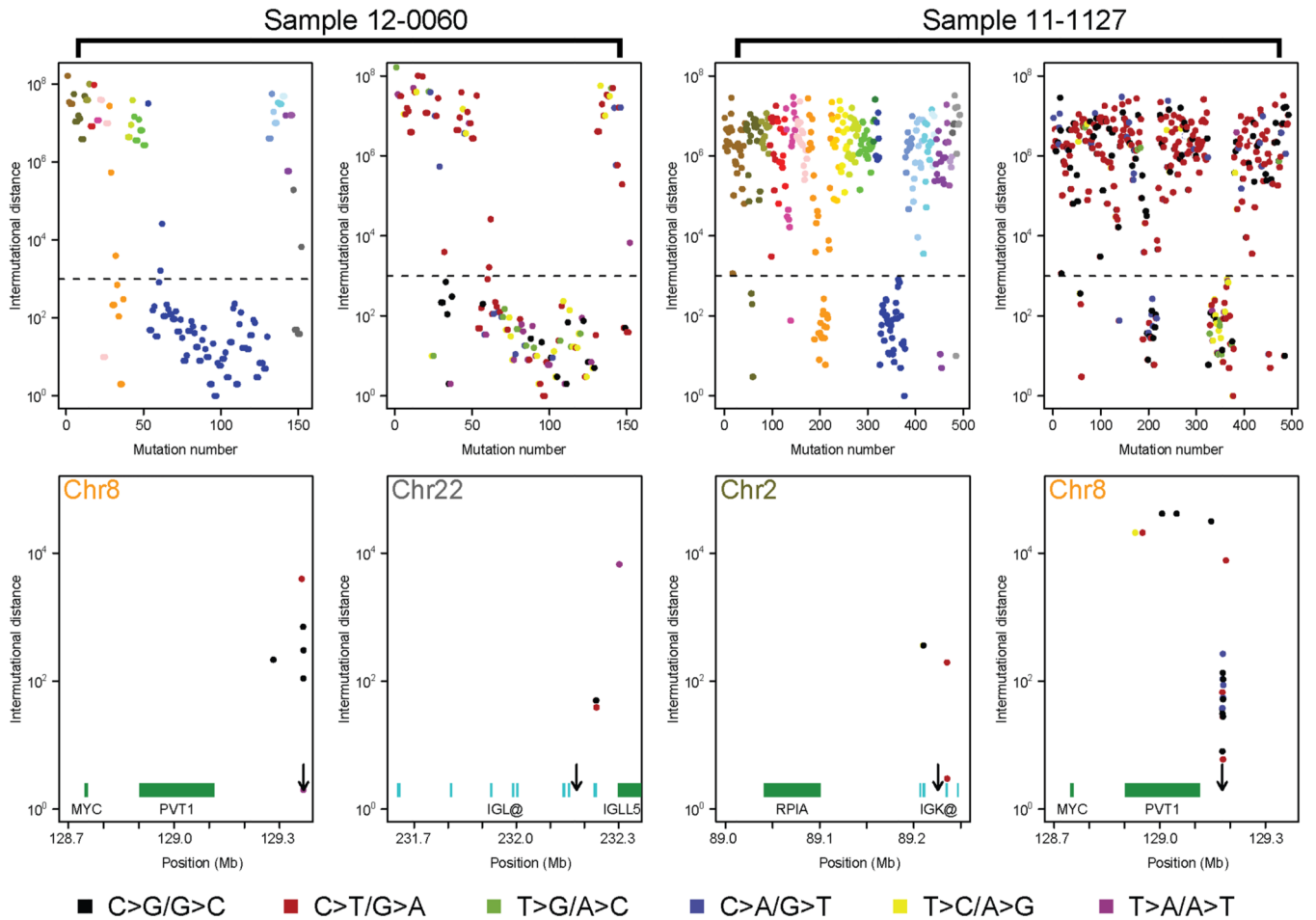
Figure 2. Kataegis in myeloma

Kataegis at 8q24 coincides with translocation breakpoints and occurs on both chromosomes where a translocation has occurred. Examples of two samples are shown. For each sample the top left plot shows the distances between mutations and are colored on a chromosomal basis according to UCSC coloring. The top right panel show the same data but is colored by mutation type, as per the key. The bottom panels show the location on chromosome 8 and the partner translocation chromosome (22 or 2, respectively) where kataegis is found with genes or Ig loci segments indicated in green or cyan, respectively. The arrows indicated the position of the translocation breakpoint.
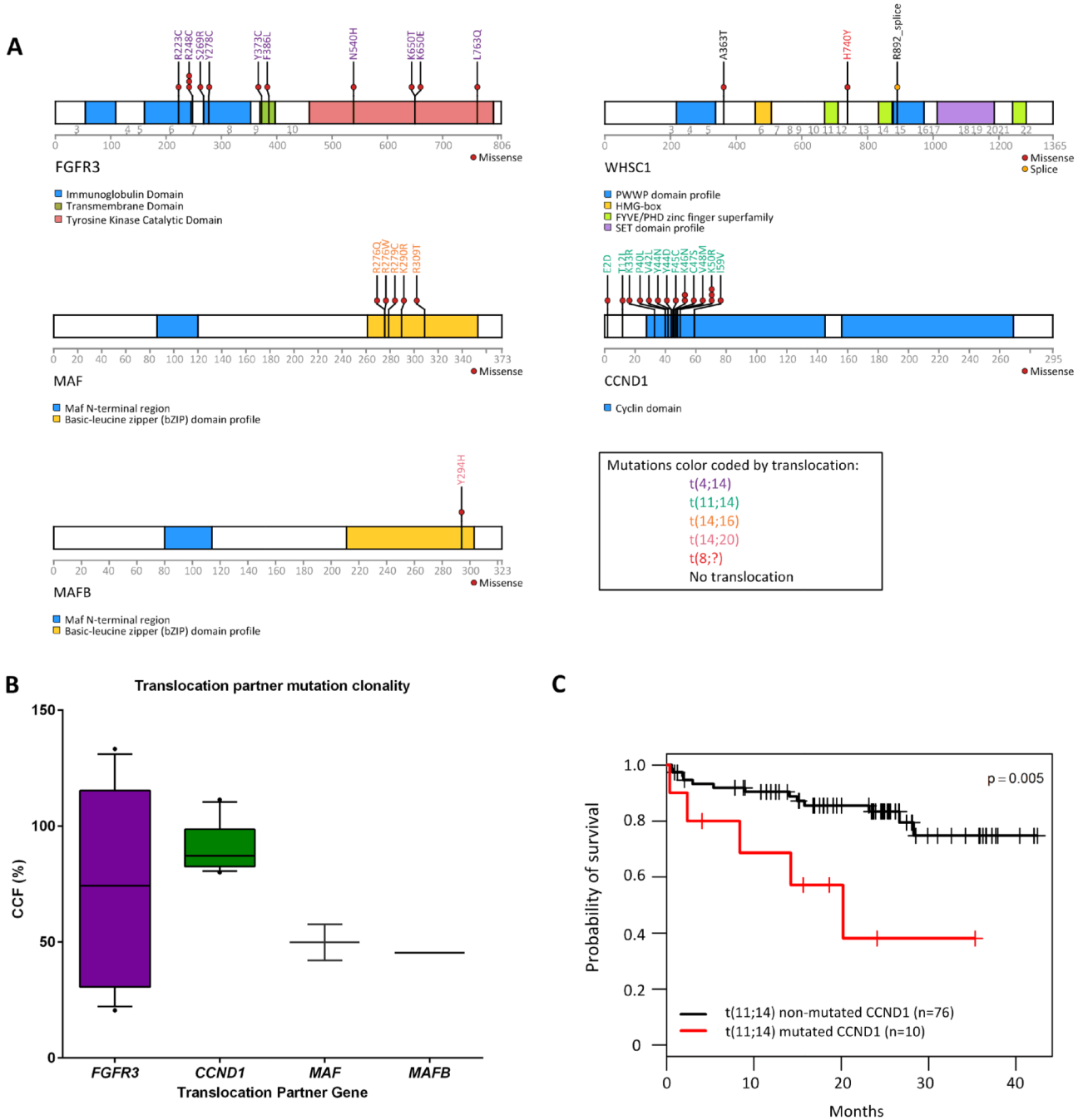
Figure 3. Mutations in translocation partner oncogenes

(A) Non-synonymous mutations in translocation partner oncogenes are depicted along with the translocation group they occur in. (B) The cancer cell fraction (CCF) in which the mutations occur. (C) Mutation of *CCND1* in t(11;14) samples results in decreased overall survival in Myeloma XI samples. The box and whisker plot shows the 10 and 90 percentiles (whiskers) and the median and 25 and 75 percentiles (box).
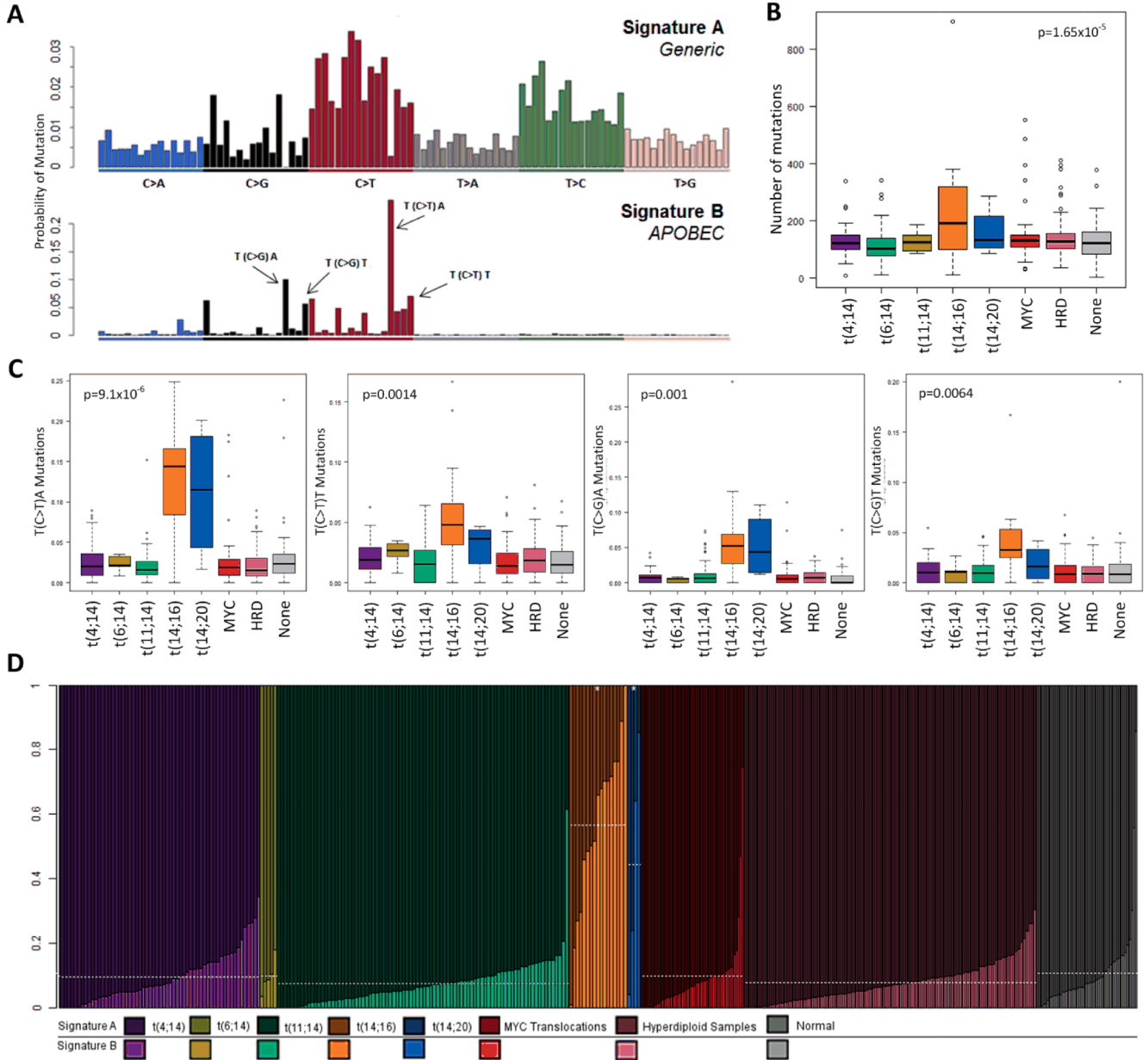
Figure 4. Analysis of mutation context identifies two signatures in myeloma

(A) Mutation context identifies two signatures in myeloma, Signature A and Signature B. (B) t(14;16) and t(14;20) samples have significantly more mutations than other translocation groups. (C) The mutational context split by translocation group identifies t(14;16) and t(14;20) with more mutations which make up Signature B. (D) Stacked bar chart showing the percentage contribution of the two signatures identified by NMF in each sample, ordered by translocation group. The box and whisker plot shows the 10 and 90 percentiles (whiskers) and the median and 25 and 75 percentiles (box). P-values determined by ANOVA.
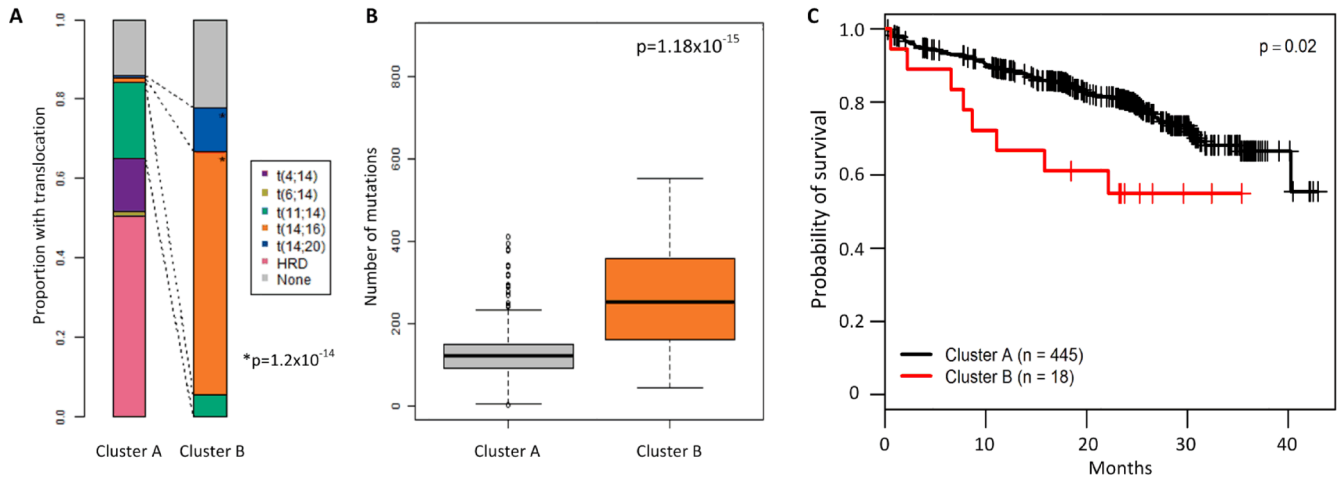
Figure 5. Myeloma mutations can be categorised as belonging to Signature A or Signature B
(A) Samples which mostly have Signature B mutations dominate Cluster B and are enriched for t(14;16) and t(14;20). P-value determined by a linear model. (B) Samples in Cluster B have more mutations than those in Cluster A, p-value determined by ANOVA. Patients in Cluster B have a significantly worse overall survival, p-value by log rank test (C). The box and whisker plot shows the 10 and 90 percentiles (whiskers) and the median and 25 and 75 percentiles (box).
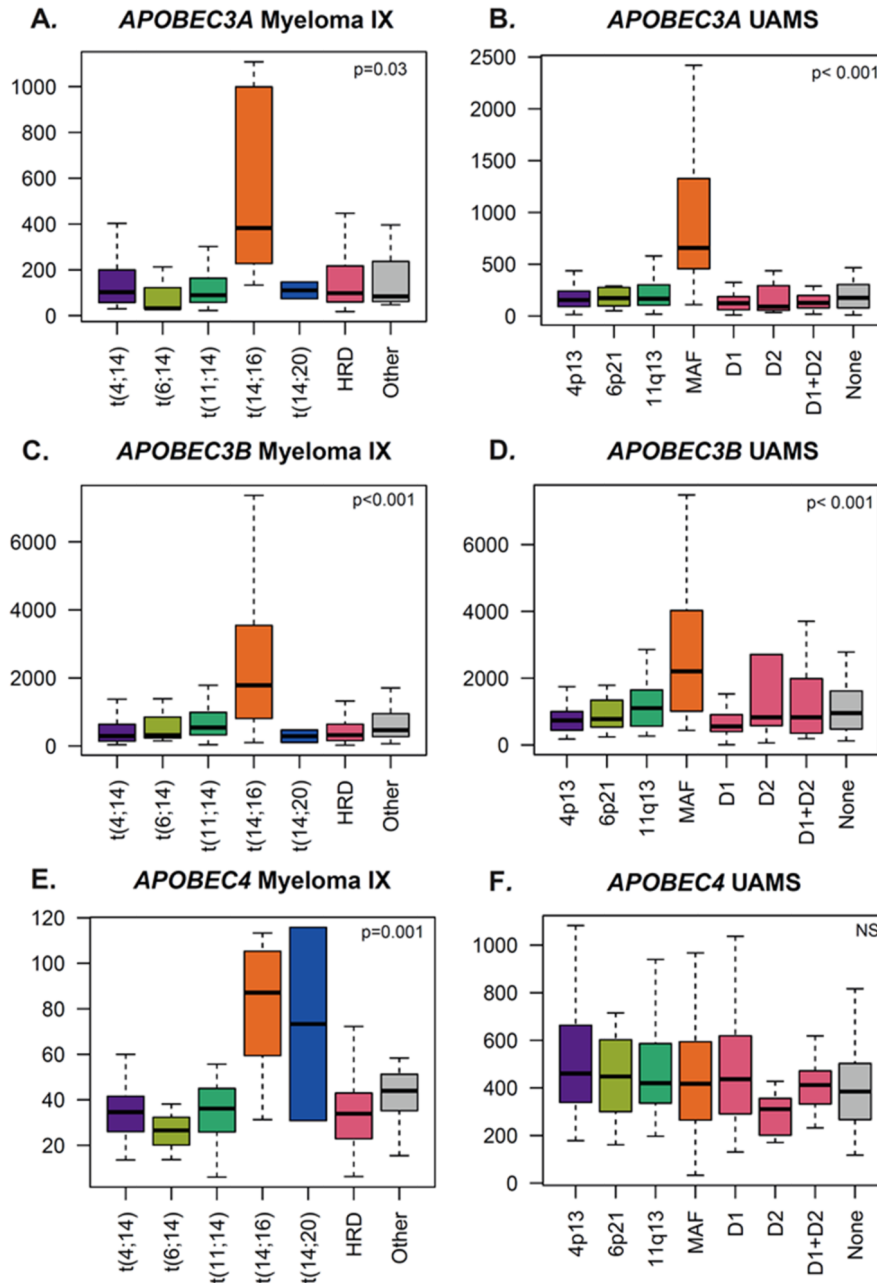
Figure 6. t(14;16)/MAF samples have increased expression of APOBEC genes in the Myeloma IX (GSE15695) and the UAMS datasets (GES4581)

APOBEC3A (210873_x_at) and *APOBEC3B* (206632_s_at) were tested for increased expression in the t(14;16)/MAF samples and the significant results shown. P-value by ANOVA.
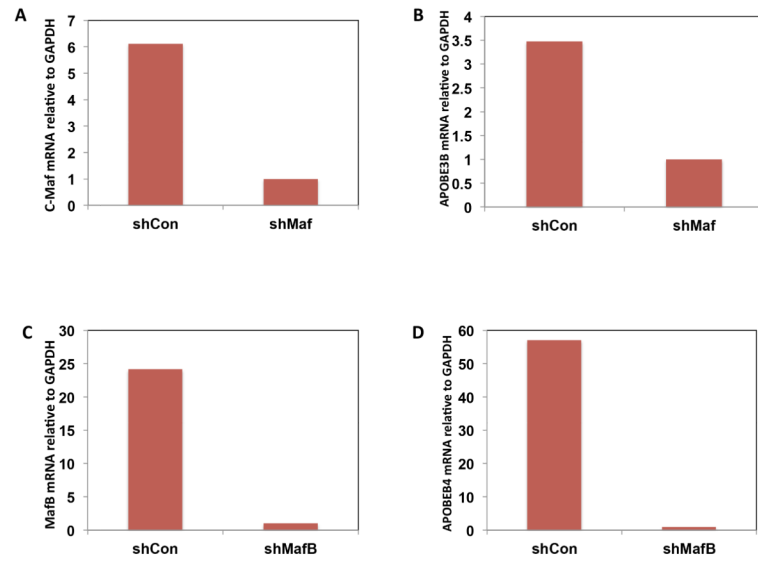
Figure 7. Knockdown of MAF and MAFB by shRNAs induced decrease in the expression of APOBECs

RPMI8226 cells (A and B) were transiently infected with supernatant containing control shRNA (shCont) or shRNA specific for c-Maf (shMaf) for 48 hours. Total RNA was then isolated and subjected to qRT-PCR analysis to detect *MAF* mRNA (A) and *APOBEC3B* mRNA (B). Sachi cells (C and B) were transiently infected with supernatant containing control shRNA (shCont) or shRNA specific for *MAFB* (shMafB) for 48hours and expression of *MAFB* (C) and *APOBEC4* (D) were analyzed.

Table 1

Patient Demographics

Variable	n=463
Median age (range)	68 (31-89)
Sex ratio (M:F)	1.4:1
Pathway	
Intensive	262 (57%)
Non-intensive	201 (43%)
Isotype	
IgA	121 (26%)
IgD	8 (1.7%)
IgG	263 (57%)
Light chain only	55 (12%)
Non-Secretors	2 (1%)
Data missing	14 (3%)
ISS stage	
I	140 (30%)
II	135 (29%)
III	159 (34%)
II+III	294 (63%)
Data missing	29 (6%)
Beta-2-microglobulin	
3.5 mg/L	153 (33%)
3.5<-<5.5 mg/L	119 (26%)
5.5 mg/L	162 (35%)
Data missing	29 (6%)
Creatinine	
104 µmol/L	162 (35%)
150 µmol/L	57 (13%)
Bone disease	324 (70%)
Hypercalcaemia 2.65 mmol/L	58 (12.5%)
Copy number abnormality (gene)	
del(1p33) (*FAF1/CDNK2C*)	39 (8.5%)
del(1p12) (*FAM46C*)	111 (24%)
gain(1q21.2) (*CKS1B*)	166 (36%)
del(13q) (*RB1*)	195 (42%)
del(17p) (*TP53*)	44 (9.5%)
Hyperdiploidy	239 (52%)
Translocations	
t(4;14)	59 (12.7%)
t(6;14)	5 (1%)
t(11;14)	86 (18.6%)
t(14;16)	17 (4%)
t(14;20)	4 (1%)