Article:
Walker, BA, Wardell, CP, Murison, A et al. (18 more authors) (2015) APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. Nature Communications, 6. ARTN 6997.

https://doi.org/10.1038/ncomms7997

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

Authors:

Brian A Walker[1], Christopher P Wardell[1], Alex Murison[1], Eileen M Boyle[1], Lorenzo Melchor[1], Charlotte Pawlyn[1], Martin F Kaiser[1], Dil B Begum[1], Nasrin B Dahir[1], Paula Proszek[1], David C Johnson[1],Ya-Wei Qiang[2], John 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], Gareth J Morgan[1].

Affiliations:

1- Division of Molecular Pathology, The Institute of Cancer Research, London, United Kingdom
2- University of Arkansas Medical School, 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

Conflicts of Interest: The authors have no relevant conflicts of interest to disclose.

Financial Support: This work was supported by a Myeloma UK program grant, a Cancer Research UK CTAAC sample collection grant and a Cancer Research UK Biomarkers and Imaging Discovery and Development grant as well as funds from the National Institute of Health Biomedical Research Centre at the Royal Marsden Hospital.

Corresponding author:

Professor Gareth Morgan, Division of Molecular Pathology, The Institute of Cancer Research, London, SM2 5NG. Email: gareth.morgan@icr.ac.uk Telephone: +44 2087224130 Fax: +44 2087224432

Abstract:

We have sequenced 463 presenting cases of myeloma entered into the UK Myeloma XI study using whole exome sequencing plus capture of the Ig heavy and light chain loci together with the region on 8q24 surrounding *MYC*. We identify frequent (19%) translocation into *MYC* which together with copy number gain and loss increases the prevalence of *MYC* abnormalities to 57.5%. *MYC* translocations are positively associated with the hyperdiploid subgroup and poor clinical outcome. In addition to these adverse prognostic translocations 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. Kataegis, a further mutational pattern associated with APOBEC deregulation, is seen at the sites of the *MYC* translocation both at 8q24 and the translocated partner chromosome. The APOBEC mutational signature seen in myeloma is, therefore, associated with poor prognosis primary and secondary translocations and the molecular mechanisms involved in generating it are potential therapeutic targets.

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.

Results

Rearrangements at MYC are the most common translocation in presenting myeloma and are associated with a poor outcome

Whole exome sequencing was performed on 463 presentation patients enrolled into the Myeloma XI trial. 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 Figure 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) (correlation=-0.13, BF=2.11) and positive correlation with the hyperdiploid (HRD) group (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 and B**).

We used copy number data, generated from exome sequencing, to identify additional abnormalities surrounding the *MYC* locus. There is an abnormality in 257 of the patients (55%), which can be subdivided into those with a translocation alone

(8 patients), those with a translocation and a copy number abnormality (77 patients), those with focal or chromosomal gains (143 patients) or deletions (29 patients). 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 1C**). Gains and deletions were not associated with significant increases in expression of *MYC*, although those with a deletion did have a median expression similar to those with a translocation (26217 vs. 28815 units) and the lack of significance is presumably due to lack of numbers in the deleted group. We looked at the clinical significance of carrying these abnormalities and show that carrying any *MYC* abnormality is associated with a poor prognosis (**Figure 1A and B**).

Kataegis mutation patterns co-localise 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 (**Table 2**). 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 2**). 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 translocation breakpoint, type of breakpoint (RAG or switch mediated) or allele of the variant associated with the t(11;14). However, there was 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, p=0.005; **Figure 3C**) in the Myeloma XI trial. 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 Figure 3**). 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 Figure 4**).

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. 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 Figure 5**).

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 at the N-terminal of the protein and do not affect the cyclin box fold domains. 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[16].

APOBEC Mutations are enriched in the '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 ($p=1.65 \times 10^{-5}$), **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, $p=9.1 \times 10^{-5}$; 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, $p=2 \times 10^{-16}$; 0.44, $p=8.26 \times 10^{-11}$, 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, $p=1.18 \times 10^{-15}$; **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 (p=0.02; **Figure 5C**). Due to the interaction of the translocation, APOBEC signature and mutational load 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.

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**). 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,[17,18] 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[19,20]. 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 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* locus enabled us to identify of a second mutational signature, kataegis. 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 no specific sequence context within which the mutations on the partner oncogenes occur 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. 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[21]. 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.

If 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[16]. 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. Mutations in *FGFR3* also included the K650 mutation, which constitutively activates the receptor[22]. We conclude, therefore, that these mutations have been positively selected for their importance in the pathogenesis of the disease and could be targeted therapeutically.

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[23]. 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,[24-26] 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[17,18,27].

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 their 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

Methods and any associated references are available in the online version of the paper.

Acknowledgements

The authors would like to thank all the patients and staff at centres throughout the UK whose participation made this study possible. The authors are grateful to all principle investigators for their dedication and commitment to recruiting patients to the study. The principal investigators at the four top recruiting centres were Dr Don Milligan (Heart of England NHS Foundation Trust), Dr Jindriska Lindsay (Kent and Canterbury Hospital), Dr Nigel Russell (Nottingham University Hospital) and Dr Clare Chapman (Leicester Royal Infirmary). The support of the Clinical Trials Research Unit at The University of Leeds was essential to the successful running of the study and the authors would like to thank all the staff including Helen Howard, Corrine Collett, Jacqueline Ouzman, Ana Quartilho and Alex Szubert. We also acknowledge The Institute of Cancer Research Tumour Profiling Unit for their support and technical expertise in this study.

Tables

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%)

Figure Legends

Figure 1: **MYC abnormalities affect survival**. A, Progression free survival. B, Overall survival. C, MYC expression in different abnormality groups.

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.

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.

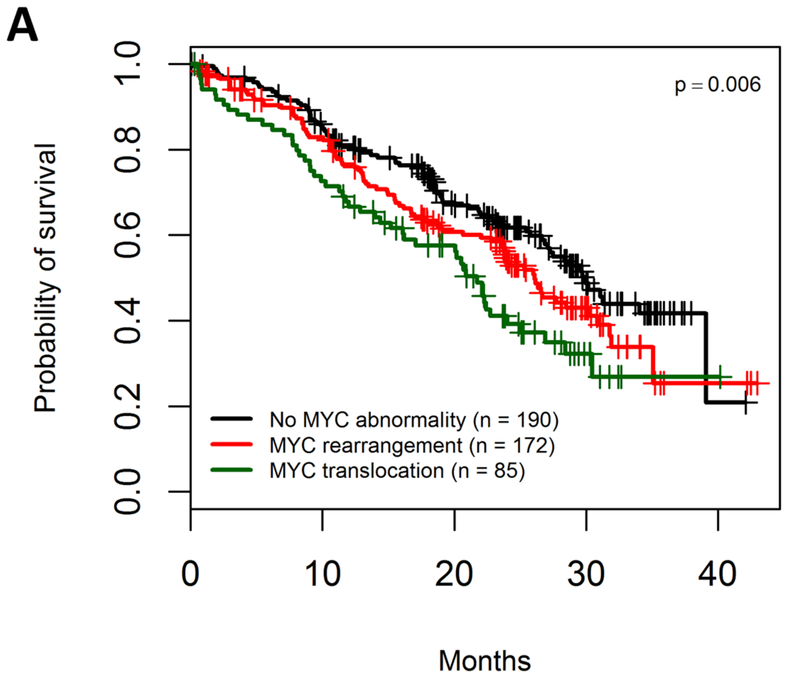
Figure 5: **Myeloma mutations can be categorised as belonging to Signature A or Signature B.** (A) Samples which mostly have Signature B mutations dominate Cluster B and are enriched for t(14;16) and t(14;20). (B) Samples in Cluster B have more mutations than those in Cluster A. Patients in Cluster B have a significantly worse overall survival (C).

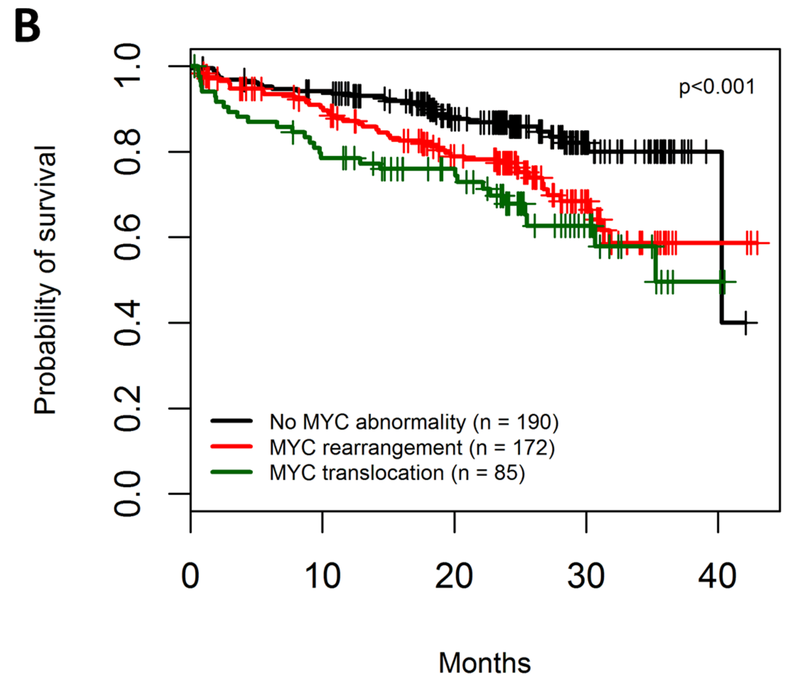
Figure 6: **t(14;16)/MAF samples have increased expression of APOBEC genes in the Myeloma IX (GSE15695) and the UAMS datasets (GES4581).** APOBEC3A (210873_x_at) and APOBEC3B (206632_s_at) were tested for increased expression in the t(14;16)/MAF samples and the significant results shown.

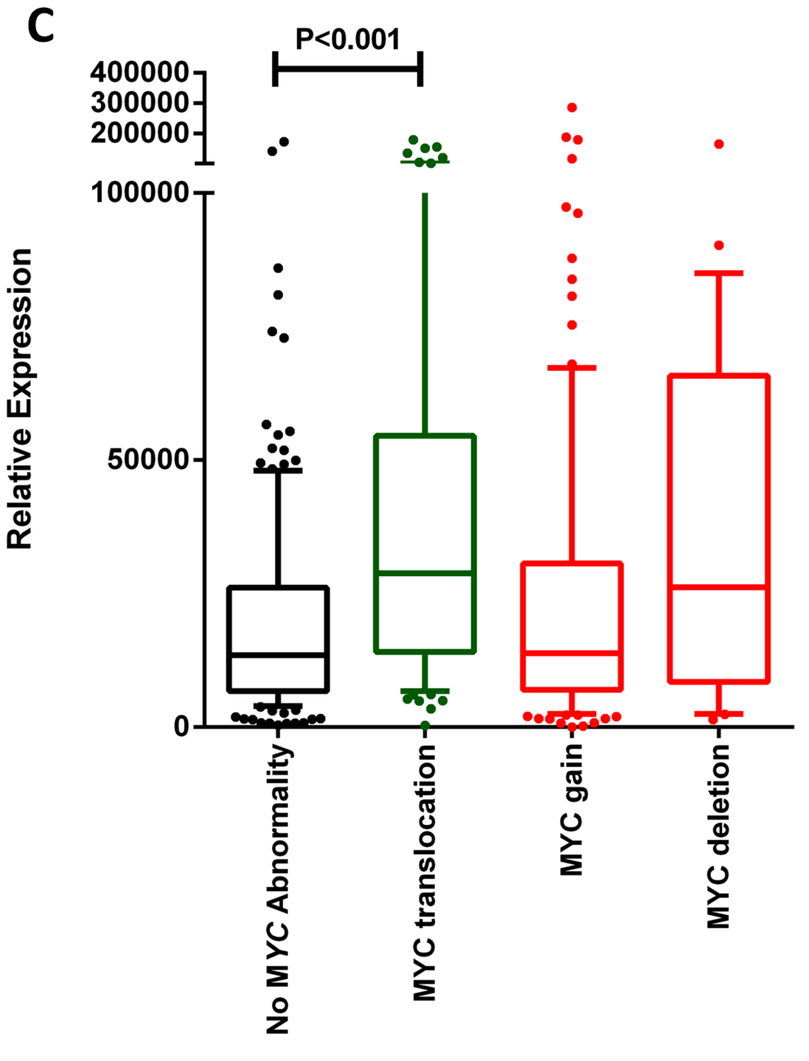
References

1. Morgan, G.J., Walker, B.A. & Davies, F.E. The genetic architecture of multiple myeloma. *Nat Rev Cancer* **12**, 335-48 (2012).
2. Kuehl, W.M. & Bergsagel, P.L. Multiple myeloma: evolving genetic events and host interactions. *Nat.Rev.Cancer* **2**, 175-187 (2002).
3. Zhan, F. *et al.* The molecular classification of multiple myeloma. *Blood* **108**, 2020-8 (2006).
4. Chesi, M. *et al.* The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. *Blood* **92**, 3025-3034 (1998).
5. Walker, B.A. *et al.* A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood* **116**, e56-65 (2010).
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* **140**, 625-34 (2008).
7. Walker, B.A. *et al.* Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood* **117**, 553-62 (2011).
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* **45**, 522-5 (2013).
9. Ross, F.M. *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* **95**, 1221-5 (2010).
10. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
11. Walker, B.A. *et al.* Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells. *Blood* **121**, 3413-9 (2013).
12. Walker, B.A. *et al.* Translocations at 8q24 juxtapose MYC with genes that harbor superenhancers resulting in overexpression and poor prognosis in myeloma patients. *Blood Cancer J* **4**, e191 (2014).
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* **149**, 979-93 (2012).
15. Chapman, M.A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-72 (2011).
16. 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).
17. Burns, M.B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366-70 (2013).
18. Burns, M.B., Temiz, N.A. & Harris, R.S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* **45**, 977-83 (2013).
19. 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).

20. Gabrea, A., Leif Bergsagel, P. & Michael Kuehl, W. Distinguishing primary and secondary translocations in multiple myeloma. *DNA Repair (Amst)* **5**, 1225-33 (2006).
21. Bea, S. *et al.* Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc Natl Acad Sci U S A* **110**, 18250-5 (2013).
22. Bellus, G.A. *et al.* Distinct missense mutations of the FGFR3 lys650 codon modulate receptor kinase activation and the severity of the skeletal dysplasia phenotype. *Am J Hum Genet* **67**, 1411-21 (2000).
23. Bacolla, A., Cooper, D.N. & Vasquez, K.M. Mechanisms of base substitution mutagenesis in cancer genomes. *Genes (Basel)* **5**, 108-46 (2014).
24. Boyd, K.D. *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* **26**, 349-55 (2012).
25. Avet-Loiseau, H. Ultra high-risk myeloma. *Hematology Am Soc Hematol Educ Program* **2010**, 489-93 (2010).
26. Mateos, M.V. *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* **11**, 934-41 (2010).
27. Leonard, B. *et al.* APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res* **73**, 7222-31 (2013).


A
p = 0.006
Probability of survival
0.0
0.2
0.4
0.6
0.8
1.0
No MYC abnormality (n = 190)
MYC rearrangement (n = 172)
MYC translocation (n = 85)
0
10
20
30
40
Months


B
p<0.001
Probability of survival
0.0
0.2
0.4
0.6
0.8
1.0
No MYC abnormality (n = 190)
MYC rearrangement (n = 172)
MYC translocation (n = 85)
0
10
20
30
40
Months


C
P<0.001
Relative Expression
400000
300000
200000
100000
50000
0
No MYC Abnormality
MYC translocation
MYC gain
MYC deletion

Sample 12-0060

Sample 11-1127

Intermutational distance

Mutation number

Chr8

Chr22

Chr2

Chr8

MYC

PVT1

IGL@

IGLL5

RPIA

IGK@

Position (Mb)

C>G/G>C

C>T/G>A

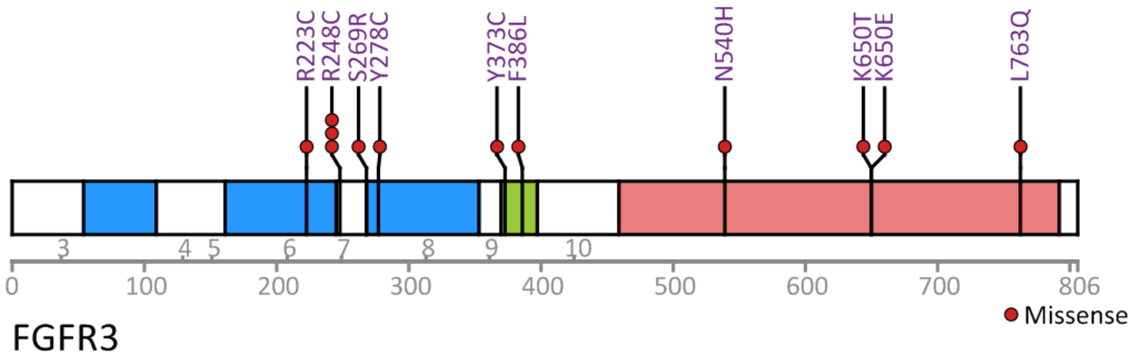
T>G/A>C

C>A/G>T

T>C/A>G

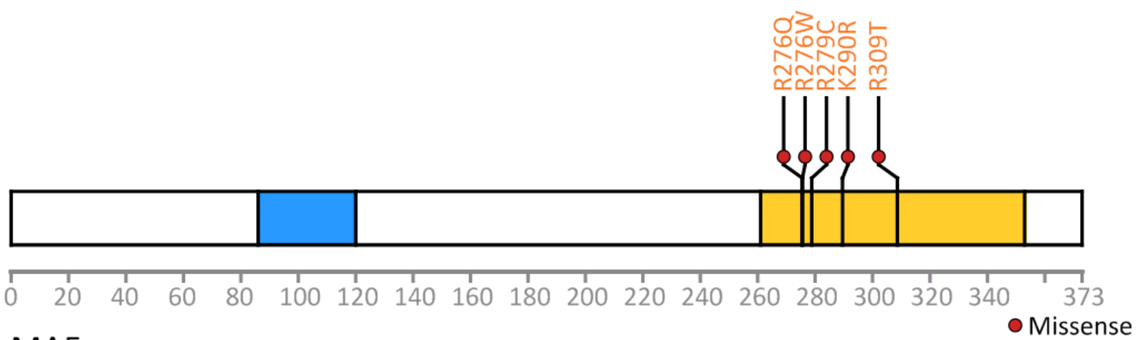
T>A/A>T

A



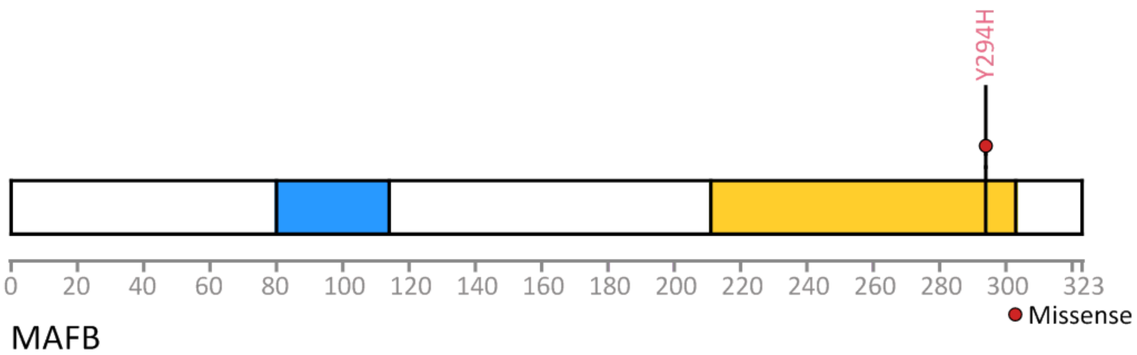
FGFR3

Immunoglobulin Domain
Transmembrane Domain
Tyrosine Kinase Catalytic Domain



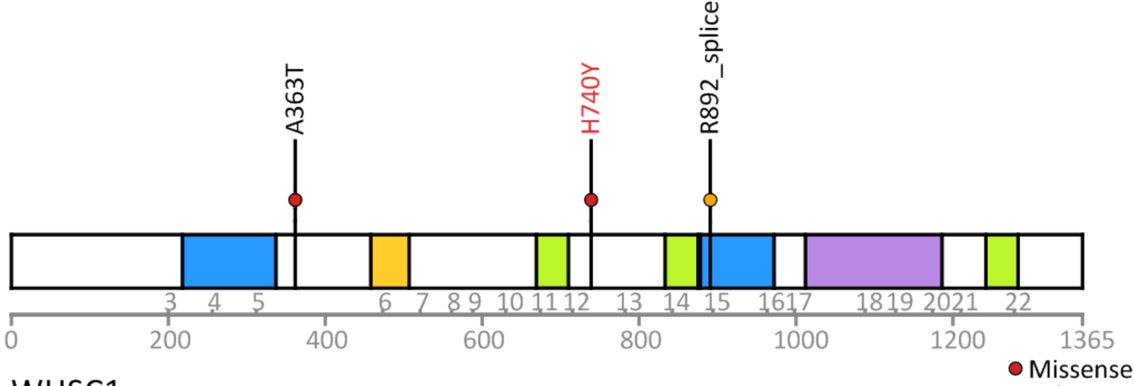
MAF

Maf N-terminal region
Basic-leucine zipper (bZIP) domain profile



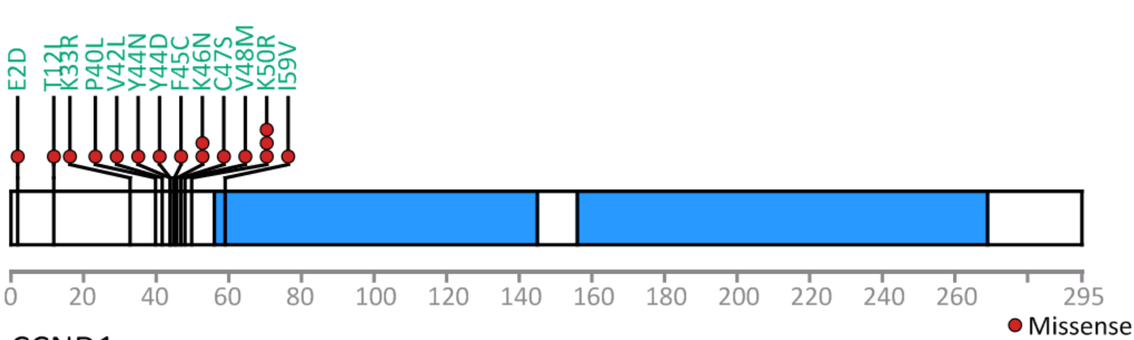
MAFB

Maf N-terminal region
Basic-leucine zipper (bZIP) domain profile



WHSC1

PWWP domain profile
HMG-box
FYVE/PHD zinc finger superfamily
SET domain profile

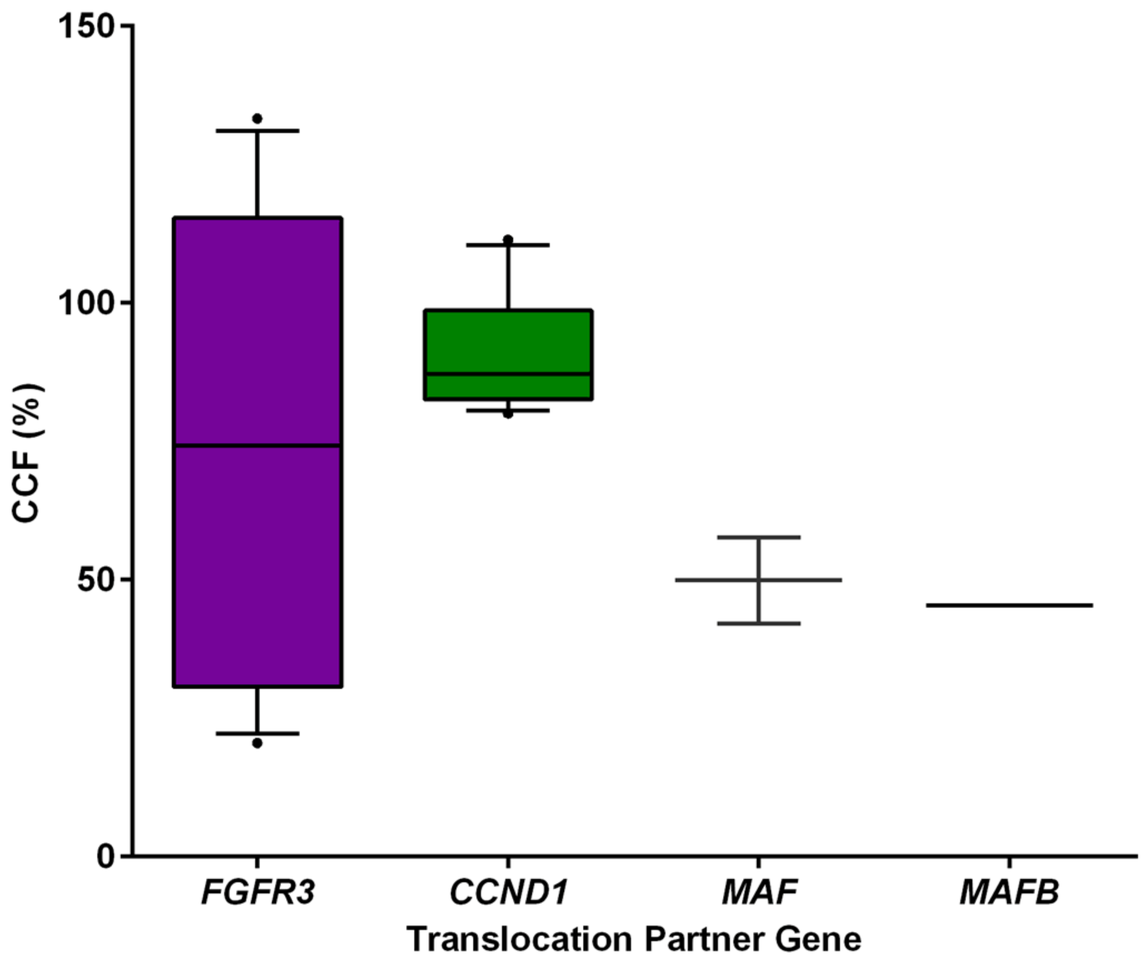


CCND1

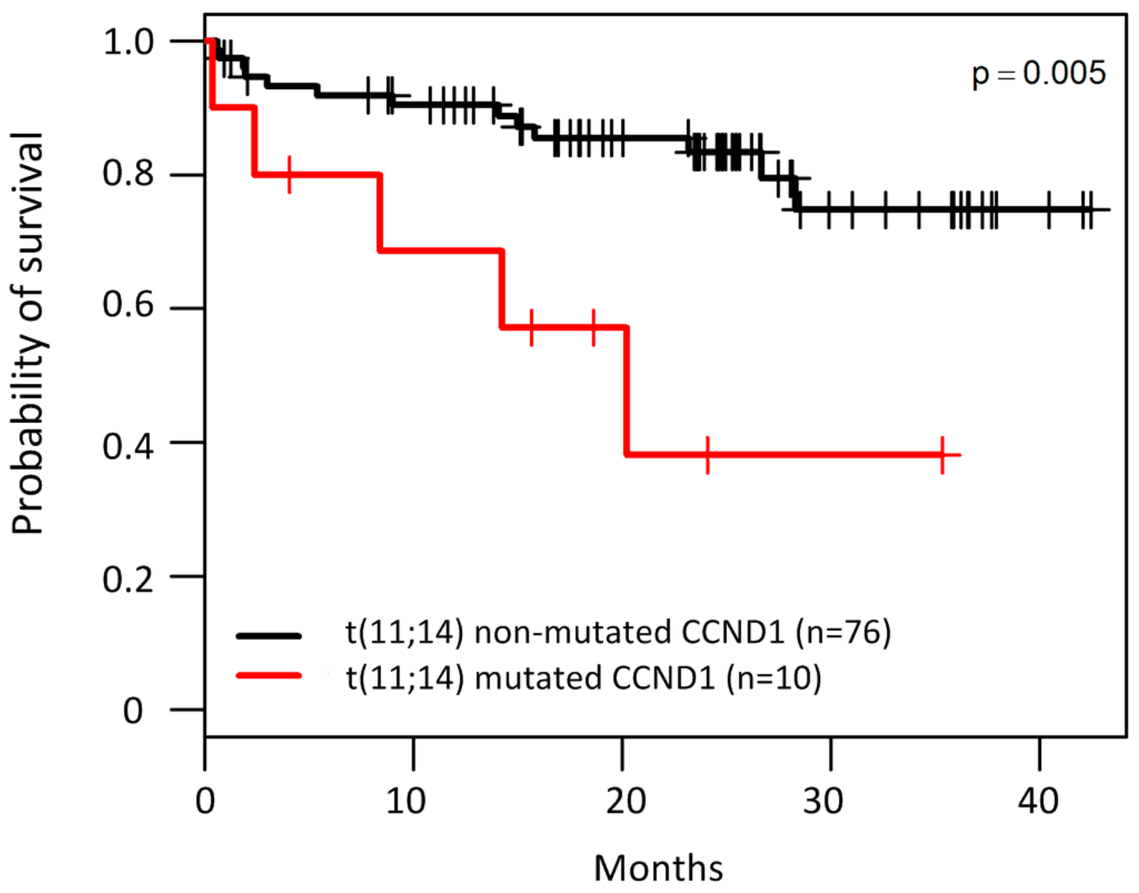
Cyclin domain

Mutations color coded by translocation:
t(4;14)
t(11;14)
t(14;16)
t(14;20)
t(8;?)
No translocation

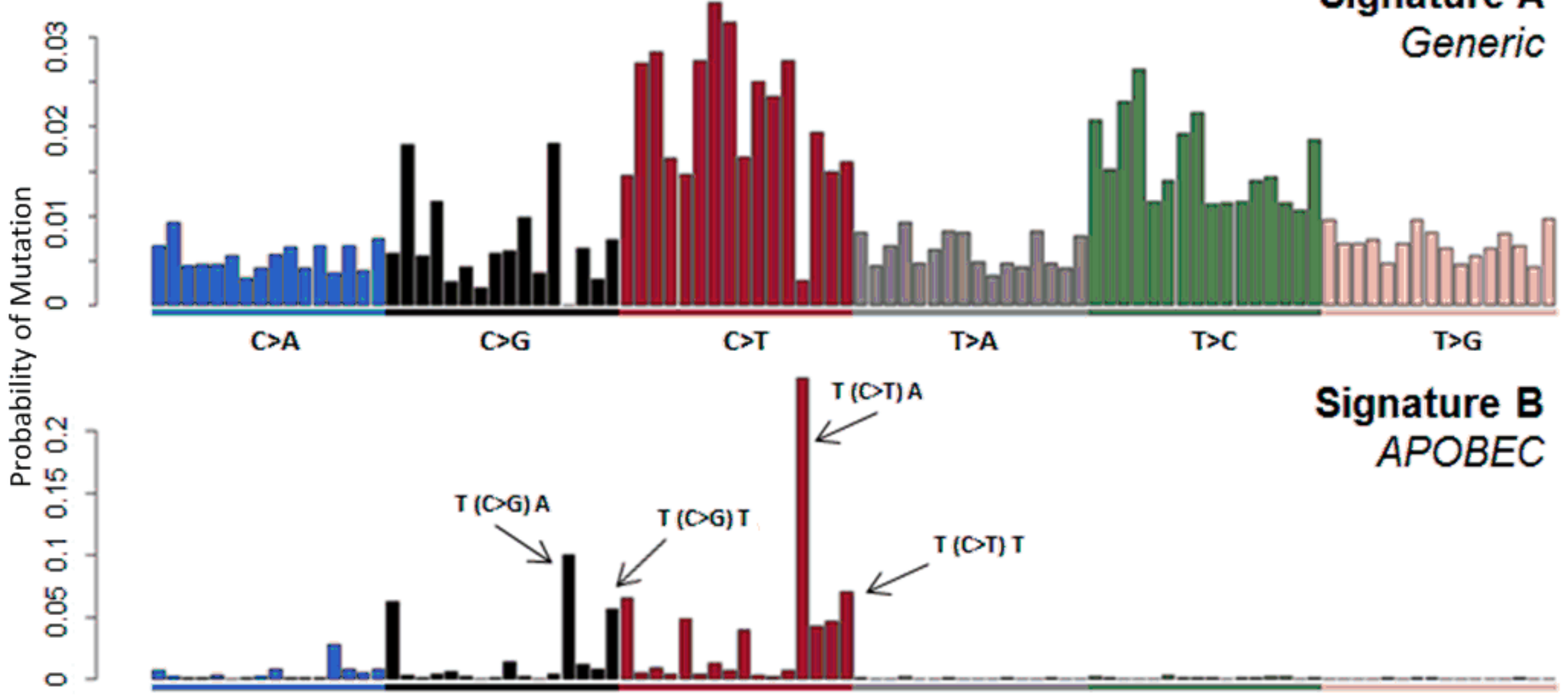
B



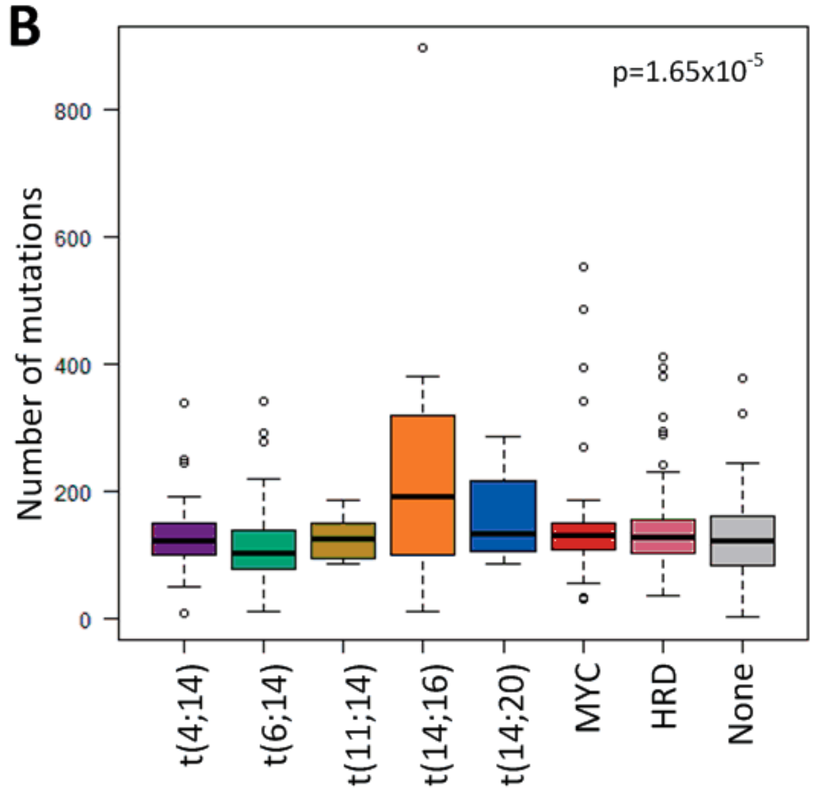
C



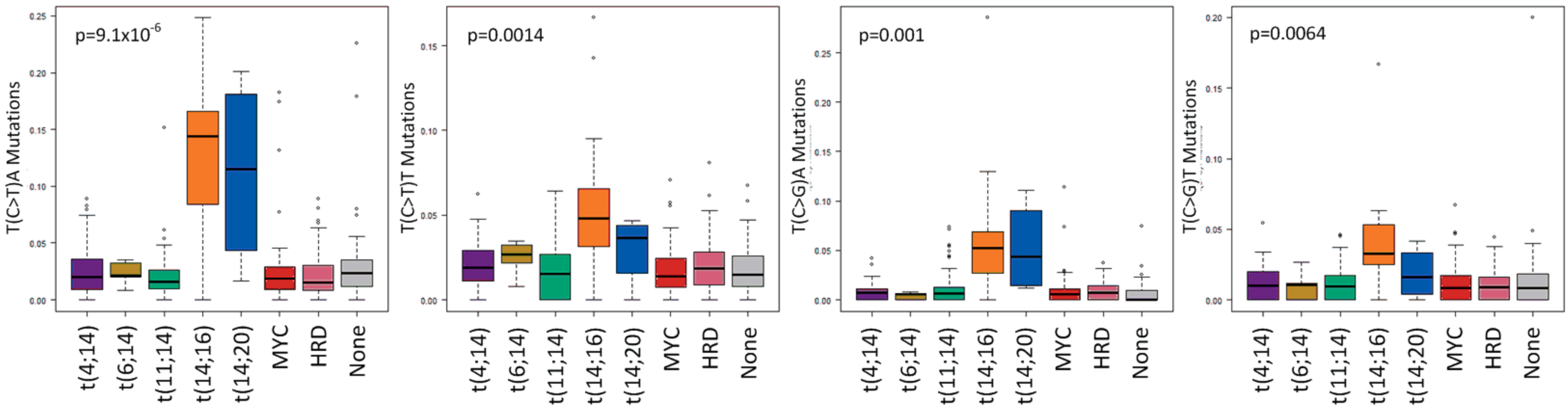
A


Signature A
Generic
Probability of Mutation
0
0.01
0.02
0.03
C>A
C>G
C>T
T>A
T>C
T>G
Signature B
APOBEC
0
0.05
0.1
0.15
0.2
T (C>T) A
T (C>G) A
T (C>G) T
T (C>T) T

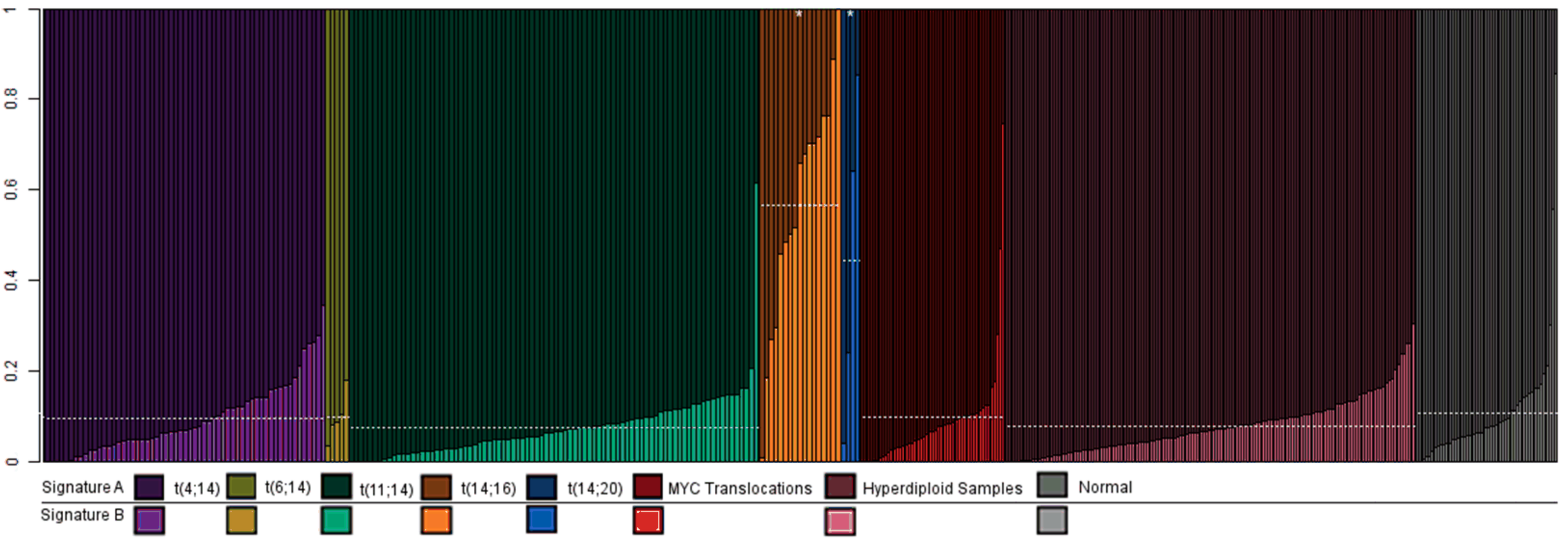
B

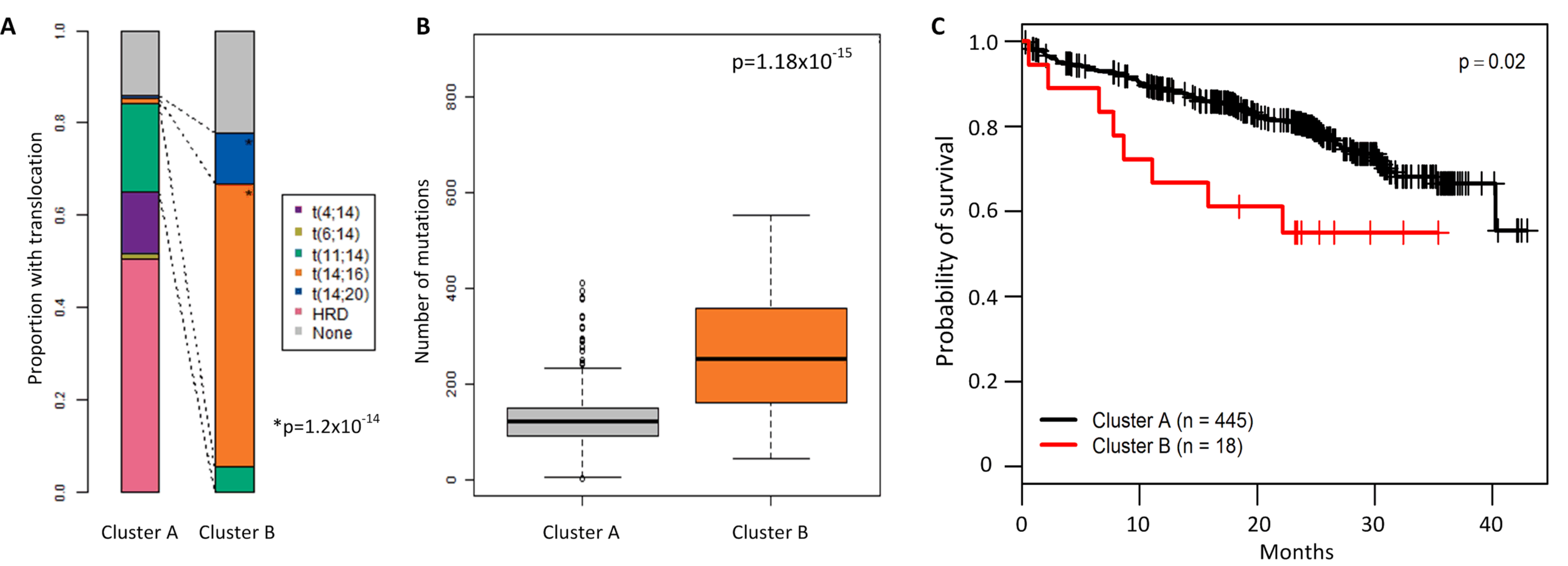

p=1.65x10-5
Number of mutations
0
200
400
600
800
t(4;14)
t(6;14)
t(11;14)
t(14;16)
t(14;20)
MYC
HRD
None

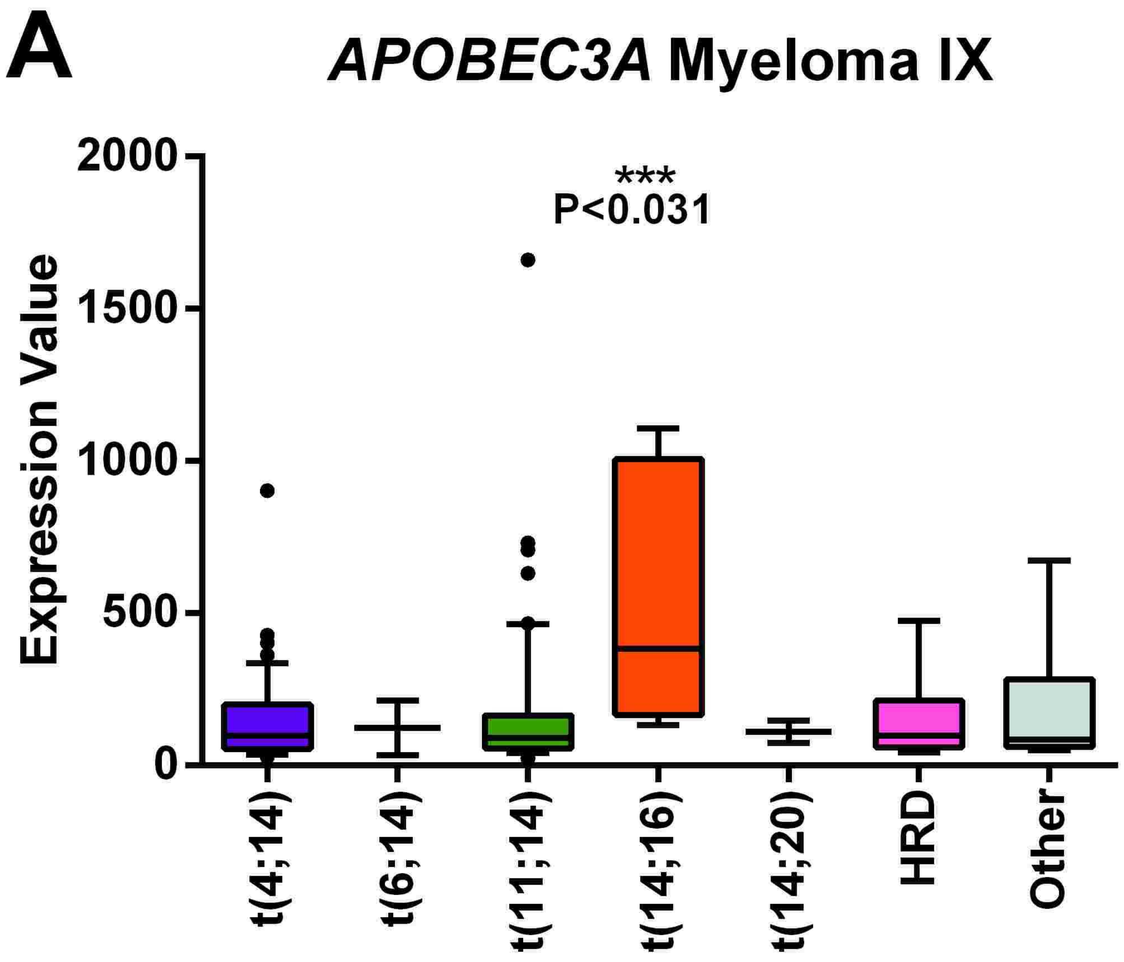
C

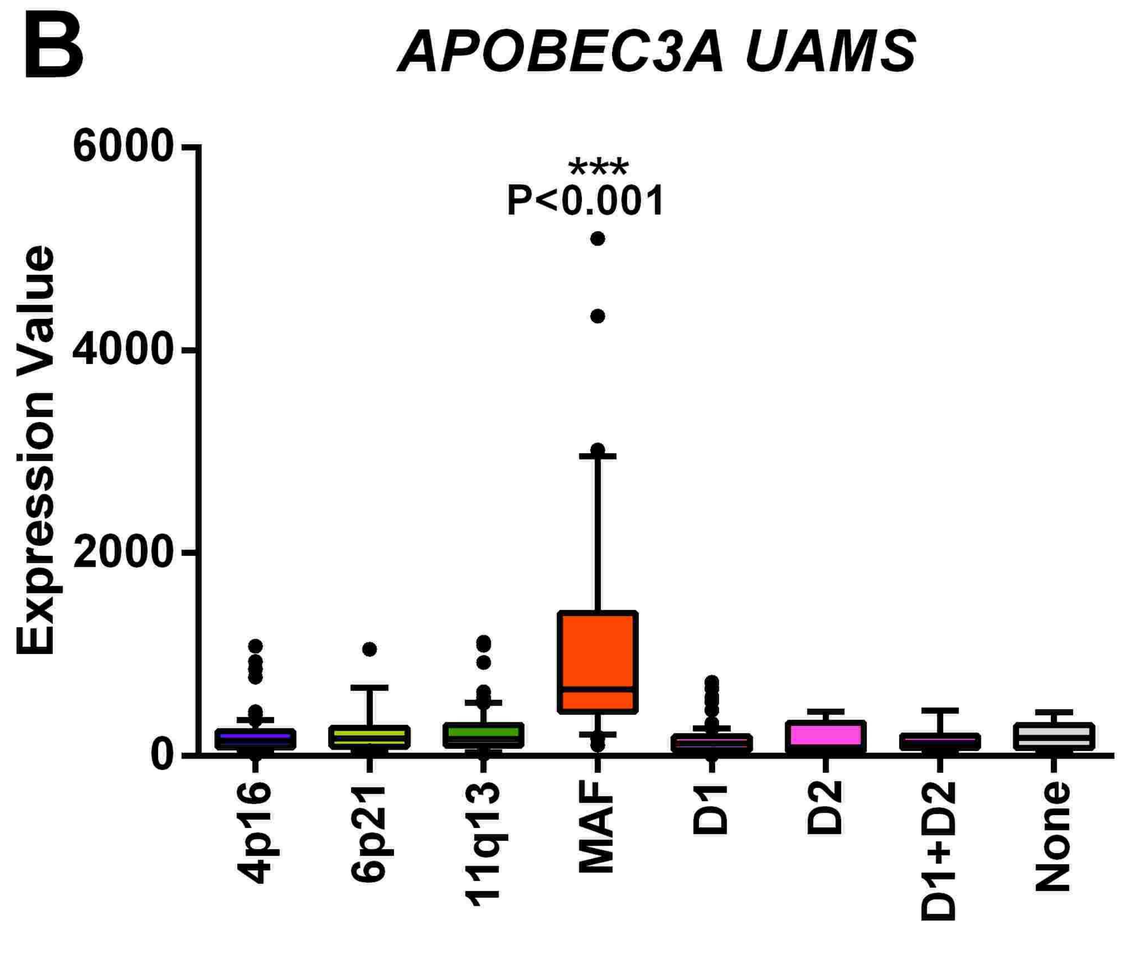

p=9.1x10-6
T(C>T)A Mutations
p=0.0014
T(C>T)T Mutations
p=0.001
T(C>G)A Mutations
p=0.0064
T(C>G)T Mutations
t(4;14)
t(6;14)
t(11;14)
t(14;16)
t(14;20)
MYC
HRD
None

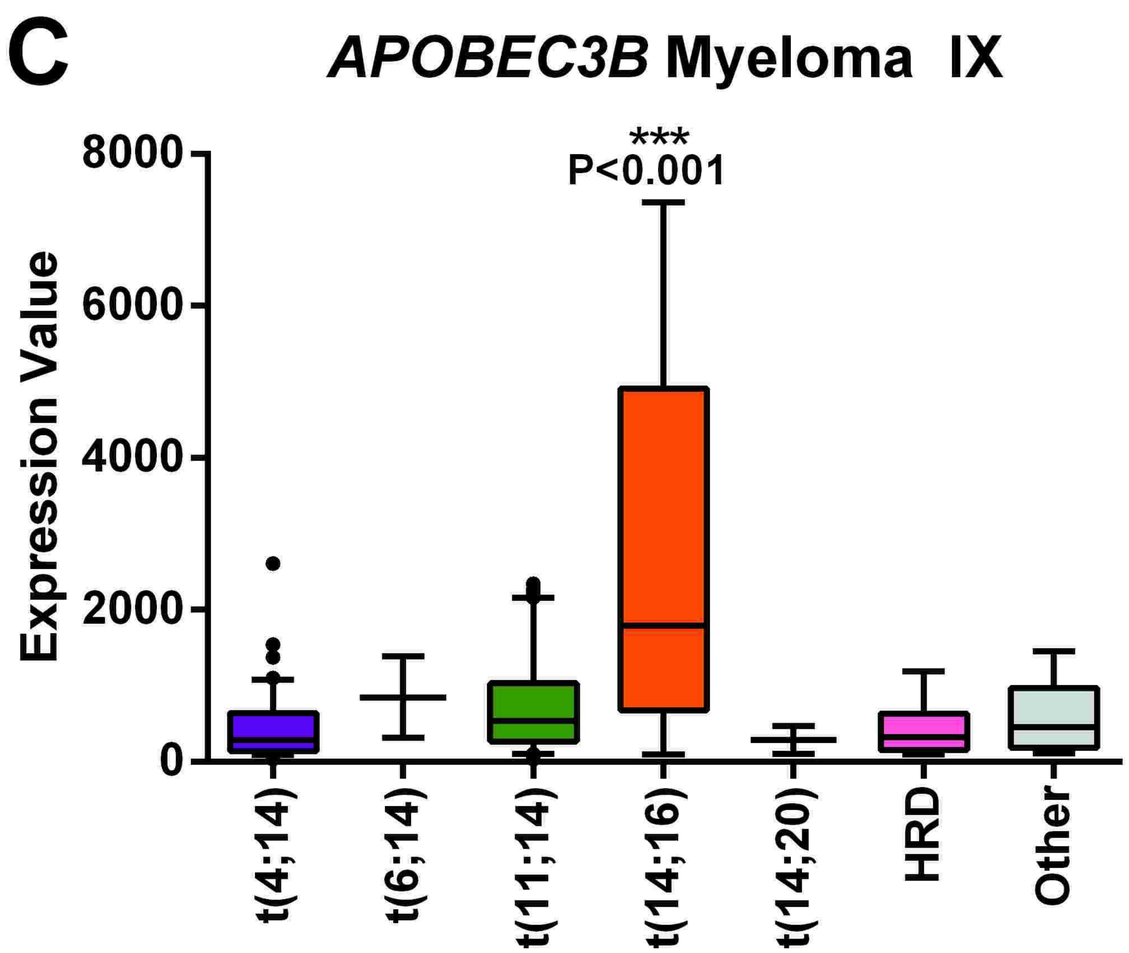
D

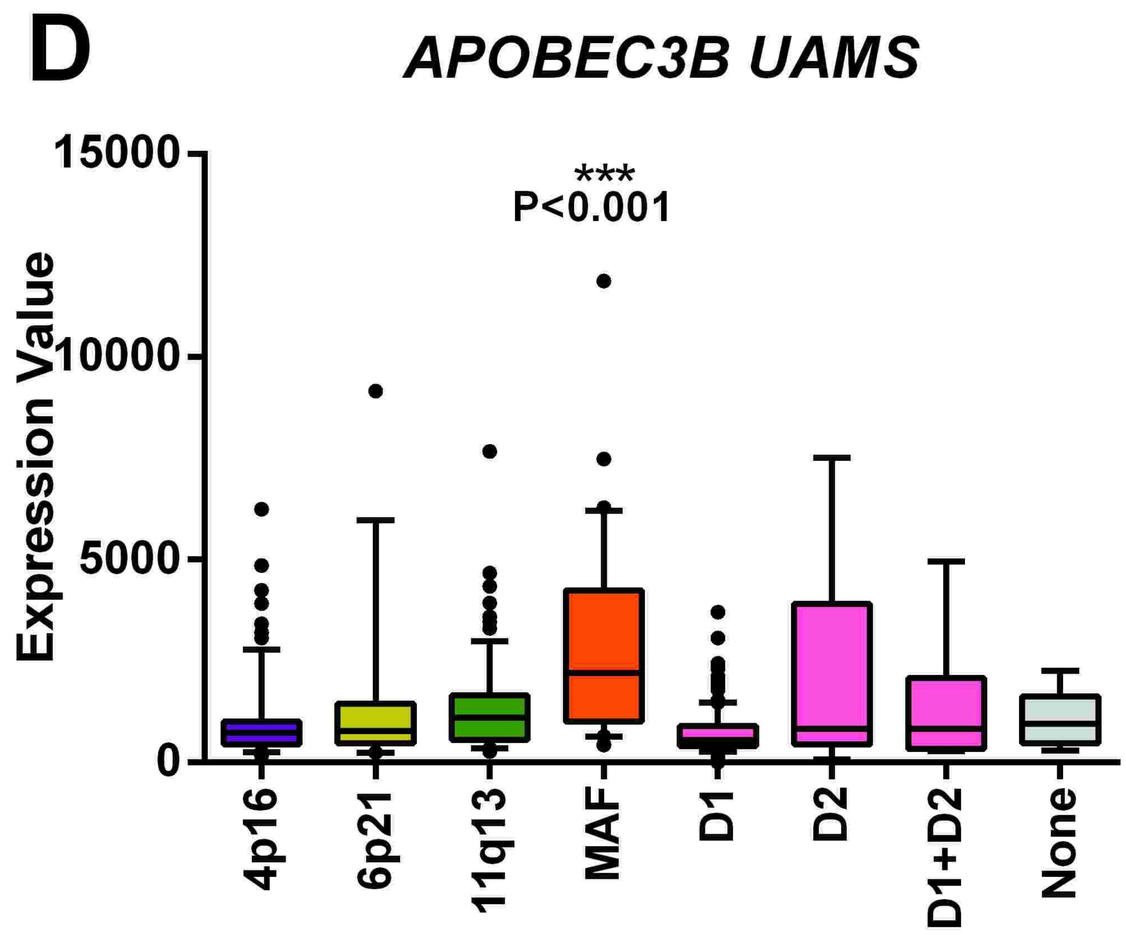

0
0.2
0.4
0.6
0.8
1
Signature A
Signature B
t(4;14)
t(6;14)
t(11;14)
t(14;16)
t(14;20)
MYC Translocations
Hyperdiploid Samples
Normal


A
Proportion with translocation
0.0
0.2
0.4
0.6
0.8
1.0
t(4;14)
t(6;14)
t(11;14)
t(14;16)
t(14;20)
HRD
None
*p=1.2x10-14
Cluster A
Cluster B
B
Number of mutations
0
200
400
600
800
p=1.18x10-15
Cluster A
Cluster B
C
Probability of survival
0
0.2
0.4
0.6
0.8
1.0
p = 0.02
Cluster A (n = 445)
Cluster B (n = 18)
0
10
20
30
40
Months


A
APOBEC3A Myeloma IX
Expression Value
2000
1500
1000
500
0
***
P<0.031
t(4;14)
t(6;14)
t(11;14)
t(14;16)
t(14;20)
HRD
Other


B
APOBEC3A UAMS
Expression Value
6000
4000
2000
0
***
P<0.001
4p16
6p21
11q13
MAF
D1
D2
D1+D2
None


C
APOBEC3B Myeloma IX
Expression Value
8000
6000
4000
2000
0
***
P<0.001
t(4;14)
t(6;14)
t(11;14)
t(14;16)
t(14;20)
HRD
Other


D
APOBEC3B UAMS
Expression Value
15000
10000
5000
0
***
P<0.001
4p16
6p21
11q13
MAF
D1
D2
D1+D2
None